ORTHOGEN: A GENERATIVE ORTHOGONAL LEARNER FOR TIME-VARYING TREATMENTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating conditional distributional potential outcomes (CDPOs) *over time* is important in medicine (e.g., to estimate patient-specific risks under different treatment sequences). However, this task is challenging because of time-varying confounding, yet existing adjustment strategies for this task are limited. In this paper, we aim to learn CDPOs under time-varying treatments using flexible generative models. Our contributions are two-fold. (1) We introduce a tailored adjustment strategy for our setting, namely, *generative recursive g-computation*. Our adjustment strategy recursively propagates full conditional outcome distributions rather than conditional means, modeling the variables of interest directly rather than full trajectories. Building on our adjustment strategy, we formulate simple generative learners for CDPO estimation. However, these learners can be sensitive to nuisance estimation errors, which motivates an orthogonal learner. (2) We thus introduce ORTHOGEN, a Neyman-orthogonal and doubly robust generative learner. Importantly, we show that ORTHOGEN further achieves rate double robustness and quasi-oracle efficiency under suitable conditions. Our learners are flexible and can be instantiated with different generative backbones (e.g., normalizing flows and diffusion models). Across experiments with synthetic, semi-synthetic and real-world datasets, we find that ORTHOGEN is highly effective. To the best of our knowledge, we are the first to propose a generative orthogonal learner for estimating CDPOs under time-varying treatments.

1 INTRODUCTION

Estimating the causal effects of time-varying treatments is important in medicine (Feuerriegel et al., 2024). For example, in cancer care, chemotherapy regimens are updated across several treatment cycles (Press et al., 2016), while, in diabetes care, insulin doses are repeatedly adjusted in response to a patient's health condition (Meneghini et al., 2007). Clinicians therefore need to understand how different treatment sequences affect the patient-level probability of future outcomes, such as tumor size, hypoglycemia, or toxicity. Here, ***we aim to learn conditional distributional potential outcomes (CDPOs) in time-varying settings***, which characterize the distribution of potential outcomes under a given treatment sequence, conditional on the observed patient history (see Fig. 1).

A key challenge in this setting is *time-varying confounding* (Robins et al., 2000; Frauen et al., 2025; Hess et al., 2026a). Patient characteristics may change over time in response to earlier treatments and, in turn, affect both subsequent treatment decisions and future outcomes. As a result, adjusting only for the patient history observed at the prediction origin is generally insufficient; rather, *adjustment strategies over time* are needed that account for the future treatment sequence. So far, the existing adjustment strategy for direct learning of CDPOs under time-varying treatments is inverse probability of treatment weighting (IPTW) (Mu et al., 2025). Thus, the available adjustment strategies for direct generative CDPO estimation in time-varying settings remain incomplete, and we therefore develop alternative adjustment strategies. Our strategies avoid full simulation of covariate trajectories and model instead the outcomes of interest directly.

Existing methods have important gaps along three dimensions. **(1)** *Static vs. time-varying treatments.* Several generative methods learn CDPOs (Yoon et al., 2018; Vanderschueren et al., 2023; Ma et al., 2024; Melnychuk & Feuerriegel, 2026; Luedtke & Fukumizu, 2025). However, these are restricted to static treatment and therefore do **not** address *time-varying* treatment sequences.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(2) *Mean outcome vs. CDPO.* Many methods for time-varying treatments learn conditional mean outcomes (e.g., Bica et al., 2020; Melnychuk et al., 2022; Frauen et al., 2025; Hess et al., 2026a), but **not** the *full* outcome distribution. **(3)** *Choice of adjustment strategy.* To the best of our knowledge, there is only one generative method for CDPOs under time-varying treatments: Mu et al. (2025) combine IPTW with an expert-guided diffusion model. However, learners for CDPOs over time based on alternative adjustment strategies, such as g-computation, or offering favorable properties such as double robustness are missing.

Here, *we aim to learn CDPOs under time-varying treatments using flexible generative models*. To address this task, our contributions are two-fold.

(1) We introduce a tailored adjustment strategy for our setting, which we call *generative recursive g-computation*. Our approach extends standard iterative g-computation (Robins, 1986) from conditional mean outcomes to *full* conditional outcome distributions. The difference is that, instead of recursively propagating conditional expectations (Hess et al., 2026a), we recursively propagate full conditional outcome distributions through a sequence of generative models. Importantly, the recursive construction allows us to directly recover the target outcome distribution at the prediction origin, without simulating complete future covariate trajectories at inference time, which is particularly efficient for longitudinal settings. Building on our adjustment strategy, we develop simple generative learners for CDPO estimation, namely, plug-in and regression-adjusted learners. While straightforward, these learners can be sensitive to nuisance estimation errors, and we thus primarily use these learners to motivate an orthogonal learner ($\rightarrow$ see (2)).

(2) We then develop ORTHOGEN, a *Neyman-orthogonal* and *doubly robust* generative learner for CDPO estimation over time. For this, we combine our generative recursive g-computation with an additional propensity-based adjustment, which makes ORTHOGEN first-order insensitive to nuisance estimation errors ($\rightarrow$ Theorem 1). ORTHOGEN is flexible and can be instantiated with different generative backbones (e.g., normalizing flows and diffusion models). We further derive two favorable theoretical properties for ORTHOGEN: (i) *quasi-oracle efficiency*, meaning that the target model can achieve the oracle learning rate even when the nuisance functions are estimated from observational data; and (ii) *rate double robustness*, meaning a slower convergence rate of one nuisance component can be compensated for by a faster convergence rate of the other, thus being robust to misspecification of either nuisance function ($\rightarrow$ Theorem 2).

In summary, in this work:[1] **(1)** We introduce generative recursive g-computation as a tailored adjustment strategy for CDPOs over time. **(2)** We propose ORTHOGEN, a doubly robust and Neyman-orthogonal generative learner for this task. We show theoretically that ORTHOGEN achieves quasi-oracle efficiency and rate double robustness. **(3)** We demonstrate the effectiveness of ORTHOGEN using different generative backbones (i.e., normalizing flows and diffusion models) across various experiments. In particular, our experiments demonstrate the benefits of ORTHOGEN over the IPTW-based learner.

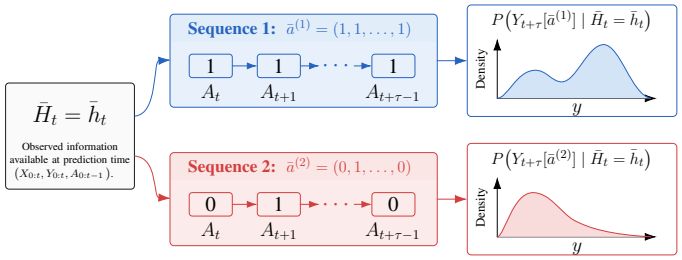


Figure 1: **Task:** Estimating the CDPOs under different treatment sequences.

2 RELATED WORK[2]

(1) Learning CDPOs in static vs. time-varying settings. Several generative methods learn CDPOs in settings with a single static treatment, using different adjustment strategies and generative backbones (Yoon et al., 2018; Vanderschueren et al., 2023; Ma et al., 2024; Luedtke & Fukumizu, 2025). This stream has recently culminated in a general class of learners for static settings called

[1] Code for review is available at `https://anonymous.4open.science/r/OrthoGen_submission-8952`. Upon acceptance, we will move the code to a public GitHub repository.

[2] We provide an extended related work in Appendix A.

GDR-learners (Melnychuk & Feuerriegel, 2026). ⇒ *Our setting is different in that we consider a sequence of treatment decisions rather than a single treatment decision, which gives rise to* ***time-varying confounding*** *(Robins et al., 2000; Frauen et al., 2025) and therefore requires tailored adjustment strategies over time.*

Table 1: Comparison of related approaches for direct learning of distributional potential outcomes over time.

Method	Adjustment	Backbone	Modeling target	Neyman-orthogonality
Wu et al. (2024)	IPTW	CVAE / DM	✗ Marginal DPO	✗
Mu et al. (2025)	IPTW	DM	✓ CDPO	✗
PI (*ours*)	G.R. g-comp.	Model agnostic	✓ CDPO	✗
RA (*ours*)	G.R. g-comp.	Model agnostic	✓ CDPO	✗
ORTHOGEN (*ours*)	DR	Model agnostic	✓ CDPO	✓

Notes: G.R. g-comp. = generative recursive g-computation; CVAE = conditional variational autoencoder; DM = diffusion model; DPO = distributional potential outcome; PI = plug-in; RA = regression-adjusted; DR = doubly robust.

(2) Time-varying treatments with mean outcome vs. CDPO. A broad literature stream develops methods for causal estimation under time-varying treatments with different adjustment strategies but primarily targets the so-called conditional average potential outcomes (CAPOs), i.e., the conditional mean of the potential outcome under a given treatment sequence (e.g., Lim, 2018; Bica et al., 2020; Melnychuk et al., 2022; Frauen et al., 2025). For CAPOs, there are some efforts to break the curse of dimensionality that is inherent to full simulation g-computation by identifying recursive procedures (Hess et al., 2026a), but these efforts are limited to CAPOs and are *not* applicable to generative approaches. ⇒ *The aforementioned methods target a different estimand: the conditional mean outcomes or treatment effects, rather than the* ***full*** *conditional distribution of potential outcomes as in our work.*

A different stream models the full longitudinal data-generating process (DGP) to enable interventional sampling by simulating complete trajectories (e.g., Li et al., 2021; Xiong et al., 2024; Deng et al., 2025; Yeom et al., 2026; Wu et al., 2026b). However, these works have a different modeling target: they aim to model and simulate the full DGP rather than directly learning the CDPO of the target outcome.

(3) Choice of adjustment strategies for CDPOs over time. Only a few methods directly learn CDPOs under time-varying treatments with explicit confounding adjustment. Wu et al. (2024) use IPTW to learn the *marginal* potential outcome distribution, which is a related but *different* estimand. Closest to our work is Mu et al. (2025), who combine IPTW with an expert-guided diffusion model to directly target CDPOs. However, IPTW can suffer from high variance and, most importantly, it is sensitive to nuisance misspecification. ⇒ *The only existing method relies on IPTW adjustment. In contrast, alternative and potentially more favorable adjustment strategies (e.g., g-computation) are missing for our task. As a result, generative learners with favorable learning properties (e.g., Neyman-orthogonality and double robustness) are also missing.*

Research gap. There are important gaps in the literature for estimating CDPOs through generative learners under time-varying treatments (see Table 1). To the best of our knowledge, ORTHOGEN is the first learner for this task that is Neyman-orthogonal and doubly robust.

3 PROBLEM SETTING

Notation. Capital letters denote random variables and lowercase letters their realizations. We use $P(Z)$ to denote the distribution of a random variable Z, whereas $P(Z = z)$ denotes its density or probability mass function evaluated at z. Expectations are denoted by $\mathbb{E}[\cdot]$. For n observations, we write $\mathbb{P}_n[f] = n^{-1} \sum_{i=1}^{n} f_i$ for the empirical average of a function f. When needed, $\mathbb{P}_0$ and $\mathbb{E}_0$ denote probability and expectation under the true observed-data distribution, and $\|f\|_{L^p(\mathbb{P}_0)} = \left(\mathbb{E}_0\left[|f|^p\right]\right)^{1/p}$ denotes the corresponding L^p norm. Estimated quantities are indicated by a hat, as in $\widehat{f}$. For time series, we write $\bar{U}_t = U_{0:t} = (U_0, \ldots, U_t)$.

Setup. We follow the standard setup for causal effect estimation with time-varying treatments (e.g., Bica et al., 2020; Melnychuk et al., 2022; Frauen et al., 2025). We observe n i.i.d. longitudinal trajectories. For patient i, the observed trajectory contains outcomes $Y_t^{(i)} \in \mathbb{R}^{d_y}$, time-varying covariates $X_t^{(i)} \in \mathbb{R}^{d_x}$, and treatments $A_t^{(i)} \in \mathcal{A}$ over time steps $t \in \{0, \ldots, T\}$. At time t, we write the observed pre-treatment history as $\bar{H}_t = \left(\bar{Y}_t, \bar{X}_t, \bar{A}_{t-1}\right)$.

Estimand. We fix a prediction origin t and a prediction horizon $\tau \geq 1$. Within the intervention window, $\delta = 0, \ldots, \tau - 1$ denotes the relative time offset, and the corresponding absolute treatment time is $t + \delta$. We aim to learn the CDPO for τ time steps ahead given the interventional treatment sequence $\bar{a}_{t:t+\tau-1} = (a_t, \ldots, a_{t+\tau-1})$. We make use of the potential outcomes framework for time-varying settings (Robins et al., 2000; Hernán & Robins, 2025). Specifically, $Y_{t+\tau}[a_{t+\delta:t+\tau-1}]$ is the potential outcome at future time step $t + \tau$ under the remaining treatment sequence from relative offset δ. Our aim is to learn the CDPO given by

$$P\left(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] \mid \bar{H}_t = \bar{h}_t\right). \tag{1}$$

Identifiability. In order to ensure identifiability, we follow the standard assumptions for time-varying treatments (Robins, 1986; Hernán & Robins, 2025; Frauen et al., 2025; Hess et al., 2026a): (i) *Recursive consistency/composition:* for every $\delta = 0, \ldots, \tau - 1$ and every treatment sequence $a_{t+\delta:t+\tau-1}$ under consideration, if $A_{t+\delta} = a_{t+\delta}$, then $Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = Y_{t+\tau}[a_{t+\delta+1:t+\tau-1}]$. In particular, if $A_{t:t+\tau-1} = a_{t:t+\tau-1}$, then $Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] = Y_{t+\tau}$. (ii) *Positivity:* for every $\bar{h}_{t+\delta}$ in the support of $\bar{H}_{t+\delta}$, $P(A_{t+\delta} = a_{t+\delta} \mid \bar{H}_{t+\delta} = \bar{h}_{t+\delta}) > 0$, $\delta = 0, \ldots, \tau - 1$. (iii) *Sequential exchangeability:* for every $\delta = 0, \ldots, \tau - 1$ and every treatment sequence $a_{t+\delta:t+\tau-1}$, $A_{t+\delta} \perp Y_{t+\tau}[a_{t+\delta:t+\tau-1}] \mid \bar{H}_{t+\delta}$.

The challenge of time-varying confounding. For a treatment sequence ($\tau > 1$), potential outcomes estimation is challenging because confounders may arise *after* the prediction origin t. A variable observed before a later treatment $A_{t+\delta}$, with $\delta > 0$, may be affected by earlier treatments A_t, while also influencing both the assignment of $A_{t+\delta}$ and the terminal outcome $Y_{t+\tau}$ (Robins et al., 2000; Frauen et al., 2025; Hess et al., 2026b). Then, adjusting only for the history $\bar{H}_t$ available at the prediction origin leaves later treatment decisions confounded. Sequential adjustment must instead account for the updated history $\bar{H}_{t+\delta}$ at each relative treatment step, even though these intermediate histories have not yet occurred when the counterfactual distribution is requested at time t.

Importantly, naıve history adjustment is biased: conditioning on $\bar{H}_t$ and selecting patients for which the observed future treatment sequence equals $\bar{a}$ generally does **not** recover the interventional distribution:

$$P\left(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] \mid \bar{H}_t = \bar{h}_t\right) \neq P\left(Y_{t+\tau} \mid \bar{H}_t = \bar{h}_t, A_{t:t+\tau-1} = a_{t:t+\tau-1}\right). \tag{2}$$

The left-hand side is our target, which does not coincide with the naïve history adjustment on the right-hand side, which is thus *biased*.

Target generative risk

To estimate the CDPO under a treatment sequence $\bar{a}$ using a generative model, we seek the best projection of the true counterfactual distribution $P(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] \mid \bar{H}_t = \bar{h}_t)$ in a predefined class of generative models $\mathcal{G}$. We denote a candidate model by $g_{\bar{a}} \in \mathcal{G}$ and define the score for an outcome y conditional on the pre-interventional history $\bar{h}_t$ as

$$\ell_{g_{\bar{a}}}(y, \bar{h}_t) = \mathbb{E}_{Z \sim \varepsilon_z(\cdot \mid y, \bar{h}_t)}\left[\log g_{\bar{a}}(y, Z \mid \bar{h}_t)\right]. \tag{3}$$

This generic notation allows us to accommodate different generative backbones, such as likelihood-based normalizing flows and diffusion models, through their respective training objectives.[3]

Accordingly, the optimal model within $\mathcal{G}$ is the one that maximizes this oracle target risk

$$\mathcal{L}_{\bar{a}}(g_{\bar{a}}) = \mathbb{E}\left[\ell_{g_{\bar{a}}}\left(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}], \bar{H}_t\right)\right]. \tag{4}$$

That is, the model in $\mathcal{G}$ that best approximates the true CDPO according to the chosen generative objective.

This oracle risk provides a target that we use for all learners below. The challenge is that it depends on the unobserved counterfactual outcome $Y_{t+\tau}[\bar{a}_{t:t+\tau-1}]$ and therefore cannot be evaluated directly from observational data.

[3]The above formulation assumes a score-maximization objective; it analogously applies to generative models formulated through loss minimization by replacing the corresponding maximization/minimization operators.

Outline. (1) Below, we first introduce *generative recursive g-computation*, which is a tailored adjustment strategy for learning CDPOs under time-varying treatments (Sec. 4). (2) Building on this adjustment strategy, we then introduce a plug-in (PI) learner and a regression-adjusted (RA) learner for CDPO estimation (Sec. 5). (3) Finally, we develop ORTHOGEN and show that it is doubly robust and Neyman-orthogonal (Sec. 6).

4 GENERATIVE RECURSIVE G-COMPUTATION

We first introduce *generative recursive g-computation*, a tailored adjustment strategy for learning CDPOs over time. It builds on g-computation (Robins, 1986), which identifies interventional outcomes by integrating over the longitudinal data-generating process under a specified treatment sequence. We extend this principle from conditional mean outcomes to the *full conditional outcome distribution* and derive a recursive generative formulation that propagates the required distributional information backward across treatment times. Rather than simulating complete future covariate trajectories, this recursion *directly* targets the distribution relevant for the final outcome.

Overview. We develop *generative recursive g-computation* in two steps. First, we define a backward recursion over conditional outcome distributions and show that it identifies the target CDPO at the prediction origin. Second, we show how this recursion can be fitted using generative models through backward pseudo-sample fitting.

***Recursive* identification of the CDPO.** Our goal is to express the target CDPO at the prediction origin through a sequence of quantities that can be identified from the observed data distribution. We proceed backward from the final treatment decision: there, the distribution of the terminal outcome is directly observed conditional on the current history and treatment. We then recursively propagate this distribution backward across earlier treatment decisions by averaging over the subsequent histories that arise under the respective treatment (see Fig. 2).

To formalize this recursion, we define $\{\xi^{\bar{a}}_{t+\delta}\}_{\delta=0}^{\tau-1}$, which we refer to as *continuation densities*. Intuitively, $\xi^{\bar{a}}_{t+\delta}$ represents the distribution of the terminal outcome that remains to be propagated backward from relative treatment step δ, incorporating the remaining treatment sequence $a_{t+\delta:t+\tau-1}$.

• *Recursion at the final treatment stage*. Let

$$\xi^{\bar{a}}_{t+\tau-1}(y \mid \bar{h}_{t+\tau-1}) = P\Big(Y_{t+\tau} = y \;\Big|\; \bar{H}_{t+\tau-1} = \bar{h}_{t+\tau-1}, A_{t+\tau-1} = a_{t+\tau-1}\Big). \tag{5}$$

The terminal density can be estimated from the observed outcomes $Y_{t+\tau}$ of patients receiving $a_{t+\tau-1}$ at time $t+\tau-1$.

• *Recursive step from* $t+\delta+1 \to t+\delta$. Starting from the final treatment time, we move backward recursively. For $\delta = 0, \ldots, \tau-2$, we define

$$\xi^{\bar{a}}_{t+\delta}(y \mid \bar{h}_{t+\delta}) = \mathbb{E}\left[\xi^{\bar{a}}_{t+\delta+1}(y \mid \bar{H}_{t+\delta+1}) \;\middle|\; \bar{H}_{t+\delta} = \bar{h}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\right]. \tag{6}$$

At each relative treatment step δ, the successor density $\xi^{\bar{a}}_{t+\delta+1}$ already incorporates the future treatment values $a_{t+\delta+1}, \ldots, a_{t+\tau-1}$. Hence, the recursion only requires conditioning on the current treatment $A_{t+\delta} = a_{t+\delta}$ and averages the successor distribution over the next history $\bar{H}_{t+\delta+1}$.

The above construction extends the recursive identification principle of longitudinal g-computation from conditional means (Hess et al., 2026a) to full conditional outcome distributions. The following result shows that each continuation density identifies the distribution of the terminal potential outcome under the remaining treatment sequence and that $\xi^{\bar{a}}_{t}$ identifies the target CDPO in Eq. 1.

Proposition 1. *Under the causal assumptions from Sec. 3, for every* $\delta = 0, \ldots, \tau-1$, *history* $\bar{h}_{t+\delta}$, *and outcome* y, $\xi^{\bar{a}}_{t+\delta}(y \mid \bar{h}_{t+\delta}) = P\left(Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = y \mid \bar{H}_{t+\delta} = \bar{h}_{t+\delta}\right)$. *In particular,* $\xi^{\bar{a}}_{t}(y \mid \bar{h}_{t}) = P\left(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] = y \mid \bar{H}_{t} = \bar{h}_{t}\right)$.

Proof. See Appendix D.2.

Fitting the continuation densities. The recursion in Eq. 6 defines the continuation densities, but does not yet provide a way to estimate them with generative models. At the final treatment time, this is straightforward because the corresponding outcomes $Y_{t+\tau}$ are observed. At earlier treatment times, however, outcomes following the remaining treatment sequence are not observed

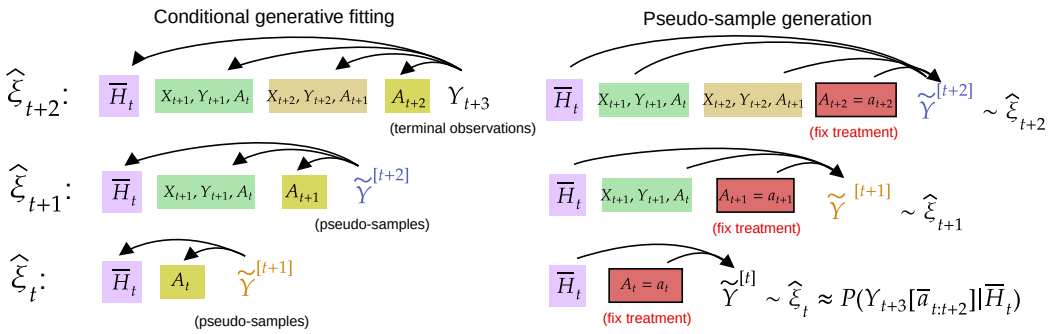


Figure 2: **Illustration of generative recursive g-computation** for $\tau = 3$. *Left:* Continuation densities are fitted through backward recursion. The terminal density is fitted from observed outcomes among patients receiving the target treatment. At each earlier treatment step δ, pseudo-outcomes are sampled from the fitted successor density and used to fit the preceding density. *Right:* Generating the pseudo-samples that will be used to fit the continuation nuisances at earlier steps.

and therefore cannot be used directly as training targets. We address this challenge recursively: after fitting $\widehat{\xi}^{\bar{a}}_{t+\delta+1}$, we draw pseudo-outcomes $\widetilde{Y}_{t+\delta+1} \sim \widehat{\xi}^{\bar{a}}_{t+\delta+1}(\cdot \mid \bar{H}_{t+\delta+1})$ and fit $\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot \mid \bar{H}_{t+\delta})$ to these pseudo-outcomes among observations with $A_{t+\delta} = a_{t+\delta}$, for $\delta = \tau-2, \ldots, 0$. This procedure yields the sequence of estimated continuation densities and, in particular, $\widehat{\xi}^{\bar{a}}_t$ as an estimate of the target CDPO. We provide pseudo-code in Appendix D.1.

5 Naïve generative learners based on recursive g-computation

We build directly on *generative recursive g-computation* and introduce two generative learners that use its estimated continuation densities in different ways: The **plug-in (PI) learner** directly uses the fitted continuation density at the prediction origin as an estimate of the CDPO, while the **regression-adjusted (RA) learner** projects this estimate onto a separate target generative model through the corresponding generative risk.

• **Plug-in (PI) learner.** Proposition 1 shows that the population continuation density at time t equals the target CDPO in Eq. 1. Therefore, the fitted continuation density obtained through generative recursive g-computation at the prediction origin can be used as an estimate of the CDPO. The plug-in (PI) learner is

$$\widehat{P}^{\bar{a}}_{\mathrm{PI}}(y \mid \bar{h}_t) = \widehat{\xi}^{\bar{a}}_t(y \mid \bar{h}_t). \tag{7}$$

• **Regression-adjusted (RA) learner.** The regression-adjusted (RA) learner uses the estimated continuation density to construct a generative learning objective over the model class $\mathcal{G}$. It follows a two-stage procedure: in Stage 1, generative recursive g-computation is used to estimate the continuation densities; in Stage 2, these estimates define a generative learning objective for fitting a model in $\mathcal{G}$. This allows the second-stage model to incorporate, for example, additional regularization, architectural constraints, or interpretability requirements.

To construct this objective, for any candidate model $g_{\bar{a}} \in \mathcal{G}$ and relative treatment step δ, we define the *continuation score*

$$m^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}) = \int_{\mathcal{Y}} \ell_{g_{\bar{a}}}(y, \bar{H}_t)\, \xi^{\bar{a}}_{t+\delta}(y \mid \bar{H}_{t+\delta})\, \mathrm{d}y, \qquad \delta = 0, \ldots, \tau-1, \tag{8}$$

which is the expected generative score of $g_{\bar{a}}$ under the continuation density $\xi^{\bar{a}}_{t+\delta}$. Further, let $\widehat{m}^{g_{\bar{a}},\bar{a}}_{t+\delta}$ denote the corresponding quantity obtained by replacing $\xi^{\bar{a}}_{t+\delta}$ with $\widehat{\xi}^{\bar{a}}_{t+\delta}$.

At the prediction origin, Proposition 1 implies that $\xi^{\bar{a}}_t$ equals the target CDPO. Hence, the oracle target risk in Eq. 4 can equivalently be written as $\mathcal{L}_{\bar{a}}(g_{\bar{a}}) = \mathbb{E}\left[m^{g_{\bar{a}},\bar{a}}_t(\bar{H}_t)\right]$. The previous identity suggests replacing the unknown continuation density with the corresponding estimate obtained via generative recursive g-computation. The RA learner is defined by the risk

$$\widehat{\mathcal{L}}^{\bar{a}}_{\mathrm{RA}}(g_{\bar{a}}) = \mathbb{P}_n\left[\widehat{m}^{g_{\bar{a}},\bar{a}}_t(\bar{H}_t)\right]. \tag{9}$$

Intuitively, the RA learner optimizes the estimated risk over $g_{\bar{a}} \in \mathcal{G}$. Thus, the key difference is that the PI learner directly returns $\widehat{\xi}_t^{\bar{a}}$, whereas the RA learner uses $\widehat{\xi}_t^{\bar{a}}$ to define the learning objective for a second-stage generative model in $\mathcal{G}$.

Thus, our PI and RA learners provide an alternative to the IPTW-based learner, which has been proposed for CDPO estimation over time (Mu et al., 2025) and which we use as another baseline (see Appendix C.1).

An important limitation of the two learners above and IPTW is that they rely on a single nuisance component, which makes the target risk sensitive to nuisance-estimation error at first order (Vansteelandt & Morzywołek, 2025). To eliminate the first-order bias, we develop a Neyman-orthogonal method called ORTHOGEN in the following section.

6 ORTHOGEN

We introduce ORTHOGEN for CDPO estimation under time-varying treatments, which, as we show later, enjoys several favorable theoretical properties (e.g., Neyman orthogonality, double robustness, quasi-oracle efficiency). First, we define the propensity score $\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) = P(A_{t+\delta} = a_{t+\delta} \mid \bar{H}_{t+\delta})$, for $\delta = 0, \ldots, \tau - 1$. We then define the one-step inverse-propensity ratio as $r_{t+\delta}^{\bar{a}} = \mathbf{1}\{A_{t+\delta} = a_{t+\delta}\}/\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})$ and the corresponding cumulative ratio as $R_{t:t+\delta}^{\bar{a}} = \prod_{\delta'=0}^{\delta} r_{t+\delta'}^{\bar{a}}$, with the empty-product convention $R_{t:t-1}^{\bar{a}} = 1$.

ORTHOGEN is defined by the risk:

$$\widehat{\mathcal{L}}_{\mathrm{OrthoGen}}^{\bar{a}}(g_{\bar{a}}) = \mathbb{P}_n \left[\widehat{R}_{t:t+\tau-1}^{\bar{a}} \ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t) + \sum_{\delta=0}^{\tau-1} \widehat{R}_{t:t+\delta-1}^{\bar{a}} \left(1 - \widehat{r}_{t+\delta}^{\bar{a}}\right) \widehat{m}_{t+\delta}^{g_{\bar{a}},\bar{a}}(\bar{H}_{t+\delta}) \right]. \tag{10}$$

Estimation. ORTHOGEN proceeds in two stages: First, we estimate the nuisance functions $\hat{\eta} = \left(\{\hat{\pi}_{t+\delta}\}_{\delta=0}^{\tau-1}, \{\hat{\xi}_{t+\delta}^{\bar{a}}\}_{\delta=0}^{\tau-1}\right)$. Second, we fit the target generative model using the objective in Eq. 10. The estimated propensity scores yield the one-step and cumulative inverse propensity ratios $\widehat{r}_{t+\delta}^{\bar{a}}$ and $\widehat{R}_{t:t+\delta}^{\bar{a}}$ defined above, while the estimated continuation densities determine the continuation scores $\widehat{m}_{t+\delta}^{g_{\bar{a}},\bar{a}}$ through Eq. 8. ORTHOGEN is model agnostic and we instantiate it with two generative models: normalizing flows and diffusion models.

6.1 THEORETICAL PROPERTIES

ORTHOGEN has several favorable asymptotic properties. First, it is Neyman-orthogonal, that is, first-order insensitive to the errors in the nuisance functions. The following theorem derives this property for ORTHOGEN:

Theorem 1 (Neyman-orthogonality)**.** *Under the causal assumptions in Sec. 3, the objective is Neyman-orthogonal at $(g_{\bar{a}}^\star, \eta)$, namely*

$$D_\eta D_{g_{\bar{a}}} \mathcal{L}_{\mathrm{OrthoGen}}^{\bar{a}}(g_{\bar{a}}^\star; \eta) \left[g_{\bar{a}} - g_{\bar{a}}^\star, \tilde{\eta} - \eta \right] = 0, \tag{11}$$

for every $g_{\bar{a}} \in \mathcal{G}$ and every admissible nuisance perturbation $\tilde{\eta}$, where $D\mathcal{L}(\cdot)[\cdot]$ denotes the pathwise derivative (see Appendix E.1 for definitions).

Proof. See Appendix E.1.

Second, ORTHOGEN is quasi-oracle efficient and doubly robust. To formalize these properties, we define the fitted second-stage target generator as $\widehat{g}_{\bar{a}} \in \mathcal{G}$ and its conditional population target-learning error as $\mathrm{Opt}_{\mathcal{G}}^{\bar{a}} := \sup_{g_{\bar{a}} \in \mathcal{G}} \mathcal{L}_{\mathrm{OrthoGen}}^{\bar{a}}(g_{\bar{a}}; \widehat{\eta}) - \mathcal{L}_{\mathrm{OrthoGen}}^{\bar{a}}(\widehat{g}_{\bar{a}}; \widehat{\eta})$. Thus, $\mathrm{Opt}_{\mathcal{G}}^{\bar{a}}$ measures how well the second-stage target generator is learned while holding the estimated nuisance functions fixed.

Theorem 2 (Quasi-oracle efficiency and double robustness)**.** *Under mild regularity conditions for the target model, the squared distance between the fitted target generator $\widehat{g}_{\bar{a}}$ and the oracle popula-*

tion maximizer $g_{\bar{a}}^{\star}$ can be upper-bounded in the following way:

$$\|\widehat{g}_{\bar{a}} - g_{\bar{a}}^{\star}\|_{\mathcal{G}}^2 \lesssim \underbrace{\mathrm{Opt}_{\mathcal{G}}^{\bar{a}}}_{\text{(I) target-model learning error}} + \underbrace{\left[\sum_{\delta=0}^{\tau-1} \|\widehat{\pi}_{t+\delta} - \pi_{t+\delta}\|_{L^4(\mathbb{P}_0)} \, \|\widehat{\xi}_{t+\delta}^{\bar{a}} - \xi_{t+\delta}^{\bar{a}}\|_{4,2}\right]^2}_{\text{(II) higher-order nuisance error}}, \tag{12}$$

where $\|\cdot\|_{\mathcal{G}}$ denotes a model-specific norm, (I) is a standard target-model learning error term, and (II) is a higher-order nuisance error term. This bound implies that ORTHOGEN *is (a)* ***doubly robust*** *and (b)* ***quasi-oracle efficient****.*

Proof. See Appendix E.2.

Theorem 2 decomposes the target-model error into a second-stage learning error and a higher-order contribution from nuisance estimation. Crucially, the latter depends on products of the stagewise propensity and continuation density errors rather than on either nuisance error alone. This product structure yields the following rate-robustness property.

Remark 1 (Rate double robustness)**.** *Let ρ_n be the learning rate of the target model, and let $\varepsilon_{\pi,t+\delta,n}$ and $\varepsilon_{\xi,t+\delta,n}$ be the convergence rates of the propensity and continuation density nuisances at relative treatment step δ, respectively. Because nuisance estimation enters through $\varepsilon_{\pi,t+\delta,n}\varepsilon_{\xi,t+\delta,n}$, faster convergence of one nuisance component can offset slower convergence of the other. In particular, if $\sum_{\delta=0}^{\tau-1} \varepsilon_{\pi,t+\delta,n}\varepsilon_{\xi,t+\delta,n} = o(\rho_n)$, then nuisance estimation is asymptotically negligible and* ORTHOGEN *attains the same first-order rate as an oracle learner with known nuisances. More generally, if the product term is $O(\rho_n)$, nuisance estimation does not worsen the learning rate for the target model; for example, two $n^{-1/4}$ nuisance rates yield an $n^{-1/2}$ product term.*

7 EVALUATION

In this section, we evaluate our learners through several numerical experiments across synthetic, semi-synthetic, and real-world datasets to assess their effectiveness.

Instantiations. We instantiate the sequence encoder using a transformer (Vaswani et al., 2017), which processes the observed pre-interventional history to produce a low-dimensional context representation fed to the nuisance and target models. For the generative models, we employ two complementary state-of-the-art approaches: **(i)** *normalizing flows* (**NFs**) (Rezende & Mohamed, 2015) and **(ii)** *diffusion models* (**DMs**) (Ho et al., 2020). See Appendix B for more details. In both cases, we use the same generative backbone for both the continuation density nuisances and the target generative model. Our transformer implementation follows an encoder-only design similar to Frauen et al. (2025) and Hess & Feuerriegel (2025). Both generative model families receive the learned context representation and generate samples conditionally on the learned history representation. Details are in Appendix G.

Table 2: **Synthetic data.**

	Method	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
NFs	PI	$\mathbf{0.308 \pm 0.004}$	0.235 ± 0.007	0.350 ± 0.015	0.249 ± 0.010	$\mathbf{0.344 \pm 0.015}$
	RA	0.320 ± 0.008	0.232 ± 0.008	0.347 ± 0.015	$\mathbf{0.247 \pm 0.008}$	0.348 ± 0.018
	IPTW	0.325 ± 0.009	0.240 ± 0.002	0.373 ± 0.015	0.305 ± 0.019	0.518 ± 0.062
	ORTHOGEN	0.312 ± 0.005	$\mathbf{0.229 \pm 0.009}$	$\mathbf{0.342 \pm 0.011}$	0.267 ± 0.015	0.401 ± 0.051
DMs	PI	$\mathbf{0.306 \pm 0.013}$	0.312 ± 0.025	0.506 ± 0.038	0.482 ± 0.015	0.612 ± 0.014
	RA	0.411 ± 0.030	0.404 ± 0.018	0.610 ± 0.032	0.537 ± 0.025	0.651 ± 0.019
	IPTW	0.341 ± 0.026	0.284 ± 0.020	0.447 ± 0.053	0.386 ± 0.048	0.522 ± 0.055
	ORTHOGEN	0.329 ± 0.014	$\mathbf{0.244 \pm 0.012}$	$\mathbf{0.378 \pm 0.017}$	$\mathbf{0.293 \pm 0.026}$	$\mathbf{0.383 \pm 0.024}$

Reported: squared 2-Wasserstein costs across treatment horizons τ.
Values are mean $\pm$ standard deviation over 5 replicates; lower is better.

Datasets. Following common practice in causal inference (Bica et al., 2020; Melnychuk et al., 2022), we use synthetic and semi-synthetic datasets for evaluation as they provide ground truths for counterfactuals, while we use a real-world dataset to assess the applicability of our learners and discuss potential insights. We thus evaluate on three datasets; see Appendix F for full details. ① A **fully synthetic benchmark** generated from a longitudinal data-generating process with treatment–confounder feedback, a scalar time-varying covariate, and a three-dimensional continuous target outcome, where the components are mixture distributions based on Gaussian and asymmetric-Laplace components and are correlated with each other. ② A **semi-synthetic benchmark** built from real patient trajectories extracted from MIMIC-III (Johnson et al., 2016). ③ The **real-world MIMIC-III** benchmark (Johnson et al., 2016) based on previous cohort construction and covariates, but where we retain the factual treatments and a one-dimensional observed outcome from the ICU records. All three datasets use 3,000 training, 1,000 validation, and 1,000 test trajectories.

Training. Our implementation follows a two-stage procedure. First, the propensity scores are estimated using a binary classifier predicting treatment assignment at each decision point from the transformer representation, while the continuation densities are learned through generative recursive g-computation. Hence, the terminal model is fitted to observed factual outcomes, and earlier continuation models are subsequently trained on pseudo-outcomes sampled from frozen successor models using Monte Carlo sampling to estimate expectations. Then, the target generator of ORTHOGEN is fitted using the objective in Eq. 10. See further details on architectures and hyperparameter selection in Appendix G.1 and G.2 and in the code. We provide runtime information in Appendix H.

Table 3: **Semi-synthetic data.**

	Method	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
NFs	PI	1.13 ± 0.39	1.46 ± 0.70	1.59 ± 0.21	1.78 ± 0.67	2.55 ± 0.58
	RA	1.22 ± 0.33	1.22 ± 0.55	1.47 ± 0.23	1.59 ± 0.44	2.22 ± 0.52
	IPTW	0.78 ± 0.11	0.67 ± 0.20	1.45 ± 0.57	0.92 ± 0.22	2.06 ± 0.87
	ORTHOGEN	$\mathbf{0.49 \pm 0.09}$	$\mathbf{0.47 \pm 0.13}$	$\mathbf{0.91 \pm 0.13}$	$\mathbf{0.73 \pm 0.22}$	$\mathbf{1.23 \pm 0.45}$
DMs	PI	0.56 ± 0.06	0.59 ± 0.13	0.81 ± 0.32	0.93 ± 0.36	0.93 ± 0.36
	RA	0.63 ± 0.08	0.74 ± 0.20	0.99 ± 0.38	1.09 ± 0.39	1.09 ± 0.41
	IPTW	0.47 ± 0.10	0.62 ± 0.27	1.43 ± 0.68	1.77 ± 0.79	2.49 ± 0.50
	ORTHOGEN	$\mathbf{0.45 \pm 0.05}$	$\mathbf{0.50 \pm 0.10}$	$\mathbf{0.74 \pm 0.21}$	$\mathbf{0.88 \pm 0.30}$	$\mathbf{0.77 \pm 0.18}$

Reported: squared 2-Wasserstein costs $\times 10^{-2}$ across treatment horizons τ.
Values are mean $\pm$ standard deviation over 5 replicates; lower is better.

Baselines. To the best of our knowledge, the **IPTW** from Mu et al. (2025) is the **only** applicable baseline for direct learning of CDPOs. Other methods solve *different* estimands (e.g., CAPO but *not* CDPOs, so that the full distributional information is lost). We thus additionally benchmark ORTHOGEN against two naïve learners: **PI** and **RA** from Sec. 5. As a result, we can isolate the benefit of the the doubly robust and Neyman-orthogonal adjustment approach in ORTHOGEN. Importantly, all baselines are instantiated with the *same* transformer architecture and generative backbones to ensure a *fair* comparison.

Metrics. For the (1) fully synthetic and (2) semi-synthetic datasets, we can obtain ground-truth interventional outcome samples from the simulation: for each test sample, we generate samples from the fitted model and matched oracle samples under the target regime. We compare the resulting predictive distributions using the squared 2-Wasserstein cost, which measures discrepancy between empirical outcome distributions. (3) For the real-world MIMIC-III dataset, no oracle counterfactual distribution is available. We thus assess factual predictive accuracy using the continuous ranked probability score (CRPS), which is a proper scoring rule for one-dimensional predictive distributions that requires only the observed factual outcome. Predictions are made under the target treatment sequence, and CRPS is computed only for held-out patients whose observed treatment sequence exactly matches the target. See Appendix F for details.

Results. Overall, our results demonstrate the theoretical benefits of ORTHOGEN across different datasets and generative backbones. (1) On the *fully synthetic* benchmark (Table 2), ORTHOGEN performs particularly strongly with DMs and remains highly competitive with NFs. Importantly, for $\tau > 1$, ORTHOGEN consistently outperforms the PI learner, with its advantage becoming increasingly pronounced at longer treatment horizons. This pattern is consistent with the robustness benefits predicted by our theoretical analysis: as the treatment horizon grows and nuisance estimation becomes more challenging, the doubly robust and Neyman-orthogonal adjustment of ORTHOGEN yields increasingly substantial gains over the alternative learners. (2) On the *semi-synthetic* benchmark (Table 3), ORTHOGEN consistently achieves the lowest Wasserstein cost across **all** treatment horizons and for **both** NFs and DMs, and the gains again become larger for longer horizons, which confirms the theoretical properties of ORTHOGEN due to Neyman-orthogonality. (3) Finally, on the *real-world* benchmark (Table 4), ORTHOGEN achieves the lowest factual CRPS across all horizons with NFs. Again, ORTHOGEN is effective across different backbones. $\Rightarrow$ *Taken together, these results demonstrate the benefits of the doubly robust and Neyman-orthogonal construction of* ORTHOGEN*, which yields substantial improvements over the alternative learners, particularly at longer treatment horizons.*

Table 4: **Real-world data.**

	Method	$\tau = 1$	$\tau = 3$	$\tau = 5$
NFs	PI	0.338 ± 0.005	0.417 ± 0.006	0.459 ± 0.014
	RA	0.329 ± 0.004	0.405 ± 0.007	0.453 ± 0.015
	IPTW	0.319 ± 0.006	0.410 ± 0.004	0.457 ± 0.012
	ORTHOGEN	$\mathbf{0.307 \pm 0.007}$	$\mathbf{0.387 \pm 0.007}$	$\mathbf{0.442 \pm 0.013}$
DMs	PI	$\mathbf{0.314 \pm 0.008}$	0.428 ± 0.005	0.500 ± 0.007
	RA	0.330 ± 0.007	0.446 ± 0.005	0.506 ± 0.005
	IPTW	0.347 ± 0.020	$\mathbf{0.426 \pm 0.019}$	$\mathbf{0.472 \pm 0.011}$
	ORTHOGEN	0.330 ± 0.017	0.432 ± 0.018	0.475 ± 0.020

Reported are factual matching CRPS values across treatment horizons τ.
Values are mean $\pm$ standard deviation over 5 replicates; lower is better.

Conclusion. We introduced ORTHOGEN, the first doubly robust and Neyman-orthogonal generative learner for CDPO estimation under time-varying treatments, built on our generative recursive g-computation strategy. We showed that ORTHOGEN achieves rate double robustness and quasi-oracle efficiency, and performs effectively across synthetic, semi-synthetic, and real-world benchmarks with different generative backbones.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript editing and presentation, literature search, code development and debugging, and discussions concerning methodological and theoretical aspects of the work. All AI-assisted content, including mathematical arguments, code, experimental design, and scientific claims, was independently reviewed and verified by the authors. The authors take full responsibility for the final content of the paper.

REPRODUCIBILITY STATEMENT

We provide the implementation of all proposed methods and experimental pipelines in an anonymized code repository. The appendix contains the complete theoretical derivations and proofs, additional details on the datasets and preprocessing, model architectures, training procedures, hyperparameters, and evaluation protocols used in our experiments. Together, these resources are intended to enable reproduction of theoretical and empirical results reported in the paper.

REFERENCES

Ioana Bica, Ahmed M Alaa, James Jordon, and Mihaela van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *International Conference on Learning Representations*, 2020.

Leon Deng, Hong Xiong, Feng Wu, Sanyam Kapoor, Soumya Gosh, Zach Shahn, and Li-wei Lehman. Uncertainty quantification for conditional treatment effect estimation under dynamic treatment regimes. In Stefan Hegselmann, Helen Zhou, Elizabeth Healey, Trenton Chang, Caleb Ellington, Vishwali Mhasawade, Sana Tonekaboni, Peniel Argaw, and Haoran Zhang (eds.), *Proceedings of the 4th Machine Learning for Health Symposium*, volume 259 of *Proceedings of Machine Learning Research*, pp. 248–266. PMLR, 15–16 Dec 2025.

Stefan Feuerriegel, Dennis Frauen, Valentyn Melnychuk, Jonas Schweisthal, Konstantin Hess, Alicia Curth, Stefan Bauer, Niki Kilbertus, Isaac S Kohane, and Mihaela van der Schaar. Causal machine learning for predicting treatment outcomes. *Nature Medicine*, 30(4):958–968, 2024.

Dennis Frauen, Konstantin Hess, and Stefan Feuerriegel. Model-agnostic meta-learners for estimating heterogeneous treatment effects over time. In *The Thirteenth International Conference on Learning Representations*, 2025.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, 2025.

Konstantin Hess and Stefan Feuerriegel. Stabilized neural prediction of potential outcomes in continuous time. In *The Thirteenth International Conference on Learning Representations*, 2025.

Konstantin Hess, Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. IGC-net for conditional average potential outcome estimation over time. In *The Fourteenth International Conference on Learning Representations*, 2026a.

Konstantin Hess, Dennis Frauen, Mihaela van der Schaar, and Stefan Feuerriegel. Overlap-weighted orthogonal meta-learner for treatment effect estimation over time. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=0Xi3WDwd5w.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):160035, 2016.

Edward H Kennedy, Sivaraman Balakrishnan, and LA Wasserman. Semiparametric counterfactual density estimation. *Biometrika*, 110(4):875–896, 2023.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

Rui Li, Stephanie Hu, Mingyu Lu, Yuria Utsumi, Prithwish Chakraborty, Daby M Sow, Piyush Madan, Jun Li, Mohamed Ghalwash, Zach Shahn, et al. G-net: a recurrent network approach to g-computation for counterfactual prediction under a dynamic treatment regime. In *Machine Learning for Health*, pp. 282–299. PMLR, 2021.

Bryan Lim. Forecasting treatment responses over time using recurrent marginal structural networks. *Advances in neural information processing systems*, 31, 2018.

Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Alex Luedtke and Kenji Fukumizu. Doublegen: Debiased generative modeling of counterfactuals. *arXiv preprint arXiv:2509.16842*, 2025.

Yuchen Ma, Valentyn Melnychuk, Jonas Schweisthal, and Stefan Feuerriegel. Diffpo: A causal diffusion model for learning distributions of potential outcomes. *Advances in Neural Information Processing Systems*, 37:43663–43692, 2024.

Valentyn Melnychuk and Stefan Feuerriegel. Gdr-learners: Orthogonal learning of generative models for potential outcomes. In *International Conference on Learning Representations*, 2026.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In *International conference on machine learning*, pp. 15293–15329. PMLR, 2022.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Normalizing flows for interventional density estimation. In *International Conference on Machine Learning*, pp. 24361–24397. PMLR, 2023.

Luigi Meneghini, Claudia Koenen, Wen Weng, and Jean-Louis Selam. The usage of a simplified self-titration dosing guideline (303 algorithm) for insulin detemir in patients with type 2 diabetes: results of the randomized, controlled predictive 303 study. *Diabetes, Obesity and Metabolism*, 9 (6):902–913, 2007. doi: 10.1111/j.1463-1326.2007.00804.x.

Wenhao Mu, Zhi Cao, Mehmed Uludag, and Alexander Rodríguez. Counterfactual probabilistic diffusion with expert models. *arXiv preprint arXiv:2508.13355*, 2025.

George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

Alvaro Parafita and Jordi Vitria. Estimand-agnostic causal query estimation with deep causal graphs. *IEEE Access*, 10:71370–71386, 2022.

Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 857–869. Curran Associates, Inc., 2020.

Oliver W. Press, Hongli Li, Heiko Schöder, David J. Straus, Craig H. Moskowitz, et al. Us intergroup trial of response-adapted therapy for stage iii to iv hodgkin lymphoma using early interim fluorodeoxyglucose-positron emission tomography imaging: Southwest oncology group s0816. *Journal of Clinical Oncology*, 34(17):2020–2027, 2016. doi: 10.1200/JCO.2015.63.1119.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

James M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7 (9-12):1393–1512, 1986.

James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

Pedro Sanchez and Sotirios A. Tsaftaris. Diffusion causal models for counterfactual estimation. In *First Conference on Causal Learning and Reasoning*, 2022. URL https://openreview.net/forum?id=LAAZLZIMN-o.

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

Sarah L Taubman, James M Robins, Murray A Mittleman, and Miguel A Hernán. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International journal of epidemiology*, 38(6):1599–1611, 2009.

Toon Vanderschueren, Jeroen Berrevoets, and Wouter Verbeke. NOFLITE: Learning to predict individual treatment effect distributions. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856.

Stijn Vansteelandt and Paweł Morzywołek. Orthogonal prediction of counterfactual outcomes. *Journal of Causal Inference*, 13(1):20240051, 2025.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Shirly Wang, Matthew BA McDermott, Geeticka Chauhan, Marzyeh Ghassemi, Michael C Hughes, and Tristan Naumann. Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. In *Proceedings of the ACM conference on health, inference, and learning*, pp. 222–235, 2020.

Dongze Wu, David I. Inouye, and Yao Xie. Flow-based generative modeling of potential outcomes and counterfactuals, 2026a. URL https://arxiv.org/abs/2505.16051.

Dongze Wu, Feng Qiu, and Yao Xie. Doflow: Flow-based generative models for interventional and counterfactual forecasting on time series. In *The Fourteenth International Conference on Learning Representations*, 2026b.

Shenghao Wu, Wenbin Zhou, Minshuo Chen, and Shixiang Zhu. Counterfactual generative models for time-varying treatments. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3402–3413, 2024.

Hong Xiong, Feng Wu, Leon Deng, Megan Su, and Li-wei H. Lehman. G-transformer: Counterfactual outcome prediction under dynamic and time-varying treatment regimes. In Kaivalya Deshpande, Madalina Fiterau, Shalmali Joshi, Zachary Lipton, Rajesh Ranganath, and Iñigo Urteaga (eds.), *Proceedings of the 9th Machine Learning for Healthcare Conference*, volume 252 of *Proceedings of Machine Learning Research*. PMLR, 16–17 Aug 2024.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM computing surveys*, 56(4):1–39, 2023.

Yoonseok Yeom, Jonghwan Kim, Taehui Yun, Juhyun Lyu, Jung-Hee Kim, Sangmin Lee, Jinseok Yang, Hyemin Jung, Woohyung Lim, and Sanghack Lee. Estimating interventional outcomes over time with causal normalizing flow. In *Forty-Second Annual Conference on Uncertainty in Artificial Intelligence*, 2026.

Jinsung Yoon, James Jordon, and Mihaela Van Der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. In *International conference on learning representations*, 2018.

APPENDIX CONTENTS

A Extended Related Work **14**

B Generative Model Instantiations **14**

B.1 Conditional normalizing flows 14

B.2 Conditional diffusion models 15

C Additional Learner Details and Risk Representations **16**

C.1 IPTW learner 16

C.2 Oracle-risk representation of the RA learner 16

C.3 Oracle-risk representation of the IPTW learner 17

D Generative Recursive G-Computation **18**

D.1 Pseudocode for generative recursive g-computation 18

D.2 Proof of Proposition 1 19

E Theoretical Properties of ORTHOGEN **20**

E.1 Neyman orthogonality 20

E.2 Double robustness and quasi-oracle efficiency 23

F Dataset Details **27**

F.1 Fully synthetic dataset 27

F.2 Semi-synthetic MIMIC-III dataset 28

F.3 Real-world MIMIC-III dataset 29

G Implementation Details **30**

G.1 Transformer history encoder 30

G.2 Generative models 30

H Computational runtime **33**

A Extended Related Work

Learning CDPOs in static settings. Most causal machine-learning methods for static treatments have traditionally targeted conditional average potential outcomes (CAPOs), or contrasts such as conditional average treatment effects (CATEs), using methods such as S-, T-, and X-learners (Künzel et al., 2019) or representation-learning approaches (Shalit et al., 2017). More recent work moves beyond these mean-based targets to learn potential-outcome distributions and thereby characterize treatment effects beyond the mean. This literature considers both marginal potential-outcome distributions and covariate-conditional potential-outcome distributions (CDPOs). At the marginal level, Kennedy et al. (2023) develop semiparametric one-step, doubly robust estimators for projections of counterfactual densities. Building on this framework, Melnychuk et al. (2023) introduce Interventional Normalizing Flows, which instantiate the corresponding bias correction using normalizing flows. For CDPOs, GANITE uses generative adversarial networks with regression adjustment (Yoon et al., 2018), while NOFLITE (Vanderschueren et al., 2023) and PO-Flow (Wu et al., 2026a) are flow-based plug-in learners. DiffPO instead combines conditional diffusion models with inverse probability of treatment weighting (IPTW) and is Neyman-orthogonal only under a realizability condition (Ma et al., 2024). A neighboring line uses plug-in learning to model the entire causal data-generating process with different generative models (Louizos et al., 2017; Pawlowski et al., 2020; Parafita & Vitria, 2022; Sanchez & Tsaftaris, 2022). Most recently, Melnychuk & Feuerriegel (2026) and Luedtke & Fukumizu (2025) introduced Neyman-orthogonal and doubly robust learners for CDPOs in the static setting.

Learning CAPOs in time-varying settings. Classical approaches for CAPO estimation under time-varying treatments include marginal structural models estimated through IPTW (Robins et al., 2000) and the parametric g-formula, or g-computation (Taubman et al., 2009). More recent methods adapt these strategies to complex longitudinal data with deep learning. Lim (2018) combines recurrent models with propensity weighting. Li et al. (2021) implement simulation-based g-computation with recurrent networks, while G-Transformer (Xiong et al., 2024) uses Transformer-based sequence models. IGC-Net instead performs iterative regression-based g-computation with a Transformer architecture (Hess et al., 2026a). Counterfactual Recurrent Networks (Bica et al., 2020) and Causal Transformer (Melnychuk et al., 2022) learn balanced representations of patient histories. Moving beyond architecture-specific estimators, Frauen et al. (2025) introduce model-agnostic meta-learners based on history adjustment, regression adjustment, propensity weighting, doubly robust correction, and inverse-variance weighting, while Hess et al. (2026b) extend the previous work by introducing an orthogonal overlap-weighted estimator that addresses the diminishing overlap associated with long treatment sequences.

B Generative Model Instantiations

We consider two conditional generative-model families as instantiations of the framework: normalizing flows and diffusion models. Following the score-based formulation of Melnychuk & Feuerriegel (2026), we describe both the probabilistic model and the population distributional criterion induced by its score. Throughout this section, C denotes the conditioning variables: $C = \bar{H}_{t+\delta}$ for a continuation model at relative step δ, and $C = \bar{H}_t$ for the target generator. We write $P_{\bar{a}}(\cdot \mid \bar{h}_t) := P(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] \in \cdot \mid \bar{H}_t = \bar{h}_t)$ for the target CDPO.

B.1 Conditional normalizing flows

Probabilistic model. A conditional normalizing flow maps a base variable $U \sim p_0$ through an invertible differentiable transformation $f_\theta(\cdot; c)$ (Papamakarios et al., 2021). For $Y = f_\theta(U; c)$, the change-of-variables formula yields

$$p_\theta(y \mid c) = p_0\big(f_\theta^{-1}(y; c)\big) \left| \det \left(\frac{\partial f_\theta^{-1}(y; c)}{\partial y} \right) \right|. \tag{13}$$

Thus, both sampling and density evaluation are available without integrating over U. In the notation of Eq. 3, the auxiliary space is degenerate, $g_{\bar{a}}(y, z \mid c) = p_\theta(y \mid c)$, and $\ell_{g_{\bar{a}}}(y, c) = \log p_\theta(y \mid c)$.

Distributional interpretation. Let $\mathcal{P}_{\mathrm{NF}}$ be the chosen class of conditional flow densities. If the target CDPO has density $p_{\bar{a}}$ and finite conditional entropy $\mathbb{H}(P_{\bar{a}}) = -\mathbb{E}[\log p_{\bar{a}}(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] \mid$

$\bar{H}_t)]$, then the population flow criterion satisfies

$$\begin{aligned}\mathcal{L}_{\bar{a}}^{\mathrm{NF}}(p_\theta) &:= \mathbb{E}\big[\log p_\theta(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] \mid \bar{H}_t)\big] \\ &= -\mathbb{H}(P_{\bar{a}}) - \mathbb{E}_{\bar{H}_t}\big[\mathrm{KL}\big(P_{\bar{a}}(\cdot \mid \bar{H}_t) \,\|\, p_\theta(\cdot \mid \bar{H}_t)\big)\big]. \end{aligned} \tag{14}$$

The entropy term does not depend on p_θ. Consequently,

$$\arg\max_{p_\theta \in \mathcal{P}_{\mathrm{NF}}} \mathcal{L}_{\bar{a}}^{\mathrm{NF}}(p_\theta) = \arg\min_{p_\theta \in \mathcal{P}_{\mathrm{NF}}} \mathbb{E}_{\bar{H}_t}\big[\mathrm{KL}\big(P_{\bar{a}}(\cdot \mid \bar{H}_t) \,\|\, p_\theta(\cdot \mid \bar{H}_t)\big)\big]. \tag{15}$$

Conditional maximum likelihood therefore returns the forward-KL projection of the CDPO onto the flow class.

B.2 Conditional diffusion models

Probabilistic model. Diffusion models specify a fixed forward noising process and learn its conditional reversal (Yang et al., 2023). Let $j = 1, \ldots, S$ index diffusion steps, separately from longitudinal time, and set $Z_0 = Y$. For a schedule $\alpha_j \in (0, 1)$ and $\bar{\alpha}_j = \prod_{r=1}^{j} \alpha_r$, the forward chain is

$$\begin{aligned} q(z_{1:S} \mid y) &= \prod_{j=1}^{S} q_j(z_j \mid z_{j-1}), \\ q_j(z_j \mid z_{j-1}) &= \mathcal{N}\big(z_j; \sqrt{\alpha_j} z_{j-1}, (1-\alpha_j) I\big). \end{aligned} \tag{16}$$

Its jth marginal can be sampled directly as

$$Z_j = \sqrt{\bar{\alpha}_j}\, Y + \sqrt{1-\bar{\alpha}_j}\, \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I). \tag{17}$$

For a schedule ending close to pure noise, Z_S is approximately standard Gaussian. The learned reverse model has joint density

$$p_\theta(y, z_{1:S} \mid c) = p_S(z_S) \prod_{j=1}^{S} p_\theta(z_{j-1} \mid z_j, c), \qquad z_0 = y, \tag{18}$$

where p_S is usually standard Gaussian. Samples are generated from p_S and passed through the reverse transitions from S down to 1.

Variational score and distributional interpretation. The marginal likelihood of Y requires integration over the noising path. To obtain a tractable score, set

$$g_{\bar{a},\theta}(y, z_{1:S} \mid c) = \frac{p_\theta(y, z_{1:S} \mid c)}{q(z_{1:S} \mid y)}, \qquad \varepsilon_z(\,\mathrm{d}z_{1:S} \mid y, c) = q(z_{1:S} \mid y)\,\mathrm{d}z_{1:S}. \tag{19}$$

Equation 3 then becomes the conditional evidence lower bound (ELBO):

$$\begin{aligned} \ell_{g_{\bar{a},\theta}}(y, c) &= \mathbb{E}_{Z_{1:S} \sim q(\cdot \mid y)}\left[\log \frac{p_\theta(y, Z_{1:S} \mid c)}{q(Z_{1:S} \mid y)}\right] \\ &= \log p_\theta(y \mid c) - \mathrm{KL}(q(Z_{1:S} \mid y) \,\|\, p_\theta(Z_{1:S} \mid y, c)) \le \log p_\theta(y \mid c). \end{aligned} \tag{20}$$

Define the average posterior (or inference gap) by

$$\mathrm{IG}_{\bar{a}}(\theta) := \mathbb{E}_{\bar{H}_t} \mathbb{E}_{Y \sim P_{\bar{a}}(\cdot \mid \bar{H}_t)}\big[\mathrm{KL}\big(q(Z_{1:S} \mid Y) \,\|\, p_\theta(Z_{1:S} \mid Y, \bar{H}_t)\big)\big]. \tag{21}$$

Averaging Eq. 20 under the target CDPO gives

$$\begin{aligned} \mathcal{L}_{\bar{a}}^{\mathrm{DM}}(\theta) &:= \mathbb{E}\big[\ell_{g_{\bar{a},\theta}}(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}], \bar{H}_t)\big] \\ &= -\mathbb{H}(P_{\bar{a}}) - \mathbb{E}_{\bar{H}_t}\big[\mathrm{KL}\big(P_{\bar{a}}(\cdot \mid \bar{H}_t) \,\|\, p_\theta(\cdot \mid \bar{H}_t)\big)\big] - \mathrm{IG}_{\bar{a}}(\theta). \end{aligned} \tag{22}$$

Hence the population diffusion target obeys

$$\arg\max_{\theta \in \Theta_{\mathrm{DM}}} \mathcal{L}_{\bar{a}}^{\mathrm{DM}}(\theta) = \arg\min_{\theta \in \Theta_{\mathrm{DM}}} \left\{ \mathbb{E}_{\bar{H}_t}\big[\mathrm{KL}\big(P_{\bar{a}}(\cdot \mid \bar{H}_t) \,\|\, p_\theta(\cdot \mid \bar{H}_t)\big)\big] + \mathrm{IG}_{\bar{a}}(\theta) \right\}. \tag{23}$$

Unlike the flow criterion, the diffusion ELBO contains the additional posterior gap. It vanishes when the fixed forward law agrees with the model posterior; more generally, it records the price of using q as a variational approximation.

For Gaussian reverse transitions with fixed variances, the terms involving their means can be written as weighted noise-prediction errors. A common practical objective samples J uniformly from $\{1, \dots, S\}$ and uses

$$\mathcal{J}_{\mathrm{DM}}(\theta) = \mathbb{E}_{Y,C,J,\varepsilon}\left[\left\|\varepsilon - \varepsilon_\theta\big(\sqrt{\bar{\alpha}_J}Y + \sqrt{1-\bar{\alpha}_J}\varepsilon, J, C\big)\right\|_2^2\right]. \tag{24}$$

The exactly weighted objective, together with its endpoint terms, is the negative ELBO in Eq. 20; the uniform-weight version in Eq. 24 is the usual simplified surrogate. In either case, the per-observation negative denoising loss can be used as the model-specific score in Eq. 3, with the sampled diffusion step and Gaussian noise serving as the auxiliary variables.

For generative recursive g-computation, these model-specific criteria are applied at each relative step δ with $C = \bar{H}_{t+\delta}$ and with either the observed terminal outcome or the generated pseudo-outcome as the response. For the target generator, $C = \bar{H}_t$; the RA, IPTW, and ORTHOGEN risks replace the inaccessible expectation under $P_{\bar{a}}$ by the observational representations developed in the main text.

C Additional Learner Details and Risk Representations

C.1 IPTW learner

IPTW learner. An alternative representation of the oracle risk is obtained through inverse probability of treatment weighting (IPTW). We define the time-specific propensity score $\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) = P(A_{t+\delta} = a_{t+\delta} \mid \bar{H}_{t+\delta})$, for $\delta = 0, \dots, \tau - 1$, and the one-step and cumulative inverse-propensity ratios, respectively, as:

$$r^{\bar{a}}_{t+\delta} = \frac{\mathbf{1}\{A_{t+\delta} = a_{t+\delta}\}}{\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}, \qquad R^{\bar{a}}_{t:t+\delta} = \prod_{\delta'=0}^{\delta} r^{\bar{a}}_{t+\delta'}, \qquad \delta = 0, \dots, \tau-1, \qquad R^{\bar{a}}_{t:t-1} = 1. \tag{25}$$

Under the identification assumptions of Sec. 3, the oracle target risk in Eq. 4 can be written as $\mathcal{L}_{\bar{a}}(g_{\bar{a}}) = \mathbb{E}\left[R^{\bar{a}}_{t:t+\tau-1}\ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t)\right]$. Let $\widehat{\pi}$, $\widehat{r}$, and $\widehat{R}$ denote the corresponding estimated quantities. The empirical IPTW risk is

$$\widehat{\mathcal{L}}^{\bar{a}}_{\mathrm{IPTW}}(g_{\bar{a}}) = \mathbb{P}_n\left[\widehat{R}^{\bar{a}}_{t:t+\tau-1}\ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t)\right]. \tag{26}$$

Unlike the RA learner, the IPTW learner does not rely on the continuation nuisances and instead requires consistent estimation of the treatment mechanism.

C.2 Oracle-risk representation of the RA learner

Lemma 1 (Population equivalence of the RA risk)**.** *Suppose that the standard causal assumptions stated in Sec. 3 hold and that the relevant quantities are integrable. Let the continuation densities $\{\xi^{\bar{a}}_{t+\delta}\}_{\delta=0}^{\tau-1}$ be defined as in Sec. 4, and let $m^{g_{\bar{a}},\bar{a}}_{t+\delta}$ be the corresponding population score responses defined in Eq. 8. Then, for every $g_{\bar{a}} \in \mathcal{G}$,*

$$\mathcal{L}_{\bar{a}}(g_{\bar{a}}) = \mathbb{E}\left[m^{g_{\bar{a}},\bar{a}}_t(\bar{H}_t)\right]. \tag{27}$$

Hence, the population risk underlying the RA learner is identical to the oracle target generative risk in Eq. 4.

Proof. Starting from Eq. 4,

$$\mathcal{L}_{\bar{a}}(g_{\bar{a}}) = \mathbb{E}\big[\ell_{g_{\bar{a}}}(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}], \bar{H}_t)\big] \tag{28}$$

$$= \mathbb{E}\big[\mathbb{E}\big[\ell_{g_{\bar{a}}}(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}], \bar{H}_t) \mid \bar{H}_t\big]\big] \tag{29}$$

$$= \mathbb{E}\left[\int_{\mathcal{Y}} \ell_{g_{\bar{a}}}(y, \bar{H}_t) P(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] = y \mid \bar{H}_t)\, dy\right] \tag{30}$$

$$= \mathbb{E}\left[\int_{\mathcal{Y}} \ell_{g_{\bar{a}}}(y, \bar{H}_t) \xi_t^{\bar{a}}(y \mid \bar{H}_t)\, dy\right] \tag{31}$$

$$= \mathbb{E}\big[m_t^{g_{\bar{a}},\bar{a}}(\bar{H}_t)\big]\,. \tag{32}$$

The second equality is the tower property, the third expresses the conditional expectation through the conditional law of $Y_{t+\tau}[\bar{a}_{t:t+\tau-1}]$, the fourth invokes Proposition 1, and the last is Eq. 8. □

C.3 Oracle-risk representation of the IPTW learner

Lemma 2 (Population equivalence of the IPTW risk)**.** *Suppose that the standard causal assumptions stated in Sec. 3 hold and that the relevant quantities are integrable. Let the cumulative inverse-propensity ratios $R^{\bar{a}}_{t:t+\delta}$ be defined as in Sec. 6. Then, for every $g_{\bar{a}} \in \mathcal{G}$,*

$$\mathcal{L}_{\bar{a}}(g_{\bar{a}}) = \mathbb{E}\left[R^{\bar{a}}_{t:t+\tau-1} \ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t)\right]. \tag{33}$$

Hence, the population risk underlying the IPTW learner is identical to the oracle target generative risk in Eq. 4.

Proof. The stated integrability condition ensures that all conditional expectations below are well defined. For every relevant integrable random variable Z, positivity and the definition of the propensity score give, almost surely,

$$\mathbb{E}\big[r^{\bar{a}}_{t+\delta} Z \mid \bar{H}_{t+\delta}\big] = \frac{1}{\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})} \mathbb{E}\big[\mathbf{1}\{A_{t+\delta} = a_{t+\delta}\} Z \mid \bar{H}_{t+\delta}\big] \tag{34}$$

$$= \frac{P(A_{t+\delta} = a_{t+\delta} \mid \bar{H}_{t+\delta})}{\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})} \mathbb{E}\big[Z \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big] \tag{35}$$

$$= \mathbb{E}\big[Z \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big]\,. \tag{36}$$

Thus Eq. 36 is an inverse-propensity weighting identity; positivity justifies every division by $\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})$.

Define

$$Q_{t+\tau}(\bar{H}_{t+\tau}) := \ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t), \tag{37}$$

$$Q_{t+\delta}(\bar{H}_{t+\delta}) := \mathbb{E}\big[Q_{t+\delta+1}(\bar{H}_{t+\delta+1}) \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big]\,, \qquad \delta = \tau-1, \dots, 0. \tag{38}$$

Because $\bar{H}_{t+\delta}$ is the pre-interventional history, it contains all treatments through time $t+\delta-1$; hence $R^{\bar{a}}_{t:t+\delta-1}$ is $\bar{H}_{t+\delta}$-measurable. Consequently, for each $\delta = 0, \dots, \tau-1$,

$$\mathbb{E}\big[R^{\bar{a}}_{t:t+\delta} Q_{t+\delta+1}(\bar{H}_{t+\delta+1})\big] \tag{39}$$

$$= \mathbb{E}\big[R^{\bar{a}}_{t:t+\delta-1} \mathbb{E}\big[r^{\bar{a}}_{t+\delta} Q_{t+\delta+1}(\bar{H}_{t+\delta+1}) \mid \bar{H}_{t+\delta}\big]\big] \tag{40}$$

$$= \mathbb{E}\big[R^{\bar{a}}_{t:t+\delta-1} \mathbb{E}\big[Q_{t+\delta+1}(\bar{H}_{t+\delta+1}) \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big]\big] \tag{41}$$

$$= \mathbb{E}\big[R^{\bar{a}}_{t:t+\delta-1} Q_{t+\delta}(\bar{H}_{t+\delta})\big]\,. \tag{42}$$

The first equality uses the tower property and the measurability of the previous-stage cumulative weight, the second uses Eq. 36, and the third uses the definition of $Q_{t+\delta}$. Peeling Eq. 42 backward from $\delta = \tau-1$ through $\delta = 0$, and using the empty-product convention $R^{\bar{a}}_{t:t-1} = 1$, yields

$$\mathbb{E}\big[R^{\bar{a}}_{t:t+\tau-1} \ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t)\big] = \mathbb{E}\big[Q_t(\bar{H}_t)\big]\,. \tag{43}$$

It remains to give the causal interpretation of Q_t. We show by backward induction that, for $\delta = 0, \dots, \tau-1$,

$$Q_{t+\delta}(\bar{H}_{t+\delta}) = \mathbb{E}\big[\ell_{g_{\bar{a}}}\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}], \bar{H}_t\big) \mid \bar{H}_{t+\delta}\big]\,. \tag{44}$$

At $\delta = \tau - 1$, the recursion in Eq. 38 gives

$$Q_{t+\tau-1}(\bar{H}_{t+\tau-1}) = \mathbb{E}\big[\ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t) \mid \bar{H}_{t+\tau-1}, A_{t+\tau-1} = a_{t+\tau-1}\big] \tag{45}$$
$$= \mathbb{E}\big[\ell_{g_{\bar{a}}}(Y_{t+\tau}[a_{t+\tau-1}], \bar{H}_t) \mid \bar{H}_{t+\tau-1}, A_{t+\tau-1} = a_{t+\tau-1}\big] \tag{46}$$
$$= \mathbb{E}\big[\ell_{g_{\bar{a}}}(Y_{t+\tau}[a_{t+\tau-1}], \bar{H}_t) \mid \bar{H}_{t+\tau-1}\big]. \tag{47}$$

The second equality is terminal recursive consistency on $A_{t+\tau-1} = a_{t+\tau-1}$, and the third is sequential exchangeability at time $t + \tau - 1$.

Now suppose Eq. 44 holds at $t + \delta + 1$, for some $\delta < \tau - 1$. Then

$$Q_{t+\delta}(\bar{H}_{t+\delta}) = \mathbb{E}\big[Q_{t+\delta+1}(\bar{H}_{t+\delta+1}) \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big] \tag{48}$$
$$= \mathbb{E}\Big[\mathbb{E}\big[\ell_{g_{\bar{a}}}\big(Y_{t+\tau}[a_{t+\delta+1:t+\tau-1}], \bar{H}_t\big) \mid \bar{H}_{t+\delta+1}\big] \;\Big|\; \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\Big] \tag{49}$$
$$= \mathbb{E}\Big[\mathbb{E}\big[\ell_{g_{\bar{a}}}\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}], \bar{H}_t\big) \mid \bar{H}_{t+\delta+1}\big] \;\Big|\; \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\Big] \tag{50}$$
$$= \mathbb{E}\big[\ell_{g_{\bar{a}}}\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}], \bar{H}_t\big) \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big] \tag{51}$$
$$= \mathbb{E}\big[\ell_{g_{\bar{a}}}\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}], \bar{H}_t\big) \mid \bar{H}_{t+\delta}\big]. \tag{52}$$

The first equality is the definition of $Q_{t+\delta}$, and the second is the induction hypothesis. For the third equality, recursive consistency/composition gives $Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = Y_{t+\tau}[a_{t+\delta+1:t+\tau-1}]$ on $A_{t+\delta} = a_{t+\delta}$. The fourth equality is the tower property, since $\bar{H}_{t+\delta+1}$ contains $(\bar{H}_{t+\delta}, A_{t+\delta})$, and the fifth is sequential exchangeability at time $t + \delta$. In particular, $\bar{H}_t$ is contained in every $\bar{H}_{t+\delta}$, so it is measurable in all these conditioning steps and sequential exchangeability for the potential outcome also applies to the displayed score transformation.

Taking $\delta = 0$ in Eq. 44 and using the tower property therefore gives

$$\mathbb{E}\big[Q_t(\bar{H}_t)\big] = \mathbb{E}\big[\ell_{g_{\bar{a}}}\big(Y_{t+\tau}[a_{t:t+\tau-1}], \bar{H}_t\big)\big] \tag{53}$$
$$= \mathbb{E}\big[\ell_{g_{\bar{a}}}(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}], \bar{H}_t)\big] = \mathcal{L}_{\bar{a}}(g_{\bar{a}}), \tag{54}$$

where $a_{t:t+\tau-1} = \bar{a}_{t:t+\tau-1}$. Combining this equality with Eq. 43 proves the claim. □

D Generative Recursive G-Computation

D.1 Pseudocode for generative recursive g-computation

Algorithm 1 presents the pseudocode for generative recursive g-computation.

Algorithm 1 Generative recursive g-computation

Require: Longitudinal training sample $\mathcal{D}$, prediction origin t, horizon τ, fixed intervention $\bar{a}_{t:t+\tau-1}$, parameter spaces $\{\Phi_{t+\delta}\}_{\delta=0}^{\tau-1}$ for the nuisance generators $\xi^{\bar{a}}_{\phi,t+\delta}$, model-specific generative score ℓ, and Monte Carlo size M.

Ensure: Fitted continuation laws $\{\widehat{\xi}^{\bar{a}}_{t+\delta}\}_{\delta=0}^{\tau-1}$.

$\mathcal{I}_{t+\tau-1} \leftarrow \{i : A^{(i)}_{t+\tau-1} = a_{t+\tau-1}\}$

Fit the final-stage nuisance generator:

$$\widehat{\phi}_{t+\tau-1} \in \arg\max_{\phi\in\Phi_{t+\tau-1}} \frac{1}{|\mathcal{I}_{t+\tau-1}|} \sum_{i\in\mathcal{I}_{t+\tau-1}} \ell_{\xi^{\bar{a}}_{\phi,t+\tau-1}}\left(Y^{(i)}_{t+\tau}, \bar{H}^{(i)}_{t+\tau-1}\right).$$

$\widehat{\xi}^{\bar{a}}_{t+\tau-1} \leftarrow \xi^{\bar{a}}_{\widehat{\phi}_{t+\tau-1},t+\tau-1}$

for $\delta = \tau-2, \tau-3, \ldots, 0$ **do**

  $\mathcal{I}_{t+\delta} \leftarrow \{i : A^{(i)}_{t+\delta} = a_{t+\delta}\}$

  Hold $\widehat{\xi}^{\bar{a}}_{t+\delta+1}$ fixed as the teacher distribution.

  for $i \in \mathcal{I}_{t+\delta}$ **do**

    for $m = 1, \ldots, M$ **do**

      Draw $\widetilde{Y}^{(m)}_{i,t+\delta+1} \sim \widehat{\xi}^{\bar{a}}_{t+\delta+1}(\cdot \mid \bar{H}^{(i)}_{t+\delta+1})$

    end for

  end for

  Fit the time-$t+\delta$ nuisance generator to the pseudo-outcomes:

$$\widehat{\phi}_{t+\delta} \in \arg\max_{\phi\in\Phi_{t+\delta}} \frac{1}{|\mathcal{I}_{t+\delta}|M} \sum_{i\in\mathcal{I}_{t+\delta}} \sum_{m=1}^{M} \ell_{\xi^{\bar{a}}_{\phi,t+\delta}}\left(\widetilde{Y}^{(m)}_{i,t+\delta+1}, \bar{H}^{(i)}_{t+\delta}\right).$$

  $\widehat{\xi}^{\bar{a}}_{t+\delta} \leftarrow \xi^{\bar{a}}_{\widehat{\phi}_{t+\delta},t+\delta}$

end for

return $\{\widehat{\xi}^{\bar{a}}_{t+\delta}\}_{\delta=0}^{\tau-1}$

D.2 Proof of Proposition 1

Proof. We first verify the desired representation at the terminal stage $t+\tau-1$ and then use the recursion in Eq. 6 to move successively from stage $t+\delta+1$ to stage $t+\delta$, for $\delta = \tau-2, \ldots, 0$. Specifically, we establish that

$$\xi^{\bar{a}}_{t+\delta}(y \mid \bar{h}_{t+\delta}) = P\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = y \mid \bar{H}_{t+\delta} = \bar{h}_{t+\delta}\big) \tag{55}$$

for every $\delta = 0, \ldots, \tau-1$ and every relevant $\bar{h}_{t+\delta}$. Throughout, positivity guarantees that the conditional distributions given $A_{t+\delta} = a_{t+\delta}$ appearing below are identified on the relevant support.

At the terminal stage, Eq. 5, recursive consistency, and sequential exchangeability give

$$\begin{aligned}
\xi^{\bar{a}}_{t+\tau-1}(y \mid \bar{h}_{t+\tau-1}) &= P\big(Y_{t+\tau} = y \mid \bar{H}_{t+\tau-1} = \bar{h}_{t+\tau-1}, A_{t+\tau-1} = a_{t+\tau-1}\big) && (56)\\
&= P\big(Y_{t+\tau}[a_{t+\tau-1}] = y \mid \bar{H}_{t+\tau-1} = \bar{h}_{t+\tau-1}, A_{t+\tau-1} = a_{t+\tau-1}\big) && (57)\\
&= P\big(Y_{t+\tau}[a_{t+\tau-1}] = y \mid \bar{H}_{t+\tau-1} = \bar{h}_{t+\tau-1}\big). && (58)
\end{aligned}$$

The first equality is the terminal condition. The next equality uses the terminal case of recursive consistency: on the event $A_{t+\tau-1} = a_{t+\tau-1}$, $Y_{t+\tau}[a_{t+\tau-1}] = Y_{t+\tau}$. The final equality removes $A_{t+\tau-1} = a_{t+\tau-1}$ by sequential exchangeability.

Now consider any $\delta < \tau - 1$ for which Eq. 55 has already been established at stage $t + \delta + 1$. For a relevant history $\bar{h}_{t+\delta}$, the recursive definition and the representation at stage $t + \delta + 1$ give

$$\xi^{\bar{a}}_{t+\delta}(y \mid \bar{h}_{t+\delta}) = \mathbb{E}\Big[\xi^{\bar{a}}_{t+\delta+1}(y \mid \bar{H}_{t+\delta+1}) \;\Big|\; \bar{H}_{t+\delta} = \bar{h}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\Big] \tag{59}$$

$$= \mathbb{E}\Big[P\big(Y_{t+\tau}[a_{t+\delta+1:t+\tau-1}] = y \mid \bar{H}_{t+\delta+1}\big) \;\Big|\; \bar{H}_{t+\delta} = \bar{h}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\Big] \tag{60}$$

$$= \mathbb{E}\Big[P\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = y \mid \bar{H}_{t+\delta+1}\big) \;\Big|\; \bar{H}_{t+\delta} = \bar{h}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\Big] \tag{61}$$

$$= P\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = y \mid \bar{H}_{t+\delta} = \bar{h}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big) \tag{62}$$

$$= P\big(Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = y \mid \bar{H}_{t+\delta} = \bar{h}_{t+\delta}\big)\,. \tag{63}$$

The first equality is the recursion in Eq. 6, and the second uses the representation already obtained at stage $t + \delta + 1$. Under the conditioning event $A_{t+\delta} = a_{t+\delta}$, recursive consistency gives $Y_{t+\tau}[a_{t+\delta:t+\tau-1}] = Y_{t+\tau}[a_{t+\delta+1:t+\tau-1}]$, which yields the third equality. Because $\bar{H}_{t+\delta+1}$ contains both $\bar{H}_{t+\delta}$ and $A_{t+\delta}$, the next equality follows from the law of iterated expectations, applied to the conditional density at y. The final equality again follows from sequential exchangeability. Starting from the terminal-stage equality and applying this recursive step for $\delta = \tau - 2, \ldots, 0$ establishes Eq. 55 at every stage. At $\delta = 0$, the remaining sequence is the full treatment sequence $a_{t:t+\tau-1} = \bar{a}_{t:t+\tau-1}$, and therefore

$$\xi^{\bar{a}}_{t}(y \mid \bar{h}_{t}) = P\big(Y_{t+\tau}[a_{t:t+\tau-1}] = y \mid \bar{H}_{t} = \bar{h}_{t}\big) \tag{64}$$

$$= P\big(Y_{t+\tau}[\bar{a}_{t:t+\tau-1}] = y \mid \bar{H}_{t} = \bar{h}_{t}\big)\,, \tag{65}$$

which is the CDPO in Eq. 1. □

E Theoretical Properties of OrthoGen

E.1 Neyman orthogonality

Let $\eta = (\pi, \xi)$ collect the true propensity-score and continuation-density nuisances. Let $\widetilde{\eta} = (\widetilde{\pi}, \widetilde{\xi})$ be any candidate pair of conditional treatment laws and continuation densities for which the path below is admissible.

For every $g_{\bar{a}} \in \mathcal{G}$, let $s \mapsto g_{\bar{a},s} \in \mathcal{G}$ be an admissible differentiable target-model path with $g_{\bar{a},0} = g^{\star}_{\bar{a}}$ and tangent direction $g_{\bar{a}} - g^{\star}_{\bar{a}}$ at $s = 0$.

For any fixed $0 < \epsilon < 1$, consider the nuisance path $\eta_u = (\pi_u, \xi_u)$, $u \in [0, \epsilon)$, where, for every $\delta = 0, \ldots, \tau - 1$,

$$\begin{aligned} \pi_{t+\delta,u}(\cdot \mid \bar{h}_{t+\delta}) &:= (1-u)\pi_{t+\delta}(\cdot \mid \bar{h}_{t+\delta}) + u\widetilde{\pi}_{t+\delta}(\cdot \mid \bar{h}_{t+\delta}), \\ \xi^{\bar{a}}_{t+\delta,u}(y \mid \bar{h}_{t+\delta}) &:= (1-u)\xi^{\bar{a}}_{t+\delta}(y \mid \bar{h}_{t+\delta}) + u\widetilde{\xi}^{\bar{a}}_{t+\delta}(y \mid \bar{h}_{t+\delta}). \end{aligned} \tag{66}$$

Positivity gives

$$\pi_{t+\delta,u}(a_{t+\delta} \mid \bar{H}_{t+\delta}) \geq (1-u)\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) > 0, \qquad 0 \leq u < \epsilon, \tag{67}$$

so the path is locally positive at $u = 0$. Define

$$r^{\bar{a}}_{u,t+\delta} := \frac{\mathbf{1}\{A_{t+\delta} = a_{t+\delta}\}}{\pi_{t+\delta,u}(a_{t+\delta} \mid \bar{H}_{t+\delta})}, \qquad R^{\bar{a}}_{u,t:t+\delta} := \prod_{\delta'=0}^{\delta} r^{\bar{a}}_{u,t+\delta'}, \qquad \delta = 0, \ldots, \tau-1, \qquad R^{\bar{a}}_{u,t:t-1} := 1. \tag{68}$$

For $g_{\bar{a}} \in \mathcal{G}$, define

$$m^{g_{\bar{a}},\bar{a}}_{t+\delta,u}(\bar{h}_{t+\delta}) := \int_{\mathcal{Y}} \ell_{g_{\bar{a}}}(y, \bar{h}_t)\xi^{\bar{a}}_{t+\delta,u}(y \mid \bar{h}_{t+\delta})\,\mathrm{d}y, \qquad \delta = 0, \ldots, \tau - 1, \tag{69}$$

and

$$m^{g_{\bar{a}},\bar{a}}_{t+\tau,u}(\bar{H}_{t+\tau}) := \ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t). \tag{70}$$

Write $m^{g_{\bar{a}},\bar{a}}_{t+\delta} := m^{g_{\bar{a}},\bar{a}}_{t+\delta,0}$ at the true nuisance.

Along this path, the population ORTHOGEN objective is

$$\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}};\eta_u) := \mathbb{E}_0\left[R^{\bar{a}}_{u,t:t+\tau-1}\ell_{g_{\bar{a}}}(Y_{t+\tau},\bar{H}_t) + \sum_{\delta=0}^{\tau-1} R^{\bar{a}}_{u,t:t+\delta-1}\left(1-r^{\bar{a}}_{u,t+\delta}\right) m^{g_{\bar{a}},\bar{a}}_{t+\delta,u}(\bar{H}_{t+\delta})\right]. \tag{71}$$

Assume that the paths are locally admissible and that the score path has a measurable right derivative and an integrable local envelope sufficient for Fubini's theorem, conditional dominated convergence, and the differentiations below. Endpoint derivatives are right directional derivatives, or Gâteaux derivatives for paths extending to a two-sided neighborhood.

Define the pathwise score derivative in the direction $g_{\bar{a}} - g^\star_{\bar{a}}$ by

$$\dot{\ell}_{g_{\bar{a}}}(y,\bar{h}_t) := D_{g_{\bar{a}}}\ell_{g^\star_{\bar{a}}}(y,\bar{h}_t)\left[g_{\bar{a}} - g^\star_{\bar{a}}\right] := \frac{d}{ds}\ell_{g_{\bar{a},s}}(y,\bar{h}_t)\Big|_{s=0}. \tag{72}$$

For $\delta = 0,\dots,\tau-1$, let

$$\begin{aligned}\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta,u}(\bar{h}_{t+\delta}) &:= \frac{d}{ds} m^{g_{\bar{a},s},\bar{a}}_{t+\delta,u}(\bar{h}_{t+\delta})\Big|_{s=0} \\ &= \int_{\mathcal{Y}} \dot{\ell}_{g_{\bar{a}}}(y,\bar{h}_t)\xi^{\bar{a}}_{t+\delta,u}(y \mid \bar{h}_{t+\delta})\,\mathrm{d}y,\end{aligned} \tag{73}$$

and define

$$\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\tau,u}(\bar{H}_{t+\tau}) := \dot{\ell}_{g_{\bar{a}}}(Y_{t+\tau},\bar{H}_t). \tag{74}$$

At the true nuisance, write

$$\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta} := \dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta,u}\Big|_{u=0}, \qquad \delta = 0,\dots,\tau. \tag{75}$$

At the true continuation nuisance, Eqs. 5 and 6 imply, for every s,

$$m^{g_{\bar{a},s},\bar{a}}_{t+\delta,0}(\bar{H}_{t+\delta}) = \mathbb{E}_0\left[m^{g_{\bar{a},s},\bar{a}}_{t+\delta+1,0}(\bar{H}_{t+\delta+1}) \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\right], \qquad \delta = 0,\dots,\tau-1. \tag{76}$$

Differentiating at $s=0$ gives

$$\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}) = \mathbb{E}_0\left[\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta+1}(\bar{H}_{t+\delta+1}) \mid \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\right], \qquad \delta = 0,\dots,\tau-1. \tag{77}$$

Define the mixed derivative by

$$D_\eta D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\left[g_{\bar{a}} - g^\star_{\bar{a}}, \widetilde{\eta} - \eta\right] := \frac{d}{du}\left\{\frac{d}{ds}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a},s};\eta_u)\Big|_{s=0}\right\}\Big|_{u=0}. \tag{78}$$

Theorem 1 (Neyman-orthogonality)**.** *Under the causal assumptions in Sec. 3, the objective is Neyman-orthogonal at $(g^\star_{\bar{a}},\eta)$, namely*

$$D_\eta D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\left[g_{\bar{a}} - g^\star_{\bar{a}}, \tilde{\eta} - \eta\right] = 0, \tag{11}$$

for every $g_{\bar{a}} \in \mathcal{G}$ and every admissible nuisance perturbation $\tilde{\eta}$, where $D\mathcal{L}(\cdot)[\cdot]$ denotes the pathwise derivative (see Appendix E.1 for definitions).

Proof. Differentiating Eq. 71 with respect to the target-model path gives

$$D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta_u)\left[g_{\bar{a}} - g^\star_{\bar{a}}\right] = \mathbb{E}_0\left[R^{\bar{a}}_{u,t:t+\tau-1}\dot{\ell}_{g_{\bar{a}}}(Y_{t+\tau},\bar{H}_t) + \sum_{\delta=0}^{\tau-1} R^{\bar{a}}_{u,t:t+\delta-1}\left(1-r^{\bar{a}}_{u,t+\delta}\right)\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta,u}(\bar{H}_{t+\delta})\right]. \tag{79}$$

Since

$$R^{\bar{a}}_{u,t:t+\delta-1} - R^{\bar{a}}_{u,t:t+\delta} = R^{\bar{a}}_{u,t:t+\delta-1}\left(1-r^{\bar{a}}_{u,t+\delta}\right), \tag{80}$$

Eq. 79 becomes

$$D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta_u)\,[g_{\bar{a}}-g^\star_{\bar{a}}] = \mathbb{E}_0\left[\dot{m}^{g_{\bar{a}},\bar{a}}_{t,u}(\bar{H}_t)+\sum_{\delta=0}^{\tau-1}R^{\bar{a}}_{u,t:t+\delta}\left\{\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta+1,u}(\bar{H}_{t+\delta+1})-\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta,u}(\bar{H}_{t+\delta})\right\}\right]. \tag{81}$$

Set

$$\dot{\Delta}^{g_{\bar{a}}}_{t+\delta} := \dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta+1}(\bar{H}_{t+\delta+1})-\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}). \tag{82}$$

By Eq. 77,

$$\mathbb{E}_0\left[\dot{\Delta}^{g_{\bar{a}}}_{t+\delta}\mid \bar{H}_{t+\delta},A_{t+\delta}=a_{t+\delta}\right]=0,\qquad \delta=0,\dots,\tau-1. \tag{83}$$

Continuation-density derivative. Define

$$\dot{m}^{g_{\bar{a}},\bar{a}}_{\xi,t+\delta}(\bar{H}_{t+\delta}) := \frac{d}{du}\dot{m}^{g_{\bar{a}},\bar{a}}_{t+\delta,u}(\bar{H}_{t+\delta})\Big|_{u=0} = \int_{\mathcal{Y}}\dot{\ell}_{g_{\bar{a}}}(y,\bar{H}_t)\left\{\widetilde{\xi}^{\bar{a}}_{t+\delta}(y\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(y\mid\bar{H}_{t+\delta})\right\}\mathrm{d}y, \tag{84}$$

Holding π fixed in Eq. 79 gives

$$D_\xi D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\left[g_{\bar{a}}-g^\star_{\bar{a}},\widetilde{\xi}-\xi\right]=\sum_{\delta=0}^{\tau-1}\mathbb{E}_0\left[R^{\bar{a}}_{t:t+\delta-1}\left(1-r^{\bar{a}}_{t+\delta}\right)\dot{m}^{g_{\bar{a}},\bar{a}}_{\xi,t+\delta}(\bar{H}_{t+\delta})\right]. \tag{85}$$

Equation 36, with $Z=1$, gives $\mathbb{E}_0[r^{\bar{a}}_{t+\delta}\mid\bar{H}_{t+\delta}]=1$. Because $R^{\bar{a}}_{t:t+\delta-1}\dot{m}^{g_{\bar{a}},\bar{a}}_{\xi,t+\delta}(\bar{H}_{t+\delta})$ is $\bar{H}_{t+\delta}$-measurable, every summand in Eq. 85 is zero. Therefore,

$$D_\xi D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\left[g_{\bar{a}}-g^\star_{\bar{a}},\widetilde{\xi}-\xi\right]=0. \tag{86}$$

Propensity derivative. For every $\delta'=0,\dots,\tau-1$,

$$\frac{d}{du}r^{\bar{a}}_{u,t+\delta'}\Big|_{u=0}=-r^{\bar{a}}_{t+\delta'}\frac{\widetilde{\pi}_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})-\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}{\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}. \tag{87}$$

The product rule gives

$$\frac{d}{du}R^{\bar{a}}_{u,t:t+\delta}\Big|_{u=0}=-R^{\bar{a}}_{t:t+\delta}\sum_{\delta'=0}^{\delta}\frac{\widetilde{\pi}_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})-\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}{\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}. \tag{88}$$

Holding ξ fixed in Eq. 81 gives

$$D_\pi D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\,[g_{\bar{a}}-g^\star_{\bar{a}},\widetilde{\pi}-\pi]=-\sum_{\delta=0}^{\tau-1}\sum_{\delta'=0}^{\delta}\mathbb{E}_0\left[R^{\bar{a}}_{t:t+\delta}\frac{\widetilde{\pi}_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})-\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}{\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}\dot{\Delta}^{g_{\bar{a}}}_{t+\delta}\right]. \tag{89}$$

For $\delta'\le\delta$, the prefactor

$$R^{\bar{a}}_{t:t+\delta-1}\frac{\widetilde{\pi}_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})-\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})}{\pi_{t+\delta'}(a_{t+\delta'}\mid\bar{H}_{t+\delta'})} \tag{90}$$

is $\bar{H}_{t+\delta}$-measurable. The IPW identity in Eq. 36 and Eq. 83 give

$$\mathbb{E}_0\left[r^{\bar{a}}_{t+\delta}\dot{\Delta}^{g_{\bar{a}}}_{t+\delta}\mid\bar{H}_{t+\delta}\right]=\mathbb{E}_0\left[\dot{\Delta}^{g_{\bar{a}}}_{t+\delta}\mid\bar{H}_{t+\delta},A_{t+\delta}=a_{t+\delta}\right]=0. \tag{91}$$

Conditioning on $\bar{H}_{t+\delta}$ therefore makes every summand in Eq. 89 zero, and

$$D_\pi D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\,[g_{\bar{a}}-g^\star_{\bar{a}},\widetilde{\pi}-\pi]=0. \tag{92}$$

Linearity in the nuisance direction and Eqs. 86–92 yield

$$D_\eta D_{g_{\bar{a}}}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)\,[g_{\bar{a}}-g^\star_{\bar{a}},\widetilde{\eta}-\eta]=0. \tag{93}$$

□

E.2 Double robustness and quasi-oracle efficiency

Let $\widehat{\eta} = (\widehat{\pi}, \widehat{\xi})$ be fitted on an auxiliary sample or complementary cross-fitting folds. We condition on the nuisance-training sample throughout. The corresponding weights $\widehat{r}^{\bar{a}}_{t+\delta}$ and $\widehat{R}^{\bar{a}}_{t:t+\delta}$ are obtained by replacing π with $\widehat{\pi}$ in Eq. 68.

For fixed $g_{\bar{a}} \in \mathcal{G}$, let $m^{g_{\bar{a}},\bar{a}}_{t+\delta}$ and $\widehat{m}^{g_{\bar{a}},\bar{a}}_{t+\delta}$ denote the score responses in Eq. 8 evaluated at $\xi^{\bar{a}}_{t+\delta}$ and $\widehat{\xi}^{\bar{a}}_{t+\delta}$, respectively, and set $m^{g_{\bar{a}},\bar{a}}_{t+\tau}(\bar{H}_{t+\tau}) := \ell_{g_{\bar{a}}}(Y_{t+\tau}, \bar{H}_t)$ as in Eq. 70. The continuation recursion implies

$$\begin{aligned} \mathcal{L}_{\bar{a}}(g_{\bar{a}}) &= \mathbb{E}_0[m^{g_{\bar{a}},\bar{a}}_t(\bar{H}_t)], \\ m^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}) &= \mathbb{E}_0\big[m^{g_{\bar{a}},\bar{a}}_{t+\delta+1}(\bar{H}_{t+\delta+1}) \,\big|\, \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big], \qquad \delta = 0, \dots, \tau-1. \end{aligned} \tag{94}$$

Let $\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta})$ denote the population objective in Eq. 71 evaluated at $\widehat{\eta}$.

Lemma 3 (Exact longitudinal doubly robust remainder)**.** *Suppose the causal assumptions in Sec. 3 hold, $\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) > 0$ almost surely for $\delta = 0, \dots, \tau-1$, and the displayed quantities are integrable. Then, for every fixed $g_{\bar{a}} \in \mathcal{G}$,*

$$\begin{aligned} &\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta}) - \mathcal{L}_{\bar{a}}(g_{\bar{a}}) \\ &\quad = \sum_{\delta=0}^{\tau-1} \mathbb{E}_0\left[\widehat{R}^{\bar{a}}_{t:t+\delta-1} \frac{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) - \pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})} \left\{\widehat{m}^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}) - m^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta})\right\}\right]. \end{aligned} \tag{95}$$

Proof. For any $\{m_{t+\delta}\}_{\delta=0}^{\tau}$,

$$\widehat{R}^{\bar{a}}_{t:t+\tau-1} m_{t+\tau} + \sum_{\delta=0}^{\tau-1} \widehat{R}^{\bar{a}}_{t:t+\delta-1}(1 - \widehat{r}^{\bar{a}}_{t+\delta}) m_{t+\delta} = m_t + \sum_{\delta=0}^{\tau-1} \widehat{R}^{\bar{a}}_{t:t+\delta}(m_{t+\delta+1} - m_{t+\delta}). \tag{96}$$

Applying this identity to the sequence $\{m^{g_{\bar{a}},\bar{a}}_{t+\delta}\}_{\delta=0}^{\tau}$, the recursion in Eq. 94 and the conditional inverse-probability identity give

$$\mathbb{E}_0\big[\widehat{r}^{\bar{a}}_{t+\delta}\left\{m^{g_{\bar{a}},\bar{a}}_{t+\delta+1}(\bar{H}_{t+\delta+1}) - m^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta})\right\} \,\big|\, \bar{H}_{t+\delta}\big] \tag{97}$$

$$= \frac{\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})} \mathbb{E}_0\big[m^{g_{\bar{a}},\bar{a}}_{t+\delta+1}(\bar{H}_{t+\delta+1}) - m^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}) \,\big|\, \bar{H}_{t+\delta}, A_{t+\delta} = a_{t+\delta}\big] = 0. \tag{98}$$

Since $\widehat{R}^{\bar{a}}_{t:t+\delta-1}$ is $\bar{H}_{t+\delta}$-measurable,

$$\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; (\widehat{\pi}, \xi)) = \mathbb{E}_0\big[m^{g_{\bar{a}},\bar{a}}_t(\bar{H}_t)\big] = \mathcal{L}_{\bar{a}}(g_{\bar{a}}). \tag{99}$$

Subtracting the two objectives gives

$$\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta}) - \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; (\widehat{\pi}, \xi)) = \sum_{\delta=0}^{\tau-1} \mathbb{E}_0\left[\widehat{R}^{\bar{a}}_{t:t+\delta-1}(1 - \widehat{r}^{\bar{a}}_{t+\delta})\left\{\widehat{m}^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta}) - m^{g_{\bar{a}},\bar{a}}_{t+\delta}(\bar{H}_{t+\delta})\right\}\right]. \tag{100}$$

The factor multiplying $1 - \widehat{r}^{\bar{a}}_{t+\delta}$ is $\bar{H}_{t+\delta}$-measurable, and

$$\mathbb{E}_0\big[1 - \widehat{r}^{\bar{a}}_{t+\delta} \,\big|\, \bar{H}_{t+\delta}\big] = 1 - \frac{\pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})} = \frac{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) - \pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}. \tag{101}$$

Conditioning each summand on $\bar{H}_{t+\delta}$ proves Eq. 95. No property of the score beyond integrability is used. □

Corollary 1 (Exact stagewise double robustness)**.** *For every fixed $g_{\bar{a}} \in \mathcal{G}$, $\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta}) = \mathcal{L}_{\bar{a}}(g_{\bar{a}})$ if, at every $\delta = 0, \dots, \tau-1$, either*

$$\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) = \pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) \quad \text{or} \quad \widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot \mid \bar{H}_{t+\delta}) = \xi^{\bar{a}}_{t+\delta}(\cdot \mid \bar{H}_{t+\delta}) \qquad \mathbb{P}_0\text{-almost surely}. \tag{102}$$

The choice may vary across stages. More generally, for a fixed $g_{\bar{a}}$, equality of the corresponding score responses $\widehat{m}^{g_{\bar{a}},\bar{a}}_{t+\delta} = m^{g_{\bar{a}},\bar{a}}_{t+\delta}$ is sufficient even when the full continuation densities differ.

Proof. Immediate from Eq. 95; equality of the continuation densities implies equality of their score integrals. □

Thus, Corollary 1 is an exact, rather than only rate, double-robustness statement for the population objective.

Let $g_{\bar{a}}^\star$ retain its definition as an oracle projection, let $\widehat{g}_{\bar{a}} \in \mathcal{G}$ denote the fitted second-stage target generator obtained using the estimated nuisances, and write $\|\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star\|_{\mathcal{G}}$ for the relevant model-dependent discrepancy. For measurable $f(y, \bar{H}_{t+\delta})$ at a relative treatment step δ, define

$$\|f\|_{4,2} := \left[\mathbb{E}_0\left\{\left(\int_{\mathcal{Y}} f(y, \bar{H}_{t+\delta})^2 \, \mathrm{d}y\right)^2\right\}\right]^{1/4}. \tag{103}$$

Assumption 1 (Target-model regularity)**.** *For the selected pair $(g_{\bar{a}}^\star, \widehat{g}_{\bar{a}})$, the following conditions hold.*

(i) There is an admissible path $s \mapsto g_{\bar{a},s}$, $s \in [0,1]$, with $g_{\bar{a},0} = g_{\bar{a}}^\star$ and $g_{\bar{a},1} = \widehat{g}_{\bar{a}}$. The map $s \mapsto \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a},s}; \widehat{\eta})$ is twice continuously differentiable, while $s \mapsto \mathcal{L}_{\bar{a}}(g_{\bar{a},s})$ is right differentiable at zero; differentiation may be interchanged with expectation and integration. First- and second-order derivative notation with bracketed direction $\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star$ is induced by this path.

(ii) Moving from the oracle projection toward the fitted target cannot increase the population objective to first order:

$$D_g \mathcal{L}_{\bar{a}}(g_{\bar{a}}^\star)[\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star] \le 0. \tag{104}$$

(iii) There is a constant $\alpha_{\mathcal{G}} > 0$ such that, at the intermediate model $\bar{g}_{\bar{a}} = g_{\bar{a},\bar{s}}$ furnished by the Taylor expansion,

$$-D_g^2 \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(\bar{g}_{\bar{a}}; \widehat{\eta})\,[\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star, \widehat{g}_{\bar{a}} - g_{\bar{a}}^\star] \ge \alpha_{\mathcal{G}} \|\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star\|_{\mathcal{G}}^2. \tag{105}$$

(iv) There is a finite constant $B_{\mathcal{G}}$ such that

$$\left\| \left[\int_{\mathcal{Y}} \left| D_g \ell_{g_{\bar{a}}^\star}(y, \bar{H}_t)[\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star] \right|^2 \, \mathrm{d}y \right]^{1/2} \right\|_{L^4(\mathbb{P}_0)} \le B_{\mathcal{G}} \|\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star\|_{\mathcal{G}}. \tag{106}$$

Interpretation of Assumption 1. In our work, we slightly generalize the target model regularity assumptions, initially suggested by Melnychuk & Feuerriegel (2026) for the cross-sectional setting. Still, our conditions (i)–(iv) are mild and can be realistically satisfied by a broad class of probabilistic models (including normalizing flows and diffusion models).

- First, our condition (i) requires smoothness only along the selected path connecting $g_{\bar{a}}^\star$ and $\widehat{g}_{\bar{a}}$; and such pathwise differentiability and interchange of differentiation with integration are standard for smoothly parameterized generative objectives.
- Further, condition (ii) is automatic whenever the path remains locally in $\mathcal{G}$, by the optimality of the oracle projection.
- Then, in condition (iii), curvature is required only in the realized direction $\widehat{g}_{\bar{a}} - g_{\bar{a}}^\star$ and only at the intermediate model arising in the one-dimensional Taylor expansion.[4] This requirement is also compatible with standard generative models:
  - For normalizing flows, Eq. 14 shows that population maximum likelihood is equivalent, up to an additive constant, to minimizing a forward Kullback–Leibler criterion (Papamakarios et al., 2021), making local quadratic curvature natural under regularity and local identification.

[4] Closely related model-dependent curvature conditions for normalizing-flow and diffusion-based doubly robust generative learners have also been formulated and shown to hold asymptotically under regularity and nuisance-consistency conditions (Melnychuk & Feuerriegel, 2026).

– For diffusion models, the ELBO terms reduce to weighted noise-prediction errors, while Eq. 24 gives the usual simplified squared-error objective (Ho et al., 2020); the criterion is therefore directly quadratic in the predicted-noise function.

Thus, condition (iii) concerns curvature in the induced function-space discrepancy and does not require global convexity of the neural-network parameterization.

- Finally, condition (iv) is a standard local continuity and moment condition: It requires directional changes in the generative score to be controlled by the chosen model discrepancy and to possess a suitable fourth-moment envelope.

To obtain the propensity-error form in Theorem 2, we also require estimated positivity: for some fixed $\varepsilon \in (0, 1]$ and every $\delta = 0, \dots, \tau - 1$,

$$\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) \geq \varepsilon \qquad \mathbb{P}_0\text{-almost surely.} \tag{107}$$

The estimated-positivity condition in Eq. 107 is the usual stability requirement for inverse-propensity weighting and can be enforced in practice by clipping estimated propensities away from zero.

In the following, whenever constants depending on the horizon are absorbed into $\lesssim$, τ is treated as fixed.

Theorem 2 (Quasi-oracle efficiency and double robustness)**.** *Under mild regularity conditions for the target model, the squared distance between the fitted target generator $\widehat{g}_{\bar{a}}$ and the oracle population maximizer $g^\star_{\bar{a}}$ can be upper-bounded in the following way:*

$$\|\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}\|^2_{\mathcal{G}} \lesssim \underbrace{\mathrm{Opt}^{\bar{a}}_{\mathcal{G}}}_{\text{(I) target-model learning error}} + \underbrace{\left[\sum_{\delta=0}^{\tau-1} \|\widehat{\pi}_{t+\delta} - \pi_{t+\delta}\|_{L^4(\mathbb{P}_0)} \, \|\widehat{\xi}^{\bar{a}}_{t+\delta} - \xi^{\bar{a}}_{t+\delta}\|_{4,2}\right]^2}_{\text{(II) higher-order nuisance error}}, \tag{12}$$

where $\|\cdot\|_{\mathcal{G}}$ denotes a model-specific norm, (I) is a standard target-model learning error term, and (II) is a higher-order nuisance error term. This bound implies that ORTHOGEN *is (a)* ***doubly robust*** *and (b)* ***quasi-oracle efficient****.*

Proof of Theorem 2. Recall the conditional-population optimization error from the main text, written in the notation of Eq. 71 as

$$\mathrm{Opt}^{\bar{a}}_{\mathcal{G}} := \sup_{g_{\bar{a}} \in \mathcal{G}} \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta}) - \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(\widehat{g}_{\bar{a}}; \widehat{\eta}). \tag{108}$$

Write

$$W^{\bar{a}}_{t+\delta} := \widehat{R}^{\bar{a}}_{t:t+\delta-1} \frac{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta}) - \pi_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}{\widehat{\pi}_{t+\delta}(a_{t+\delta} \mid \bar{H}_{t+\delta})}. \tag{109}$$

A second-order Taylor expansion along the selected path gives, for some intermediate $\bar{g}_{\bar{a}}$,

$$\begin{aligned}\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(\widehat{g}_{\bar{a}}; \widehat{\eta}) = \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}}; \widehat{\eta}) + D_g \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}}; \widehat{\eta})[\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}] \\ + \frac{1}{2} D^2_g \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(\bar{g}_{\bar{a}}; \widehat{\eta}) \, [\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}, \widehat{g}_{\bar{a}} - g^\star_{\bar{a}}] \, .\end{aligned} \tag{110}$$

Condition (iii) and $g^\star_{\bar{a}} \in \mathcal{G}$ imply

$$\frac{\alpha_{\mathcal{G}}}{2} \|\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}\|^2_{\mathcal{G}} \leq \mathrm{Opt}^{\bar{a}}_{\mathcal{G}} + D_g \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}}; \widehat{\eta})[\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}]. \tag{111}$$

Here the first term is bounded by $\mathrm{Opt}^{\bar{a}}_{\mathcal{G}}$ because $g^\star_{\bar{a}} \in \mathcal{G}$. To isolate the nuisance perturbation, add and subtract the true-nuisance population objective:

$$\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta}) = \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \eta) + \left\{ \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \widehat{\eta}) - \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \eta) \right\}. \tag{112}$$

Lemma 3 gives $\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g_{\bar{a}}; \eta) = \mathcal{L}_{\bar{a}}(g_{\bar{a}})$. Hence, condition (ii) yields

$$D_g \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}}; \widehat{\eta})[\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}] \leq \left| D_g \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}}; \widehat{\eta})[\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}] \; - D_g \mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}}; \eta)[\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}] \right| . \tag{113}$$

Differentiating Eq. 95 in the same direction yields

$$D_g\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\widehat{\eta})[\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}] - D_g\mathcal{L}^{\bar{a}}_{\mathrm{OrthoGen}}(g^\star_{\bar{a}};\eta)[\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}]$$
$$= \sum_{\delta=0}^{\tau-1}\mathbb{E}_0\left[W^{\bar{a}}_{t+\delta}\int_{\mathcal{Y}} D_g\ell_{g^\star_{\bar{a}}}(y,\bar{H}_t)[\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}]\left\{\widehat{\xi}^{\bar{a}}_{t+\delta}(y\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(y\mid\bar{H}_{t+\delta})\right\}\mathrm{d}y\right]. \quad (114)$$

By Cauchy–Schwarz in y, Hölder's inequality with exponents $(2,4,4)$, and condition (iv),

$$\left|\mathbb{E}_0\left[W^{\bar{a}}_{t+\delta}\int_{\mathcal{Y}} D_g\ell_{g^\star_{\bar{a}}}(y,\bar{H}_t)[\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}]\left\{\widehat{\xi}^{\bar{a}}_{t+\delta}(y\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(y\mid\bar{H}_{t+\delta})\right\}\mathrm{d}y\right]\right| \quad (115)$$
$$\le \|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)}\left\|\left[\int_{\mathcal{Y}}\left|D_g\ell_{g^\star_{\bar{a}}}(y,\bar{H}_t)[\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}]\right|^2\mathrm{d}y\right]^{1/2}\right\|_{L^4(\mathbb{P}_0)}\left\|\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})\right\|_{4,2} \quad (116)$$
$$\le B_{\mathcal{G}}\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|_{\mathcal{G}}\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)}\left\|\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})\right\|_{4,2}. \quad (117)$$

Combining this bound with Eqs. 111 and 113 gives

$$\frac{\alpha_{\mathcal{G}}}{2}\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|^2_{\mathcal{G}} \le \mathrm{Opt}^{\bar{a}}_{\mathcal{G}} + B_{\mathcal{G}}\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|_{\mathcal{G}}\left[\sum_{\delta=0}^{\tau-1}\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)}\left\|\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})\right\|_{4,2}\right]. \quad (118)$$

Young's inequality yields

$$B_{\mathcal{G}}\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|_{\mathcal{G}}\left[\sum_{\delta=0}^{\tau-1}\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)}\left\|\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})\right\|_{4,2}\right] \quad (119)$$
$$\le \frac{\alpha_{\mathcal{G}}}{4}\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|^2_{\mathcal{G}} + \frac{B^2_{\mathcal{G}}}{\alpha_{\mathcal{G}}}\left[\sum_{\delta=0}^{\tau-1}\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)}\left\|\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})\right\|_{4,2}\right]^2. \quad (120)$$

Rearranging yields the explicit intermediate bound

$$\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|^2_{\mathcal{G}} \le \frac{4}{\alpha_{\mathcal{G}}}\mathrm{Opt}^{\bar{a}}_{\mathcal{G}} + \frac{4B^2_{\mathcal{G}}}{\alpha^2_{\mathcal{G}}}\left[\sum_{\delta=0}^{\tau-1}\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)}\left\|\widehat{\xi}^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})-\xi^{\bar{a}}_{t+\delta}(\cdot\mid\bar{H}_{t+\delta})\right\|_{4,2}\right]^2. \quad (121)$$

It remains to control the stagewise weight error. Under Eq. 107, each estimated one-step weight is at most ε^{-1}, and therefore

$$0 \le \widehat{R}^{\bar{a}}_{t:t+\delta-1} \le \varepsilon^{-\delta}, \qquad \|\widehat{R}^{\bar{a}}_{t:t+\delta-1}\|_{L^4(\mathbb{P}_0)} \le \varepsilon^{-\delta}. \quad (122)$$

Thus, Hölder's inequality gives

$$\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)} \le \varepsilon^{-1}\|\widehat{R}^{\bar{a}}_{t:t+\delta-1}\|_{L^4(\mathbb{P}_0)}\|\widehat{\pi}_{t+\delta}(a_{t+\delta}\mid\bar{H}_{t+\delta})-\pi_{t+\delta}(a_{t+\delta}\mid\bar{H}_{t+\delta})\|_{L^4(\mathbb{P}_0)} \quad (123)$$
$$\le \varepsilon^{-(\delta+1)}\|\widehat{\pi}_{t+\delta}(a_{t+\delta}\mid\bar{H}_{t+\delta})-\pi_{t+\delta}(a_{t+\delta}\mid\bar{H}_{t+\delta})\|_{L^4(\mathbb{P}_0)}. \quad (124)$$

In particular,

$$\|W^{\bar{a}}_{t+\delta}\|_{L^2(\mathbb{P}_0)} \le C_\delta\|\widehat{\pi}_{t+\delta}-\pi_{t+\delta}\|_{L^4(\mathbb{P}_0)}, \qquad C_\delta := \varepsilon^{-(\delta+1)}, \quad (125)$$

where the propensity functions in the norm are evaluated at $(a_{t+\delta},\bar{H}_{t+\delta})$. Substitution into Eq. 121 gives

$$\|\widehat{g}_{\bar{a}}-g^\star_{\bar{a}}\|^2_{\mathcal{G}} \le \frac{4}{\alpha_{\mathcal{G}}}\mathrm{Opt}^{\bar{a}}_{\mathcal{G}} + \frac{4B^2_{\mathcal{G}}}{\alpha^2_{\mathcal{G}}}\left[\sum_{\delta=0}^{\tau-1}C_\delta\|\widehat{\pi}_{t+\delta}-\pi_{t+\delta}\|_{L^4(\mathbb{P}_0)}\|\widehat{\xi}^{\bar{a}}_{t+\delta}-\xi^{\bar{a}}_{t+\delta}\|_{4,2}\right]^2. \quad (126)$$

The additional cumulative-weight condition used in the previous formulation is therefore implied by estimated positivity and is not needed separately.

Finally, fixed τ gives $\max_{0\le\delta\le\tau-1} C_\delta \le \varepsilon^{-\tau}$. Absorbing $\alpha_{\mathcal{G}}^{-1}$, $B_{\mathcal{G}}$, ε^{-1}, and fixed-horizon factors into $\lesssim$ yields

$$\|\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}\|^2_{\mathcal{G}} \lesssim \mathrm{Opt}^{\bar{a}}_{\mathcal{G}} + \left[\sum_{\delta=0}^{\tau-1} \|\widehat{\pi}_{t+\delta} - \pi_{t+\delta}\|_{L^4(\mathbb{P}_0)} \|\widehat{\xi}^{\bar{a}}_{t+\delta} - \xi^{\bar{a}}_{t+\delta}\|_{4,2}\right]^2, \tag{127}$$

which is Eq. 12. True propensity positivity remains part of the causal identification assumptions, but only estimated positivity is used in the stagewise weight bound above. □

Rate interpretation. Suppose the preceding conditions hold with probability tending to one, with $\alpha_{\mathcal{G}}$ bounded away from zero, $B_{\mathcal{G}}$ bounded above, fixed $\varepsilon > 0$, and fixed τ. For every $\delta = 0, \dots, \tau - 1$, let

$$\|\widehat{\pi}_{t+\delta} - \pi_{t+\delta}\|_{L^4(\mathbb{P}_0)} = O_{\mathbb{P}}(\varepsilon_{\pi,t+\delta,n}), \tag{128}$$

$$\|\widehat{\xi}^{\bar{a}}_{t+\delta} - \xi^{\bar{a}}_{t+\delta}\|_{4,2} = O_{\mathbb{P}}(\varepsilon_{\xi,t+\delta,n}). \tag{129}$$

Matching the main-text convention, ρ_n is a rate for the target-generator norm, so the corresponding squared-scale condition is $\mathrm{Opt}^{\bar{a}}_{\mathcal{G}} = O_{\mathbb{P}}(\rho_n^2)$. Theorem 2 then gives

$$\|\widehat{g}_{\bar{a}} - g^\star_{\bar{a}}\|^2_{\mathcal{G}} = O_{\mathbb{P}}\left(\rho_n^2 + \left[\sum_{\delta=0}^{\tau-1} \varepsilon_{\pi,t+\delta,n}\varepsilon_{\xi,t+\delta,n}\right]^2\right). \tag{130}$$

Equivalently, the target-generator norm is $O_{\mathbb{P}}(\rho_n + \sum_{\delta=0}^{\tau-1} \varepsilon_{\pi,t+\delta,n}\varepsilon_{\xi,t+\delta,n})$.

Exact double robustness is supplied by Lemma 3 and Corollary 1: at each stage, correctness of either nuisance component makes the corresponding product remainder vanish. *Rate double robustness* follows because nuisance effects enter through products, so a slower rate for one nuisance can be offset by a faster rate for the other. Finally, if

$$\sum_{\delta=0}^{\tau-1} \varepsilon_{\pi,t+\delta,n}\varepsilon_{\xi,t+\delta,n} = o(\rho_n), \tag{131}$$

then nuisance estimation is asymptotically negligible at the target-learning scale: the norm and squared norm are respectively $O_{\mathbb{P}}(\rho_n)$ and $O_{\mathbb{P}}(\rho_n^2)$. This is the stated *quasi-oracle efficiency*, since the learner attains the rate of the corresponding oracle learner with known nuisances.

F Dataset Details

F.1 Fully synthetic dataset

Overview. We use a fully synthetic benchmark to evaluate joint distributional learning when the outcome is multivariate and the covariate process itself responds to treatment. Each trajectory contains a scalar time-varying state X_t, a binary treatment A_t, and a three-dimensional continuous outcome $Y_t = (Y_{t1}, Y_{t2}, Y_{t3})^\top$. The data are generated directly in the timing convention of Sec. 3, $\bar{H}_t \to A_t \to (X_{t+1}, Y_{t+1})$. We simulate trajectories of length 20 and use disjoint outer partitions of 3,000 training, 1,000 validation, and 1,000 test patients. As in the semi-synthetic experiment, the training partition is divided evenly between nuisance and target-model fitting. Each outcome coordinate is standardized using its mean and standard deviation in the outer training partition, and this transformation is kept fixed for validation, test, and oracle samples. The equations below describe the raw DGP scale; the coordinate-wise outcome transformation is applied only after complete trajectories have been generated.

On the raw DGP scale, the baseline variables are initialized independently as

$$X_0 \sim \mathcal{N}(0, 0.8^2), \qquad Y_0 \sim \mathcal{N}_3(0, D_0 R_0 D_0), \qquad X_0 \perp Y_0, \tag{132}$$

where $D_0 = \mathrm{diag}(0.60, 0.66, 0.57)$ and

$$R_0 = \begin{pmatrix} 1 & 0.32 & 0.20 \\ 0.32 & 1 & 0.28 \\ 0.20 & 0.28 & 1 \end{pmatrix}. \tag{133}$$

This Gaussian law for Y_0 initializes the recursion only; the conditional Gaussian/asymmetric-Laplace mixture in Eq. 136 governs each post-baseline outcome Y_{t+1} for $t = 0, \ldots, 18$. The complete numerical configuration—including all coefficients, scales, clipping constants, and correlation matrices—is provided in the code. We describe the structure of the data-generating process below.

State and treatment processes. The next state follows

$$X_{t+1} = \alpha_X X_t + \beta_Y s^\top Y_t + \beta_A(A_t - 0.5) + \epsilon^X_{t+1}, \qquad \epsilon^X_{t+1} \sim \mathcal{N}(0, \sigma^2_X), \tag{134}$$

where s is a fixed weighted summary of the three outcomes. For treatment assignment, g_X and g_Y denote clipped and scaled versions of the current state and outcomes. The observational propensity is

$$P(A_t = 1 \mid \bar{H}_t) = \text{sigmoid}\big\{\gamma_0 + \gamma_X g_X(X_t) + \gamma_Y r^\top g_Y(Y_t) + \gamma_A(A_{t-1} - 0.5)\big\}, \tag{135}$$

where r combines all three outcome coordinates and $A_{-1} = 0.5$ is used at the first decision. Clipping bounds the treatment logits and avoids exact-zero or exact-one propensities. Since the current state and outcomes affect treatment, while treatment subsequently affects both the next state and next outcome, the process has explicit treatment–confounder feedback.

Joint outcome law. Conditional on the current history and treatment, each outcome coordinate has a Gaussian/asymmetric-Laplace mixture marginal,

$$Y_{t+1,j} \mid \bar{H}_t, A_t \sim w_{t,j}\mathcal{N}(\mu^G_{t,j}, \sigma^2_{G,t,j}) + (1 - w_{t,j})\,\text{ALD}(\mu^L_{t,j}, b_{t,j}, q_{t,j}), \tag{136}$$

where $\mu^L_{t,j}$ is the $q_{t,j}$ quantile of the asymmetric-Laplace component. The Gaussian location contains a stable, near-diagonal vector autoregression together with coordinate-specific sinusoidal state terms and centered treatment terms. The asymmetric-Laplace location equals the Gaussian location plus a coordinate-specific hyperbolic-tangent shift in the state. Both component scales, the mixture probability $w_{t,j}$, and the asymmetry parameter $q_{t,j}$ vary with the clipped state, the corresponding current outcome coordinate, and the current treatment. The three coordinates are not conditionally independent. Cross-coordinate dependence uses the same configured correlation matrix for the multivariate Gaussian innovation and for Gaussian-copula latent variables underlying each of the two exponential vectors that form the asymmetric-Laplace draw and the mixture indicators. This preserves the specified marginal mixtures while inducing contemporaneous dependence across outcomes. All numerical parameters are given in the code.

Interventional ground truth. For a requested regime $\bar{a}_{t:t+\tau-1}$, the oracle fixes the factual history through t and then recursively simulates both future states and future outcomes. We impose the alternating sequence specified in the main evaluation section; for example, $\bar{a}_{t:t+\tau-1} = (1)$ when $\tau = 1$, $\bar{a}_{t:t+\tau-1} = (1, 0)$ when $\tau = 2$, and $\bar{a}_{t:t+\tau-1} = (1, 0, 1, 0, 1)$ when $\tau = 5$. For each horizon we set $t = 19 - \tau$, so all methods target the outcome at the final state time. Evaluation uses 300 draws from the fitted model and 300 oracle draws for each test trajectory. We compare the resulting three-dimensional empirical distributions using the squared 2-Wasserstein transport cost W_2^2, averaged across test trajectories; the optimal transport is computed on complete outcome vectors.

F.2 Semi-synthetic MIMIC-III dataset

Cohort and covariates. Our semi-synthetic benchmark follows the construction of Hess et al. (2026b): it combines real covariate trajectories from MIMIC-III (Johnson et al., 2016), represented as hourly records from the MIMIC-Extract pipeline (Wang et al., 2020), with simulated treatments and outcomes. We draw a training cohort of 1,500 patients for each stage of our sample-split procedure. Specifically, from patients with at least 20 hourly records, we sample 5,000 without replacement and retain the first 20 hours of each trajectory. The resulting patient-level partitions contain 3,000 training (1500 for each stage), 1,000 validation, and 1,000 test trajectories.

The time-varying covariates are heart rate, red blood cell count, sodium, mean blood pressure, systemic vascular resistance, glucose, urine chloride, Glasgow Coma Scale total, hematocrit, set positive end-expiratory pressure, and respiratory rate. Age and gender are included as static covariates. To match the released data pipeline, the selected hourly vital and laboratory channels are forward-filled and then backward-filled before patients are sampled. We standardize the selected continuous channels, standardize age, and one-hot encode gender. Static features are repeated over

time and concatenated with the eleven dynamic measurements when evaluating the data-generating mechanism. After simulation, covariates and generated outcomes are standardized using means and standard deviations from the outer training partition; the same transformation is then applied to validation and test patients. The whole-frame filling and selected-cohort preprocessing are kept to reproduce the reference benchmark; all model fitting takes place only after this construction step.

Treatment and outcome mechanism. Let $\widetilde{X}_t \in \mathbb{R}^d$ contain the eleven measurements and the repeated encoded static covariates at hour t. We define

$$q_t = \frac{1}{d}\sum_{j=1}^{d} \sin(\widetilde{X}_{tj}), \qquad c_t = \frac{1}{d}\sum_{j=1}^{d} \cos(\widetilde{X}_{tj}). \tag{137}$$

The initial binary treatment B_0 and scalar outcome $\widetilde{Y}_0$ are drawn as

$$B_0 \sim \text{Bernoulli}\{\text{sigmoid}(q_0)\}, \qquad \widetilde{Y}_0 = 2(B_0 - 1/2) + \epsilon_0, \qquad \epsilon_0 \sim \mathcal{N}(0, 0.1^2). \tag{138}$$

For $t = 1, \ldots, 19$, factual treatments are sampled according to

$$\begin{aligned} \eta_t &= \sin(\widetilde{Y}_{t-1}) + q_{t-1} - B_{t-1} + \nu_t, & \nu_t &\sim \mathcal{N}(0, 0.2^2), \\ B_t &\sim \text{Bernoulli}\{\text{sigmoid}(\eta_t)\}. \end{aligned} \tag{139}$$

Writing $D_t = 2(B_t - 1/2) \in \{-1, 1\}$, the outcome then evolves as

$$\widetilde{Y}_t = \widetilde{Y}_{t-1} + D_t \exp\left\{\frac{D_t c_t}{t}\sin(\widetilde{Y}_{t-1})\right\} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, 0.1^2). \tag{140}$$

The innovations are drawn independently. Thus, previous outcomes affect later treatment assignment, while treatment, current covariates, and the previous outcome jointly determine the next outcome.

Timing, splitting, and ground truth. The reference generator assigns B_t and $\widetilde{Y}_t$ at the same raw hour. We align it with our pre-interventional history convention by setting $A_t = B_{t+1}$ and retaining B_0 in the baseline history. For a horizon τ, we set $t = 19 - \tau$ and intervene with the all-treated sequence $a_{t+\delta} = 1$ for $\delta = 0, \ldots, \tau - 1$. Hence, the regimes for $\tau = 1, \ldots, 5$ are (1), $(1, 1)$, $(1, 1, 1)$, $(1, 1, 1, 1)$, and $(1, 1, 1, 1, 1)$, respectively, and the terminal outcome is always $\widetilde{Y}_{19}$.

The 3,000 training patients are split once, by patient, into 1,500 patients for nuisance estimation and 1,500 for fitting the target generator; validation and test patients remain separate. For evaluation, the oracle starts from each test patient's realized outcome at the prediction origin, keeps that patient's covariate path fixed, imposes the all-treated treatment sequence, and resamples the subsequent outcome innovations. The reported squared 2-Wasserstein costs compare 100 samples from each fitted model with 100 such oracle draws per test patient.

F.3 Real-world MIMIC-III dataset

Cohort and covariates. Our real-world benchmark uses the same MIMIC-III cohort construction and preprocessing pipeline as the semi-synthetic experiment, but keeps the factual treatments and factual outcomes observed in the data. Concretely, we use hourly records from MIMIC-III (Johnson et al., 2016) extracted with the MIMIC-Extract pipeline (Wang et al., 2020), retain patients with at least 20 hourly records, and keep the first 20 hours of each selected trajectory. The dataset is split at the patient level into 3,000 training, 1,000 validation, and 1,000 test trajectories, with the training partition further divided into 1,500 patients for nuisance estimation and 1,500 for fitting the final target generator.

The covariates are exactly those used in the semi-synthetic benchmark: eleven time-varying clinical measurements together with static age and gender. Specifically, the dynamic channels are heart rate, red blood cell count, sodium, mean blood pressure, systemic vascular resistance, glucose, urine chloride, Glasgow Coma Scale total, hematocrit, set positive end-expiratory pressure, and respiratory rate. As above, selected hourly channels are forward-filled and backward-filled before patient sampling; continuous covariates and outcomes are standardized using statistics from the outer training partition, and gender is one-hot encoded.

Treatment, outcome, and evaluation setup. Unlike the semi-synthetic benchmark, no simulated treatment or outcome model is introduced here: the observed treatment path and observed outcome trajectory are taken as the real-world data source. Specifically, the observed treatment A_t is the binary mechanical-ventilation indicator (`vent`) at raw hourly index $t+1$, whereas the observed outcome Y_t is diastolic blood pressure at raw hourly index t; this one-step reindexing aligns the data with our convention that A_t is applied after the history through time t. For evaluation horizons $\tau \in \{1, 3, 5\}$, we set $t = 9 - \tau$ and predict the distribution of the factual terminal outcome under the all-treated sequence $a_{t+\delta} = 1$ for $\delta = 0, \ldots, \tau - 1$. This choice keeps the terminal evaluation time fixed at $Y_{t+\tau} = Y_9$ (hour 9) across horizons. Thus, the intervention sequences are (1) for $\tau = 1$, $(1, 1, 1)$ for $\tau = 3$, and $(1, 1, 1, 1, 1)$ for $\tau = 5$. Because counterfactual ground truth is unavailable in the real-world setting, the squared Wasserstein cost used for the synthetic experiments is not available here: it requires paired empirical samples from a known oracle distribution over complete trajectories, which we do not have. Instead, we evaluate factual prediction quality directly against the observed outcome with the continuous ranked probability score (CRPS), which is a proper scoring rule for one-dimensional predictive distributions and only requires the observed factual outcome. For this evaluation, predictions are made under each target treatment sequence, and CRPS is computed against the observed factual outcome only for held-out patients whose observed factual treatment sequence exactly matches that target sequence. The CRPS results are reported in Table 4.

G Implementation Details

G.1 Transformer history encoder

Architecture. All propensity, continuation, and second-stage target models use the same Transformer specification to encode the observed pre-interventional history. We follow the compact encoder-only design of Frauen et al. (2025), which was also adopted by the overlap-weighted learner of Hess et al. (2026b). At decision time t, each history token concatenates the contemporaneous covariates and outcome with the preceding treatment as $[X, V, Y, A_{\text{prev}}]$, where the static covariates V are repeated along the time axis and A_{-1} is represented by a zero sentinel. A linear input map projects each token to the Transformer dimension, after which we add a fixed sinusoidal positional encoding. A causal attention mask prevents look-ahead, while a separate padding mask accommodates histories of different lengths. The output at the last valid time point is passed through a linear layer and a ReLU activation to obtain the context vector supplied to the propensity head or conditional generator. The encoder is trained jointly with the head to which it belongs; there is no separate representation-pretraining step. The propensity nuisance is one time-shared model evaluated at each decision point, whereas the stage-specific continuation models and the final target learners have separate fitted parameters but use the same encoder specification.

Hyperparameters. We fixed the complete Transformer profile in Table 5 before running the experiments. All encoder-specific values were inherited from the longitudinal Transformer configurations of Frauen et al. (2025) and Hess et al. (2026b). We kept them fixed across datasets, treatment horizons, generative learners, and the normalizing-flow and diffusion backends; none was tuned for the experiments in this paper. This choice deliberately separates the comparison of causal objectives from a search over history-encoder architectures.

G.2 Generative models

We implement the framework with conditional normalizing flows and diffusion models. Their mathematical definitions and population objectives are given separately in Appendix B; this subsection records only the architectures, optimization choices, and hyperparameter selection used in the experiments. Every generator receives the 20-dimensional context produced by the history encoder above. A continuation model at relative step δ is conditioned on $\bar{H}_{t+\delta}$, whereas a target generator is conditioned on $\bar{H}_t$. The intervention is fixed for each fitted model and is therefore not included again in the conditioning input. All outcome coordinates are represented on the standardized scale described in Appendix F; any additional affine normalization internal to a generator is inverted before returning samples on that evaluation scale.

Table 5: Fixed hyperparameters of the Transformer history encoder. None of these values was tuned specifically for our experiments.

Hyperparameter	**Value**	**Role**
Transformer blocks	1	Causal encoder block
Model dimension d_{model}	30	Token and attention representation width
Attention heads	3	Head width $d_{\text{model}}/3 = 10$
Feed-forward dimension	20	Hidden width inside the encoder block
Dropout probability	0.1	Applied to the input embedding and encoder block
Readout dimension	20	Dimension of the context passed to the downstream head
Activation	ReLU	Encoder feed-forward and readout nonlinearity
Normalization	Post-norm	Layer normalization after residual updates
Positional encoding	Fixed sinusoidal	Defined up to the padded history length

G.2.1 NORMALIZING-FLOW ARCHITECTURE

Our implementation uses a standard Gaussian base, followed by a trainable coordinate-wise affine map and one conditional monotone quadratic-spline transform. For the scalar MIMIC-III outcome, a one-hidden-layer ELU network maps the history context to the spline parameters. For the three-dimensional synthetic outcome, the base is multivariate Gaussian and the final transform is a conditional autoregressive spline whose one-hidden-layer autoregressive network produces the parameters for all three coordinates. The latter is a joint flow and hence models contemporaneous dependence between outcome coordinates. The spline acts on the standardized outcome scale. Its symmetric bound was fixed at 3, whereas the number of bins was tuned as described below. Optional Gaussian perturbations of the context and outcome are used only as training regularizers; validation NLLs and generated samples always use the unperturbed model.

G.2.2 DIFFUSION-MODEL ARCHITECTURE

We use a linear β schedule. The diffusion step is represented by a fixed sinusoidal embedding and concatenated with the noisy outcome. The denoiser is a two-layer, per-observation MLP: a one-hidden-layer ELU hypernetwork maps the history context to all weights and biases of this MLP, whose hidden width is reported below and whose output has the same dimension as the outcome. Thus, the complete denoising map adapts to the patient history. Training samples a diffusion step uniformly, adds Gaussian noise, and minimizes Eq. 24, averaged over outcome coordinates. After fitting, samples are obtained from standard Gaussian noise by ordinary ancestral DDPM sampling through all S reverse transitions. Because this model does not provide a tractable exact conditional density, we use its denoising objective for validation.

G.2.3 TRAINING WITHIN THE LONGITUDINAL LEARNERS

The history encoder and its generative head are optimized jointly with Adam; there is no separate encoder pretraining. Training pools all eligible rolling origins, with continuation models indexed by relative regime stage and each input restricted to the corresponding absolute-time history. At the terminal step of generative recursive g-computation, the model is fitted to the observed terminal outcome among factual rows whose last treatment agrees with the regime. Earlier continuation models are then fitted backward to detached pseudo-outcomes drawn from the already fitted, frozen successor model; these pseudo-outcomes are redrawn once per epoch. We use 16 Monte Carlo draws for NFs and 64 for DMs. The IPTW, RA, and ORTHOGEN target generators are subsequently trained with their respective objectives on the target-model split. For numerical stability, the cumulative inverse-propensity ratios used by IPTW and ORTHOGEN are clipped. Their validation objective is checked after a 10-epoch warm-up, with patience 20 and minimum improvement 10^{-3}; the parameters with the best validation objective are restored. For RA and ORTHOGEN, an exponential moving average with decay 0.995 is updated after each optimizer step and used for validation and final sampling.

Table 6: Consolidated hyperparameter search space for the terminal nuisance. Only parameters that were varied are shown. σ_C and σ_Y denote training-only Gaussian noise applied to the history context and standardized outcome, respectively.

Family	Hyperparameter	Values considered
NF	Conditioner width	$\{20, 32, 64\}$
	Spline bins	$\{5, 10, 15, 20\}$
	Learning rate	$\{3\times10^{-4}, 10^{-3}\}$
	Batch size	$\{32, 64\}$
	Weight decay	$\{0, 10^{-4}\}$
	Context noise σ_C	$\{0, 0.01, 0.05\}$
	Outcome noise σ_Y	$\{0, 0.01, 0.05\}$
	Maximum epochs	$\{30, 50, 100\}$
DM	Diffusion steps S	$\{50, 100, 150\}$
	Denoiser hidden width	$\{10, 20\}$
	Context noise σ_C	$\{0, 0.01\}$
	Learning rate	$\{3\times10^{-4}, 10^{-3}, 5\times10^{-3}\}$
	Weight decay	$\{0, 10^{-4}\}$

Table 7: Selected generative-model hyperparameters for each dataset.

Dataset	Family	Selected architecture	Selected optimization and regularization
Semi-synthetic MIMIC-III	NF	Width 20; 10 bins	Learning rate 10^{-3}; batch 64; weight decay 10^{-4}; 100 epochs; $(\sigma_C, \sigma_Y) = (0.01, 0)$
	DM	$S = 100$; denoiser width 20	Learning rate 3×10^{-4}; weight decay 10^{-4}; $\sigma_C = 0.01$
Real-world MIMIC-III	NF	Width 20; 10 bins	Learning rate 10^{-3}; batch 64; weight decay 10^{-4}; 100 epochs; $(\sigma_C, \sigma_Y) = (0.01, 0)$
	DM	$S = 100$; denoiser width 20	Learning rate 3×10^{-4}; weight decay 10^{-4}; $\sigma_C = 0.01$
Fully synthetic	NF	Width 20; 10 bins	Learning rate 10^{-3}; batch 64; weight decay 10^{-4}; 30 epochs; $(\sigma_C, \sigma_Y) = (0, 0)$
	DM	$S = 150$; denoiser width 20	Learning rate 3×10^{-4}; weight decay 10^{-4}; $\sigma_C = 0$

G.2.4 HYPERPARAMETER SELECTION

We tuned each generative family separately on each dataset by isolating the terminal nuisance. Specifically, each candidate fits only the conditional law of $Y_{t+\tau}$ given $\bar{H}_{t+\tau-1}$ on factual rows with $A_{t+\tau-1} = a_{t+\tau-1}$. This is direct generative fitting of the outcome conditional on the history available at the last treatment step: it does not require backward pseudo-outcomes and does not fit a propensity model or a second-stage target learner. Within the nuisance-estimation split, patients were divided into 80% tuning-training and 20% validation subsets. The same five replicate datasets and patient splits were reused for every candidate, and candidates were ranked by their mean validation generative loss. For NFs, this criterion was the clean conditional NLL (the joint NLL for the three-dimensional outcome). For DMs, it was the fixed-seed denoising loss in Eq. 24. Neither test outcomes nor counterfactual oracle samples entered model selection.

Table 6 gives one consolidated search space. When the dataset-specific grids differed, we report the broader set containing all candidate values considered across the datasets. Only parameters that were actually varied are included. The NF spline bound was fixed at 3. For DMs, we fixed the sinusoidal step-embedding width at 16, used a linear schedule with $(\beta_1, \beta_S) = (10^{-4}, 0.02)$, a batch size of 64, and a maximum of 100 epochs. Selection was nevertheless performed independently for the semi-synthetic MIMIC-III, real-world MIMIC-III, and fully synthetic datasets.

The selected configurations are summarized in Table 7. The selected generator architecture was used for all continuation and target models of the corresponding backend and dataset. The listed optimization values are those of the terminal nuisance.

Table 8: Average end-to-end wall-clock runtime per replicate for a representative treatment horizon $\tau = 3$. Each replicate includes nuisance estimation, fitting of PI, IPTW, RA, and ORTHOGEN, and evaluation.

Dataset	**NFs**	**DMs**
Fully synthetic	56 min	122 min
Semi-synthetic MIMIC-III	110 min	176 min
Real-world MIMIC-III	27 min	63 min

H COMPUTATIONAL RUNTIME

We report wall-clock runtimes to provide a representative measure of the computational cost of our experiments. Table 8 reports the average end-to-end runtime of one replicate for the intermediate treatment horizon $\tau = 3$, using the same configurations as in the main experiments. Each replicate includes data preprocessing, nuisance estimation, fitting of all four learners (PI, IPTW, RA, and ORTHOGEN), and evaluation. The reported values are averages over five replicates and are rounded to the nearest minute. All experiments use $n_{\text{train}} = 3000$ and were run on an Amazon SageMaker `ml.g4dn.xlarge` instance equipped with one NVIDIA T4 GPU (16 GB GPU memory), 4 vCPUs, and 16 GiB of system memory.

The computational requirements of the individual learners differ according to their adjustment strategy. PI requires fitting the sequence of τ continuation models, whereas IPTW requires one time-shared propensity model and one target generator. RA adds a target generator on top of the continuation sequence, while ORTHOGEN combines the time-shared propensity model and continuation sequence with a second-stage target generator. In our implementation, nuisance estimates are shared across the corresponding learners; hence, we report the end-to-end runtime of the complete experiment. As expected, the DM instantiations are computationally more demanding than the NF instantiations due to their iterative denoising procedure.