ESTIMATION OF LONG-TERM DOSE-RESPONSE VIA LONGITUDINAL SURROGACY UNDER TIME-VARYING CONFOUNDERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating long-term treatment effects is critical in many domains, yet randomized experiments with long follow-up periods are often costly and ethically challenging. Such studies also typically involve dynamic treatment and evolving disease states, and single-visit observation provides little predictive signal for the chronic outcome, such as 90-day mortality given daily treatment for a chronic disease. Existing methods either average the dose, ignore time-dependent confounders, or rely on stringent assumptions. In this paper, we propose a Window-based Sequential Causal Framework (WSCF) that exploits longitudinal surrogates and the treatment path to identify long-term effects across sequential analysis windows. To address time-dependent confounding bias, we learn representations of the history to balance current treatment assignment, while aligning long-term observational data with auxiliary experimental data via optimal transport weighting to adjust for unmeasured confounding. Theoretically, we derive sequential g-formula for identification under a mean exchangeability assumption weaker than prior work, control the rollout error due to estimated surrogates, and establish a generalization bound on counterfactual prediction error. Experiments on synthetic and semi-synthetic datasets from real-world healthcare and online industrial data demonstrate both the effectiveness and long-depth robustness.

1 INTRODUCTION

Estimating long-term treatment effects is critical yet challenging. While randomized experiments serve as the gold-standard for analysis, extended follow-up makes them prohibitively costly and ethically burdensome. Long-term observational studies are more accessible but vulnerable to unmeasured confounding (Hernán & Robins, 2020). Furthermore, given dynamic treatments and evolving subject states inherent to these settings, single-time-point observations offer limited predictive power for final long-term outcomes (Moraffah et al., 2021; Schomaker et al., 2024).

Prior works on time-varying treatments span causal g-computation (Robins, 1986; Li et al., 2021), surrogate bridging (Kallus & Mao, 2025; Battocchi et al., 2021), sequence models (Lim et al., 2018; Bica et al., 2020b), and continuous-time dynamics (Soleimani et al., 2017; Seedat et al., 2022), but are mostly confined to binary treatments, short-term feedback or unconfoundedness. For work on long-term effects, a classic way is indirect inference via a short-term surrogate outcome effect transportable between a long-term cohort and a small short-term trial. Early works naïvely average all time-varying covariates (Chen & Ritzwoller, 2023), later works extend to longitudinal surrogates but stick to static treatment (Yang et al., 2024b; Cai et al., 2025) or ignore time-varying confounders (Huang et al., 2026). Among these methods, observed confounding is adjusted by propensity weighting (Robins et al., 2000; Liu et al., 2020), regression (Athey et al., 2020) or balancing representation (Bica et al., 2020b); unobserved confounding by surrogacy (Hu et al., 2022), equi-confounding (Ghassami et al., 2022), identifiable latent representations (Cai et al., 2024), optimal transport (Yang et al., 2024b), or instrumental variables (Wang et al., 2025). None addresses long-term effect estimation under time-varying continuous treatments and confounders in the data fusion framework.

This unexplored intersection faces pressing demand in studies of chronic disease and online platform maintenance, etc., where treatment and subject conditions co-evolve. Our investigation is motivated

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.

by findings from a real-world industrial longitudinal study for dynamic incentive effect on long-term user activeness (detailed in AppendixD). The study shows that single/average incentive barely correlates with future 7-day activity, but its oscillation matters. Furthermore, accuracy saturates beyond including a moderate history depth and surrogates, suggesting that long-term estimation can therefore rely on manageable trajectory windows, which raises the central question: how to systematically leverage manageable surrogate history to robustly estimate long-term dose-responses, when a single treatment carries limited predictive signal under time-varying confounding bias?

For chronic and insensitive feedback, long-term dose response given a treatment path needs to borrow surrogates that carry stronger signal and causally bridge intervention and outcome (Athey et al., 2026), together with recursive histories. Debiasing is harder. Propensity-based methods fail on the countless counterfactual path space, all the more so for continuous doses (Yang et al., 2024b). Time-varying observed and unobserved confounders evolve with treatment and corrupt both assignment and outcome (Hernán et al., 2000; Bica et al., 2020b), so temporal transportability is untenable without external validity and unconfoundedness (Bareinboim & Pearl, 2016; Dahabreh et al., 2020). At decision time, nearest surrogates may still be unobserved, such as delayed disease progression for overall survival, and need extrapolation, so single-step bias compounds exponentially through sequential phase transition into rollout error (Ross et al., 2011; Janner et al., 2019).

To address these challenges, we propose the Window-based Sequential Causal Framework (WSCF). We extend data fusion to dynamic settings with evolving continuous doses and subject status, robustly estimating long-term dose-responses across fixed-width surrogate windows via sequential g-formulation under weakened transportability assumptions. For debiasing in this temporal structure, we adapt representation balancing (Bica et al., 2020b) for history confounders and optimal transport (OT) reweighting to align transition distributions across domains for unmeasured ones (Villani et al., 2009). We further introduce Jacobian Spectral Regularization (JSR) to bound exponential error accumulation from sequential inference on estimated surrogate to a provably milder magnitude. Our main contributions are summarized as follows:

- To our knowledge, this is the first data fusion study to estimate long-term dose-path-response across sequential analysis windows with longitudinal surrogates and time-varying confounders, given weak single-treatment signal; it couples a window-based sequential inference with temporal representation balancing and optimal-transport reweighting.
- For identification, we propose a sequential g-formula under weakened assumptions, establishing an OT reweighting method with theoretical guarantees to eliminate unobserved confounding via cross-domain joint distributional alignment.
- We develop Jacobian Spectral Regularization to suppress exponential rollout error due to nested conditioning on estimated surrogates and derive the generalization bound for counterfactual prediction error under reweighted distributions and JSR constraints.
- We validate the effectiveness of our proposed method through extensive experiments on synthetic and semi-synthetic cohorts built from real healthcare and online industrial data.

2 Related Work

Long-term treatment effect estimation proceeds either from a single source or by data fusion, pairing a short-term trial with long-term observational data. Single-source methods include g-computation (Robins, 1997; Li et al., 2021), surrogate bridging (Prentice, 1989; Frangakis & Rubin, 2002), recurrent networks (Lim et al., 2018; Bica et al., 2020b; Melnychuk et al., 2022), continuous-time models (Soleimani et al., 2017; Seedat et al., 2022), and reinforcement and policy learning (Tran et al., 2024; Wu et al., 2024; Yang et al., 2024a). These rely on sequential exchangeability, focus on discrete treatments, adjust only for observed confounding, and omit long-term outcome definitions. Data-fusion approaches vary by temporal granularity, from single decisions with static surrogates (Athey et al., 2026; Ghassami et al., 2022; Cai et al., 2024) to multi-period surrogates under invariant treatments (Battocchi et al., 2021; Imbens et al., 2025); continuous doses are studied by optimal transport (Yang et al., 2024b) and representation balancing (Cheng et al., 2021), but with time-invariant treatments.

Confounding is addressed across baseline and sequential settings. In data fusion, baseline observed confounding is adjusted by propensity weighting (Hu et al., 2022; Chen & Ritzwoller, 2023), re-

gression (Athey et al., 2020) or representation balancing (Yang et al., 2024b; Bica et al., 2020b); single-source methods rely on sequential weighting (Robins et al., 2000; Liu et al., 2020), which suffers from variance explosion, or on balanced representations (Bica et al., 2020b;a). No data-fusion method, however, performs step-wise balancing under continuous treatments, and single-source methods lack long-term outcomes. For unobserved confounding, identification uses surrogacy (Athey et al., 2026), equi-confounding (Ghassami et al., 2022), latent representations (Cai et al., 2024) or optimal transport (Yang et al., 2024b), while proximal and instrumental approaches relax unconfoundedness (Wang et al., 2025; Imbens et al., 2025); in dynamic settings, existing studies assume latent confounders away (Battocchi et al., 2021) or extrapolate the bias without correction (Chen et al., 2025). The intersection of time-varying continuous dosing, long-term outcomes, time-varying observed and unobserved confounding, and data fusion remains unexplored.

3 PRELIMINARY

Let $t = 0, 1, \ldots, T$ index the visited time of intervention, with T the total/targeted follow-up time. Let $A_t \in \mathbb{R}$ be the continuous dose assigned at time t, $L_t \in \mathbb{R}^q$ the observed covariates, $U_t \in \mathbb{R}^d$ the unmeasured confounders and $S_t \in \mathbb{R}^m$ the short-term surrogate. Following the causal inference framework (Bica et al., 2020b), write $\bar{A}_t = (A_0, \ldots, A_t)$ for a dosing path, with the same convention for $\bar{L}_t, \bar{S}_t$; $S_t(\bar{a}_t), Y_t(\bar{a}_t)$ are the potential outcome given the whole path, $\bar{S}_{k:j}(\bar{a}_{k:j}) = \{\bar{S}_k(\bar{a}_k), \ldots, \bar{S}_j(\bar{a}_j)\}, k \leq j$; $\bar{H}_t = (\bar{A}_{t-1}, \bar{L}_t)$ the collected history, encoded by $Z_t = \phi(\bar{H}_t)$.

Following the data-fusion paradigm (Athey et al., 2026), let $G \in \{0, 1\}$ denote the data source, with $G = 0$ for observational and $G = 1$ for trial data without Y collection: $\mathcal{D}_0 = \{\bar{A}_{i,t}, \bar{L}_{i,t}, \bar{U}_{i,t}, \bar{S}_{i,t}, \bar{Y}_{i,t}, G_i = 0\}_{i=1}^{n_0}$ and $\mathcal{D}_1 = \{\bar{A}_{i,t}, \bar{L}_{i,t}, \bar{S}_{i,t}, G_i = 1\}_{i=1}^{n_1}$, with samples sizes $n_1 \ll n_0$, and $\mathcal{T}_0, \mathcal{T}_1$ the sets of follow-up periods in 2 studies. $\hat{S}, \hat{Y}$ are estimated via model g, h. Our target is the long-term dose-path-response under dosing trajectory given history information,

$$\mu(\bar{a}_t, \bar{h}_t) \;:=\; \mathbb{E}\big[\, Y(\bar{a}_t) \;\big|\; \bar{A}_{t-1} = \bar{a}_{t-1},\; \bar{L}_t = \bar{l}_t \,\big] = \mathbb{E}\big[\, Y(\bar{a}_t) \;\big|\; \bar{H}_t = \bar{h}_t \,\big]. \tag{1}$$

It cannot be estimated robustly from any source alone, since Y is absent in D_1, while D_0 carries U.

3.1 IDENTIFICATION ASSUMPTIONS

We follow the standard causal inference framework for time-varying treatments and for long-term outcome estimation in the data-fusion setting (Hu et al., 2022; Bica et al., 2020b). The causal graphs corresponding to the identification assumptions below are shown in Appendix Figure4.

Assumption 1 (Consistency). *If $\bar{A}_t = \bar{a}_t$ for a given subject, then $Y_t = Y(\bar{a}_t)$ and $S_t = S(\bar{a}_t)$.*

Assumption 2 (Positivity). *For any t, there exists a constant $0 < \epsilon < 1/2$ such that, (i) $\epsilon < P(A_t = a \mid \bar{A}_{t-1}, \bar{S}_{(t-\delta):(t-1)}, \bar{L}_t, G) < 1 - \epsilon, G = 0, 1$; (ii) $\epsilon < P(G = 0 \mid \bar{A}_t = \bar{a}, \bar{S}_{t-\delta:t} = \bar{s}, \bar{L}_t = \bar{l}) < 1 - \epsilon, \forall(\bar{a}, \bar{s}, \bar{l})$ with $P(\bar{a}, \bar{s}, \bar{l}|G = 1) > 0$.*

A2(i) makes every dose level reachable, and A2(ii) bounds the selection probability away from extremes, so the two-sources alignment weight in Section3.2 is well defined and uniformly bounded.

Assumption 3 (Sequential exchangeability). $Y_t(\bar{a}_t),\ S_t(\bar{a}_t) \perp A_t \mid \bar{A}_{t-1}, \bar{L}_t, G = 1,\ \forall t$.

Assumption 4 (Mean exchangeability). $\mathbb{E}[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t), \bar{A}_t = \bar{a}_t, \bar{L}_t, G = 1] = \mathbb{E}[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t), \bar{A}_t, \bar{L}_t, G = 0],\ \forall t$.

A3 holds by randomization in trial data Its observational counterpart would additionally condition on U_t, which is the bias we aim to address. A4 says that the conditional mean of the long-term outcome is transportable across sources, the conditional version of the comparability requirement underlying surrogate-based transport (Hu et al., 2022). Compared with previous work (Yang et al., 2024b), Proposition 1 shows that classic but more stringent assumptions, conditional external validity and unconfoundedness, are jointly relaxed by A4 under A3, as detailed in the Appendix.

Assumption 5 (Markov assumption). *There exists a constant step δ such that $S_t \perp S_{t'} \mid S_{t-\delta}, \bar{A}_t, \bar{L}_t, G = 1$, and $Y_t \perp S_{t'} \mid S_{t-\delta}, \bar{A}_t, \bar{L}_t, G \in \{0, 1\}$, for any $t' < t - \delta$.*

Assumption 6 (No surrogate feedback). $S_t(\bar{a}_t) \perp A_{t:\cdot}, L_{t+1:\cdot} \mid \bar{S}_{t-1}(\bar{a}_{t-1}), \bar{A}_{t-1}, \bar{L}_t, G = 1$.

A5 states that the pharmacological memory spans δ periods, conditioned on surrogate at δ steps back, the earlier surrogate history no longer affects the current outcome directly. Inspired by Huang et al. (2026), A6 states that conditional on the previous history, the surrogate does not feed forward into A and L directly, where $t:\cdot$ indicates the future from t.

3.2 Optimal Transport Weights

Optimal Transport (OT) (Villani et al., 2009) measures distribution discrepancy by a minimum transport cost and is widely used to learn weights that align heterogeneous domains. To bridge D_1 and D_0, we target the Radon-Nikodym derivative of the alignment block $\chi_t = (\bar{S}_{(t-\delta):t}, \bar{A}_t, \bar{L}_t)$:

$$\tilde{w}_t := \frac{dP\big(S_t,\ \bar{S}_{(t-\delta):(t-1)},\ \bar{A}_t,\ \bar{L}_t \mid G=1\big)}{dP\big(S_t,\ \bar{S}_{(t-\delta):(t-1)},\ \bar{A}_t,\ \bar{L}_t \mid G=0\big)}, \qquad w_t^\star := \frac{\tilde{w}_t}{E_{G=0}[\tilde{w}_t]}, \tag{2}$$

which is fixed by the data-generating mechanisms rather than a batch of free parameters, and yields the joint alignment $w_t^\star\, dP\big(\chi_t \mid G=0\big) = dP\big(\chi_t \mid G=1\big)$. For candidate weights $w \in \mathcal{F}$ we write $\mathrm{d}P^w := w\,\mathrm{d}P(\cdot \mid G=0)$ for the reweighted observational measure and $\mathbb{E}^w[\cdot \mid G=0]$ for its mean, and learn w window by window, each transporting the observational law of χ_t onto the trial one. The learning of the dynamic optimal weight $\hat{w}_t$ is formulated as: $\hat{w}_t = \arg\min_{w\in\mathcal{F}} \mathrm{OT}_t\big(P^w(\bar{S}_{(t-\delta):t},\ A_t,\ Z_t \mid G=0),\ P(\bar{S}_{(t-\delta):t},\ A_t,\ Z_t \mid G=1)\big)$, where $\mathcal{F}$ is the weight-function class and OT_t the 1-Wasserstein distance under the metric d_t.

4 Methodology

We formulate long-term dose-path-response estimation as a Window-based Sequential Causal Framework (WSCF), which reads the follow-up period through finite analysis windows sliding along time: each Y_t is estimated from a δ-width local window holding the preceding surrogates $\bar{S}_{(t-\delta):t}$, the dosing and covariate path $\bar{A}_t, \bar{L}_t$, so estimation at each visit is one instance of the same local problem. The path functional is assembled window by window together with propagated error, and the three subsections below address identification, reweighting alignment and error control.

4.1 Identification of Long-Term Dose Response

Under A1-A6, we formulate the estimand as a nested expectation computable from observed data. We transport the window-conditional law of the surrogate and the conditional mean with the weights w of Eq3, and assemble the two nested sequential g-formulas for S, Y identification. Reweighting the observational measure to align the joint law of χ_t also aligns the conditional law of S_t, since integrating $P(\chi_t|G)$ over s_t returns the denominator of Eq2, leading to the transition kernel within each δ- width window:

Lemma 1 (Single-step cross-domain transport)**.** *Under A1-3, A5 and6, given a set of weights w for observation unit with $P^w(\chi_t|G=0) = P(\chi_t|G=1)$, then for any t, we have*

$$P\big(S_t(\bar{a}_t) \mid \bar{S}_{t-1}(\bar{a}_{t-1}),\ \bar{A}_{t-1},\ \bar{L}_t,\ G=1\big) = P^w\big(S_t \mid \bar{S}_{(t-\delta):(t-1)},\ \bar{A}_t,\ \bar{L}_t,\ G=0\big)$$
$$\mathbb{E}\big[S_t(\bar{a}_t) \mid \bar{S}_{(t-\delta):(t-1)}(\bar{a}_{(t-\delta):(t-1)}),\ \bar{A}_t,\ \bar{L}_t,\ G=1\big] = \mathbb{E}^w\big[S_t \mid \bar{S}_{(t-\delta):(t-1)},\ \bar{A}_t,\ \bar{L}_t,\ G=0\big].$$

Then the estimand of short-term surrogate, $\mu^S(\bar{a}_t \mid \bar{h}_t) := \mathbb{E}[S_t(\bar{a}_t) \mid \bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t]$, is identifiable through nested sequential g-formula, stated by 1 in Appendix, with $\mathbb{E}^w_{S_k|\cdot}$ the conditional expectation of S_k on the reweighted window $P^w(S_k \mid \bar{S}_{(k-\delta):(k-1)}, \bar{A}_k, \bar{L}_k,\ G=0)$. Proposition1 & A5 extent the mean exchangeability to the truncated surrogate path:

Lemma 2 (Terminal cross-domain transport)**.** *Under A1-5, for any t,*

$$\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_{(t-\delta):t}(\bar{a}_{(t-\delta):t}),\ \bar{A}_t,\ \bar{L}_t,\ G=1\big] = \mathbb{E}\big[Y_t \mid \bar{S}_{(t-\delta):t},\ \bar{A}_t,\ \bar{L}_t,\ G=0\big].$$

Theorem 2 (Identification of long-term outcome)**.** *Under Theorem1, Lemma2, and A1-6, the long-term outcome is identifiable via below nested sequential g-formula*

$$\mathbb{E}^w_{S_1|\cdot}\Big[\cdots\ \mathbb{E}^w_{S_t|\cdot}\big[\ \mathbb{E}\big[Y_t \mid \bar{S}_{(t-\delta):t},\ \bar{A}_t,\ \bar{L}_t,\ G=0\big]\ \big]\ \cdots\Big].$$

$\mu(\bar{a}_t, \bar{h}_t)$ is identifiable, with one more g-formula layer than Theorem1, the extra layer being current S_t. All t layers carry w, since the state-trajectory distribution is what is transported, whereas the innermost regression carries none by A4 transportable mechanism. All proofs are in the Appendix.

4.2 Estimation of Optimal Transport Weights

Following Yang et al. (2024b), we estimate OT weights by windowwise Kantorovich transport with entropic Sinkhorn iterations (Cuturi, 2013) via the alignment block $\chi_t = (S_t, R_t, A_t)$, $R_t = (\bar{S}_{(t-\delta):(t-1)}, Z_t)$. Let $P_{g,t}$ denote χ_t distribution in source g, and the true alignment ratio $w_t^\star$ in Eq.2, the estimation-layer oracle $w_t^\sharp := \mathrm{d}P_{1,t}/\mathrm{d}P_{0,t}$, and its estimation $\hat{w}_t$ normalized by $\mathbb{E}_{P_{0,t}}[\hat{w}_t] = 1$. A2(ii) ensures a uniform bound on $w_t^\star \le \bar{u} := \frac{1-\epsilon}{\epsilon}\frac{P(G=0)}{P(G=1)} < \infty$. Empirical, we minimize the windowwise 1-Wasserstein distance

$$\min_{\substack{w\in\mathcal{F}_t \\ \mathbb{E}_{P_{0,t}}[w]=1}} \mathrm{OT}_t\big(P_{0,t}^w, P_{1,t}\big), \qquad \mathrm{d}P_{0,t}^w = w\,\mathrm{d}P_{0,t}, \tag{3}$$

where $\mathcal{F}_t$ is the weight-function class. Following Yang et al. (2024b) Theorem2, our Lemma3 in Appendix shows that the oracle-weighted observational empirical measure then converges weakly, almost surely, to $P_{1,t}$. Identification, however, requires the conditional $P^w\big(S_t \mid \cdot\big)$ alignment rather than χ_t. Its estimation-layer counterpart averages the conditional transport cost over the experimental history-dose marginal $P_{1,t}^{R_t,A_t}$: $\mathrm{OT}_t^{\mathrm{con}}(w) := \int \mathrm{OT}_t\big(P_{0,t}^w(S_t \mid r_t, a_t), P_{1,t}(S_t \mid r_t, a_t)\big)\, P_{1,t}^{R_t,A_t}(\mathrm{d}(r_t, a_t))$, with $P_{g,t}^{R_t,A_t}$ is the (R_t, A_t)-marginal of $P_{g,t}$. Minimising it entails infinitely many subproblems over continuous, multidimensional histories, so we optimise the joint transport empirically. But Lemma3's oracle convergence alone does not deliver conditional alignment. The transfer rests on the regularity below, through the ratio of joint to marginal densities; write total variation distance (TV) as $\mathrm{TV}(P, Q) = \sup_D \big|\, P(D) - Q(D) \,\big| = \frac{1}{2}\|p - q\|_{L^1}$, over measurable events D, and

Condition (R1) [weight function regularity]. Every $f \in \mathcal{F}_t$ satisfies $0 < \underline{\omega}_t \le f \le \overline{\omega}_t < \infty$ and $\mathrm{Lip}(f) \le K_{u,t} < \infty$, with fixed constants independent of sample sizes. The support $\mathcal{X}_t$ of χ_t is compact, and $P_{0,t}$ has full support on it. Normalization bounds learned weights by $\overline{\omega}_t/\underline{\omega}_t$, separately from the oracle bound $\bar{u}$.

Theorem 3 (Alignment transfer)**.** *Suppose (R1) holds and* $\mathrm{OT}_t(P_{0,t}^{\hat{w}_t}, P_{1,t}) \to 0$ *as* $n_0, n_1 \to \infty$. *Define the joint error* $\Delta_t := \mathrm{TV}(P_{0,t}^{\hat{w}_t}, P_{1,t}) = \frac{1}{2}\|\hat{w}_t - w_t^\sharp\|_{L^1(P_{0,t})}$ *and the integrated conditional error* $\nabla_t := \int \mathrm{TV}\big(P_{0,t}^{\hat{w}_t}(S_t \mid r_t, a_t), P_{1,t}(S_t \mid r_t, a_t)\big)\, \nu(\mathrm{d}(r_t, a_t))$, *where* $\nu \in \big\{(P_{0,t}^{\hat{w}_t})^{R_t,A_t},\ P_{1,t}^{R_t,A_t}\big\}$. *Then* $\Delta_t \to 0$, *and, with* $D_{S,t}$ *the surrogate-support diameter,*

$$\nabla_t \le 2\Delta_t \to 0, \quad \mathbb{E}_{P_{1,t}^{R_t,A_t}}\Big[\mathrm{OT}_t\Big(P_{0,t}^{\hat{w}_t}(S_t \mid R_t, A_t), P_{1,t}(S_t \mid R_t, A_t)\Big)\Big] \le 2D_{S,t}\Delta_t \to 0,$$

using the surrogate metric for conditional transport.

Yang et al. (2024b) bounds the conditional transport cost by the joint one, and its Theorem 2 states that the conditional convergence rate is upper bounded by the joint rate. That bound and proof are questionable: a weight oscillating along the surrogate direction drives the joint cost to zero while the conditional error stays bounded away from zero. Theorem3 under (R1) instead supplies the conditional rate $\nabla_t \le 2\Delta_t \to 0$ from the weights' L^1 error. Transport plans provide weights only at sampled points, so out-of-sample evaluation requires a weight function satisfying (R1); applying the alignment to Lemma1 additionally requires Z_t to retain what the history contributes to predicting S_t. Additional theorem discussion is in the Appendix.

4.3 Jacobian Spectral Regularization for Rollout Eror

Short-term surrogates are not always observed at the analysis period, so they must be predicted and a single-step model error then propagates through the sequential rollout. Write $\hat{S}_t^{tf} = g(\bar{S}_{(t-\delta):(t-1)}, Z_t)$ and $\hat{Y}_t^{tf} = h(\bar{S}_{(t-\delta):t}, Z_t)$ for the true observations and $\hat{S}_t$, $\hat{Y}_t$ if using $\hat{\bar{S}}$ under rollout, with per-trajectory errors $e_t := \|\hat{S}_t - S_t\|_2 \le E_t$, and $e_t^{(Y)} := \|\hat{Y}_t - Y_t\|_2$; let

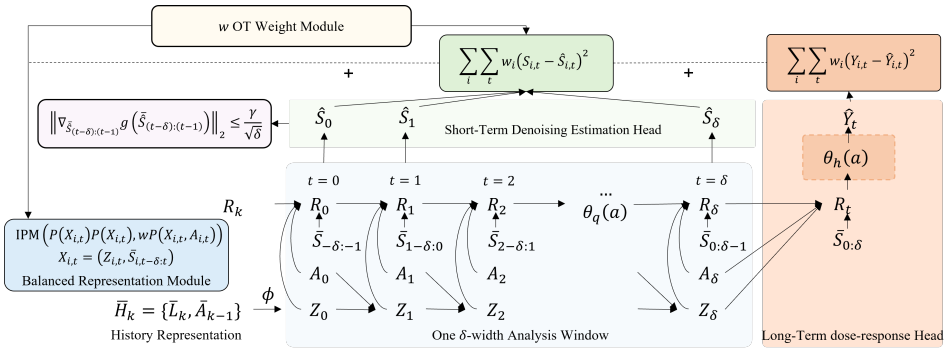


Figure 1: Model architecture of WSCF.

$|\epsilon| \le C_\epsilon, |\epsilon_Y| \le C_{\epsilon,Y}$ be the random errors; L_g and L_h are the Lipschitz constants of $g(\cdot)$ and $h(\cdot)$, and set $\kappa := L_g\sqrt{\delta}$. Lemma4 in Appendix shows that the upper-bound rollout error sequence of e_t is $E_t = \kappa E_{t-1} + C_\epsilon$, and the unconstrained recursion grows as $E_t = O(\kappa^t)$ and $E_t^{(Y)} = O(\sqrt{\delta+1}\,\kappa^t)$, hence exponentially in the depth when $\kappa \ge 1$, concerning the inference error bound. We therefore impose Jacobian Spectral Regularization (JSR) on the estimated generator, based on the Jacobian of g with respect to $\bar{\hat{S}}_{(t-\delta):(t-1)}$,

$$\hat{J}_t := \nabla_{\bar{\hat{S}}_{(t-\delta):(t-1)}} g\big(\bar{\hat{S}}_{(t-\delta):(t-1)}\big) \in \mathbb{R}^{m\times(\delta m)}, \qquad \|\hat{J}_t\|_2 \le \frac{\gamma}{\sqrt{\delta}}, \qquad 0<\gamma<1, \tag{4}$$

γ bounds the sensitivity of the generated state to the whole window. Then the finite-depth bounds:

Theorem 5 (Controlled rollout bias)**.** *Given bounded single-step model error $\epsilon \le C_\epsilon, \epsilon_Y \le C_{\epsilon,Y}$, and g, h are L_g, L_h-Lipschitz continuous, then, with $\|\hat{J}_t\|_2 \le \gamma/\sqrt{\delta}, \gamma \in (0,1)$,*

(i) $e_t \le \frac{C_\epsilon}{1-\gamma} = \mathcal{O}\left(\frac{1}{1-\gamma}\right)$, *(ii)* $e_t^{(Y)} \le L_h\sqrt{\delta+1}\,\frac{C_\epsilon}{1-\gamma} + C_{\epsilon,Y} = \mathcal{O}\left(\frac{\sqrt{\delta+1}}{1-\gamma}\right)$.

The JSR regularizer frees the prediction error from blowing up and confines it to a bounded regime, with asymptotic convergence rate $O(1/(1-\gamma))$. At $\delta = 0$ the the roll out error reduces to single step, and the rollout bias of the overall training is derived with Theorem6. All proofs in the Appendix.

5 WSCF Model Architecture

We instantiate the framework model with 5 modules, shown in Figure1: a shared balanced representation encoder, a per-window reweighting module, a short- and long-term head pivoted on the surrogate path, and the JSR constraint.

OT Reweighting Module. This module learns the alignment weights $\mathbf{w}$, by solving the Kantorovich problem. Adapted from Yang et al. (2024a) to temporal structure, our transport problem is solved per window over the alignment block χ_t, with a batch of observational trajectories against the full experimental sample: $\min_{\mathbf{w}} \min_{\Gamma\in\Pi(\mathbf{w},\boldsymbol{\mu})}\langle\Gamma, C_t\rangle$, $\Pi(\mathbf{w},\boldsymbol{\mu}) = \left\{\Gamma \in \mathbb{R}^{n_0\times n_1} \mid \Gamma\mathbf{1}_{n_1} = \mathbf{w},\, \Gamma^\top\mathbf{1}_{n_0} = \boldsymbol{\mu},\, \Gamma_{ij}\in[0,1]\right\}$, where $\mathbf{w} = (w_{1,t},\ldots,w_{n_0,t})$ for observational, uniform weights $\boldsymbol{\mu}$ for the trial, and cost matrix $(C_t)_{ij} = d_t(\chi_{i,t},\chi_{j,t})$. The weights need to come from a bounded Lipschitz function of χ_t for Theorem3 (R1).

Balanced Representation Module. We design this module for observed time-varying confounders based on A3, by mapping L, A history path to balanced representation $Z_t = \phi(\bar{H}_t)$ with respect to A_t, through a MLP $\phi(\cdot;\theta_\phi)$. By globally minimizing the integral probability metric (IPM) as Wasserstein distance $\mathrm{IPM}(P(\mathbf{X})P(\mathbf{A}), P^w(\mathbf{X},\mathbf{A}))$, $X_t = (\bar{S}_{t:\delta:t}, Z_t)$, the proof in Theorem6 shows that the global solution also upper bound the local IPM_t, mitigating confounding at each timepoint.

Short-&Long-Term Outcome Estimation Modules. These modules estimate the counterfactual outcome. The short-term head feeds $Z_t, \bar{\hat{S}}_{t-\delta:t-1}$ to a GRU $q(\cdot)$ to obtain a latent representation R_t through an MLP $g(\cdot;\theta_g)$, or directly to an alternative LSTM. The long-term head uses the same

Markov step by default, so the window $\hat{\bar{S}}_{t-\delta:t}$ of length $\delta+1$ enters $h(\cdot;\theta_h)$ for each Y_t. We either consider an LSTM, or leverage the attention mechanism $f(\cdot;\theta_f)$ to construct R_t from $\bar{R}_{t-\delta:t-1}$, which then fed to an MLP $h(\cdot;\theta_h)$ (Nie et al., 2021).

JSR Module. At inference $\hat{S}_t$ is rolled forward and the sequence passed to the terminal head. This autoregressive use is the main source of the protocol gap v.s. the teacher-forced evaluation (Janner et al., 2019). The module penalizes the window Jacobian $\hat{J}_{i,t}$ through the hinge $\big(\|\hat{J}_{i,t}\|_F^2-\gamma^2/\delta\big)_+$.

Training Objective. The objective corresponds to the above modules are

$$\begin{aligned}\mathcal{L}(\theta) =& \frac{1}{|\mathcal{T}_0|\,|\mathcal{D}_0|}\sum_{i\in\mathcal{D}_0}\Big[\mathcal{L}_Y\big(Y_{i,t},\hat{Y}_{i,t}\big)+\lambda_S\sum_{t\in\mathcal{T}_0}w_{i,t}\,\mathcal{L}_S\big(S_{i,t},\hat{S}_{i,t}\big)\Big]\\ &+\frac{1-\lambda_S}{|\mathcal{T}_1|\,|\mathcal{D}_1|}\sum_{j\in\mathcal{D}_1}\sum_{t\in\mathcal{T}_1}\mathcal{L}_S\big(S_{j,t},\hat{S}_{j,t}\big)+\frac{\lambda_{\mathrm{JSR}}}{|\mathcal{D}_0|\,|\mathcal{T}_0|}\sum_{i\in\mathcal{D}_0}\sum_{t\in\mathcal{T}_0,\,t\geq\delta}\Big(\big\|\hat{J}_{i,t}\big\|_F^2-\frac{\gamma^2}{\delta}\Big)_+ \qquad (5)\\ &+\frac{\lambda_{\mathrm{IPM}}}{|\mathcal{T}_0|}\,\mathrm{IPM}_{\mathcal{M}}\Big(\big\{X_{i,t},\tilde{A}_{i,t}\big\}_{i\times t}^{n_0\times|\mathcal{T}_0|},\ \big\{w_{i,t}X_{i,t},w_{i,t}A_{i,t}\big\}_{i\times t}^{n_0\times|\mathcal{T}_0|}\Big),\end{aligned}$$

where $\lambda_S,\lambda_{\mathrm{IPM}},\lambda_{\mathrm{JSR}}\in 0,1$ regulates penalization of corresponding modules, $X_t=(\bar{S}_{t-\delta:t},Z_t)$, and $\tilde{A}_i$ is the samples obtained by randomly permuting the originals doses.

Generalization Bound. Although the long-term effect is identifiable, its counterfactual error remains large, as training only on factual trajectories, infinite counterfactual paths under continuous dose path, and time-varying confounding bias. Let $\phi:\bar{\mathcal{H}}\to\mathcal{Z}$, $g:\mathcal{Z}\times\mathcal{A}\times\bar{\mathcal{S}}_{t-\delta:t-1}\to\mathcal{S}$ and $h:\mathcal{Z}\times\mathcal{A}\times\bar{\mathcal{S}}_{t-\delta:t}\to\mathcal{Y}$ be the encoder and two heads. The pointwise losses $\ell_{\phi,g,h}(\bar{h}_t,\bar{a}_t)=\ell^s_{\phi,g}(\bar{h}_t,\bar{a}_t)+\ell^y_{\phi,h}(\bar{h}_t,\bar{a}_t)$, where $\ell^s_{\phi,g}(\bar{h}_t,\bar{a}_t)=\int_{\mathcal{S}}\mathcal{L}(s_t(\bar{a}_t),g(\phi(\bar{h}_t),\bar{a}_t))P(s_t(\bar{a})\mid\bar{h}_t)ds_t(\bar{a})$, $\ell^y_{\phi,h}(\bar{h}_t,\bar{a}_t)=\int_{\mathcal{S}\times\mathcal{Y}}\mathcal{L}(y_t(\bar{a}_t),h(\phi(\bar{h}_t),\bar{a}_t,s_t(\bar{a}_t)))P(y_t(\bar{a}_t),s_t(\bar{a}_t)\mid\bar{h}_t)ds(\bar{h}_t)dy_t(\bar{h}_t)$. We denote the OT reweighted loss at trajectory level as $\ell^{\mathbf{w}}(\theta)$ with expectation $\mathcal{E}^{\mathbf{w}}_f(\theta)=\mathbb{E}_{P^{\phi}_f}[\ell^{\mathbf{w}}_{\phi,g,h}(\theta)]$, and counterfactual risk $\mathcal{E}_{cf}(\theta)=\mathbb{E}_{P^{\phi}_{cf}}[\ell_{\phi,g,h}(\theta)]$ and empirical risk $\hat{\mathcal{E}}_n(\theta)=(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T\ell^{\mathbf{w}}_i(\theta)$ over $|\mathcal{T}_0|$ factual trajectories. Let $\mathcal{M}$ be global class on the pooled window block with $\mathrm{IPM}_{\mathcal{M}}(P,Q)=\sup_{m\in\mathcal{M}}|\int m\,\mathrm{d}P-\int m\,\mathrm{d}Q|$. Write $\mathcal{M}_w=\{\ell^{\mathbf{w}}(\theta):\theta\in\Theta\}$, and fix Θ before sampling. Given the $\ell^{\mathbf{w}}$ upper bound M, and the rollout-and-terminal prediction error bound B, the max protocol gap between inference via $\hat{S}$ vs S. The generalization bound for long-term outcome estimation (experimental loss or weighting bias are not discussed here):

Theorem 6 (End-to-end generalization bound)**.** *Given bounds $w_t\leq\bar{u},|\epsilon|\leq C_\epsilon,|\epsilon_Y|\leq C_{\epsilon,Y}$, measurable ϕ, and $\|J_t\|_2\leq\gamma/\sqrt{\delta},\gamma\in(0,1)$, assuming g,h and both losses are $L_g,L_h,L^S_\ell,L^Y_\ell$-Lipschitz continuous, then we have the max protocol gap as*

$$B=\frac{\lambda_S\gamma L^S_\ell}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}E_{t-1}+\frac{L^Y_\ell L_h}{|\mathcal{T}_0|}\sqrt{\sum_{k=0}^{\delta}E^2_{T-\delta+k}}+\frac{L^Y_\ell C_{\epsilon,Y}}{|\mathcal{T}_0|},$$

and $M=\lambda_S\bar{u}L^S_\ell C_\epsilon+L^Y_\ell C_{\epsilon,Y}/|\mathcal{T}_0|$. If constant $\Lambda_{\mathrm{bal}}>0$ exists for $\lambda_S\ell^s_{\phi,g}/\Lambda_{\mathrm{bal}}\in\mathcal{M}$, and Rademacher complexity $\mathfrak{R}_n$, with probability at least $1-\eta,\eta\in(0,1)$, then

$$\mathcal{E}_{cf}-\hat{\mathcal{E}}_n\ \leq\ \Lambda_{\mathrm{bal}}\,\mathrm{IPM}_{\mathcal{M}}\big(P_{cf}(\mathbf{X})P_{cf}(\mathbf{A}),\ \mathbf{w}P_f(\mathbf{X},\mathbf{A})\big)+B+2\mathfrak{R}_n(\mathcal{M}_w)+M\sqrt{\frac{\log(2/\eta)}{2n}},$$

6 Experiments

We conduct synthetic and semi-synthetic experiments based on real healthcare and industrial online platform data to validate method effectiveness, addressing the following research questions (RQs):

RQ1: Does WSCF improve long-term dose-response estimation compared to other baselines?

RQ2: Do the designed modules address time-varying confounding bias and rollout error?

RQ3: Is the model robust to the temporal structure of analysis window length & prediction depth?

RQ4: How sensitive is model performance to the penalty hyperparameters?

6.1 EXPERIMENTAL SETUP

Dataset Generation. Since true counterfactual outcomes and unmeasured confounders are unavailable in real data, we evaluate the method on a synthetic time-varying cohort and 2 semi-synthetic cohorts built on real longitudinal panels. The synthetic cohort adapts the tumour-growth PK-PD benchmark of Geng et al. (2017) to a chemotherapeutic agent with a daily continuous dose, treating tumour volume growth as S and survival time as Y. We simulate $n_0 = 6000$ observational and $n_1 = 1500$ trial data, with outcomes depending on the current and previous $\delta = 4$ visits and $\beta_U \in \{1, 1.5, 2, 2.5, 3\}$: $\log S_t = \sum_{j=1}^{\delta} w_j \big[(1-\rho)\log S_{t-j} + \rho \log K_V - \beta_c \operatorname{Emax}(C_{t-j}) + \alpha_s^\top L_{t-j} + \gamma_s^\top U_{t-j} + e_{t-j}\big], Y_t = \beta_{Y0} + \frac{1}{\delta}\sum_{j=0}^{\delta}\big(\beta_{YL,j}^\top L_{t-j} + \beta_{YS,j} S_{t-j}\big) + \epsilon_{Y,t}$, with accumulated concentration $C_t = \gamma_h C_{t-1} + A_t$, decay weights w_j, and random noise e, ϵ, and saturating $\operatorname{Emax}(C) = C/(C + C_{50})$ term makes the timing of the dosing. The dose is $A_t = A_{\max}\sigma\big(0.7\tilde{m}_t + 0.35\,\beta_U^\top U_t\big) D_t$, with $\tilde{m}_t$ assignment score on history H_t and D_t a piecewise-constant interruption mask for trial data, and $\beta_u \in \{1, 1.5, 2, 2.5, 3\}$ controls the strength of U imapct. The first semi-synthetic cohort uses MIMIC-III (Johnson et al., 2016) with titrated vasopressors as A, mean arterial pressure as S, and 90-day mortality as Y. The second is built from a real-world industrial online platform (Industrial), with a dynamic incentive score as A and future 3-, 7-day users active time as S, Y. For confidentiality, the cohort undergoes privacy-preserving reconstruction following Battocchi et al. (2021) via resampled moments. Both split units 9:1 across sources, treating part of the covariates as U. Further details are in Appendix.

Baselines. We compare with static direct-estimation methods: naive A-Y regression, DRNet (Zhang et al., 2020) ; surrogate-based methods, including static "A,S-Y" (Hu et al., 2022), time varying Longitudinal-S (Huang et al., 2026), and LTEE (Cheng et al., 2021), LIRL (Cai et al., 2025), and LEARN (Yang et al., 2024b) for time-varying surrogates but static treatment/covaraites; and time-varying-dose methods without surrogate pivot: RMSN (Lim et al., 2018), CRN (Bica et al., 2020b), G-Net (Li et al., 2021), DCRN (Berrevoets et al., 2021). Ablation variants drop individual components (WSCF-noOT/noJSR/noIPM), all components (WSCF-None), with WSCF-oracle using true surrogates as a skyline. WSCF-LSTM use the same method but different LSTM model, while WSCF use GRU sequential model.

Metrics. The observational cohort is split 6:2:2 into training/validation/test with no unit crossing, and report the mean integrated squared error (MISE) over potential dosing paths $\mathcal{A}$: $\mathrm{MISE} = (n|\mathcal{A}|)^{-1}\sum_{i=1}^{n}\sum_{a\in\mathcal{A}}\big(\hat{\mu}_i(a) - \mu_i(a)\big)^2$, averaged and standard deviation (SD) are derive by 3 independent replications.

Implementation. We use Adam at learning rate of 10^{-3} and batch size 512, $\lambda_{\mathrm{IPM}} = \lambda_{\mathrm{JSR}} = 0.05, \lambda_S = 0.5$, with early stopping on observational validation loss. Further details is in Appendix.

6.2 RESULTS

Overall Performance (RQ1). We evaluate WSCF performance on long-term dose-response estimation against the baselines on the synthetic cohort with $\beta_U \in \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$, $\delta = 4$, and two semi-synthetic cohorts. Table1 shows the results. Adding a surrogate helps in both static (A-Y vs. A,S-Y) and time-varying (RMSN vs. longitudinal-S) regimes; sequential models without surrogates fail to beat static robust methods, whereas models with longitudinal surrogates improve substantially. DRNet outperforming naive regression further motivates causal debiasing, and under time-varying confounding our family is the most robust, particularly against surrogate-based methods that ignore it (longitudinal-S) or address only the baseline (LEARN). Within WSCF family, the inclusion of debiasing modules improves as β_U increases, and from $\beta_U = 2.0$, WSCF outperforms all variants and baselines, same for both semi-synthetic cohorts, indicating the improvement of WSCF. More results and running time comparisons are shown in the Appendix.

Module Contribution (RQ2). We ablate WSCF on the synthetic cohort ($\beta_U = 3.5$, $\delta = 4$) over 2^3 switch combinations of the OT, IPM, JSR modules and four surrogate representations (window average, the nearest, no surrogate, and longitudinal version). Table 1 and 3 show that removing any module leads to degraded performance, as does any naive surrogate representation; Figure 2 shows that JSR controls rollout error as more predicted surrogates enter with growing depth T ($T = \delta + 1$ by default). Together confirms the necessity of every module and of longitudinal surrogates.

Table 1: Overall performance evaluation on the synthetic cohort over increasing β_U and two semi-synthetic cohorts. The best is bolded and the second best is underlined except the oracle.

Method	Synthetic						MIMIC-III	Industrial
MISE(SD)	$\beta_U = 1.0$	$\beta_U = 1.5$	$\beta_U = 2.0$	$\beta_U = 2.5$	$\beta_U = 3.0$	$\beta_U = 3.5$		
A-Y	6.27(1.188)	6.36(1.207)	6.38(1.217)	6.40(1.225)	6.41(1.238)	6.43(1.252)	0.985(0.024)	0.928(0.004)
A,S-Y	4.12(0.089)	4.07(0.033)	3.81(0.141)	3.81(0.127)	3.70(0.130)	3.64(0.111)	2.77(0.203)	0.806(0.007)
VCNet	3.62(0.283)	3.98(0.401)	4.27(0.188)	4.36(0.352)	4.34(0.028)	4.54(0.221)	7.65(2.094)	0.489(0.016)
DRNet	2.44(0.124)	2.46(0.150)	2.39(0.101)	2.40(0.105)	2.36(0.105)	2.41(0.082)	1.50(0.066)	0.463(0.006)
RMSN	3.38(0.061)	3.39(0.141)	3.29(0.080)	3.26(0.156)	3.21(0.100)	3.08(0.116)	2.92(0.175)	0.705(0.013)
DCRN	3.31(0.036)	3.40(0.221)	3.39(0.197)	3.15(0.170)	3.10(0.204)	3.08(0.255)	3.00(0.201)	0.706(0.017)
G-Net	3.70(0.288)	3.59(0.353)	3.55(0.301)	3.47(0.207)	3.17(0.295)	3.13(0.241)	2.88(0.214)	0.698(0.022)
CRN	3.43(0.084)	3.55(0.251)	3.45(0.327)	3.25(0.231)	3.21(0.282)	3.18(0.223)	2.88(0.139)	0.705(0.018)
LIRL	0.574(0.048)	0.563(0.090)	0.564(0.049)	0.568(0.077)	0.543(0.089)	0.535(0.041)	0.553(0.077)	0.655(0.008)
LEARN	0.567(0.042)	0.589(0.081)	0.573(0.055)	0.580(0.052)	0.573(0.075)	0.554(0.036)	0.585(0.066)	0.653(0.007)
LTEE	0.539(0.053)	0.540(0.058)	0.533(0.051)	0.519(0.036)	0.537(0.056)	0.528(0.032)	0.541(0.070)	0.657(0.008)
longitudinal-S	0.533(0.039)	0.543(0.039)	0.562(0.051)	0.569(0.055)	0.582(0.062)	0.602(0.060)	0.511(0.047)	0.658(0.006)
WSCF-None	**0.506(0.029)**	0.550(0.043)	0.617(0.044)	0.699(0.072)	0.688(0.052)	0.719(0.052)	0.546(0.045)	0.463(0.021)
WSCF-noOT	0.513(0.055)	**0.503(0.041)**	0.537(0.085)	0.594(0.127)	0.575(0.081)	0.617(0.136)	0.475(0.084)	0.466(0.027)
WSCF-noJSR	0.555(0.042)	0.544(0.030)	0.558(0.042)	0.546(0.039)	0.529(0.028)	0.517(0.024)	0.499(0.025)	0.461(0.056)
WSCF-noIPM	0.527(0.065)	0.528(0.020)	0.543(0.051)	0.537(0.038)	0.517(0.020)	0.506(0.048)	0.514(0.026)	0.468(0.028)
WSCF-LSTM	0.522(0.026)	0.517(0.039)	0.528(0.037)	0.519(0.031)	0.528(0.024)	0.510(0.036)	0.475(0.027)	0.484(0.032)
WSCF (Ours)	0.528(0.022)	0.506(0.042)	**0.521(0.052)**	**0.496(0.028)**	**0.500(0.028)**	**0.504(0.027)**	**0.470(0.025)**	**0.433(0.025)**
WSCF-oracle	0.262(0.027)	0.254(0.026)	0.253(0.025)	0.253(0.030)	0.245(0.024)	0.244(0.026)	0.310(0.013)	0.172(0.004)

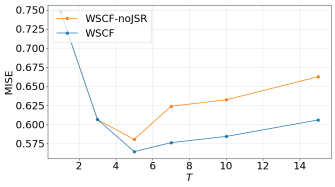
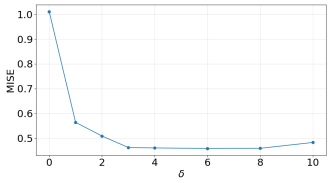
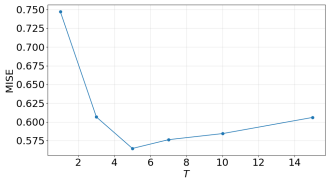


Figure 2: JSR module for rollout stability.



(a) Surrogate window



(b) Prediction Depth

Figure 3: Robustness Analysis of Temporal Structure.

Robustness to the Window and the depth (RQ3). We check robustness to the temporal structural of surrogate window depth δ and prediction depth T on synthetic datasets ($\beta_U = 3.5, \delta = 4$). In Figure2, accuracy improves with δ up to a point then levels off, so finite history suffices for robust prediction; it improves as T increases to a point then deteriorates, suggesting a roll-out error issue.

Hyperparameter Sensitivity (RQ4). Varying the three loss weights one at a time on the synthetic cohort ($\beta_U = 3.5$), with the other two at their defaults (Appendix Figure5), each weight has an interior optimum, suggesting the trade-offs of modules in the objective.

7 Conclusion

In this paper, we investigate long-term dose-path-response estimation with time-varying continuous treatments, longitudinal surrogates, and time-dependent confounding, a dynamic data-fusion problem largely overlooked by existing work. To identify the target, we introduce a window-based sequential g-formula under a mean-exchangeability condition weaker than conventional transportability and unconfoundedness assumptions. For latent confounding, we learn window-specific weights by aligning transition distributions between observational and experimental data through optimal transport; for observed confounding, we learn balanced history representations that reduce dependence on current treatment. We further propose Jacobian Spectral Regularization to prevent surrogate-prediction errors from accumulating exponentially during sequential rollout. Theoretically, we establish conditional alignment guarantees for the learned weights and derive an end-to-end generalization bound for counterfactual prediction. Building on these results, we develop WSCF to estimate long-term responses from finite surrogate windows. Experiments on synthetic and semi-synthetic healthcare and industrial cohorts demonstrate the effectiveness and stability of WSCF, especially under strong latent confounding and long prediction horizons.

AI USE STATEMENT

In this work, we used the generative-AI assistant DeepSeek-V4.1-Flash and ChatGPT-6 for editing and polishing parts of the manuscript, formatting tables and references, and drafting visualization scripts for experimental figures. We have not used generative-AI tools for problem formulation, theoretical analysis, experimental design and implementation, or interpretation of results. All AI-assisted text and artifacts were reviewed, verified, and revised by the authors. We take responsibility for the final content of this work, including any text or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This study employs a synthetic cohort and two semi-synthetic cohorts built from pre-existing sources: MIMIC-III, a publicly-available de-identified critical-care database, plus an anonymized order-level industrial-platform panel. No new human-subject data collection or human-subject interventions are conducted. Learned estimates inherit selection biases from source datasets, which must be accounted for when applying this method to other contexts. These outputs are not a validated clinical tool; real-world deployment for treatment or targeting decisions introduces fairness and manipulation concerns outside the scope of this work. Some of the authors received research funding from industry.

REPRODUCIBILITY STATEMENT

Sections 4 and 5 describe the framework and the network, and the AppendixC states the assumptions, theorems, and the results that the main text abbreviates, along with complete proofs; Appendix C.3 also gives the synthetic data-generation parameters, semi-synthetic data processing details, the cohort filters and splits, the training settings, and the implementation details. We will release the code and the trained heads upon publication.

REFERENCES

Susan Athey, Raj Chetty, and Guido Imbens. Combining experimental and observational data to estimate treatment effects on long term outcomes. *arXiv preprint arXiv:2006.09676*, 2020.

Susan Athey, Raj Chetty, Guido W Imbens, and Hyunseung Kang. The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *Review of Economic Studies*, 93(4):2284–2312, 2026.

Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

Keith Battocchi, Eleanor Dillon, Maggie Hei, Greg Lewis, Paul Oka, Miruna Oprescu, and Vasilis Syrgkanis. Estimating the long-term effects of novel treatments. In *Advances in Neural Information Processing Systems*, 2021.

Jeroen Berrevoets, Alicia Curth, Ioana Bica, Eoin McKinney, and Mihaela van der Schaar. Disentangled counterfactual recurrent networks for treatment effect inference over time. *arXiv preprint arXiv:2112.03811*, 2021.

Ioana Bica, Ahmed Alaa, and Mihaela Van Der Schaar. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International conference on machine learning*, 2020a.

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*, 2020b.

Ioana Bica, James Jordon, and Mihaela van der Schaar. Estimating the effects of continuous-valued interventions using generative adversarial networks. In *Advances in Neural Information Processing Systems*, 2020c.

Ruichu Cai, Weilin Chen, Zeqin Yang, Shu Wan, Chen Zheng, Xiaoqing Yang, and Jiecheng Guo. Long-term causal effects estimation via latent surrogates representation learning. *Neural Networks*, 176:106336, 2024.

Ruichu Cai, Junjie Wan, Weilin Chen, Zeqin Yang, Zijian Li, Peng Zhen, and Jiecheng Guo. Long-term individual causal effect estimation via identifiable latent representation learning. In *International Joint Conference on Artificial Intelligence*, 2025.

Jiafeng Chen and David M. Ritzwoller. Semiparametric estimation of long-term treatment effects. *Journal of Econometrics*, 237(2):105545, 2023.

Weilin Chen, Ruichu Cai, Yuguang Yan, Zhifeng Hao, and José Miguel Hernández-Lobato. Long-term causal inference via modeling sequential latent confounding. *arXiv preprint arXiv:2502.18994*, 2025.

Lu Cheng, Ruocheng Guo, and Huan Liu. Long-term effect estimation with surrogate representation. In *ACM International Conference on Web Search and Data Mining*, 2021.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.

Issa J Dahabreh, Sarah E Robertson, Jon A Steingrimsson, Elizabeth A Stuart, and Miguel A Hernan. Extending inferences from a randomized trial to a new target population. *Statistics in Medicine*, 39(14):1999–2014, 2020.

Constantine E. Frangakis and Donald B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

Changran Geng, Harald Paganetti, and Clemens Grassberger. Prediction of treatment response for combined chemo- and radiation therapy for non-small cell lung cancer patients using a biomathematical model. *Scientific Reports*, 7(1):13542, 2017.

AmirEmad Ghassami, Chang Liu, Alan Yang, David Richardson, Ilya Shpitser, and Eric Tchetgen Tchetgen. Combining experimental and observational data for identification and estimation of long-term causal effects. *arXiv preprint arXiv:2201.10743*, 2022.

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

Miguel A. Hernán, Babette Brumback, and James M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5):561–570, 2000.

Wenjie Hu, Xiaohua Zhou, and Peng Wu. Identification and estimation of treatment effects on long-term outcomes in clinical trials with external observational data. *Statistica Sinica*, 2022.

Shan Huang, Chen Wang, Yuan Yuan, Jinglong Zhao, and Brocco Zhang. Estimating effects of long-term treatments. *Management Science*, 2026.

Guido Imbens, Nathan Kallus, Xiaojie Mao, and Yuhao Wang. Long-term causal inference under persistent confounding via data combination. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(2):362–388, 2025.

Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. When to trust your model: Model-based policy optimization. In *Advances in Neural Information Processing Systems*, 2019.

Fredrik Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *International Conference on Machine Learning*, 2016.

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*, 3:160035, 2016.

Nathan Kallus and Xiaojie Mao. On the role of surrogates in the efficient estimation of treatment effects with limited outcome data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, pp. 480–509, 2025.

Amirreza Kazemi and Martin Ester. Adversarially balanced representation for continuous treatment effect estimation. In *AAAI Conference on Artificial Intelligence*, 2024.

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*, 2021.

Bryan Lim, Ahmed Alaa, and Mihaela van der Schaar. Forecasting treatment responses over time using recurrent marginal structural networks. In *Advances in Neural Information Processing Systems*, 2018.

Ruoqi Liu, Changchang Yin, and Ping Zhang. Estimating individual treatment effects with time-varying confounders. In *IEEE International Conference on Data Mining*, 2020.

Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In *International Conference on Machine Learning*, 2022.

Raha Moraffah, Paras Sheth, Mansooreh Karami, Anchit Bhattacharya, Qianru Wang, Anique Tahir, Adrienne Raglin, and Huan Liu. Causal inference for time series analysis: Problems, methods and evaluation. *Knowledge and Information Systems*, 63(12):3041–3085, 2021.

Lizhen Nie, Mao Ye, Qiang Liu, and Dan Nicolae. VCNet and functional targeted regularization for learning causal effects of continuous treatments. In *International Conference on Learning Representations*, 2021.

Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.

Ross L. Prentice. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8(4):431–440, 1989.

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. Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality*, pp. 69–117. Springer, 1997.

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

Stéphane Ross, Geoffrey J. Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *International Conference on Artificial Intelligence and Statistics*, 2011.

Michael Schomaker, Helen McIlleron, Paolo Denti, and Iván Díaz. Causal inference for continuous multiple time point interventions. *Statistics in Medicine*, 43(28):5380–5400, 2024.

Nabeel Seedat, Fergus Imrie, Alexis Bellot, Zhaozhi Qian, and Mihaela van der Schaar. Continuous-time modeling of counterfactual outcomes using neural controlled differential equations. In *International Conference on Machine Learning*, 2022.

Hossein Soleimani, Adarsh Subbaswamy, and Suchi Saria. Treatment-response models for counterfactual reasoning with continuous-time, continuous-valued interventions. In *Conference on Uncertainty in Artificial Intelligence*, 2017.

Bharath K. Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, and Gert R. G. Lanckriet. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6:1550–1599, 2012.

Allen Tran, Aurélien Bibaut, and Nathan Kallus. Inferring the long-term causal effects of long-term treatments from short-term experiments. In *International Conference on Machine Learning*, 2024.

Cédric Villani et al. *Optimal Transport: Old and New*, volume 338. Springer, 2009.

Xin Wang, Shengfei Lyu, Xingyu Wu, Tianhao Wu, and Huanhuan Chen. Generalization bounds for estimating causal effects of continuous treatments. In *Advances in Neural Information Processing Systems*, 2022.

Yingrong Wang, Anpeng Wu, Baohong Li, Ziyang Xiao, Ruoxuan Xiong, Qing Han, and Kun Kuang. Sequential treatment effect estimation with unmeasured confounders. *arXiv preprint arXiv:2505.09113*, 2025.

Peng Wu, Ziyu Shen, Feng Xie, Zhongyao Wang, Chunchen Liu, and Yan Zeng. Policy learning for balancing short-term and long-term rewards. In *International Conference on Machine Learning*, 2024.

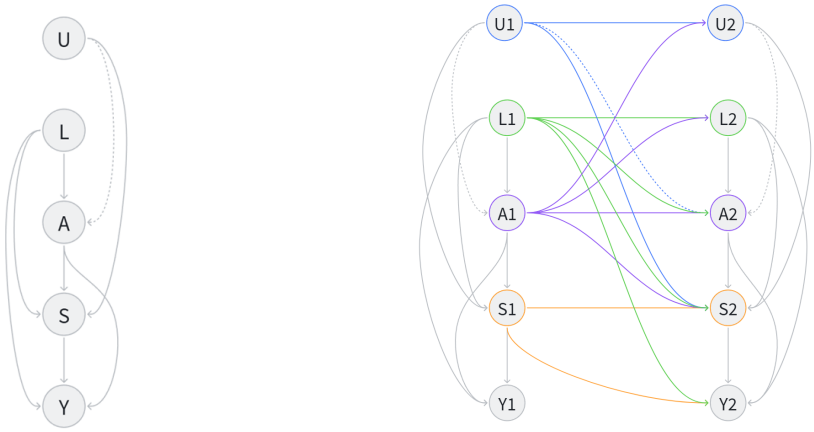
Yuguang Yan, Zhihao Yang, Wei Chen, Ruichu Cai, Zhifeng Hao, and Michael Kwok-Po Ng. Exploiting geometry for treatment effect estimation via optimal transport. In *AAAI Conference on Artificial Intelligence*, 2024.

Qinwei Yang, Xueqing Liu, Yan Zeng, Ruocheng Guo, Yang Liu, and Peng Wu. Learning the optimal policy for balancing short-term and long-term rewards. In *Advances in Neural Information Processing Systems*, 2024a.

Zeqin Yang, Weilin Chen, Ruichu Cai, Yuguang Yan, Zhifeng Hao, Zhipeng Yu, Zhichao Zou, Peng Zhen, and Jiecheng Guo. Estimating long-term heterogeneous dose-response curve: Generalization bound leveraging optimal transport weights. *arXiv preprint arXiv:2406.19195*, 2024b.

Yao Zhang, Alexis Bellot, and Mihaela Schaar. Learning overlapping representations for the estimation of individualized treatment effects. In *International Conference on Artificial Intelligence and Statistics*, 2020.

A Causal Diagrams



(a) Single-time-point DAG

(b) Longitudinal time-varying DAG

Figure 4: Causal diagrams for surrogacy under static and time-varying treatment. (a) In the static setting, or within each time step of the longitudinal setting, the skeleton $L \to A \to S \to Y$ holds and the dashed arrow is the confounding edge that randomization controls in the experimental study but not in the observational one; (b) for each transition, with grey edges for the relationships within a period and coloured intertemporal edges for history propagation, blue from U, green from L, purple from A and orange from S across successive periods.

B Statement of Theoretical Results

This appendix states the results of Sections 4.1 to 4.3 that the main text does not give in full and derives all of them, the weights $w_t^\star$, $w_t^\sharp$, $\hat{w}_t$ and the laws $P_{g,t}$ being as in Sections 3.2 and 4.2; a symbol not defined there is defined where it first appears here.

B.1 Distributional Conditions and the Mean-Level Condition

Assumption 7 (Conditional external validity)**.** *For all t, $G \perp \{Y_t(\bar{a}_t), S_t(\bar{a}_t)\} \mid \bar{A}_{t-1}, \bar{L}_t$.*

Assumption 8 (Latent unconfoundedness)**.** *For all t, $Y_t(\bar{a}_t) \perp A_t \mid \bar{S}_t(\bar{a}_t), \bar{A}_{t-1}, \bar{L}_t, G = 0$.*

The first is the conditional form of the comparability requirement underlying surrogate-based transport (Hu et al., 2022); the second removes, in the observational study alone, the residual association between the current dose and the long-term outcome once the whole surrogate path is conditioned on. Neither is invoked directly: they enter through the proposition below.

Proposition 1 (Mean exchangeability from the distributional conditions)**.** *Under A3, A4 is implied by Assumptions 7 and 8.*

Proof of Proposition 1. The chain runs through the intermediate mean that drops the current dose in *both* studies. In the experimental study A3 makes $\{Y_t(\bar{a}_t), S_t(\bar{a}_t)\}$ independent of A_t given $(\bar{A}_{t-1}, \bar{L}_t, G = 1)$, so weak union enlarges the conditioning set to contain the potential surrogate path and $A_t = a_t$ may be removed,

$$\begin{aligned}&\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t) = \bar{s}_t,\ \bar{A}_t = \bar{a}_t,\ \bar{L}_t = \bar{l}_t,\ G = 1\big]\\ &\quad = \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t) = \bar{s}_t,\ \bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t,\ G = 1\big].\end{aligned} \tag{6}$$

A7 gives the same independence across studies, so decomposition to $Y_t(\bar{a}_t)$ followed by weak union to the potential window moves the statement to $G = 0$,

$$\begin{aligned}&\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t) = \bar{s}_t,\ \bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t,\ G = 1\big]\\ &\quad = \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t) = \bar{s}_t,\ \bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t,\ G = 0\big],\end{aligned} \tag{7}$$

and A8 restores the current dose on the observational side,

$$\begin{aligned}&\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t) = \bar{s}_t,\ \bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t,\ G = 0\big]\\ &\quad = \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t) = \bar{s}_t,\ \bar{A}_t = \bar{a}_t,\ \bar{L}_t = \bar{l}_t,\ G = 0\big].\end{aligned} \tag{8}$$

The left-hand side of the first display is the right-hand side of the third, and the intermediate mean is the same object in both, which is A4. Every conditioning event has positive probability by A2(i), and A5 takes no part in the chain. □

B.2 The Identification Layer

Lemma 1 is stated in Section 4.1; its hypothesis supplies only the joint alignment of χ_t, the marginal identity, the kernel and the conditional mean being derived here, the first being what Theorems 1 and 2 integrate out and the second the innermost regression of their recursions.

Proof of Lemma 1. Abbreviate $\bar{s}_{(t-\delta):(t-1)}$ by $\bar{s}$ and $(\bar{A}_{t-1} = \bar{a}_{t-1}, \bar{L}_t = \bar{l}_t)$ by $\bar{h}_t$; write $w := w_t^\star$ for the weight of Eq.2 and let P^w be the reweighted observational law of χ_t. Only the joint alignment $w_t^\star \, \mathrm{d}P(\chi_t \mid G = 0) = \mathrm{d}P(\chi_t \mid G = 1)$ written below Eq.2 is given.

(i) The kernel. Adding the current dose to the conditioning set of the left-hand side leaves the law unchanged by A3, with A2(i) keeping the event of positive probability; A1 replaces the potential outcome and history by their observed counterparts; and A5, applied period by period, truncates the path to the window,

$$\begin{aligned}&P\big(S_t(\bar{a}_t) \mid \bar{S}_{t-1}(\bar{a}_{t-1}) = \bar{s}_{t-1},\ \bar{A}_{t-1},\ \bar{L}_t,\ G = 1\big)\\ &\quad \overset{\text{A3, 2(i)}}{=} P\big(S_t(\bar{a}_t) \mid \bar{S}_{t-1}(\bar{a}_{t-1}) = \bar{s}_{t-1},\ \bar{a}_t,\ \bar{l}_t,\ G = 1\big)\\ &\quad \overset{\text{A1}}{=} P\big(S_t \mid \bar{S}_{t-1} = \bar{s}_{t-1},\ \bar{a}_t,\ \bar{l}_t,\ G = 1\big)\\ &\quad \overset{\text{A5}}{=} P\big(S_t \mid \bar{s},\ \bar{a}_t,\ \bar{l}_t,\ G = 1\big).\end{aligned} \tag{9}$$

Integrating that identity over s_t gives the marginal identity

$$P^w\big(\bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=0\big) = P\big(\bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=1\big), \tag{10}$$

so after reweighting the two studies also agree on the law of the window, the dose and the covariates, the left-hand side being positive by A2(i). Writing the $G=1$ conditional probability as a ratio of joint probabilities, replacing numerator and denominator by the reweighted measure through the joint alignment, and using Eq.10 for the denominator gives the kernel,

$$\begin{aligned} P\big(S_t = s_t \mid \bar{s},\ \bar{a}_t,\ \bar{l}_t,\ G=1\big) &\overset{\text{def.}}{=} \frac{P\big(S_t = s_t,\ \bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=1\big)}{\displaystyle\int_{s_t} P\big(S_t = s_t,\ \bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=1\big)\,\mathrm{d}s_t} \\ &\overset{2}{=} \frac{P^w\big(S_t = s_t,\ \bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=0\big)}{\displaystyle\int_{s_t} P^w\big(S_t = s_t,\ \bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=0\big)\,\mathrm{d}s_t} \\ &= \frac{P^w\big(S_t = s_t,\ \bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=0\big)}{P^w\big(\bar{s},\ \bar{a}_t,\ \bar{l}_t \mid G=0\big)} \\ &\overset{\text{def.}}{=} P^w\big(S_t = s_t \mid \bar{s},\ \bar{a}_t,\ \bar{l}_t,\ G=0\big), \end{aligned} \tag{11}$$

the second line replacing numerator and denominator alike through the joint alignment, and the third being the step from the joint to the marginal, in which the conditioned variable s_t is integrated out of numerator and denominator alike, so that the denominator is the left-hand side of Eq.10 and the whole chain is an identity rather than a bound.

(ii) The conditional mean. A6 makes $S_t(\bar{a}_t)$ independent of $\{A_{t:}, L_{t+1:\cdot}\}$ jointly, so weak union moves $A_t = a_t$ into the conditioning set; A1 replaces the potential outcome and window by the observed ones; and integrating the kernel of (i) over s_t returns the conditional mean under P^w,

$$\begin{aligned} &\mathbb{E}\big[S_t(\bar{a}_t) \mid \bar{S}_{(t-\delta):(t-1)}(\bar{a}_{(t-\delta):(t-1)}) = \bar{s},\ \bar{A}_{t-1},\ \bar{L}_t,\ G=1\big] \\ &\quad\overset{\text{A6, 2(i)}}{=} \mathbb{E}\big[S_t(\bar{a}_t) \mid \bar{S}_{(t-\delta):(t-1)}(\bar{a}_{(t-\delta):(t-1)}) = \bar{s},\ \bar{a}_t,\ \bar{l}_t,\ G=1\big] \\ &\quad\overset{\text{A1}}{=} \mathbb{E}\big[S_t \mid \bar{s},\ \bar{a}_t,\ \bar{l}_t,\ G=1\big] \\ &\quad\overset{\text{(i), 2}(i)}{=} \mathbb{E}^w\big[S_t \mid \bar{s},\ \bar{a}_t,\ \bar{l}_t,\ G=0\big], \end{aligned} \tag{12}$$

A2(i) making the weighted conditional law well defined. The recursions below repeat the same reduction at every layer. □

Theorem 1 (Identification of surrogate)**.** *Under Lemma 1 and Assumptions 1–6, $\mu_t^S(\bar{a}_t \mid \bar{h}_t)$ equals the nested expectation*

$$\mathbb{E}^w_{S_1|\cdot}\Big[\cdots \mathbb{E}^w_{S_{t-1}|\cdot}\Big[\,\mathbb{E}^w\big[S_t \mid \bar{S}_{(t-\delta):(t-1)},\ \bar{a}_t,\ \bar{l}_t,\ G=0\big]\,\Big]\cdots\Big],$$

where $\mathbb{E}^w_{S_k|\cdot}$ integrates S_k against the reweighted window kernel $P^w(S_k \mid \bar{S}_{(k-\delta):(k-1)}, \bar{A}_k, \bar{L}_k, G=0)$.

The window provides the truncation, Lemma 1 the transport, and A6 the dimension reduction that makes each factor a window kernel.

Proof of Theorem 1. Every assumption holds at every period, so no compatibility condition across periods is required. Write $\mu_t^{\mathrm{S}}(\bar{a}_t \mid \bar{h}_t) := \mathbb{E}[S_t(\bar{a}_t) \mid \bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t,\ G=1]$, and abbreviate the conditioning pair $(\bar{A}_{t-1} = \bar{a}_{t-1},\ \bar{L}_t = \bar{l}_t)$ by $\bar{h}_t$.

Step 1 (law of total expectation). The potential trajectory is introduced and integrated out,

$$\begin{aligned} \mu_t^{\mathrm{S}}(\bar{a}_t \mid \bar{h}_t) = \int_{\bar{s}_{t-1}} &\mathbb{E}\big[S_t(\bar{a}_t) \mid \bar{S}_{t-1}(\bar{a}_{t-1}) = \bar{s}_{t-1},\ \bar{h}_t,\ G=1\big] \\ &\times P\big(\bar{S}_{t-1}(\bar{a}_{t-1}) = \bar{s}_{t-1} \mid \bar{h}_t,\ G=1\big)\,\mathrm{d}\bar{s}_{t-1}. \end{aligned} \tag{13}$$

Step 2 (only the inner factor moves). The inner expectation is reduced by A6, A2(i) and A5, and the potential-outcome notation is then dropped by A1,

$$\begin{aligned}
&\overset{\text{A6, 2(i), 5}}{=} \int_{\bar{s}_{t-1}} \mathbb{E}\big[S_t(\bar{a}_t) \mid \bar{S}_{(t-\delta):(t-1)}(\bar{a}_{(t-\delta):(t-1)}) = \bar{s}_{(t-\delta):(t-1)}, \\
&\qquad \bar{A}_t = \bar{a}_t,\ \bar{L}_t = \bar{l}_t,\ G = 1\big]\, P\big(\bar{S}_{t-1}(\bar{a}_{t-1}) = \bar{s}_{t-1} \mid \bar{h}_t,\ G = 1\big)\, \mathrm{d}\bar{s}_{t-1} \\
&\overset{\text{A1}}{=} \int_{\bar{s}_{t-1}} \mathbb{E}\big[S_t \mid \bar{s}_{(t-\delta):(t-1)},\ \bar{a}_t,\ \bar{l}_t,\ G = 1\big]\, P\big(\bar{S}_{t-1}(\bar{a}_{t-1}) = \bar{s}_{t-1} \mid \bar{h}_t,\ G = 1\big)\, \mathrm{d}\bar{s}_{t-1}.
\end{aligned} \tag{14}$$

Step 3 (chain rule, the conditioning set held fixed).

$$\begin{aligned}
&\overset{\text{chain rule}}{=} \int_{\bar{s}_{t-1}} \mathbb{E}\big[S_t \mid \bar{s}_{(t-\delta):(t-1)},\ \bar{a}_t,\ \bar{l}_t,\ G = 1\big] \\
&\qquad \times \prod_{k=1}^{t-1} P\big(S_k(\bar{a}_k) = s_k \mid \bar{S}_{k-1}(\bar{a}_{k-1}) = \bar{s}_{k-1},\ \bar{h}_t,\ G = 1\big)\, \mathrm{d}\bar{s}_{t-1}.
\end{aligned} \tag{15}$$

Step 4 (dimension reduction, by A6 alone). Each factor of the product carries the surplus future, $\bar{A}_{k+1:t-1}$ and $\bar{L}_{k+1:t}$, so it is not yet a window kernel. Reading A6 at $t = k$,

$$S_k(\bar{a}_k) \ \perp\ \big\{A_{k:\cdot},\ L_{k+1:\cdot}\big\} \ \mid\ \bar{S}_{k-1}(\bar{a}_{k-1}),\ \bar{A}_{k-1},\ \bar{L}_k,\ G = 1, \tag{16}$$

weak union moves A_k into the conditioning set, since A_k lies among the independent variables $A_{k:\cdot}$,

$$S_k(\bar{a}_k) \ \perp\ \big\{A_{k+1:\cdot},\ L_{k+1:\cdot}\big\} \ \mid\ \bar{S}_{k-1}(\bar{a}_{k-1}),\ \bar{A}_k,\ \bar{L}_k,\ G = 1, \tag{17}$$

so that the conditioning set is exactly the target one, and decomposition then drops the surplus $\bar{A}_{k+1:t-1}$ and $\bar{L}_{k+1:t}$, both of which lie in $\{A_{k+1:\cdot}, L_{k+1:\cdot}\}$. Factor by factor,

$$\begin{aligned}
&\prod_{k=1}^{t-1} P\big(S_k \mid \bar{S}_{k-1},\ \bar{h}_t,\ G = 1\big) \ \overset{\text{A6}}{=} \\
&\qquad \prod_{k=1}^{t-1} P\big(S_k(\bar{a}_k) = s_k \mid \bar{S}_{k-1}(\bar{a}_{k-1}) = \bar{s}_{k-1},\ \bar{A}_k = \bar{a}_k,\ \bar{L}_k,\ G = 1\big),
\end{aligned} \tag{18}$$

the current dose being kept rather than reduced because the conditional expectation of Lemma 1 needs $\bar{A}_k = \bar{a}_k$.

Step 5 (per-factor truncation and transport). Each factor is truncated to the window, stripped of the potential-outcome notation and transported across studies, factor by factor since what is deleted is $S_{k-\delta}$ while the target is S_t,

$$\begin{aligned}
&P\big(S_k(\bar{a}_k) = s_k \mid \bar{S}_{k-1} = \bar{s}_{k-1},\ \bar{A}_k = \bar{a}_k,\ \bar{L}_k = \bar{l}_k,\ G = 1\big) \\
&\quad \overset{\text{A5}}{=} P\big(S_k(\bar{a}_k) = s_k \mid \bar{S}_{(k-\delta):(k-1)} = \bar{s}_{(k-\delta):(k-1)},\ \bar{A}_k = \bar{a}_k,\ \bar{L}_k = \bar{l}_k,\ G = 1\big) \\
&\quad \overset{\text{A1}}{=} P\big(S_k = s_k \mid \bar{s}_{(k-\delta):(k-1)},\ \bar{a}_k,\ \bar{l}_k,\ G = 1\big) \ \overset{\text{Lemma 1, } 2(i),\ Eq.2}{=} \\
&\quad P^w\big(S_k = s_k \mid \bar{s}_{(k-\delta):(k-1)},\ \bar{a}_k,\ \bar{l}_k,\ G = 0\big).
\end{aligned} \tag{19}$$

Step 6 (inner transport). The innermost factor is an output regression in S_t, the target rather than an integration variable and therefore never integrated out,

$$\begin{aligned}
&\mathbb{E}\big[S_t \mid \bar{S}_{(t-\delta):(t-1)} = \bar{s}_{(t-\delta):(t-1)}, \bar{A}_t = \bar{a}_t, \bar{L}_t = \bar{l}_t, G = 1\big] \\
&\qquad \overset{\text{Lemma 1, } 2(i),\ Eq.2}{=} \mathbb{E}^w\big[S_t \mid \bar{s}_{(t-\delta):(t-1)}, \bar{a}_t, \bar{l}_t, G = 0\big].
\end{aligned} \tag{20}$$

Step 7 (assembly). Substituting Steps 5 and 6 into Step 3 gives the nested expectation of the theorem,

$$\begin{aligned}
\mu_t^{\mathrm{S}}(\bar{a}_t \mid \bar{h}_t) = &\int_{\bar{s}_{t-1}} E^w\big[S_t \mid \bar{S}_{(t-\delta):(t-1)} = \bar{s}_{(t-\delta):(t-1)},\ \bar{a}_t,\ \bar{l}_t,\ G = 0\big] \\
&\times \prod_{k=1}^{t-1} P^w\big(S_k = s_k \mid \bar{S}_{(k-\delta):(k-1)} = \bar{s}_{(k-\delta):(k-1)},\ \bar{a}_k,\ \bar{l}_k,\ G = 0\big)\, \mathrm{d}\bar{s}_{t-1},
\end{aligned} \tag{21}$$

an iteration of conditional expectations with $t - 1$ layers. For $k \le \delta$ the window degenerates to $\bar{S}_{1:(k-1)}$, and at $k = 1$ the outermost layer carries no window argument. □

Lemma 2 is stated in Section 4.1. The outcome mechanism is transported by A4 and carries no reweighting, which is why the outer layers of Theorem 2 carry w and its innermost regression does not.

Proof of Lemma 2. The window has length $\delta + 1$ and contains the current surrogate, whereas A4 is stated over the full path $\bar{S}_t(\bar{a}_t)$, so the proof makes one excursion, window to full path and back,

$$\begin{aligned}
&\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{s}_{(t-\delta):t}(\bar{a}_{(t-\delta):t}),\ \bar{A}_{t-1},\ \bar{L}_t,\ G=1\big] \\
&\quad \overset{\text{A3, 2(i)}}{=} \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{s}_{(t-\delta):t}(\bar{a}_{(t-\delta):t}),\ \bar{A}_t=\bar{a}_t,\ \bar{L}_t=\bar{l}_t,\ G=1\big] \\
&\quad \overset{\text{A5}}{=} \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t)=\bar{s}_t,\ \bar{A}_t=\bar{a}_t,\ \bar{L}_t=\bar{l}_t,\ G=1\big] \\
&\quad \overset{\text{A4}}{=} \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t)=\bar{s}_t,\ \bar{A}_t=\bar{a}_t,\ \bar{L}_t=\bar{l}_t,\ G=0\big] \\
&\quad \overset{\text{A1}}{=} \mathbb{E}\big[Y_t \mid \bar{S}_t=\bar{s}_t,\ \bar{A}_t=\bar{a}_t,\ \bar{L}_t=\bar{l}_t,\ G=0\big] \\
&\quad \overset{\text{A5, 2(i)}}{=} \mathbb{E}\big[Y_t \mid \bar{s}_{(t-\delta):t},\ \bar{a}_t,\ \bar{l}_t,\ G=0\big].
\end{aligned} \tag{22}$$

It is the Y side of A3 that enlarges the conditioning set, not A6, whose object is S_t rather than Y_t; A4 is then applied verbatim, both sides carrying $Y_t(\bar{a}_t)$ and $\bar{S}_t(\bar{a}_t)$ so that the conditioning sets match exactly, which is why A1 must follow it rather than precede it; and the last step truncates back to the window, A5 holding in either study. No weight appears anywhere in the chain: what is transported here is the outcome mechanism, not the state-trajectory distribution. □

Theorem 2 is stated in Section 4.1.

Proof of Theorem 2. Write $\mu_t(\bar{a}_t \mid \bar{h}_t) := \mathbb{E}[Y_t(\bar{a}_t) \mid \bar{h}_t,\ G=1]$ with $\bar{h}_t := (\bar{A}_{t-1}=\bar{a}_{t-1},\ \bar{L}_t=\bar{l}_t)$, one layer deeper than the surrogate of Theorem 1.

Step 1 (law of total expectation).

$$\begin{aligned}
\mu_t(\bar{a}_t \mid \bar{h}_t) = \int_{\bar{s}_t} &\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t)=\bar{s}_t,\ \bar{h}_t,\ G=1\big] \\
&\times P\big(\bar{S}_t(\bar{a}_t)=\bar{s}_t \mid \bar{h}_t,\ G=1\big)\,\mathrm{d}\bar{s}_t.
\end{aligned} \tag{23}$$

Step 2 (inner factor, handed over in potential-outcome form). The inner factor is taken to the window and handed to Lemma 2 *before* A1 is applied, since the object still carries $Y_t(\bar{a}_t)$ and the potential window,

$$\begin{aligned}
&\mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t)=\bar{s}_t,\ \bar{h}_t,\ G=1\big] \\
&\quad \overset{\text{A3, 2(i)}}{=} \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{S}_t(\bar{a}_t)=\bar{s}_t,\ \bar{A}_t=\bar{a}_t,\ \bar{h}_t,\ G=1\big] \\
&\quad \overset{\text{A5}}{=} \mathbb{E}\big[Y_t(\bar{a}_t) \mid \bar{s}_{(t-\delta):t}(\bar{a}_{(t-\delta):t}),\ \bar{A}_t=\bar{a}_t,\ \bar{h}_t,\ G=1\big] \\
&\quad \overset{\text{Lemma 2}}{=} \mathbb{E}\big[Y_t \mid \bar{s}_{(t-\delta):t},\ \bar{a}_t,\ \bar{l}_t,\ G=0\big].
\end{aligned} \tag{24}$$

Step 3 (trajectory density). The chain rule factorises the density with its conditioning set held fixed,

$$\begin{aligned}
&P\big(\bar{S}_t(\bar{a}_t)=\bar{s}_t \mid \bar{h}_t,\ G=1\big) \\
&\quad \overset{\text{chain rule}}{=} \prod_{k=1}^{t} P\big(S_k(\bar{a}_k)=s_k \mid \bar{S}_{k-1}(\bar{a}_{k-1})=\bar{s}_{k-1},\ \bar{h}_t,\ G=1\big),
\end{aligned} \tag{25}$$

and each factor is treated as in Steps 4 and 5 of Theorem 1, giving one layer further

$$= \prod_{k=1}^{t} P^{w}\big(S_k=s_k \mid \bar{S}_{(k-\delta):(k-1)}=\bar{s}_{(k-\delta):(k-1)},\ \bar{a}_k,\ \bar{l}_k,\ G=0\big); \tag{26}$$

at $k=t$ the decomposition half of A6 is not needed, the conditioning set already containing no future dose, but the current dose still has to be moved in by weak union.

Step 4 (assembly). Every integration variable is a surrogate and Y_t appears only in the innermost layer, acted on by no kernel: the t layers all carry w, because what is transported there is the state-trajectory distribution, whereas the innermost regression carries none, since placing w there would assert that the outcome mechanism itself needs reweighting. Substituting Steps 2 and 3 into Step 1 gives the nested expectation of the theorem. □

B.3 The Estimation Layer

The programme of Section 4.2 is solved on the atoms of χ_t with the negative-entropy term $\Omega(w) = \sum_j w_j(\log w_j - 1)$ of coefficient $\varepsilon > 0$ and Sinkhorn iterations (Cuturi, 2013; Peyré & Cuturi, 2019) under the single metric d_t, a barycentric projection (Yan et al., 2024) reducing the plan to one weight per trajectory and window. Two consequences are used later: the plan supplies values only at its atoms, so the weight must be a bounded Lipschitz *function* of χ_t out of sample, which is what condition (R1) parameterises; and the regularisation leaves an alignment floor, so either $\varepsilon \to 0$ is declared or $O(\varepsilon \log(1/\varepsilon))$ enters the error budget.

Proposition 2 (Existence and uniform boundedness of the weight)**.** *Let A2(i) hold. Then for every period t the law of χ_t under $G = 1$ is absolutely continuous with respect to its law under $G = 0$, the density ratio $w_t^\star$ of Eq.2 is well defined, and*

$$0 < w_t^\star \le \bar{u} := \frac{1-\epsilon}{\epsilon} \cdot \frac{P(G=0)}{P(G=1)} < \infty$$

uniformly in t.

Proof of Proposition 2. Fix t and write p_g for the density of χ_t in study g. The conditioning set of A2(i) is exactly this block, so its lower bound gives absolute continuity and the Radon–Nikodym derivative Eq.2 exists. By Bayes' rule,

$$w_t^\star(\chi_t) = \frac{p_1(\chi_t)}{p_0(\chi_t)} = \frac{P(G=1 \mid \chi_t)}{P(G=0 \mid \chi_t)} \cdot \frac{P(G=0)}{P(G=1)} \le \frac{1-\epsilon}{\epsilon} \cdot \frac{P(G=0)}{P(G=1)} =: \bar{u} < \infty, \tag{27}$$

whose conditional probability has the same conditioning set, so its two bounds apply with ϵ the overlap constant of A2(ii): the lower bound gives $w_t^\star > 0$ and the upper bound the uniform one. The block must contain the current dose, since otherwise the factorisation would carry an additional likelihood ratio in the dose direction, which is unbounded. □

Lemma 3 (Oracle weighting)**.** *Let A2(i) hold, let $P_{g,t}$ be the law of χ_t in study g, let $\chi_{i,t} := (s_{i,t}, r_{i,t}, a_{i,t})$ be the alignment triple of the i-th observational trajectory at period t, and let*

$$w_t^\sharp := \frac{\mathrm{d}P_{1,t}}{\mathrm{d}P_{0,t}}, \qquad \tilde{w}_{i,t}^\sharp := w_t^\sharp(\chi_{i,t}), \qquad w_{i,t}^\sharp := \frac{\tilde{w}_{i,t}^\sharp}{\sum_j \tilde{w}_{j,t}^\sharp},$$

$$\alpha_{w_t^\sharp} := \sum_{i=1}^{n_0} w_{i,t}^\sharp \, \delta_{\chi_{i,t}}, \qquad \beta_t := n_1^{-1} \sum_{j=1}^{n_1} \delta_{\chi_{j,t}}$$

be the estimation-layer oracle ratio, its unnormalised and self-normalised sample values, the reweighted empirical measure and the experimental empirical measure. Then for every bounded continuous f,

$$\sum_{i=1}^{n_0} w_{i,t}^\sharp \, f(\chi_{i,t}) \xrightarrow{\text{a.s.}} \mathbb{E}_{P_{1,t}}[f], \qquad n_0 \to \infty,$$

so that $\alpha_{w_t^\sharp}$ converges weakly to $P_{1,t}$, and if $n_1 \to \infty$ in addition, then to β_t almost surely.

The statement covers the oracle weights only; the proximity of the learned $\hat{w}_t$ to $w_t^\sharp$ is the subject of Theorem 3.

Proof of Lemma 3. The values $\tilde{w}_{i,t}^\sharp f(\chi_{i,t})$ are i.i.d. sample function values of a fixed function, the oracle weight being given rather than estimated, so the law of large numbers gives

$$n_0^{-1} \sum_i \tilde{w}_{i,t}^\sharp f(\chi_{i,t}) \xrightarrow{\text{a.s.}} \mathbb{E}_{P_{0,t}}[w_t^\sharp f] = \int f \, \mathrm{d}P_{1,t}, \tag{28}$$

the equality being the definition of the Radon–Nikodym derivative. Self-normalisation leaves the limit unchanged, the denominator converging by the same law of large numbers,

$$n_0^{-1} \sum_i \tilde{w}_{i,t}^{\sharp} \xrightarrow{\text{a.s.}} \mathbb{E}_{P_{0,t}}[w_t^{\sharp}] = 1, \tag{29}$$

so that the normalised weighted average has the same limit and $\alpha_{w_t^{\sharp}}$ converges weakly to $P_{1,t}$. The second statement needs $n_1 \to \infty$ as well, since β_t is built from n_1 random points: the weighted measure converges weakly to $P_{1,t}$, which is itself the almost sure weak limit of β_t once both sample sizes grow. □

Theorem 3 is stated in Section 4.2. Two clauses of condition (R1) carry weight: without uniform boundedness and the Lipschitz bound, a weight oscillating at high frequency destroys the compactness Arzelà–Ascoli needs. With P uniform on $[0,1]^2$ and Q_n of density $q_n(s,x) = 1 + a\sin(2\pi n x)(2s-1)$, $0 < a < 1$, one has $W_1(Q_n, P) \le a/(\pi n) \to 0$ and identical marginals, yet

$$\int_0^1 W_1\big(Q_n(S \mid x), P(S \mid x)\big)\,\mathrm{d}x = a/(3\pi) > 0$$

for every n, its Lipschitz constants growing like $2\pi n a$. Only the integrated statement is claimed, and it does not give convergence of the conditional laws for almost every (r_t, a_t).

Proof of Theorem 3. Write $Q := P_{0,t}^{\hat{w}_t}$, a probability measure by $\mathbb{E}_{P_{0,t}}[\hat{w}_t] = 1$, write $b_t := (r_t, a_t)$ and $\nu \in \{Q^{R_t,A_t}, P_{1,t}^{R_t,A_t}\}$; by A2(i) the joint error has the density form $\Delta_t = \mathrm{TV}(Q, P_{1,t}) = \frac{1}{2}\|\hat{w}_t - w_t^{\sharp}\|_{L^1(P_{0,t})}$.

Step 1 (weak convergence). Part (3) of condition (R1) makes the domain compact, and there the transport cost metrises weak convergence, so $\mathrm{OT}_t(Q, P_{1,t}) \to 0$ gives $Q \xrightarrow{d} P_{1,t}$.

Step 2 (upgrade to L^1). Parts (1) and (2) of condition (R1) are the two hypotheses of Arzelà–Ascoli for the class $\mathcal{F}_t$, part (3) supplying the compact domain. Index the sequence by $n := (n_0, n_1)$. Arzelà–Ascoli gives a subsequence with $\|\hat{w}_t^{(n_k)} - w_t^{\dagger}\|_\infty \to 0$ for some continuous $w_t^{\dagger}$; it delivers relative compactness only, so the whole sequence is not yet covered. Uniform convergence gives $P_{0,t}^{\hat{w}_t^{(n_k)}} \xrightarrow{d} P_{0,t}^{w_t^{\dagger}}$; comparing with Step 1 and using uniqueness of limits for W_1 gives $P_{0,t}^{w_t^{\dagger}} = P_{1,t}$, and reading this through the densities with respect to $P_{0,t}$, which are the two weights, gives $w_t^{\dagger} = w_t^{\sharp}$ almost everywhere, by almost everywhere uniqueness of the Radon–Nikodym derivative, so no continuous version of $w_t^{\sharp}$ is needed in advance. Hence $\Delta_t^{(n_k)} \to 0$; and since this holds for any subsequence, every subsequence of $\{\Delta_t^{(n)}\}$ has a further subsequence converging to 0, which forces $\Delta_t^{(n)} \to 0$.

Step 3 (the ratio rule). The conditional law is the joint divided by its (R_t, A_t)-marginal. For the denominator layer, marginalisation cannot increase a total variation distance, an event involving only (R_t, A_t) being also a joint event, so $\mathrm{TV}(Q^{R_t,A_t}, P_{1,t}^{R_t,A_t}) \le \Delta_t$. For the quotient layer, let K_b, K'_b be the two conditional kernels, take $\nu := Q^{R_t,A_t}$ and $M(\mathrm{d}s_t, \mathrm{d}b_t) := K'_{b_t}(\mathrm{d}s_t)\,\nu(\mathrm{d}b_t)$, which shares its marginal with Q and its kernel with $P_{1,t}$. Two measures that share a marginal differ in total variation exactly by the integral of their conditional distances against that marginal, so

$$\nabla_t = \mathrm{TV}(Q, M) \le \mathrm{TV}(Q, P_{1,t}) + \mathrm{TV}(P_{1,t}, M) = \Delta_t + \mathrm{TV}(P_{1,t}^{R_t,A_t}, \nu) \le 2\Delta_t;$$

the mirrored auxiliary measure built on K and $P_{1,t}^{R_t,A_t}$ gives the same bound. On a bounded set the transport cost is at most $D_{S,t}$ times the total variation distance, so taking expectations of the conditional bound under the fixed marginal $P_{1,t}^{R_t,A_t}$ gives the third statement. □

The experimental sample is small and the observational one large, so the transport is not solved on the two samples at once: the observational sample of size $n_0 = Ib$ is split at random, independently of the observed values, into I disjoint batches $\mathcal{B}_1, \ldots, \mathcal{B}_I$ of size b, each of them aligned against the *full* experimental sample under the cost $d_t(\chi_{i,t}, \chi_{j,t})$, the weights being learned within the batch alone.

Theorem 4 (Mini-batch upper bound). *With within-batch weights $\hat{q}_{m,t} = \{\hat{q}_{i,m,t}\}_{i\in\mathcal{B}_m}$ normalised to $\sum_{i\in\mathcal{B}_m} \hat{q}_{i,m,t} = 1$, let*

$$c_{m,t} := \langle \hat{\pi}_{m,t}, \mathbf{D}_{m,t}\rangle, \qquad (\mathbf{D}_{m,t})_{ij} := d_t(\chi_{i,t}, \chi_{j,t}), \qquad m\text{-OT}_t := I^{-1}\sum_{m=1}^{I} c_{m,t},$$

$$\hat{\alpha}_t := I^{-1}\sum_{m=1}^{I}\sum_{i\in\mathcal{B}_m} \hat{q}_{i,m,t}\,\delta_{\chi_{i,t}}$$

be the cost of the optimal plan $\hat{\pi}_{m,t}$ of batch m against β_t of Lemma 3, its cost matrix, the aggregate weighted measure and the batch-averaged cost, and let $\tilde{\pi}_{m,t}$ be the padded and $\bar{\pi}_t := I^{-1}\sum_{m=1}^{I}\tilde{\pi}_{m,t}$ the averaged plan, the zero-padding filling the rows outside $\mathcal{B}_m$. Then

$$\bar{\pi}_t \in \Pi\big(\hat{\alpha}_t, \beta_t\big), \qquad \textit{hence} \qquad \mathrm{OT}_t\big(\hat{\alpha}_t, \beta_t\big) \le m\text{-OT}_t.$$

The averaged plan is a legitimate coupling and the bound follows from the minimality of OT_t, no convexity of the cost matrix being used.

Proof of Theorem 4. Write $\hat{\alpha}_{m,t} := \sum_{i\in\mathcal{B}_m} \hat{q}_{i,m,t}\,\delta_{\chi_{i,t}}$, so that $\hat{\alpha}_t = I^{-1}\sum_{m=1}^{I}\hat{\alpha}_{m,t}$, and let $\mathbf{D}_t$ be the cost matrix whose restriction to the rows of $\mathcal{B}_m$ is $\mathbf{D}_{m,t}$. The rows of $\tilde{\pi}_{m,t}$ outside $\mathcal{B}_m$ are zero, so $\tilde{\pi}_{m,t}\mathbf{1}_{n_1} = \hat{\alpha}_{m,t}$ and

$$\bar{\pi}_t\,\mathbf{1}_{n_1} = I^{-1}\sum_{m=1}^{I}\tilde{\pi}_{m,t}\,\mathbf{1}_{n_1} = I^{-1}\sum_{m=1}^{I}\hat{\alpha}_{m,t} = \hat{\alpha}_t, \tag{30}$$

the batches being disjoint and covering the observational sample; and each $\hat{\pi}_{m,t}$ is a plan to β_t, so

$$\bar{\pi}_t^{\mathsf{T}}\mathbf{1}_{n_0} = I^{-1}\sum_{m=1}^{I}\hat{\pi}_{m,t}^{\mathsf{T}}\mathbf{1}_{|\mathcal{B}_m|} = I^{-1}\sum_{m=1}^{I}\beta_t = \beta_t. \tag{31}$$

Hence $\bar{\pi}_t \in \Pi(\hat{\alpha}_t, \beta_t)$, and its cost is

$$\langle \bar{\pi}_t, \mathbf{D}_t\rangle = I^{-1}\sum_{m=1}^{I}\langle \tilde{\pi}_{m,t}, \mathbf{D}_t\rangle = I^{-1}\sum_{m=1}^{I}\langle \hat{\pi}_{m,t}, \mathbf{D}_{m,t}\rangle = I^{-1}\sum_{m=1}^{I} c_{m,t} = m\text{-OT}_t, \tag{32}$$

the middle step holding because the rows outside $\mathcal{B}_m$ carry no mass. Since $\mathrm{OT}_t(\hat{\alpha}_t, \beta_t)$ minimises $\langle \pi, \mathbf{D}_t\rangle$ over $\Pi(\hat{\alpha}_t, \beta_t)$, the averaged plan is admissible and the bound follows. □

B.4 The Rollout Layer

The errors e_t, $e_t^{(Y)}$, the constants $C_\epsilon, C_{\epsilon,Y}, \kappa$ and the estimated generators $\hat{g}$, $\hat{h}$ are as in Section 4.3, where $C_\epsilon, C_{\epsilon,Y}$ bound the model error on a *fixed* input only, discarding the history beyond the window being a separate matter.

Lemma 4 (Unconstrained rollout error). *Under a matched initial window ($\hat{S}_j = S_j$ for $j = 1-\delta, \dots, 0$), bounded single-step model error and Lipschitz generators,*

$$e_t \le \kappa \max_{0\le k\le \delta-1} e_{t-\delta+k} + C_\epsilon, \qquad e_t^{(Y)} \le L_h\sqrt{\delta+1}\max_{0\le k\le \delta} e_{t-\delta+k} + C_{\epsilon,Y},$$

and consequently $e_t \le C_\epsilon(\kappa^t - 1)/(\kappa - 1)$ and $e_t^{(Y)} \le L_h\sqrt{\delta+1}\,C_\epsilon(\kappa^t-1)/(\kappa-1) + C_{\epsilon,Y}$ for $\kappa > 1$, while $\kappa = 1$ degenerates to linear growth $e_t \le C_\epsilon t$.

The lemma is a necessity argument only; its envelopes enter no final bound.

Proof of Lemma 4. By the triangle inequality, $e_t \le \big\|\hat{g}(\bar{\hat{S}}_{(t-\delta):(t-1)}) - \hat{g}(\bar{S}_{(t-\delta):(t-1)})\big\|_2 + \|\epsilon_t\|_2$, where the second term is bounded by C_ϵ and the first compares the same estimated generator on two histories. The estimate is L_g-Lipschitz for the full window input, the quantity Eq.4 controls, and

$$\big\|\bar{\hat{S}}_{(t-\delta):(t-1)} - \bar{S}_{(t-\delta):(t-1)}\big\|_2 = \sqrt{\sum_{k=0}^{\delta-1} e_{t-\delta+k}^2} \le \sqrt{\delta}\max_{0\le k\le\delta-1} e_{t-\delta+k},$$

with $\sqrt{\delta}$ the Euclidean length of the δ window blocks, so

$$e_t \le \kappa \max_{0\le k\le \delta-1} e_{t-\delta+k} + C_\epsilon, \tag{33}$$

a running-maximum rather than a first-order recursion. With $E_j = 0$ for $j \le 0$ and $E_t = \kappa E_{t-1} + C_\epsilon$ for $t \ge 1$, rolling out gives $E_t = C_\epsilon(\kappa^t - 1)/(\kappa - 1)$ for $\kappa > 1$ and $E_t = C_\epsilon t$ at $\kappa = 1$, and $e_t \le E_t$ follows by induction: at $t = 1$ the matched initial window gives $e_j = 0 \le E_j$ for $j \le 0$, hence $e_1 \le C_\epsilon = E_1$; and assuming $e_j \le E_j$ for $j < t$, monotonicity of E ($E_j - E_{j-1} = C_\epsilon \kappa^{j-1} \ge 0$) gives $\max_{0\le k\le \delta-1} e_{t-\delta+k} \le E_{t-1}$, so Eq.33 gives $e_t \le \kappa E_{t-1} + C_\epsilon = E_t$. Reducing the running-maximum recursion to first order therefore uses both the exact initialization at the rollout start and the monotonicity of E. The same decomposition applies to $\hat{Y}_t = \hat{h}(\hat{S}_{(t-\delta):t})$ against $Y_t = h(\bar{S}_{(t-\delta):t})$, and the L_h-Lipschitz property of the estimated terminal head gives

$$e_t^{(Y)} \le L_h \sqrt{\sum_{k=0}^{\delta} e_{t-\delta+k}^2} + C_{\epsilon,Y} \le L_h \sqrt{\delta+1} \max_{0\le k\le \delta} e_{t-\delta+k} + C_{\epsilon,Y}, \tag{34}$$

the terminal window carrying $\delta + 1$ entries because it includes the current state; with $e_j \le E_j$ this gives $e_t^{(Y)} \le L_h\sqrt{\delta+1}\, E_t + C_{\epsilon,Y}$. For $\kappa \ge 1$ neither envelope is bounded uniformly in t, so the bound enters only as the necessity argument for Eq.4, not as a final bound. □

Theorem 5 is stated in Section 4.3. The alternative to assuming a contraction is to impose it on the estimated generator, through the norm condition Eq.4, $\|\hat{J}_t\|_2 \le \gamma/\sqrt{\delta}$ with $0 < \gamma < 1$, where $\hat{J}_t = \nabla_{\bar{S}_{(t-\delta):(t-1)}} \hat{g} \in \mathbb{R}^{m\times\delta m}$ is a *rectangular* matrix, required to hold at all relevant periods and on the true input, the predicted input and the segment joining them, the last so that the mean value inequality applies. Its integral form along that segment gives

$$\begin{aligned} &\big\|\hat{g}(\bar{\hat{S}}_{(t-\delta):(t-1)}) - \hat{g}(\bar{S}_{(t-\delta):(t-1)})\big\|_2 \\ &\quad\le \Big(\sup_{\text{segment}} \big\|\hat{J}_t\big\|_2\Big)\, \big\|\bar{\hat{S}}_{(t-\delta):(t-1)} - \bar{S}_{(t-\delta):(t-1)}\big\|_2 \\ &\le \frac{\gamma}{\sqrt{\delta}} \sqrt{\sum_{k=0}^{\delta-1} e_{t-\delta+k}^2} \\ &\quad\le \frac{\gamma}{\sqrt{\delta}} \cdot \sqrt{\delta} \max_{0\le k\le \delta-1} e_{t-\delta+k} = \gamma \max_{0\le k\le \delta-1} e_{t-\delta+k}, \end{aligned} \tag{35}$$

so in the error decomposition of Lemma 4 the amplification factor $L_g\sqrt{\delta} = \kappa$ is replaced by $\gamma < 1$: the unconstrained factor grows with the window while the condition is stated as $\gamma/\sqrt{\delta}$, the same $\sqrt{\delta}$ appearing in both with opposite sense, so that the two cancel into γ. The penalty is not the condition, as noted in Section 5, and γ bounds the *last block row* $\hat{J}_t$, not the spectral radius of the window-advance operator, for which only $\rho_{\mathrm{JSR}} \le \gamma^{1/\delta}$ is available. One protocol convention is also part of the statement: the terminal head takes the model output under both protocols rather than the true S_T, otherwise the difference would absorb the single-step error at the current period and charge to rollout a cost that is not its own.

Proof of Theorem 5. Rollout and teacher forcing read one and the same model on two windows, the generated one and the true one, $\hat{S}_t = \hat{g}(\bar{\hat{S}}_{(t-\delta):(t-1)})$ against $\hat{S}_t^{\mathrm{tf}} = \hat{g}(\bar{S}_{(t-\delta):(t-1)})$, with a matched initial window. By 35 the first term of the error decomposition of Lemma 4 is bounded by $\gamma \max_{0\le k\le\delta-1} e_{t-\delta+k}$ and the second by C_ϵ, so

$$e_t \le \gamma \max_{0\le k\le \delta-1} e_{t-\delta+k} + C_\epsilon. \tag{36}$$

With $E_j = 0$ for $j \le 0$ and $E_t = \gamma E_{t-1} + C_\epsilon$ for $t \ge 1$, rolling out gives $E_t = C_\epsilon(1-\gamma^t)/(1-\gamma)$, nondecreasing as in Lemma 4, and $e_t \le E_t$ follows by the same induction, so that $e_t \le E_t \le C_\epsilon/(1-\gamma)$ since $1 - \gamma^t \le 1$. By 34 and $e_j \le E_j$, $e_t^{(Y)} \le L_h\sqrt{\sum_{k=0}^{\delta} E_{t-\delta+k}^2} + C_{\epsilon,Y} \le L_h\sqrt{\delta+1}\, E_t + C_{\epsilon,Y}$, the last inequality because $\sum_{k=0}^{\delta} E_{t-\delta+k}^2 \le (\delta+1)E_t^2$. The bricks are read off

the same comparison directly, on the difference of the two protocols, and never by subtracting their two error envelopes, since $a \le A$ and $b \le B$ do not imply $a - b \le A - B$: with $\hat{Y}_T = \hat{h}(\hat{\bar{S}}_{(T-\delta):T})$ against $\hat{Y}_T^{\mathrm{tf}} = \hat{h}(\bar{S}_{(T-\delta):T})$, and $\hat{h}$ taking the model output under both protocols so that its window carries $\delta + 1$ entries,

$$\big\|\hat{S}_t - \hat{S}_t^{\mathrm{tf}}\big\|_2 \le \gamma E_{t-1}, \qquad \big|\hat{Y}_T - \hat{Y}_T^{\mathrm{tf}}\big| \le L_h \sqrt{\sum_{k=0}^{\delta} E_{T-\delta+k}^2}. \tag{37}$$

Both bricks hold at a fixed horizon and are proportional to E_t, hence to C_ϵ, so when the model is recovered they vanish together even for $\delta \neq 0$, and the first is identically zero at $\delta = 0$. Finally C_ϵ splits as $C_\epsilon = C_\epsilon^{\mathrm{approx}} + C_\epsilon^{\mathrm{est}}(n)$, approximation and finite-sample estimation error, the estimation part being counted once only, in the statistical gap of Theorem 6. □

The loss $\ell^{\mathbf{w},\mathrm{tf}}$, the risks and the pooled laws P_{cf}^{ϕ}, $P_f^{\mathbf{w},\phi}$ are as in Section 5; $\hat{S}_t^{\mathrm{tf}}$, $\hat{Y}_T^{\mathrm{tf}}$ are the teacher-forced versions of $\hat{S}_t$, $\hat{Y}_T$, in which each head reads the true window where rollout feeds it the generated one. Reached from the raw window record through the history-compression map $\Phi_{\theta,t}(\bar{h}_t, \bar{s}_{(t-\delta):t}, a_t) = ((\bar{s}_{(t-\delta):t}, \phi_\theta(\bar{h}_t)), a_t) = (x_t, a_t)$, which needs no inverse, the two pooled laws are, for measurable m, equal-weight mixtures over the evaluation periods,

$$\int m \,\mathrm{d}P_{cf}^{\phi} = \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \mathbb{E}_{P_{cf}}[m(X_t, A_t)], \qquad \int m \,\mathrm{d}P_f^{\mathbf{w},\phi} = \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \mathbb{E}_{P_f}[w_t \, m(X_t, A_t)],$$

the second a probability measure by $\mathbb{E}_{P_f}[w_t] = 1$, with w_t multiplying the factual mass on the record it was solved from. Two constants are derived below rather than assumed: the loss upper bound, read off the teacher-forced evaluation loss, in which neither head is fed its own prediction and the weight multiplies the short-term terms alone, $w_t \le \bar{u}$ by Proposition 2,

$$0 \le \ell^{\mathbf{w},\mathrm{tf}}(\theta) \le M := \lambda_S \bar{u} L_\ell^S C_\epsilon + \frac{L_\ell^Y C_{\epsilon,Y}}{|\mathcal{T}_0|}, \tag{38}$$

and the protocol gap, assembled in part (2) of the proof from the two input-substitution bricks 37 and the terminal distribution difference,

$$B := \frac{\lambda_S \gamma L_\ell^S}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} E_{t-1} + \frac{L_\ell^Y L_h}{|\mathcal{T}_0|} \sqrt{\sum_{k=0}^{\delta} E_{T-\delta+k}^2} + \frac{L_\ell^Y C_{\epsilon,Y}}{|\mathcal{T}_0|}, \qquad E_t = C_\epsilon \frac{1 - \gamma^t}{1 - \gamma}. \tag{39}$$

Proof of Theorem 6. The three risks of the statement differ only in the population measure and in the protocol, so, exactly and without further assumption, and by telescoping,

$$\mathcal{E}_{cf}(\theta) - \hat{\mathcal{E}}_n(\theta) = \underbrace{\big[\mathcal{E}_{cf}^{\mathrm{tf}}(\theta) - \mathcal{E}_f^{\mathbf{w}}(\theta)\big]}_{\text{(1) distributional gap}} + \underbrace{\big[\mathcal{E}_{cf}(\theta) - \mathcal{E}_{cf}^{\mathrm{tf}}(\theta)\big]}_{\text{(2) error}} + \underbrace{\big[\mathcal{E}_f^{\mathbf{w}}(\theta) - \hat{\mathcal{E}}_n(\theta)\big]}_{\text{(3) statistical gap}},$$

the first bracket changing the measure but not the protocol and the second the protocol but not the measure; the two must not be merged.

(1) The distributional gap. Only the short-term part of the two teacher-forced risks is carried by the metric, the terminal part being absorbed by group (2). Integrating the difference of the short-term parts over the evaluation periods,

$$\begin{aligned} &\frac{\lambda_S}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \Big\{ \mathbb{E}_{P_{cf}} \mathcal{L}_S(S_t, \hat{S}_t^{\mathrm{tf}}) - \mathbb{E}_{P_f}\big[w_t \, \mathcal{L}_S(S_t, \hat{S}_t^{\mathrm{tf}})\big] \Big\} \\ &\qquad = \frac{\lambda_S}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \int \ell^s_{\phi_\theta, \hat{g}}(\bar{h}_t, \bar{s}_{(t-\delta):t}, a_t) \big[\mathrm{d}P_{cf}(\bar{h}_t, \bar{s}_{(t-\delta):t}, a_t) - w_t \,\mathrm{d}P_f(\bar{h}_t, \bar{s}_{(t-\delta):t}, a_t)\big], \end{aligned}$$

the raw laws of the window record at period t being meant on the right. For a test function $\mathbf{1}_C$ of the image space the defining property of the pushforward gives $\int \mathbf{1}_C \,\mathrm{d}[(\Phi_{\theta,t})_{\#}P] = P(\Phi_{\theta,t}^{-1}(C)) = \int \mathbf{1}_C \circ \Phi_{\theta,t} \,\mathrm{d}P$, where $\Phi_{\theta,t}^{-1}(C)$ is a set preimage and not an inverse map; linearity extends the

identity to integrable m, and the factorisation of the loss through $\phi_\theta(\bar{H}_t)$ lets the loss itself serve as the integrand,

$$\frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\int \ell^s_{\phi_\theta,\hat{g}}(\bar{h}_t,\bar{s}_{(t-\delta):t},a_t)\,\mathrm{d}P_{cf}(\bar{h}_t,\bar{s}_{(t-\delta):t},a_t)$$
$$=\frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\int_{\mathcal{X}\times\mathcal{A}} \ell^s_{\phi_\theta,\hat{g}}(x_t,a_t)\,\mathrm{d}P^{\phi}_{cf}(x_t,a_t),$$
$$\frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\int w_t\,\ell^s_{\phi_\theta,\hat{g}}(\bar{h}_t,\bar{s}_{(t-\delta):t},a_t)\,\mathrm{d}P_{f}(\bar{h}_t,\bar{s}_{(t-\delta):t},a_t)$$
$$=\frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\int_{\mathcal{X}\times\mathcal{A}} \ell^s_{\phi_\theta,\hat{g}}(x_t,a_t)\,w_t\,\mathrm{d}P_{f}(x_t,a_t).$$

The weight is not hidden by the change of variables, w_t multiplying the factual mass on the record it was solved from, and neither is the average $1/|\mathcal{T}_0|$, compressing the history into the block not turning an average over periods into a sum over them.

The supremum runs over one class common to all periods, so the periods must be merged into a single measure first. The only inequality available,

$$\sup_{m\in\mathcal{M}}\Big|\frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\int m\big[\mathrm{d}P_{cf}(x_t,a_t)-w_t\,\mathrm{d}P_f(x_t,a_t)\big]\Big|$$
$$\le\ \frac{1}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\sup_{m\in\mathcal{M}}\Big|\int m\big[\mathrm{d}P_{cf}(x_t,a_t)-w_t\,\mathrm{d}P_f(x_t,a_t)\big]\Big|,$$

runs the other way and cannot be reversed: two periods carrying distribution differences of opposite sign cancel in the mixture while their per-period suprema keep adding, so a sum of per-period metrics does not bound the pooled one. Merging first, and letting $1/|\mathcal{T}_0|$ pass into the definitions of the two pooled laws rather than dropping it,

$$\frac{\lambda_S}{|\mathcal{T}_0|}\sum_{t\in\mathcal{T}_0}\int_{\mathcal{X}\times\mathcal{A}} \ell^s_{\phi_\theta,\hat{g}}(x_t,a_t)\big[\mathrm{d}P_{cf}(x_t,a_t)-w_t\,\mathrm{d}P_f(x_t,a_t)\big]$$
$$=\int_{\mathcal{X}\times\mathcal{A}}\lambda_S\ell^s_{\phi_\theta,\hat{g}}(x,a)\big[\mathrm{d}P^{\phi}_{cf}(X,A)-\mathrm{d}P^{\mathbf{w},\phi}_{f}(X,A)\big]\ \le\ \Lambda_{\mathrm{bal}}\,\mathrm{IPM}_{\mathcal{M}}\big(P^{\phi}_{cf},\ P^{\mathbf{w},\phi}_{f}\big),$$

the equality using the pooled definitions of the two laws and the inequality $\lambda_S\ell^s_{\phi_\theta,\hat{g}}/\Lambda_{\mathrm{bal}}\in\mathcal{M}$ together with the dual definition of the metric: one and the same pointwise loss on both sides is what allows the periods to be merged without assuming their laws equal. The supremum sits outside the integral, acts on the pooled block of a single window and is never exchanged with the sum over periods, and the terminal conditional loss is deliberately not in $\mathcal{M}$, its distributional part being carried by the third block of group (2).

(2) The error. The gap between the two protocols under one and the same counterfactual measure is assembled from three blocks, obtained by direct inequalities and never by subtracting two upper bounds. By the first brick Eq.37 the generated state differs between the two protocols by at most γE_{t-1} in norm, and the short-term loss is L^S_ℓ-Lipschitz and vanishes at a correct prediction, so the *short-term input substitution* block contributes $\lambda_S\gamma L^S_\ell|\mathcal{T}_0|^{-1}\sum_{t\in\mathcal{T}_0}E_{t-1}$; by Eq.37 again, with the terminal loss L^Y_ℓ-Lipschitz, the *long-term input substitution* block contributes $L^Y_\ell L_h|\mathcal{T}_0|^{-1}\sqrt{\sum_{k=0}^{\delta}E^2_{T-\delta+k}}$; and under teacher forcing the terminal head is applied to the true window, so its prediction error is the single-step error $C_{\epsilon,Y}$ itself and the loss is at most $L^Y_\ell C_{\epsilon,Y}$, while the observational side is non-negative, hence the *terminal distribution difference* is a conservative upper bound of $L^Y_\ell C_{\epsilon,Y}/|\mathcal{T}_0|$, retained because A4 transports a conditional mean rather than the unweighted population loss. Substituting $E_t=C_\epsilon(1-\gamma^t)/(1-\gamma)$, with the terminal index shifted over the window $T-\delta,\dots,T$ and the terms at non-positive indices omitted because $E_0=0$, the three blocks combine into B of Eq.39.

(3) The complexity and concentration term. Under teacher forcing neither head is fed its own prediction, so the error that enters is the single-step one, giving $0\le\ell^{\mathbf{w},\mathrm{tf}}\le M$

as in Eq.38. With σ_i independent Rademacher variables independent of the sample and $\mathfrak{R}_n(\mathcal{M}_w) := \mathbb{E}_{\mathcal{D}_0}\mathbb{E}_{\boldsymbol{\sigma}}[\sup_{\theta\in\Theta} n^{-1}\sum_{i=1}^n \sigma_i \ell_i^{\mathbf{w},\mathrm{tf}}(\theta)|\mathcal{D}_0]$, an auxiliary sample of the same law (primed), $\sup_\theta \mathbb{E}[\cdot] \le \mathbb{E}[\sup_\theta(\cdot)]$, the exchangeability of each original and auxiliary observation, the subadditivity of the supremum and $\sigma \overset{d}{=} -\sigma$ give $\mathbb{E}\sup_\theta[\mathcal{E}_f^{\mathbf{w}}(\theta) - \hat{\mathcal{E}}_n(\theta)] \le 2\mathfrak{R}_n(\mathcal{M}_w)$, and the same computation with the two terms interchanged gives the same bound in the other direction. Replacing one complete observational trajectory changes the value of a fixed candidate by at most M/n, because the weights are held at their population values and the replacement does not re-solve the transport problem, so $\sum_i c_i^2 = M^2/n$ and McDiarmid's inequality applied to the supremum, with failure probability at most $\eta/2$ in each direction and a union bound, gives, with probability at least $1-\eta$, $\sup_{\theta\in\Theta}|\mathcal{E}_f^{\mathbf{w}}(\theta) - \hat{\mathcal{E}}_n(\theta)| \le 2\mathfrak{R}_n(\mathcal{M}_w) + M\sqrt{\log(2/\eta)/(2n)}$.

Assembly. Groups (1) and (2) are deterministic and hold pointwise in θ, so on the high-probability event of group (3) the three are added and the decomposition gives the statement; only the statistical component shrinks with the sample size. □

Taken together, the assumptions and condition (R1) give Lemmas 1 and 2, hence Theorems 1 and 2; Lemma 3, Theorem 3 and Theorem 4 close the alignment chain; and Lemma 4 with the norm condition Eq.4 and Theorem 5 turn the rollout error into a depth-uniform envelope, which Theorem 6 collects into the end-to-end bound.

C Experimental Details

C.1 Implementation Details

The reweighting solves each window's entropically regularised transport problem of Section 4.2 with a Sinkhorn solver (Cuturi, 2013; Peyré & Cuturi, 2019) under the metric d_t, and reduces the plan to one weight per observational trajectory and window by a barycentric projection (Yan et al., 2024); the balancing penalty is a Wasserstein distance through its Kantorovich dual with a Lipschitz critic (Sriperumbudur et al., 2012; Johansson et al., 2016). The encoder is the MLP ϕ, the short-term head the GRU q with the emission map g, and the terminal head h over the surrogate window, which at inference is filled one period at a time with the value just generated; the backbone also instantiates as an LSTM (Lim et al., 2018; Bica et al., 2020a). All analyses use Adam at a learning rate of 10^{-3} and batch size 512, for up to 120 epochs with early stopping (patience 30) on the observational validation loss, with $\lambda_S = 0.5$; λ_{IPM} and λ_{JSR} are selected once per cohort on that same validation loss, with $\lambda_S = 0.5$ and normalized penalty weights $\lambda_{\mathrm{IPM}} = \lambda_{\mathrm{JSR}} = 0.05$ on every cohort. All results are averaged over three independent replications. Runs use NVIDIA A100 GPUs, under Python 3.11.5 and torch 2.1.1.

C.2 Data Generation Details

C.2.1 Synthetic cohort

The surrogate and outcome paths are generated, as stated in Section 6, by

$$\log S_t = \sum_{j=1}^{\delta} w_j \big[(1-\rho)\log S_{t-j} + \rho \log K_V - \beta_c \,\mathrm{Emax}(C_{t-j}) + \alpha_s^\top L_{t-j} + \gamma_s^\top U_{t-j} + e_{t-j}\big],$$
$$Y_t = \beta_{Y0} + \frac{1}{\delta}\sum_{j=0}^{\delta}\big(\beta_{YL,j}^\top L_{t-j} + \beta_{YS,j} S_{t-j}\big) + \epsilon_{Y,t}, \tag{40}$$

whose coefficients are fixed below. Baseline covariates are drawn as $V^o \sim \mathcal{N}(0.1\mathbf{1}_p, I_p)$ observationally and $V^e \sim \mathcal{N}(0.2\mathbf{1}_p, I_p)$ experimentally, with $p = 15$, and the unobserved confounder follows a stationary AR(1) process,

$$\eta_t = \rho_U \eta_{t-1} + \sqrt{1-\rho_U^2}\,\epsilon_t, \qquad U_t = 0.25 + \eta_t, \qquad \epsilon_t \sim \mathcal{N}(0, I_d), \qquad \rho_U = 0.3,\ d = 3. \tag{41}$$

The observed covariates are $L_t = [L_t^{\mathrm{phys}}; \mathtt{lag_S}_t]$ with $q = 15$, where $L_t^{\mathrm{phys}} = 0.5 L_{t-1}^{\mathrm{phys}} + \beta_{la} A_{t-1} + \phi_v V + \epsilon_l$ and $\mathtt{lag_S}_t = \frac{1}{6}\sum_{k=10}^{15} S_{t-k}/K_V$ rescales the historical surrogate to the

range of the physiological block, and the surrogate starts from $S_0 = K_V \sigma(\beta_V^\top V + \beta_U^\top U_0 + \varepsilon)$ with $K_V = 1150$.

The assignment score $m_t = \beta_{AL}^\top L_t + \beta_{AA}(A_{t-1} - A_{\max}/2) + f(V) + \epsilon_A$, with $\|\beta_{AL}\| = 5$ and f the baseline feature map of Yang et al. (2024b), is standardized into $\tilde{m}_t$, and the dose is

$$A_t = A_{\max}\, \sigma\big(0.7\tilde{m}_t + 0.35\, \beta_U^\top U_t \mathbb{1}\{G{=}0\}\big)\, D_t, \qquad A_{\max} = 10, \tag{42}$$

where the mask $D_t \in \{0, 0.5, 1\}$ is piecewise constant within a trajectory and induces treatment interruptions. The confounder enters on the observational side alone, so the sources overlap by construction and a density ratio exists, whereas a pure correlation would be separable and collapse the weights to uniform; β_U is swept over six levels.

In Eq.40, $\gamma_h = 0.5$ is the concentration retention, $C_{50} = 6$, and $w_j = (1-\gamma)\gamma^{j-1}$ normalized to $\sum_j w_j = 1$ with $\gamma = 0.5$, a separate quantity from γ_h; the mean-reversion speed ρ and the dose sensitivity β_c vary across units through V. The counterfactual truth under a target path $a^*_{0:T}$ holds the covariate trajectory and S_0 at their factual values, restarts the concentration at zero and marginalizes U over $n_{mc} = 30$ Monte Carlo draws with all noise set to zero,

$$\hat{\mu}_t(a^*) = \frac{1}{n_{mc}} \sum_{m=1}^{n_{mc}} Y_t^{(m)}(a^*) \;\longrightarrow\; \mathbb{E}\big[Y_t(a^*) \mid V, L, S_0\big], \tag{43}$$

since a single draw would condition the truth on one realisation of a hidden variable that no method can recover. The evaluation grid comprises eleven constant doses from zero to the 99th percentile and seven time-varying paths, of which the step-up and step-down paths share a mean and differ only in order; a burn-in of twenty periods precedes $t = 0$.

C.2.2 MIMIC-III COHORT

The extract holds 291,425 dose administrations from 5,520 patients; after the length and first-ICU filters the analysis cohort comprises 4054 patients, of which 3367 have at least eighteen administrations. The extract is the MIMIC-III Clinical Database v1.4, publicly available at `https://physionet.org/content/mimiciii/1.4/`. The cohort is indexed by dose administration rather than by calendar day, with $T = 15$ and $\delta = 1$; real are V, L_t (26 columns), the marginal of S and the mortality level, while the confounder retains ten real covariates orthogonalized against the observed history, hence real yet not linearly recoverable from it. The dose follows a two-part model, an indicator gate shared across the two sources, which preserves positivity by construction, and a log-magnitude term that carries the confounding and is clipped to ± 1.5. The surrogate follows mean-reverting dynamics in a logit scale over $[30, 140]$ mmHg, $\ell_t = a + \rho\ell_{t-1} + \beta_c \,\mathrm{Emax}(C_t) + \alpha_S^\top L_t - \gamma_S^\top U_t + e_t$, with $\gamma_h = 0.9$; the negative sign reflects that a higher pressure is better. The outcome is a logistic model on $(\bar{L}, \bar{S}, S_{\min})$, fitted block by block with its intercept calibrated to the real mortality. Three quantities are testbed design rather than fitted: the dose-to-surrogate magnitude, which titration renders unidentifiable and which is fixed at the clinical prior of a 15 mmHg rise from a constant dose of 2 against 0; the confounder strength, fixed so that U accounts for 30% of the variance on each path; and the surrogate-to-outcome channel, rescaled to a target variance share. Ground truth uses the marginalization of equation 43.

C.2.3 INDUSTRIAL-PLATFORM COHORT

The raw panel is an order-level record of ∼210k orders placed by ∼13k users, indexed by the order sequence rather than by calendar time, and the released cohort is built from it. Following Battocchi et al. (2021) we estimate, on that panel, the full covariance of (L_t, U_t) and lassoCV models of A_t, S_t and Y on their own histories; resample the coefficients as $\theta(1 + 0.25\xi)$ with $\xi \sim \mathcal{N}(0, 1)$ and the covariance through a Bartlett-decomposed Wishart draw with eigenvalue clipping; forward-simulate all variables period by period with an accumulated-exposure channel $C_t = \gamma_c C_{t-1} + A_t$ of the same Emax form, which is what makes the order of the dosing path matter; match the marginals to the real ones by quantile mapping; and split 9:1 with no user crossing the sources, reassigning the experimental dose from the resampled model. Names are anonymized and identifiers, locations and timestamps dropped, so the released cohort is a panel of ∼7k users over five consecutive orders each, with 21 covariate and 6 confounder columns, preserving the statistical structure of the panel and retaining no real record. The evaluation grid comprises doses restricted to $[p_{05}, p_{95}]$ of the real

Table 3: Ablations on the synthetic cohort at $\beta_U = 3.5$. The best is bolded and the second best is underlined except the oracle.

Variant	Transport	Balancing	Regularizer	MISE
WSCF	✓	✓	✓	**0.479**
WSCF-noJSR	✓	✓		0.538
WSCF-noOT		✓	✓	0.548
WSCF-noIPM	✓		✓	0.598
WSCF-noOT-noJSR		✓		0.557
WSCF-noIPM-noJSR	✓			0.560
WSCF-None				0.770
WSCF-noOT-noIPM			✓	0.819
WSCF-average	window average of the surrogate			<u>0.512</u>
WSCF-single	current surrogate only			0.727
WSCF-noSurrogate	no surrogate			2.965
WSCF-oracle	counterfactual surrogates			0.215

incentive score; a wider grid would extend below the support and violate positivity. Ground truth uses the marginalization of Eq.43. The raw panel itself enters only the Motivation Example.

C.3 Additional Experiment Results

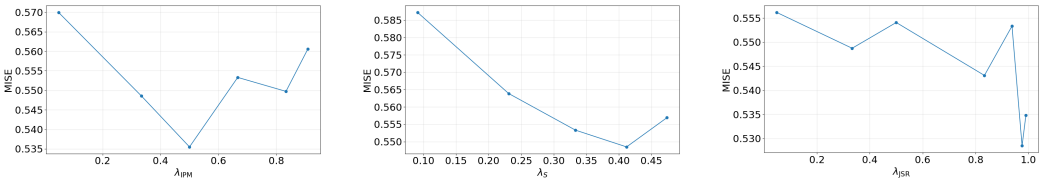
Baselines. We compare with static direct-estimation methods: naive A-Y regression, DRNet (Zhang et al., 2020), SCIGAN (Bica et al., 2020c), VCNet (Nie et al., 2021), ADMIT (Wang et al., 2022), ACFR (Kazemi & Ester, 2024); surrogate-based methods, including static "A,S-Y" (Hu et al., 2022), time varying Longitudinal-S (Huang et al., 2026), and LTEE (Cheng et al., 2021), LIRL (Cai et al., 2025), and LEARN (Yang et al., 2024b) for time-varying surrogates but static treatment/covaraites; and time-varying-dose methods without surrogate pivot: RMSN (Lim et al., 2018), CRN (Bica et al., 2020b), G-Net (Li et al., 2021), DCRN (Berrevoets et al., 2021). Ablation variants drop individual components (WSCF-noOT/noJSR/noIPM), all components (WSCF-None), with WSCF-oracle using true surrogates as a skyline. WSCF-LSTM use the same method but a different LSTM model, while WSCF use GRU sequential model.

Method	Synthetic						MIMIC-III	Industrial
MISE(SD)	$\beta_U = 1.0$	$\beta_U = 1.5$	$\beta_U = 2.0$	$\beta_U = 2.5$	$\beta_U = 3.0$	$\beta_U = 3.5$		
A-Y	6.27(1.188)	6.36(1.207)	6.38(1.217)	6.40(1.225)	6.41(1.238)	6.43(1.252)	0.985(0.024)	0.928(0.004)
A,S-Y	4.12(0.089)	4.07(0.033)	3.81(0.141)	3.81(0.127)	3.70(0.130)	3.64(0.111)	2.77(0.203)	0.806(0.007)
VCNet	3.62(0.283)	3.98(0.401)	4.27(0.188)	4.36(0.352)	4.34(0.028)	4.54(0.221)	7.65(2.094)	0.489(0.016)
ACFR	3.50(0.175)	3.53(0.267)	3.54(0.174)	3.47(0.180)	3.43(0.036)	3.35(0.198)	1.53(0.188)	0.565(0.018)
ADMIT	3.11(0.232)	2.96(0.157)	2.95(0.216)	2.95(0.141)	2.86(0.170)	2.87(0.127)	1.73(0.264)	0.562(0.013)
SCIGAN	2.74(0.238)	2.71(0.200)	2.64(0.180)	2.68(0.074)	2.79(0.189)	2.76(0.040)	5.46(3.301)	0.581(0.038)
DRNet	2.44(0.124)	2.46(0.150)	2.39(0.101)	2.40(0.105)	2.36(0.105)	2.41(0.082)	1.50(0.066)	0.463(0.006)
RMSN	3.38(0.061)	3.39(0.141)	3.29(0.080)	3.26(0.156)	3.21(0.100)	3.08(0.116)	2.92(0.175)	0.705(0.013)
DCRN	3.31(0.036)	3.40(0.221)	3.39(0.197)	3.15(0.170)	3.10(0.204)	3.08(0.255)	3.00(0.201)	0.706(0.017)
G-Net	3.70(0.288)	3.59(0.353)	3.55(0.301)	3.47(0.207)	3.17(0.295)	3.13(0.241)	2.88(0.214)	0.698(0.022)
CRN	3.43(0.084)	3.55(0.251)	3.45(0.327)	3.25(0.231)	3.21(0.282)	3.18(0.223)	2.88(0.139)	0.705(0.018)
LIRL	0.574(0.048)	0.563(0.090)	0.564(0.049)	0.568(0.077)	0.543(0.089)	0.535(0.041)	0.553(0.077)	0.655(0.008)
LEARN	0.567(0.042)	0.589(0.081)	0.573(0.055)	0.580(0.052)	0.573(0.075)	0.554(0.036)	0.585(0.066)	0.653(0.007)
LTEE	0.539(0.053)	0.540(0.058)	0.533(0.051)	0.519(0.036)	0.537(0.056)	0.528(0.032)	0.541(0.070)	0.657(0.008)
longitudinal-S	0.533(0.039)	0.543(0.039)	0.562(0.051)	0.569(0.055)	0.582(0.062)	0.602(0.060)	0.511(0.047)	0.658(0.006)
WSCF-None	**0.506(0.029)**	0.550(0.043)	0.617(0.044)	0.699(0.072)	0.688(0.052)	0.719(0.052)	0.546(0.045)	0.463(0.021)
WSCF-noOT	<u>0.513(0.055)</u>	**0.503(0.041)**	0.537(0.085)	0.594(0.127)	0.575(0.081)	0.617(0.136)	0.475(0.084)	0.466(0.027)
WSCF-noJSR	0.555(0.042)	0.544(0.030)	0.558(0.042)	0.546(0.039)	0.529(0.028)	0.517(0.024)	0.499(0.025)	<u>0.461(0.056)</u>
WSCF-noIPM	0.527(0.065)	0.528(0.020)	0.543(0.051)	0.537(0.038)	<u>0.517(0.020)</u>	<u>0.506(0.048)</u>	0.514(0.026)	0.468(0.028)
WSCF-LSTM	0.522(0.026)	0.517(0.039)	<u>0.528(0.037)</u>	<u>0.519(0.031)</u>	0.528(0.024)	0.510(0.036)	<u>0.475(0.027)</u>	0.484(0.032)
WSCF (Ours)	0.528(0.022)	<u>0.506(0.042)</u>	**0.521(0.052)**	**0.496(0.028)**	**0.500(0.028)**	**0.504(0.027)**	**0.470(0.025)**	**0.433(0.025)**
WSCF-oracle	0.262(0.027)	0.254(0.026)	0.253(0.025)	0.253(0.030)	0.245(0.024)	0.244(0.026)	0.310(0.013)	0.172(0.004)

Table 2: Overall performance evaluation on the synthetic cohort over increasing β_U and two semi-synthetic cohorts. All baseline results are included. The best is bolded and the second best is underlined except the oracle.

Table 4: Training and evaluation cost, in wall-clock seconds per run on the synthetic cohort, at the three confounding levels at which the runs were timed.

Method	$\beta_U = 0.0$	2.0	3.5
A,S-Y	38	41	37
A-Y	19	18	18
VCNet	11	13	14
ACFR	13	15	14
ADMIT	11	12	12
SCIGAN	13	15	16
DRNet	10	12	13
DCRN	20	19	18
G-Net	20	18	18
CRN	23	24	22
longitudinal-S	33	25	22
LIRL	66	43	39
LEARN	68	43	38
LTEE	67	41	37
WSCF-None	33	25	22
WSCF-noOT	72	89	57
WSCF-noJSR	56	73	50
WSCF-noIPM	40	49	35
WSCF-LSTM	76	92	59
WSCF	78	89	64
WSCF-oracle	66	105	62



(a) λ_{IPM} (b) λ_S (c) λ_{JSR}

Figure 5: Sensitivity Analysis of Hyperparameters.

D Motivation Example

This appendix documents the study that motivates the framework. It uses an order-level panel from a real-world industrial platform, covering $\sim$210k orders placed by $\sim$13k users over a certain time of interest. For confidentiality, the cohort undergoes privacy-preserving reconstruction following Battocchi et al. (2021) via resampled moments and prediction models.

Three quantities are recorded for every order. The *dynamic incentive intervention score* is a per-order adjustment the platform applies, and it acts as the dose. *active time* is the accumulated time a user spends online, measured in seconds; the 3-day active time is therefore the total active time over the three days following an order, and it is the surrogate that the platform can observe early. The 7-day active time is the corresponding long-term outcome, since user retention is what the platform ultimately cares about. Where a name carries the suffix `first-start`, the window opens at the fourth order before the current one rather than at the current order itself, so that the history feeding the window is fully observed.

Figure6 traces the test error of the three sequence models against the length of the history made available: accuracy improves and then saturates, and the complete model is the only one of the three that improves at all, which motivates a windowed formulation rather than an unbounded history.

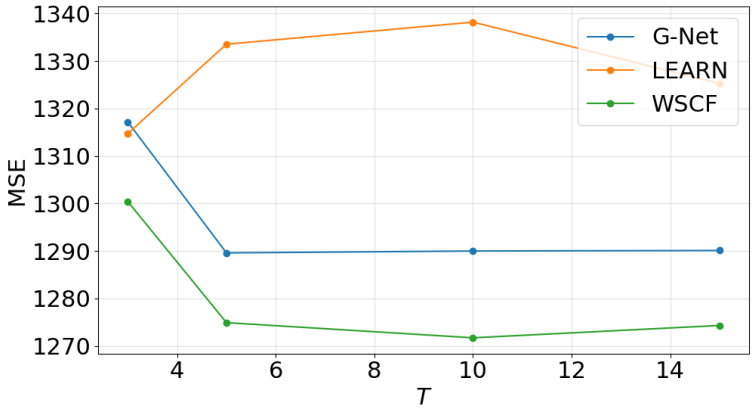


Figure 6: Test error against the length of the available sequence, for G-Net, LEARN and WSCF. Accuracy improves as more history is made available and then saturates near ten periods, and the complete model is the only one of the three that improves at all.

Table 5: The shape of the dosing trajectory rather than its total. Each statistic is computed over the last five orders, with the mean incentive score held in the middle quintile so that the units compared carry a comparable average dose. The trend and the last-minus-first difference are indistinguishable from zero, whereas the dispersion is not, so what carries signal is the amplitude of the oscillation and neither its direction nor its total.

Statistic	Outcome	Spearman
incentive-last5order-trend	3-day-active	0.0051
incentive-last5order-trend	7-day-active	0.0020
incentive-last5order-last-minus-first	3-day-active	0.0063
incentive-last5order-last-minus-first	7-day-active	0.0029
incentive-last5order-std	3-day-active	−0.1284
incentive-last5order-std	7-day-active	−0.1298

Table 6: Bridge score. The dose-to-surrogate and surrogate-to-outcome links are multiplied, so that a candidate must be movable and informative at once. The active time measures lead by a wide margin and the negative-response indicator is last, which is the screen used to fix the surrogate.

Rank	Candidate	$\mathrm{corr}(A, S)$	$\mathrm{corr}(S, Y)$	Bridge score
1	3day-active-first-start	0.1296	0.8532	0.1106
2	1day-active-first-start	0.1245	0.6370	0.0793
3	5day-active-mean	0.1298	0.6012	0.0780
4	1day-active	0.1138	0.6132	0.0698
5	5order-active	0.1416	0.3853	0.0546
6	negative-response-rate	0.1859	−0.1364	0.0254
7	negative-response-indicator	0.1023	−0.0853	0.0087