LEARNING UNDER COUNTERFACTUAL SHIFTS IN IMAGE ACQUISITION

Anonymous authors
Paper under double-blind review

ABSTRACT

A change in image acquisition can alter the optimal image-only predictor even when class proportions and the overall image distribution remain unchanged, because the class probabilities associated with a given image can change under the deployed policy. We study classification under a prespecified change in the acquisition policy using historical data alone. The intervention preserves historical assignment support, avoiding an additional assignment-positivity requirement. Under conditional exchangeability and missing-at-random labels, we identify the target Brier risk and its Bayes predictor and derive an efficient-influence-function-corrected training criterion. With linear conditional-risk regression on metadata, the criterion reduces exactly to a Brier objective with precomputed record weights, enabling gradient-based training without repeated regression fitting. We establish a uniform excess-risk bound with product-form nuisance errors and a squared propensity-error term. Simulations and a semi-synthetic Camelyon17 study show that targeting the planned policy provides most of the improvement, while the additional benefit of correction varies with the predictor and sample size.

1 INTRODUCTION

The way images are acquired often changes between training a classifier and deploying it. A pathology laboratory, for example, may plan to send more specimens to a second staining protocol that it already uses. No diagnosis changes, yet a classifier trained on past images can lose accuracy: the same tissue looks different under the two protocols, so the classifier learns a blend of their appearance–class relations, weighted by how often each protocol was used. Moving specimens between protocols changes these weights, and an appearance that indicated one class can come to indicate another. New expert labels would expose this only after deployment. We instead learn, from historical data alone, the classifier that is best for the planned policy.

We study the image-only setting, where acquisition provenance and pre-assignment metadata may be available during training but unavailable or unreliable at prediction time. We represent the planned change by an incremental propensity intervention on acquisition assignment, which multiplies the odds of one environment while preserving historical support (Kennedy, 2019). Under the causal and missing-label conditions in Section 2, historical data identify the prediction risk under this policy. Our target is the image-only classifier minimizing this counterfactual acquisition risk before the policy is deployed.

This target differs from domain adaptation, which uses target-domain samples (Ganin et al., 2016), and from covariate-shift reweighting, which assumes that the conditional label distribution given the image remains fixed (Shimodaira, 2000). It also differs from domain generalization and robust objectives that hedge against unspecified shifts (Gulrajani & Lopez-Paz, 2021; Sagawa et al., 2020; Krueger et al., 2021): here the planned policy determines the risk to minimize. Related work matches objects or learns invariant features (Mahajan et al., 2021; Lv et al., 2022); Oberst et al. (2021) target specified intervention moments in linear models with proxies.

Counterfactual prediction methods instead address runtime confounding, learn classifiers for treatment-specific outcomes, or use counterfactual constructions for general out-of-distribution generalization (Coston et al., 2020; Kim et al., 2022; Yang et al., 2026). Logged-bandit policy learning also uses weighting and doubly robust scores under a change of action policy (Swaminathan

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.

& Joachims, 2015; Dudík et al., 2011). Here, acquisition changes the images themselves, and we learn an image-only classifier for a specified acquisition policy.

Our main contributions are threefold. First, we identify the acquisition-policy risk and its image-only Bayes predictor with selectively observed labels, and derive the efficient influence function by combining incremental-intervention estimation (Kim et al., 2021), missing-label augmentation (Tsiatis, 2006), and cross-fitted orthogonal learning (Chernozhukov et al., 2018; Foster & Syrgkanis, 2023). Second, we construct a cross-fitted training criterion whose nuisance bias consists of product terms and a squared propensity-error term; with linear conditional-risk regression, the criterion and its gradient reduce exactly to a Brier objective with precomputed record weights, enabling standard gradient-based training without repeated regression fitting. Third, we establish a uniform excess-risk guarantee relative to the best predictor in the chosen class.

In simulations, objectives that ignore the planned policy remain suboptimal even at their population optima, while correction can reduce sensitivity to propensity estimation. In a semi-synthetic Camelyon17/WILDS study on held-out slides, targeting the planned policy accounts for most of the improvement; with estimated propensities, correction provides additional gains for a four-parameter head and a multilayer perceptron on the same inputs, although not uniformly across settings.

2 Counterfactual target and identification

We now formalize the setting. Let C be pre-assignment covariates, $A \in \{0, 1\}$ the acquisition environment, $H \in \mathcal{H}$ the resulting image, and $Y \in \{0, \dots, K-1\}$ a class fixed before acquisition, with one-hot vector $\mathbf{Y}$. Let R indicate label observation, so the observed data are $O = (C, A, H, R, R\mathbf{Y})$. The deployed predictor $f : \mathcal{H} \to \Delta^{K-1}$ uses only H, where Δ^{K-1} is the probability simplex. Let H^a denote the potential image under environment a.

Write $\pi_0(c) = P(A = 1 \mid C = c)$ and $d_\delta(c) = 1 + (\delta - 1)\pi_0(c)$, with

$$q_{\delta,1}(c) = \frac{\delta \pi_0(c)}{d_\delta(c)}, \qquad q_{\delta,0}(c) = \frac{1 - \pi_0(c)}{d_\delta(c)}. \tag{1}$$

For fixed $0 < \delta < \infty$, this multiplies the conditional odds of environment 1 by δ. Moreover, $q_{\delta,a}(C) > 0$ exactly where $P(A = a \mid C) > 0$, so the intervention preserves historical assignment support. The policy is indexed by the odds multiplier δ, while its assignment probabilities depend on the historical propensity π_0.

Let $A^{q_\delta} \sim q_\delta(\cdot \mid C)$, independently of (Y, H^0, H^1) given C, and set $H^{q_\delta} = H^{A^{q_\delta}}$. For a loss ℓ, the target risk is

$$\mathcal{R}_\delta(f) = \mathbb{E}[\ell\{f(H^{q_\delta}), Y\}] .$$

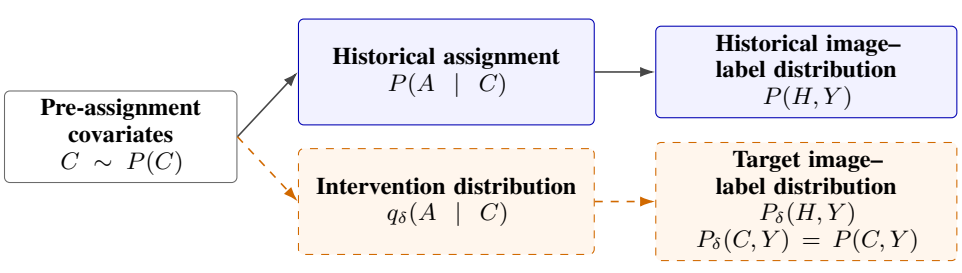

Figure 1: Changing the acquisition policy. The intervention law P_δ replaces $P(A \mid C)$ with $q_\delta(A \mid C)$, preserving $P(C, Y)$ while allowing the image–label distribution to change.

We assume throughout that A indexes well-defined and feasible acquisition environments for a common source population, with common label semantics and no interference in image formation. Identification then relies on the following four conditions, standard for stochastic interventions and missing-data problems (Kennedy, 2019; Tsiatis, 2006).

Assumption 1 (Consistency)**.** If $A = a$, then $H = H^a$. The observed image is the potential image under the assigned environment.

Assumption 2 (Conditional exchangeability)**.** $(H^a, Y) \perp A \mid C$, for $a \in \{0, 1\}$. Within strata of C, assignment carries no additional information about the potential image–label pair.

Assumption 3 (Missing at random (MAR))**.** $R \perp Y \mid (A, C, H)$. Given assignment, covariates, and the image, label availability does not further depend on the class.

Assumption 4 (Label positivity)**.** $P(R = 1 \mid A, C, H) > 0$ almost surely. Labels can be observed throughout the observed-data support.

Conditional exchangeability is imposed on the potential image–label pair (H^a, Y) because acquisition changes the image used for prediction. Note that no separate assignment-positivity condition is required, because the incremental intervention preserves historical assignment support.

For later use, define the label-observation probability and class-probability regression by

$$e_0(A, C, H) = P(R = 1 \mid A, C, H), \quad \mu_0(A, C, H) = \mathbb{E}(\mathbf{Y} \mid A, C, H, R = 1).$$

Both may depend on the post-acquisition image H.

Proposition 1. *Under consistency and conditional exchangeability, the intervention law is*

$$p_\delta(c, a, h, y) = p(c)q_{\delta,a}(c)p(h, y \mid a, c) = p(c, y)q_{\delta,a}(c)p(h \mid a, c, y).$$

Hence $P_\delta(C, Y) = P(C, Y)$ and $P_\delta(Y) = P(Y)$, while $P_\delta(H)$ and $P_\delta(Y \mid H)$ may change. Under MAR and label positivity, $P(Y \mid A, C, H) = P(Y \mid A, C, H, R = 1)$, identifying $\mathcal{R}_\delta(f)$ from O.

Thus the intervention can preserve the class distribution while changing the conditional class law given the image, and hence the image-only Bayes rule. A simple example makes this explicit. Let C be constant, $Y \sim \text{Bernoulli}(1/2)$, and A be independent of Y with $P(A = 1) = 1/4$. Suppose $H^0 = h_Y$ and $H^1 = h_{1-Y}$, with $h_0 \neq h_1$. The intervention with $\delta = 9$ changes $P(A = 1)$ from $1/4$ to $3/4$. Both $P(Y)$ and $P(H)$ remain uniform, but $P(Y = 1 \mid H = h_1)$ reverses from $3/4$ to $1/4$, so the image-only Bayes rule changes. Figure 1 summarizes the change from the historical to the intervention distribution.

3 Corrected risk learning

We now turn the policy risk identified in Section 2 into a criterion for training an image-only classifier. We account for missing labels and propensity estimation, give a way to optimize the criterion, and bound the learned classifier's risk relative to the best predictor in the chosen class.

The multiclass Brier loss is $L_f = \|f(H) - \mathbf{Y}\|_2^2$ (Brier, 1950), and $\mathcal{R}_{\delta,B}(f)$ denotes its risk under the planned policy. To express this risk using historical observations, define the intervention weight on the observed support by

$$w_\delta(C, A) = \frac{q_{\delta,A}(C)}{P(A \mid C)} = \frac{\delta^A}{1 + (\delta - 1)\pi_0(C)}. \tag{2}$$

It satisfies $\mathbb{E}(w_\delta \mid C) = 1$ and $\min(\delta, \delta^{-1}) \leq w_\delta \leq W_\delta := \max(\delta, \delta^{-1})$, without strict assignment positivity. With complete labels, Proposition 1 gives $\mathcal{R}_{\delta,B}(f) = \mathbb{E}(w_\delta L_f)$.

With selectively observed labels, L_f is unavailable when $R = 0$. We replace $\mathbf{Y}$ in the linear term of the Brier loss by the augmented inverse-probability weighted (AIPW) pseudo-outcome

$$Z_0 = \mu_0(A, C, H) + \frac{R\{\mathbf{Y} - \mu_0(A, C, H)\}}{e_0(A, C, H)}. \tag{3}$$

When $R = 0$, this equals μ_0; observed labels supply the correction. For candidate label models $(\bar{e}, \bar{\mu})$ with $\bar{e} > 0$, replacing (e_0, μ_0) by these models gives $Z(\bar{e}, \bar{\mu})$, which satisfies

$$\mathbb{E}\{Z(\bar{e}, \bar{\mu}) \mid A, C, H\} - \mu_0 = \frac{\bar{e} - e_0}{\bar{e}}(\bar{\mu} - \mu_0), \tag{4}$$

so the conditional bias is zero when either $\bar{e} = e_0$ or $\bar{\mu} = \mu_0$. We do not clip the fitted pseudo-outcomes, since clipping generally breaks Equation (4).

Under Proposition 1 and Assumptions 3 and 4, the target Brier risk is therefore

$$\mathcal{R}_{\delta,B}(f) = \mathbb{E}\big[w_\delta\{\|f(H)\|_2^2 - 2f(H)^T Z_0 + 1\}\big].$$

For each class k, the image-only Bayes predictor f^*_δ and its excess-risk identity are

$$f^*_{\delta,k}(h) = \frac{\mathbb{E}\{w_\delta Z_{0,k} \mid H = h\}}{\mathbb{E}(w_\delta \mid H = h)},$$

$$\mathcal{R}_{\delta,B}(f) - \mathcal{R}_{\delta,B}(f^*_\delta) = \mathbb{E}_\delta\|f(H) - f^*_\delta(H)\|_2^2, \tag{5}$$

where $\mathbb{E}_\delta$ denotes expectation under the intervention law.

We use Brier loss because its minimizer gives the target class probabilities and its excess risk has the squared-distance form in Equation (5). For each finite pseudo-outcome, its quadratic loss also remains bounded as $f(H)$ varies over the probability simplex; a log loss with negative pseudo-outcome coordinates can instead be unbounded below.

The weighting above still uses the unknown historical propensity π_0. Simply substituting an estimate can introduce first-order bias. The correction must account for how both the weights and policy probabilities change with π_0. A fixed-policy AIPW score treats the policy as given independently of π_0, so it omits the latter contribution. To construct a correction, we use conditional mean losses. For a fixed predictor f, let

$$\eta_f(A, C, H) = \mathbb{E}(L_f \mid A, C, H) = 1 + \|f(H)\|_2^2 - 2f(H)^T \mu_0(A, C, H).$$

The label-corrected loss T_f and its policy-weighted conditional mean $m_{q_\delta,f}$ are

$$T_f = \eta_f + \frac{R}{e_0}(L_f - \eta_f) = 1 + \|f(H)\|_2^2 - 2f(H)^T Z_0, \qquad m_{q_\delta,f}(C) = \mathbb{E}(w_\delta T_f \mid C). \tag{6}$$

On the observed support, $m_{a,f}(c) = \mathbb{E}(T_f \mid A = a, C = c)$ is the expected loss in environment a given metadata c; its policy average is $m_{q_\delta,f}(c) = \sum_a q_{\delta,a}(c) m_{a,f}(c)$.

Proposition 2. *Under the preceding conditions, and assuming additionally that* $\mathbb{E}\{1/e_0(A, C, H)\} < \infty$*, the efficient influence function (EIF) for the risk of a fixed measurable* f *in the unrestricted MAR observed-data model is*

$$D_{\delta,f}(O) = m_{q_\delta,f}(C) - \mathcal{R}_{\delta,B}(f) + w_\delta\{T_f - m_{q_\delta,f}(C)\}. \tag{7}$$

With complete labels, Equation (7) is the incremental-intervention influence function of Kennedy (2019) applied to the candidate loss; the pseudo-outcome extends it to MAR labels (Appendices D and E).

3.1 Corrected learning with precomputed weights

We train f by minimizing an estimate of the target Brier risk. Since the influence function in Equation (7) has mean zero, this risk is the expectation of $w_\delta T_f + (1 - w_\delta) m_{q_\delta,f}(C)$. We estimate the unknown functions on separate data folds and rewrite the resulting criterion using record weights computed before classifier training.

Propensity and label models. Split n independent records into three folds: S_η for fitting propensity and label models, S_m (of size n_m) for conditional-risk regression, and S_ℓ (of size n_ℓ) for averaging the corrected scores. If a patient contributes several images, keep them in one fold to prevent patient overlap across folds. On S_η, fit (i) $\widehat{\pi}(c)$ for the propensity $\pi_0(c) = P(A = 1 \mid C = c)$, from the metadata C only; (ii) $\widehat{e}(a, c, h)$ for the label-observation probability e_0, bounded below by a fixed $\epsilon > 0$ (e.g., by truncation); and (iii) $\widehat{\mu}(a, c, h)$, a vector of class probabilities for μ_0, from the labeled records. The label models use (A, C, H), whereas the deployed classifier uses only H. The fitted models are then held fixed: they are evaluated, not refitted, on S_m and S_ℓ. Substituting them for π_0, e_0, μ_0 in Equations (1) to (3) and (6) gives, for each record, the estimated policy, weight, label pseudo-outcome, and loss of a candidate f:

$$\widehat{d} = 1 + (\delta - 1)\widehat{\pi}, \quad \widehat{w}_\delta = \delta^A/\widehat{d}, \quad \widehat{q}_{\delta,1} = \delta\widehat{\pi}/\widehat{d}, \quad \widehat{q}_{\delta,0} = (1 - \widehat{\pi})/\widehat{d},$$

$$\widehat{Z} = \widehat{\mu} + R(\mathbf{Y} - \widehat{\mu})/\widehat{e}, \qquad \widehat{T}_f = 1 + \|f(H)\|_2^2 - 2f(H)^T \widehat{Z}.$$

The policy probabilities, weights, and label pseudo-outcomes remain fixed during training; $\widehat{T}_f$ changes with f. Up to a term free of f, $\widehat{T}_f = \|f(H) - \widehat{Z}\|_2^2$; with complete labels, $\widehat{e} \equiv 1$ gives $\widehat{Z} = \mathbf{Y}$. Because $\widehat{w}_\delta$ always lies between δ^{-1} and δ, $\widehat{\pi}$ needs no truncation.

Conditional-risk regressions (fold S_m). For a candidate classifier f, $m_{a,f}(c)$ is its expected Brier loss in environment a given metadata c. We estimate it on S_m, separately for each environment, by ridge regression of $\widehat{T}_f$ on bounded J-dimensional metadata features $\psi(C)$, such as splines or stratum indicators (Appendix H.4). The responses and fitted risks change with f. Ridge predictions are linear in their responses, which will let us compute the correction without refitting these regressions at every training update:

$$\widehat{m}_{a,f}(c) = \psi(c)^T \widehat{\beta}_{a,f}, \qquad \widehat{\beta}_{a,f} = (\widehat{G}_a + \lambda I)^{-1} \frac{1}{n_m} \sum_{j \in S_m} \mathbf{1}(A_j = a)\, \psi(C_j)\, \widehat{T}_{f,j}, \tag{8}$$

where $\widehat{G}_a = n_m^{-1} \sum_{j \in S_m} \mathbf{1}(A_j = a)\, \psi(C_j)\psi(C_j)^T$ and $\lambda > 0$ is the ridge penalty; I is the J-dimensional identity matrix. Under the planned policy, the fitted conditional risk is $\widehat{m}_{q_\delta,f} = \sum_a \widehat{q}_{\delta,a} \widehat{m}_{a,f}$. The features, λ, and the predictor class $\mathcal{F}_n$ (e.g., image classifiers with softmax outputs) are fixed in advance or chosen using S_η only.

Corrected criterion (fold S_ℓ). We average the estimated weighted loss and correction over S_ℓ:

$$\widehat{\mathcal{J}}_\delta(f) = \frac{1}{n_\ell} \sum_{i \in S_\ell} \widehat{\Gamma}_{\delta,f}(O_i), \qquad \widehat{\Gamma}_{\delta,f} = \widehat{w}_\delta \widehat{T}_f + (1 - \widehat{w}_\delta)\, \widehat{m}_{q_\delta,f}(C). \tag{9}$$

We minimize $\widehat{\mathcal{J}}_\delta(f)$ over $\mathcal{F}_n$, allowing an optimization error $\varepsilon^{\text{opt}}_{\text{aug}} \geq 0$. The returned classifier $\widehat{f}^{\text{aug}}_\delta \in \mathcal{F}_n$ satisfies $\widehat{\mathcal{J}}_\delta(\widehat{f}^{\text{aug}}_\delta) \leq \inf_{f \in \mathcal{F}_n} \widehat{\mathcal{J}}_\delta(f) + \varepsilon^{\text{opt}}_{\text{aug}}$.

Precomputed weights. Holding the conditional-risk regressions fixed while updating f would give an incorrect gradient. Instead, we use their linearity in the responses. Each fitted value $\widehat{m}_{a,f}(c)$ is a weighted sum of the responses $\widehat{T}_{f,j}$, $j \in S_m$, with weights free of f. Substituting this into the second term of Equation (9) and exchanging the order of summation moves that term onto S_m, with record weights

$$\omega_j = \psi(C_j)^T (\widehat{G}_{A_j} + \lambda I)^{-1} \frac{1}{n_\ell} \sum_{i \in S_\ell} (1 - \widehat{w}_{\delta,i})\, \widehat{q}_{\delta,A_j}(C_i)\, \psi(C_i), \qquad j \in S_m, \tag{10}$$

which need two J-dimensional linear solves and no knowledge of f. Hence, exactly for every f,

$$\widehat{\mathcal{J}}_\delta(f) = \frac{1}{n_\ell} \sum_{i \in S_\ell} \widehat{w}_{\delta,i} \widehat{T}_{f,i} + \frac{1}{n_m} \sum_{j \in S_m} \omega_j \widehat{T}_{f,j}. \tag{11}$$

We therefore update only the classifier parameters, using fixed record weights $\widehat{w}_{\delta,i}/n_\ell$ on S_ℓ and ω_j/n_m on S_m. Both folds contribute to training; the loss terms on S_m retain the dependence of the conditional-risk regressions on f. The ω_j may be negative and should not be clipped, which would break Equation (11). Differentiating this weighted objective gives the exact gradient (Appendix G). At $\delta = 1$, $\widehat{w}_\delta \equiv 1$ and $\omega_j = 0$, so training reduces to the unweighted Brier loss on S_ℓ with $\widehat{Z}$ in place of $\mathbf{Y}$.

Cross-fitted learning. For equal-sized S_m and S_ℓ, we exchange their roles and average the two criteria. For $K_{\text{cf}} \geq 3$ folds, we rotate all three roles and average the criteria with weights proportional to evaluation-fold sizes. We train one shared classifier from a common class fixed before observing the evaluation folds. Each score uses fits from outside its evaluation fold; Propositions 5 and 6 give the constructions and their guarantees.

3.2 Uniform excess-risk guarantee

We now bound the target risk of the learned classifier relative to the best predictor in $\mathcal{F}_n$. To examine the effect of propensity estimation, fix f and the fitted functions and first ignore label-model estimation error, so that $\widehat{T}_f = T_f$. Given C, the weighted loss and correction have means $(d_\delta/\widehat{d}) m_{q_\delta,f}$ and $(1 - d_\delta/\widehat{d}) \widehat{m}_{q_\delta,f}$. Their sum is $m_{q_\delta,f} + (1 - d_\delta/\widehat{d})(\widehat{m}_{q_\delta,f} - m_{q_\delta,f})$, where $1 - d_\delta/\widehat{d} = (\delta - 1)(\widehat{\pi} - \pi_0)/\widehat{d}$. Policy weighting omits the conditional-risk correction (Appendix F)

and has bias proportional to the propensity error. The corrected score's bias is a product of propensity and conditional-risk errors. The latter also depends on $\widehat{\pi}$ through $\widehat{q}_\delta$, giving a squared propensity-error term. With missing labels, $\widehat{Z}$ adds the product of label errors in Equation (4).

Let P be the historical law, with marginals P_C, P_{ACH}. Write $Pg = \mathbb{E}_P[g(O)]$ for expectation over a fresh observation, with fitted functions fixed, and $P_m g, P_\ell g$ for averages over S_m, S_ℓ. We use the usual scalar L_2 norms; for vector v, $\|v\|^2_{L_2(P_{ACH};\ell_2)} = \mathbb{E}_P\|v(A,C,H)\|_2^2$, where $\|\cdot\|_2$ is Euclidean. Let $q_a = q_{\delta,a}$, and define

$$\begin{aligned} r_\pi &= \|\widehat{\pi} - \pi_0\|_{L_2(P_C)}, & r_e &= \|\widehat{e} - e_0\|_{L_2(P_{ACH})}, \\ r_\mu &= \|\widehat{\mu} - \mu_0\|_{L_2(P_{ACH};\ell_2)}, & \mathcal{E}_\Gamma &= \sup_{f\in\mathcal{F}_n} |(P_\ell - P)\widehat{\Gamma}_{\delta,f}|, \\ r_{\mathrm{arm}} &= \sup_{f\in\mathcal{F}_n} \left\| \sum_a q_a(m_{a,f} - \widehat{m}_{a,f}) \right\|_{L_2(P_C)}, & M_\Delta &= \sup_{f\in\mathcal{F}_n} \|\widehat{m}_{1,f} - \widehat{m}_{0,f}\|_\infty. \end{aligned} \tag{12}$$

Here r_{arm} measures error in the policy-averaged conditional risk, uniformly over f; M_Δ bounds the fitted risk difference between environments.

Set $B_\delta = |\delta - 1|/\min(1,\delta)$ and $C_\delta = \delta|\delta - 1|/\min(1,\delta)^3$.

Theorem 1. *Under Assumptions 1–4, assume that all units across S_η, S_m, S_ℓ are independent and identically distributed (i.i.d.). Conditional on S_η, assume the fitted functions satisfy $\widehat{\pi} \in [0,1]$, $\widehat{e} \in [\epsilon, 1]$, and $\widehat{\mu} \in \Delta^{K-1}$ almost surely. Let $\sup_c \|\psi(c)\|_2 \le B$, $\lambda > 0$, and let $\mathcal{F}_n$ be a nonempty measurable simplex-valued class fixed conditional on S_η and before observing S_m, S_ℓ. Then*

$$\mathcal{R}_{\delta,B}(\widehat{f}^{\mathrm{aug}}_\delta) - \inf_{f\in\mathcal{F}_n} \mathcal{R}_{\delta,B}(f) \le 2\mathcal{E}_\Gamma + 2B_\delta r_\pi r_{\mathrm{arm}} + 2C_\delta M_\Delta r_\pi^2 + \frac{4W_\delta}{\epsilon} r_e r_\mu + \varepsilon^{\mathrm{opt}}_{\mathrm{aug}}. \tag{13}$$

Moreover $M_\Delta \le 2B^2/(\epsilon\lambda)$. The infimum need not be attained and $\mathcal{F}_n$ need not be convex.

The bound separates sampling and optimization error from the bias of the fitted models. A persistent conditional-risk error thus leaves first-order sensitivity to propensity error, and the bias reduction does not imply smaller finite-sample excess risk than weighting alone: with complete labels and a convex or well-specified class, Brier curvature can also make the excess risk of weighting quadratic in the propensity error (Proposition 4). Appendix H gives the proof, a decomposition of r_{arm}, and a Rademacher bound on $\mathcal{E}_\Gamma$ (Appendix H.3).

Corollary 1. *For fixed δ, ϵ, if for some $\rho_n \downarrow 0$,*

$$\mathcal{E}_\Gamma + \varepsilon^{\mathrm{opt}}_{\mathrm{aug}} = O_p(\rho_n), \qquad r_\pi r_{\mathrm{arm}} + M_\Delta r_\pi^2 + r_e r_\mu = o_p(\rho_n), \tag{14}$$

then

$$\mathcal{R}_{\delta,B}(\widehat{f}^{\mathrm{aug}}_\delta) - \inf_{f\in\mathcal{F}_n} \mathcal{R}_{\delta,B}(f) \le 2\mathcal{E}_\Gamma + \varepsilon^{\mathrm{opt}}_{\mathrm{aug}} + o_p(\rho_n), \tag{15}$$

and the class-relative excess is $O_p(\rho_n)$.

Thus nuisance errors of $o_p(\rho_n^{1/2})$ suffice when $M_\Delta = O_p(1)$, as for stratum indicators, although unequal product rates are allowed. In particular, $o_p(n^{-1/4})$ nuisance errors give $o_p(n^{-1/2})$ bias under this boundedness condition. Correct conditional-risk regressions alone do not remove the r_π^2 term, so the assignment correction, unlike the label correction, is not doubly robust.

4 Experiments

We compare predictors trained for the historical distribution or the planned acquisition policy, then assess whether correction improves learning for that policy. We use a simulation and a semi-synthetic experiment based on Camelyon17 images and metadata distributed through WILDS (Bándi et al., 2019; Koh et al., 2021), with evaluation on patches from held-out slides. The identification conditions hold by design in these experiments; they are not established here for observational acquisition systems. The primary experiments use $\delta = 9$, with 400 independent replications per configuration. We report Bonferroni-adjusted simultaneous 95% confidence intervals within prespecified contrast families; full protocols and additional sensitivity analyses are in Appendix A.

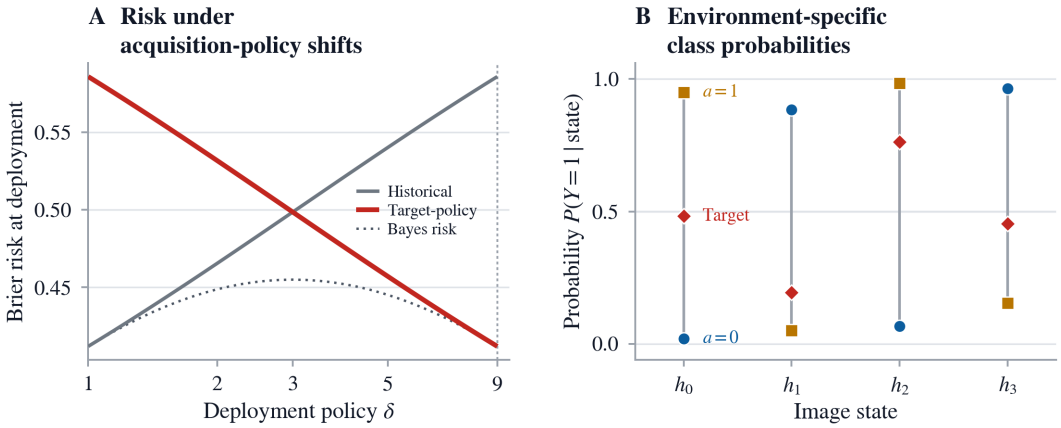

Figure 2: A planned acquisition policy changes the optimal predictor. **A**: risks of two fixed predictors as δ varies; the vertical line marks the planned policy $\delta = 9$. **B**: class probabilities under each environment and the target policy; $h_0, \ldots, h_3$ are the four binary-feature combinations. All quantities are population values.

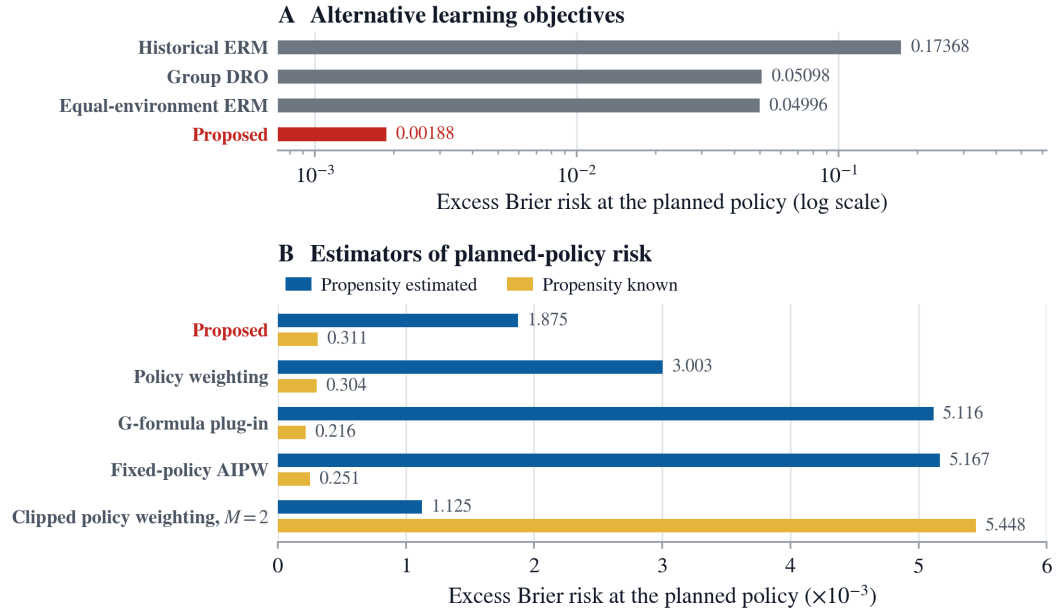

Figure 3: Simulation results at $\delta = 9$. Excess Brier risks are relative to the planned-policy optimum in the shared predictor class. **A**: alternative objectives' population optima (gray) and the proposed learner's mean finite-sample excess risk (red), on a log scale. **B**: learners using estimated (blue) or known (yellow) propensities; $M = 2$ is the weight-clipping threshold. ERM denotes empirical risk minimization; DRO denotes distributionally robust optimization.

4.1 Simulation Study

We first consider a classification problem with two binary input features. Changing the acquisition environment reverses the association between one feature and the binary label. A predictor assigns a probability of $Y = 1$ to each of the four input combinations. The mean probability of assignment to environment $A = 1$ rises from 0.325 under the historical policy to 0.675 under the intervention. The historical and planned optima differ because the policy changes $P(Y \mid H)$ (Figure 2). Each replication contains 8,192 observations and uses five-fold cross-fitting (Proposition 6). Excess risk is measured relative to the best predictor in this class under the planned policy.

The alternative objectives minimize historical risk, the equally weighted average of the two environment risks, or their maximum (group DRO). All three are optimized under the population law over the same predictor class. Even these population optima have higher planned-policy excess risk than the proposed learner fitted to 8,192 observations (Figure 3A).

Correction at a fixed target. Figure 3B compares estimators of the same planned-policy risk. With estimated propensities, correction reduces mean excess risk by 37.6% relative to policy weighting. The fixed-policy AIPW comparator omits the contribution from the policy's dependence on the historical propensity (Proposition 2). Its excess risk with estimated propensities is close to that of the G-formula plug-in, which minimizes the average fitted conditional risk under the estimated target policy. No estimator is uniformly best in these comparisons: the ranking changes between estimated and known propensities, and capping the intervention weights at 2 increases excess risk when the propensity is known. Appendix Table 1 reports the estimator and clipping results.

Nuisance errors and missing labels. Separate population calculations examine how errors in the fitted models affect the learned predictor (Appendix Figure 5). In the label experiment, either a correct label-observation model or a correct class-probability regression removes the excess risk. In the propensity experiment, correct conditional-risk regressions alone do not suffice.

We then hide labels according to an MAR mechanism and compare with a reference learner that observes every label. Label augmentation accounts for most of the recovery relative to complete-case policy weighting, with correction providing a further reduction in excess risk. Together, augmentation and correction close 98.2% of the gap between complete-case policy weighting and the full-label reference (Appendix Table 2).

4.2 Semi-synthetic Camelyon17 study

The Camelyon17 experiment combines real tissue morphology with a simulated acquisition process. The fixed training and held-out pools contain 12,288 patches from 23 whole-slide images (WSIs) and 4,096 patches from four disjoint WSIs, respectively. Each replication samples training patches with replacement and generates synthetic labels from their tumor annotations. Acquisition reverses the association between a binary red–blue color cue V and the synthetic label while preserving grayscale morphology (Figure 4A). Evaluation uses the fixed held-out pool; the reported results and intervals condition on both image pools.

On S_η, we fit the propensity model and form a morphology score g from a ridge prediction of synthetic Y based on principal component analysis (PCA) of fixed ResNet-18 features of grayscale images (He et al., 2016). A four-parameter head interpolates class probabilities over (g, V), allowing exact optimization and population-risk evaluation. The guarantee in Theorem 1 applies to this S_η-fitted representation and predictor class under the distribution of synthetic records generated from the fixed training image pool, not to a raw-image Bayes rule or the held-out population.

Each replication generates 8,192 synthetic records: 4,096 form S_η, and two disjoint folds S_1, S_2 of 2,048 records alternate as S_m and S_ℓ (Proposition 5). Correction and matched weighting train on the same 4,096 $S_1 \cup S_2$ records and use the same S_η-fitted propensity and morphology score.

To assess policy targeting, historical ERM and full-sample policy weighting train on all 8,192 records with the same morphology scores. The policy-set minimax comparator minimizes the largest empirical policy-weighted risk over $\delta \in \{1, 3, 9\}$.

In this subsection, excess risk is measured relative to the raw-image Bayes risk under the known semi-synthetic law on the fixed held-out image pool. Relative to historical ERM, policy weighting trained on all 8,192 records reduces mean held-out Brier risk by 29.8%. In the matched comparison with estimated propensities, correction reduces mean excess risk by 9.38%. Figure 4C and Appendix Table 3 report contrasts and confidence intervals.

Neural-network predictors. To evaluate the criterion with a neural-network predictor, we fit a one-hidden-layer multilayer perceptron (MLP) to the S_η-fitted 64-dimensional PCA features and cue V. The predictor uses $n \in \{4096, 8192\}$ training records, with another n records in S_η. Within each MLP configuration, all criteria share training records, initialization, optimizer, and a weight penalty selected by cross-validation for estimated-propensity weighting. Policy weighting outperforms

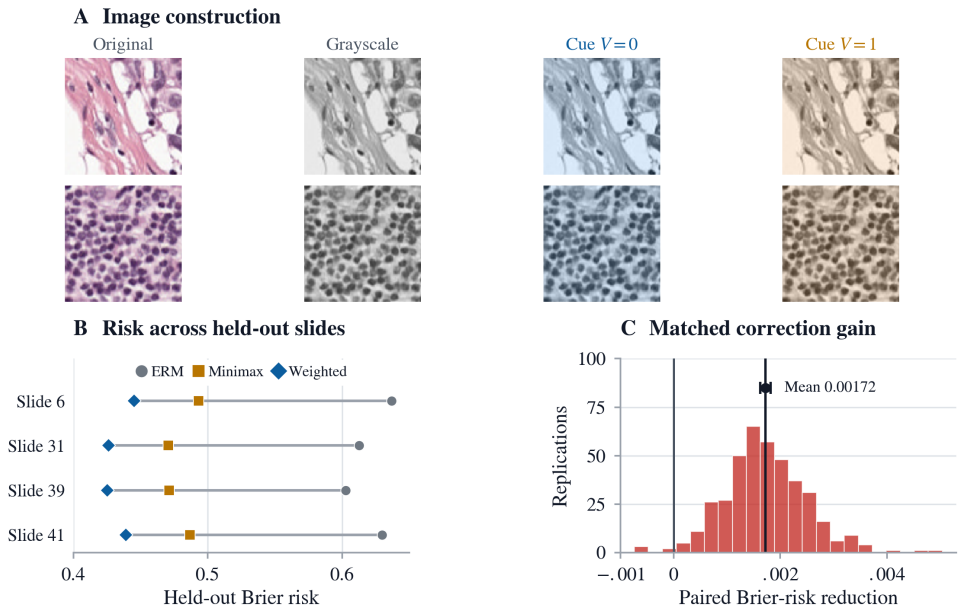

Figure 4: Camelyon17/WILDS four-parameter experiment. **A**: an original hematoxylin-and-eosin (H&E)-stained patch, grayscale morphology, and two synthetic renderings with a binary color cue V, preserving that morphology. **B**: mean Brier risks on four fixed held-out whole-slide images (WSIs; WILDS slide indices), with 8,192 training records per head. **C**: matched weighting-minus-corrected risk reductions over 400 replications. The point and bar show the mean and Bonferroni-adjusted simultaneous 95% confidence interval.

historical ERM at both sizes. With estimated propensities, correction increases mean excess risk by 1.67% at $n = 4096$ and reduces it by 3.83% at $n = 8192$, relative to matched weighting (Appendix Table 4).

To compare with the four-parameter predictor using the same input information, we also train an MLP on (g, V) at $n = 4096$. With estimated propensities, correction reduces mean excess risk by 8.35% relative to matched weighting (Appendix Table 5).

Comparisons between predictors do not isolate flexibility because their regularization and optimization also differ. The regularized MLP fits do not directly test Theorem 1; Appendix Table 6 reports further comparisons at $\delta = 3$ and with intercept-only conditional-risk regressions.

5 Discussion and Limitations

When to correct. Correction improves several comparisons with estimated propensities, but is not uniformly beneficial. It increases risk when Gaussian covariates are omitted from the simulation propensity model (Appendix A). Its benefit should therefore be assessed against matched weighting for the intended nuisance specification, predictor, and sample size.

Identification requires the conditions in Section 2 and a stable law of (C, Y, H^0, H^1). Dataset or center identifiers alone do not establish consistency, exchangeability, or missing-at-random label observation. Small label-observation probabilities can inflate variance; truncating inverse label-observation weights may introduce bias and cannot restore identification in strata where labels are never observed.

The semi-synthetic results and intervals condition on the fixed training and held-out image pools; generalization to clinical populations or other sites requires further study. End-to-end image training and two-fold training with negative combined per-record weights remain unevaluated.

AI USE STATEMENT

Generative AI tools assisted with literature search, implementation of analysis and figure-generation code, and manuscript drafting and language editing. The authors take responsibility for the final text, mathematical claims, proofs, code, and numerical results.

ETHICS STATEMENT

The semi-synthetic experiment uses publicly available histopathology images; we collected no new patient data. Its results do not establish clinical validity. Any use in patient care would require validation in the intended population, including assessment of performance across patient groups.

REPRODUCIBILITY STATEMENT

The identification assumptions are stated in Section 2, and the learning procedure and its guarantee are given in Sections 3.1 and 3.2. The appendices provide the identification and influence-function derivations, complete proofs, and the experimental data-generating processes and analysis procedures. The supplementary source package contains configurations, implementations, and saved per-replication outputs, with instructions for checking the reported results, regenerating the experimental figures and tables, and rerunning the experiments.

REFERENCES

Péter Bándi, Oscar Geessink, Quirine Manson, et al. From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge. *IEEE Transactions on Medical Imaging*, 38(2):550–560, 2019. doi: 10.1109/TMI.2018.2867350.

Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018. doi: 10.1111/ectj.12097.

Amanda Coston, Edward H. Kennedy, and Alexandra Chouldechova. Counterfactual predictions under runtime confounding. In *Advances in Neural Information Processing Systems*, volume 33, pp. 4150–4162, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2b64c2f19d868305aa8bbc2d72902cc5-Abstract.html.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, 2011. URL https://icml.cc/2011/papers/554_icmlpaper.pdf.

Dylan J. Foster and Vasilis Syrgkanis. Orthogonal statistical learning. *The Annals of Statistics*, 51(3): 879–908, 2023. doi: 10.1214/23-AOS2258.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=lQdXeXDoWtI.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. URL https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

Edward H. Kennedy. Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656, 2019. doi: 10.1080/01621459.2017.1422737.

Kwangho Kim, Edward H. Kennedy, and Ashley I. Naimi. Incremental intervention effects in studies with dropout and many timepoints. *Journal of Causal Inference*, 9(1):302–344, 2021. doi: 10.1515/jci-2020-0031.

Kwangho Kim, Edward H. Kennedy, and José R. Zubizarreta. Doubly robust counterfactual classification. In *Advances in Neural Information Processing Systems*, volume 35, pp. 34831–34845, 2022. doi: 10.52202/068431-2524.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M. Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5637–5664, 2021. URL https://proceedings.mlr.press/v139/koh21a.html.

David Krueger, Ethan Caballero, Joern-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Remi Le Priol, and Aaron Courville. Out-of-distribution generalization via risk extrapolation (REx). In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5815–5826, 2021.

Fangrui Lv, Jian Liang, Shuang Li, Bin Zang, Chi Harold Liu, Ziteng Wang, and Di Liu. Causality inspired representation learning for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. doi: 10.1109/CVPR52688.2022.00788.

Divyat Mahajan, Shruti Tople, and Amit Sharma. Domain generalization using causal matching. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7313–7324, 2021. URL https://proceedings.mlr.press/v139/mahajan21b.html.

Michael Oberst, Nikolaj Thams, Jonas Peters, and David Sontag. Regularizing towards causal invariance: Linear models with proxies. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8260–8270, 2021. URL https://proceedings.mlr.press/v139/oberst21a.html.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=ryxGuJrFvS.

Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000. doi: 10.1016/S0378-3758(00)00115-4.

Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 814–823, 2015. URL https://proceedings.mlr.press/v37/swaminathan15.html.

Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, 2006. doi: 10.1007/0-387-37345-4.

Yanhua Yang, Muli Yang, Jiahua Li, Henan Wang, Cheng Deng, and Hongyuan Zhu. Counterfactual risk minimization for out-of-distribution generalization. *IEEE Transactions on Image Processing*, 35:4253–4268, 2026. doi: 10.1109/TIP.2026.3683291.

A Experimental details and supplementary results

In these binary experiments, $f_\theta(H)$ denotes the class-1 probability; the corresponding probability vector is $(1 - f_\theta(H), f_\theta(H))$, with Brier loss $2(f_\theta(H) - Y)^2$.

Simulation protocol. Each of 400 independent replications has $N = 8192$ observations. Draw S uniformly from $\{-1, 1\}$, independently draw a 384-dimensional Gaussian covariate $U \sim N(0, I_{384})$, and put $C = (S, U)$. Generate

$$P(Y = 1 \mid S) = 0.5 + 0.4S,$$
$$\pi_0(S, U) = \operatorname{expit}\left(-\log 3 + 0.8S + 1.5 \frac{\sum_{j=1}^{384} U_j}{\sqrt{384}} \right),$$
$$X = Y \oplus E_X, \quad E_X \sim \text{Bernoulli}(0.35),$$
$$B = Y \oplus E_B \oplus A, \quad E_B \sim \text{Bernoulli}(0.05), \qquad H = 2X + B.$$

Here $\oplus$ denotes addition modulo two. The Bernoulli errors are independently drawn, and the four image states in Figure 2 correspond to $(X, B) \in \{0, 1\}^2$. For conditional-risk regression we use $\psi(C) = (1, S)^T$, which spans every function of S. The true arm risks in this simulation depend on C only through S; ridge regularization and sampling still contribute error. We set the arm-risk ridge penalty to $\lambda = 5/8192$. The propensity is fitted by logistic regression on S and standardized U, penalizing only the U coefficients. The penalty is selected by three-fold cross-validated log loss from $\{0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. The proposed estimator uses an observation-weighted $K_{\text{cf}} = 5$ rotation with fold sizes 1639/1639/1638/1638/1638; nuisance estimation, conditional-risk regression, and learning rotate across folds, as formalized by Proposition 6. We set $\delta = 9$. Reported target risks are evaluated under the population law using Gauss–Hermite quadrature for the Gaussian covariates, avoiding test-set Monte Carlo error. We report Bonferroni-adjusted intervals for prespecified contrast families. Relative reductions compare replication-averaged risks: for baseline b and method m, the reduction is $100(\bar{R}_b - \bar{R}_m)/\bar{R}_b$ percent. For class-relative excess risk, each replication's class optimum is subtracted before averaging. These are ratios of means, not means of replication-wise ratios; the reported confidence intervals concern absolute paired differences.

Objective and estimator comparisons. The objective comparisons in Figure 3A optimize pooled historical risk, the equally weighted average of the two A-specific risks, or their maximum (group DRO) over the same four-state predictor class. Each population optimum is evaluated at the $\delta = 9$ target law. Their excess risks range from 0.0500 for equal environment weighting to 0.1737 for pooled historical ERM, compared with mean excess 0.00188 for the proposed finite-sample learner. The fixed-policy AIPW comparator in Figure 3B minimizes the cross-fitted score $\widehat{w}_\delta(L_f - \widehat{m}_{A,f}) + \sum_a \widehat{q}_{\delta,a}\widehat{m}_{a,f}$, treating $\widehat{q}_\delta$ as fixed and omitting its propensity-derivative term. This differs from the corrected score in Equation (9). The G-formula plug-in minimizes the sample average of $\sum_a \widehat{q}_{\delta,a}(C)\widehat{m}_{a,f}(C)$, using out-of-fold propensity and conditional-risk estimates. The proposed learner uses the five-fold role rotation above; comparator propensity and arm-risk predictions are fitted out of fold.

Table 1: Selected absolute excess risks (Monte Carlo standard errors), both in units of 10^{-3}, over 400 replications for the estimators in Figure 3B. Clipped policy weighting replaces each intervention weight $\widehat{w}_\delta$ by $\min(\widehat{w}_\delta, M)$.

Estimator	Estimated propensity	Known propensity
Proposed corrected learner	1.875 (0.037)	0.311 (0.011)
Policy weighting	3.003 (0.058)	0.304 (0.011)
G-formula plug-in	5.116 (0.095)	0.216 (0.008)
Fixed-policy AIPW (cross-fitted)	5.167 (0.095)	0.251 (0.009)
Clipped policy weighting, $M = 5$	2.863 (0.059)	0.327 (0.013)
Clipped policy weighting, $M = 2$	1.125 (0.034)	5.448 (0.065)

No estimator is uniformly best across the two propensity regimes. With estimated propensities, clipping at $M = 2$ reduces the squared bias of the fitted predictor from 0.00268 to 0.00070, while increasing its variance from 0.00033 to 0.00043. Under known propensities, the same threshold introduces squared bias of 0.00517. In this design, the prediction shifts associated with propensity estimation and clipping partly offset, and the advantage reverses when the propensity is known.

An exploratory sensitivity omitted all U coordinates from the fitted propensity while keeping the same arm-risk feature class. Across 400 replications, the 8,192-record weighted-minus-corrected

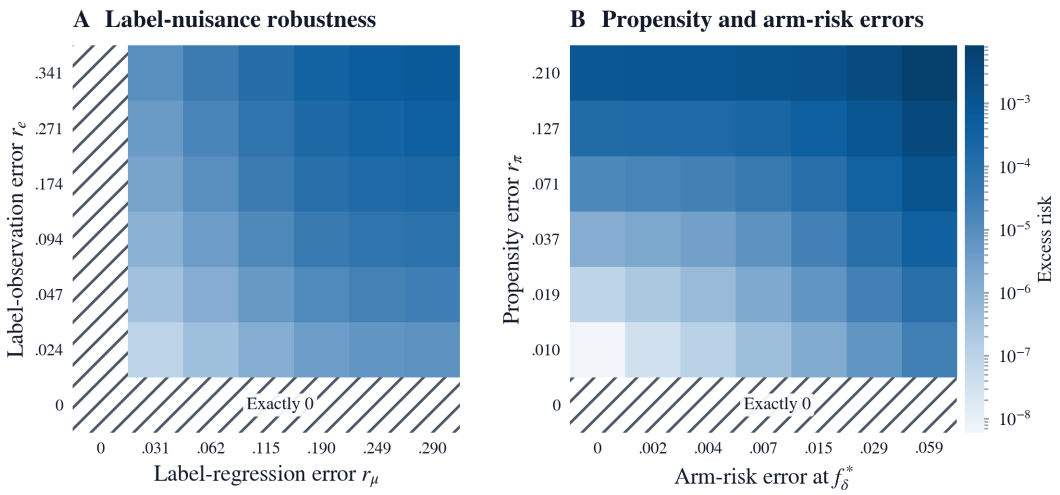

Figure 5: Population excess risk under controlled nuisance misspecification. **A**: errors in the label-observation model and label regression; either correct model removes the excess. **B**: propensity and arm-risk errors, the latter evaluated at f_δ^*. A correct propensity removes the excess for any regression error, but not conversely.

contrast was -0.000077, with a descriptive 95% interval $[-0.000087, -0.000066]$ (88 of 400 positive contrasts). The arm-risk feature class $\psi = (1, S)^T$ remains correctly specified for this simulation; correction need not improve risk under substantial propensity-model misspecification.

Nuisance perturbations. We perturb the nuisance functions along fixed directions and evaluate population excess risk by numerical quadrature on the grids shown in Figure 5. For the propensity, we add a scalar multiple of a fixed coefficient direction: 0.5 for the intercept, 0.6 for the S coefficient, and $0.9\widehat{\sigma}_j/\sqrt{384}$ for each standardized U_j coefficient, where $\widehat{\sigma}_j$ is the empirical standard deviation used to standardize U_j. The arm-risk regression is perturbed by shrinking its S coefficient toward zero. Its horizontal coordinate in Figure 5B is

$$r_{\mathrm{arm},*} = \left\| \sum_a q_{\delta,a}(m_{a,f_\delta^*} - \bar{m}_{a,f_\delta^*}) \right\|_{L_2(P_C)},$$

where $\bar{m}$ is the perturbed arm regression. This fixed-predictor error differs from the class supremum r_{arm} in Equation (12). The label-regression coordinate in Figure 5A uses the probability-vector norm r_μ, which is $\sqrt{2}$ times the scalar class-1 regression error. For the MAR surface, we perturb the observation probabilities and true label probabilities in each of the 16 (S, A, H) cells on the logit scale. Each perturbation is a scalar multiple of a linear combination of $(1, S, A, X, B)$, with coefficient vectors $(0.5, -0.6, 0.7, -0.4, 0.5)^T$ for the observation model and $(1.2, 2.1, -1.5, 1.8, -0.9)^T$ for the label regression. With either label nuisance correct, the population excess is exactly zero; near the origin it is quadratic in the product of the two label-nuisance errors. Propensity and conditional-risk-regression errors are asymmetric because a propensity error also changes the regression target. With the conditional-risk regression correct, score bias is second order in propensity error. In this saturated class, Brier curvature converts second-order displacement of the minimizer into fourth-order population excess. The fitted exponent with respect to r_π is 3.98 for population excess risk. Across replications, the exponent for the estimated squared mean prediction change from the unperturbed fit is 3.82. The exact-zero axes follow from Equation (28); the uniform bound in Theorem 1 alone does not yield fourth-order excess.

MAR simulation. Conditional on (A, S, H), labels are independently observed with

$$e_0(A, S, H) = \mathrm{expit}(0.6 + 0.4S - 0.5A + 0.3X - 0.2B), \qquad X = \lfloor H/2 \rfloor, \quad B = H \bmod 2.$$

We fit e_0 and μ_0 by 16-cell proportions on (S, A, H), with Beta$(1, 1)$ smoothing, inside each nuisance-fitting rotation. Evaluation uses the same full-label target law before masking.

Table 2: Excess risk in units of 10^{-3} under MAR labels over 400 replications. The mean label-observation probability is 0.62. The full-label learner uses outcomes before masking and is a reference, not an available method under missingness.

Method	Excess risk
Complete cases, unweighted	215.20
Complete cases, policy-weighted	5.32
Augmented labels, policy-weighted	3.06
Augmented labels, corrected	**1.94**
Full labels (reference)	1.88
True nuisances (oracle)	0.36

Label augmentation and correction together close 98.2% of the gap between complete-case policy weighting and the full-label reference. The corrected learner's paired excess over that reference is 0.000061, with a Bonferroni-adjusted simultaneous 95% confidence interval of $[0.000014, 0.000107]$ over the four prespecified MAR contrasts.

Camelyon17/WILDS source images and split. We use image patches and metadata from Camelyon17 as distributed through WILDS (Bándi et al., 2019; Koh et al., 2021). Camelyon17's patient codes identify artificial cases assembled from slides of different biological patients (Bándi et al., 2019). We use these codes to group the folds for morphology-model cross-validation. The training pool contains 12,288 patches from 23 WSIs across 19 artificial cases. The held-out pool contains 4,096 patches from four other WSIs (slide indices 6, 31, 39, and 41), balanced at 2,048 original tumor-negative and 2,048 original tumor-positive patches. The pools share no artificial case code, slide, image ID, dataset index, or decoded RGB content. Biological patient identity across the two pools is not available from these codes.

Semi-synthetic data generation. For a source patch with original tumor indicator B, independently for each sampled record we draw

$$Y = B \oplus E_Y, \quad E_Y \sim \text{Bernoulli}(0.35), \qquad P(S = 1 \mid Y) = 0.1 + 0.8Y,$$

where $S \in \{-1, 1\}$, and draw $U \sim N(0, I_{384})$. We then set $T = Y \oplus E_T$, $E_T \sim \text{Bernoulli}(0.05)$, and

$$P(A = 1 \mid S, U) = \text{expit}\left\{-\log 3 + 0.8S + 1.5\,\mathbf{1}^\mathsf{T}U/\sqrt{384}\right\}.$$

The potential rendering H^a retains a deterministic grayscale version of the source morphology and encodes the color cue $V^a = T \oplus a$. Thus acquisition changes the cue but not the morphology. Each replication samples source patches with replacement and gives repeated patches fresh draws of (Y, S, U, T, A). Conditional on the fixed training image pool, the synthetic record is the sampling unit. Source-group splitting within morphology cross-validation prevents source-group overlap between fitting and validation; nuisance, risk-regression, and learning roles are assigned at the synthetic-record level.

Four-parameter head. We extract fixed ImageNet-pretrained ResNet-18 features from grayscale morphology. On the nuisance-training subset S_η, PCA reduces these features to 64 dimensions, and a ridge regression selected by group-wise cross-validation produces a score $g \in [0, 1]$. Specifically, if r is the fitted ridge prediction of the synthetic Y, we use $g = \text{clip}\{(r - 0.35)/0.30, 0, 1\}$; the calibration uses the known synthetic label-noise rate 0.35. The image-only predictor is

$$f_\theta(H) = \phi(g, V)^\mathsf{T}\theta, \qquad \phi(g, V) = ((1-g)(1-V), (1-g)V, g(1-V), gV), \quad \theta \in [0, 1]^4.$$

For the four-parameter experiment, each replication contains 8,192 records allocated as $|S_\eta| = 4096$ and $|S_1| = |S_2| = 2048$. The propensity and morphology models use S_η; conditional-risk regression and head learning exchange S_1 and S_2. The arm-risk regressions use $\psi(C) = (1, S)^T$ and ridge penalty $\lambda = 1/2048$. The sample-size-matched weighted comparator trains its head on the corrected

learner's 4,096 $S_1 \cup S_2$ records and uses the S_η-fitted propensity. The full-sample weighted comparator trains its head on all 8,192 records with out-of-fold propensity predictions. The policy-set minimax comparator minimizes the largest empirical policy-weighted risk over $\delta \in \{1, 3, 9\}$. Box-constrained quadratic objectives are solved globally, and risks are evaluated by numerical quadrature under the known semi-synthetic law on the held-out image pool. The four full-sample heads (historical ERM, weighting with estimated or true propensity, and policy-set minimax) use group-wise out-of-fold morphology scores on S_η. Three outer group folds each fit PCA on the other groups and select a ridge penalty using group-fold validation within those training groups; the outer validation fold's labels are never used to fit its scores. Records in $S_1 \cup S_2$ and the held-out image pool instead use the morphology model fitted on all S_η. The corrected and sample-size-matched weighted heads use only $S_1 \cup S_2$. Historical ERM and full-sample weighting use the same 8,192 records and morphology scores. Comparing full-sample weighting with the corrected learner instead changes both head-training size and the use of out-of-fold scores; that contrast is descriptive, not a matched estimate of the correction's benefit.

Table 3: Held-out paired risk reductions over 400 replications. Means and interval bounds are in units of 10^{-3} Brier risk. Intervals are Bonferroni-adjusted simultaneous 95% confidence intervals (CIs) over the four prespecified contrasts, conditional on the fixed training and held-out image pools; positive values favor the second method. Full-sample rows use group-wise out-of-fold S_η morphology scores.

Contrast	Mean	95% simultaneous CI	Positive runs
Historical ERM − 8,192-record weighting	182.729	[181.850, 183.609]	400/400
Policy-set minimax − two-fold corrected learner	47.625	[47.333, 47.916]	400/400
4,096-record weighting − corrected learner	1.718	[1.621, 1.815]	395/400
8,192-record weighting − corrected learner	1.279	[1.128, 1.431]	347/400

The historical and 8,192-record weighted risks are 0.612637 and 0.429908, respectively; the corresponding paired contrast is reported in Table 3. In the matched comparison, weighting and correction have mean risks 0.430347 and 0.428629. The mean optimum over their shared four-parameter class is 0.425188, giving mean class-relative excess risks 0.005159 and 0.003441. With true propensities, the 4,096-record weighted-minus-corrected contrast is -0.000028, with descriptive interval $[-0.000075, 0.000018]$; the experiment therefore does not support general dominance of the correction.

Slide 39 (center 3) contributes 2,324 of the 4,096 held-out patches, including 1,812 of the 2,048 tumor-positive patches; center and slide composition therefore differ between pools. The intervals quantify synthetic-record variation conditional on both the fixed training and held-out image pools, rather than uncertainty over new slides or patients.

For the percentage reduction in Camelyon excess risk reported in the main text, we use the raw-image Bayes risk $R_{\mathrm{Bayes},\delta}$ on the fixed held-out image pool. This reference is shared across predictor classes and differs from the four-parameter class optimum above. At $\delta = 9$, it is 0.412033. Writing $\bar{R}_W$ and $\bar{R}_C$ for the mean risks of matched weighting and correction, respectively, the percentage reduction is

$$100 \frac{\bar{R}_W - \bar{R}_C}{\bar{R}_W - R_{\mathrm{Bayes},\delta}}.$$

We calculate this ratio from unrounded replication means. The confidence intervals remain on the absolute paired-risk scale.

MLP on PCA features. We also evaluate the matched comparison with a one-hidden-layer MLP (32 tanh units and a sigmoid output). Its inputs are the S_η-fitted 64-dimensional PCA features, standardized using the means and standard deviations on $S_1 \cup S_2$, and the rendered cue. Unlike the four-parameter head, it does not receive the S_η-supervised morphology score g. We consider two head-training sizes, $n = |S_1 \cup S_2| \in \{4096, 8192\}$, with $|S_\eta| = n$ and $|S_1| = |S_2| = n/2$, using 400 replications per size. Both configurations sample synthetic records with replacement from the same fixed training image pool and evaluate on the same four held-out slides.

PCA and propensity models are fitted on the corresponding S_η, and the two-fold correction uses the same arm-risk features and ridge penalty $\lambda = 1/2048$ at both sizes. Within each replication, all five

criteria use the same training records. We train by full-batch Adam for 1,500 steps with learning rate 0.01, using a common initialization.

Table 4: Held-out results for the PCA-feature MLP over 400 replications per head-training size n, with $|S_\eta| = n$. Intervals are Bonferroni-adjusted simultaneous 95% intervals over the three contrasts within each sample size, conditional on the fixed training image pool and four fixed held-out slides. Positive contrasts favor the second method. Both methods in each weighting-minus-corrected contrast use the indicated propensity. The upper panel reports risks in their original units; means and interval bounds in the lower panel are in units of 10^{-3} Brier risk. The raw-image Bayes risk is a reference, not the MLP class optimum.

Criterion	$n = 4096$	$n = 8192$
Historical ERM	0.565674	0.562918
Matched weighting, estimated propensity	0.432864	0.431194
Corrected, estimated propensity	0.433213	0.430460
Matched weighting, true propensity	0.437873	0.432536
Corrected, true propensity	0.437861	0.432481
Raw-image Bayes reference	0.412033	0.412033

Contrast	Mean	95% simultaneous CI	Positive runs
$n = 4096$			
ERM − weighting (estimated propensity)	132.809	[131.805, 133.813]	400/400
Weighting − corrected (estimated propensity)	−0.348	[−0.434, −0.263]	101/400
Weighting − corrected (true propensity)	0.011	[−0.074, 0.097]	194/400
$n = 8192$			
ERM − weighting (estimated propensity)	131.725	[131.200, 132.249]	400/400
Weighting − corrected (estimated propensity)	0.734	[0.660, 0.808]	361/400
Weighting − corrected (true propensity)	0.056	[0.004, 0.108]	203/400

Writing W_1 and w_2 for the input and output weights, the objective is

$$\frac{1}{n}\sum_{i\in S_1\cup S_2}\alpha_i\, 2\{f(H_i)-Y_i\}^2+\lambda_{\mathrm{MLP}}\big(\|W_1\|_F^2+\|w_2\|_2^2\big),$$

where α_i is one, an intervention weight, or a combined corrected weight, according to the criterion. Bias parameters are unpenalized. The classifier penalty λ_{MLP} is selected from $\{10^{-3}, 3\times10^{-3}, 10^{-2}, 3\times10^{-2}, 10^{-1}\}$ by twofold S_1/S_2 cross-validation: both fitting and validation use the estimated-propensity weighted Brier criterion. The selected value is shared by all five criteria. This is a controlled comparison with a shared penalty, not a comparison of separately optimized hyperparameters for each method. Within each head, weighting and correction use the same representation and training records. At $n = 4096$, λ_{MLP} is 10^{-2} in 398 replications and 3×10^{-2} in the other two; at $n = 8192$, the penalty is 10^{-2} in all 400 replications. All combined per-record weights are positive at both sample sizes. Risks are evaluated under the same known held-out law as for the four-parameter head. Comparisons across the two heads therefore reflect differences in representation, supervised feature fitting, regularization, and optimization as well as predictor flexibility.

MLP on the morphology score and cue. To compare the two heads with common input information, we give an MLP exactly the raw pair (g, V) used by the four-parameter head, without standardizing either coordinate. We reuse the same 400 synthetic-record draws, fitted morphology and propensity models, and 4,096 $S_1 \cup S_2$ training records as the original matched experiments. The MLP retains 32 tanh hidden units, a sigmoid output, and 1,500 full-batch Adam steps at learning rate 0.01. Its four criteria are weighting and correction with estimated or true propensity. Within this

representation, all four use the same initialization and the same penalty selected by the estimated-propensity weighted cross-validation rule above. Thus the true-propensity comparisons replace the propensity in the objective but retain this common penalty-selection rule.

The MLP on (g, V) and the MLP on PCA features use the same penalty grid and tuning budget. Their input dimensions and initial parameter arrays differ, and each representation selects its own penalty. For (g, V), the selected penalty is 10^{-3} in 394 replications and 3×10^{-3} in six, compared with 10^{-2} or 3×10^{-2} for the PCA-feature MLP. The four-parameter head retains its unpenalized, globally solved box-quadratic objective. This comparison therefore tests whether the correction effect persists with common input information; it does not separately identify the effects of predictor complexity, regularization, and optimization.

The primary comparison is at $\delta = 9$ with arm-risk features $\psi(C) = (1, S)^T$ and ridge penalty $1/2048$. Table 5 reports absolute risks and paired contrasts. With estimated propensity, correction improves the MLP on (g, V), and its gain exceeds that of the PCA-feature MLP. With true propensity, the intervals for both the gain and the difference between the two MLP gains contain zero.

Table 5: Matched comparisons at $\delta = 9$ over the same 400 synthetic-record replications, conditional on the fixed training and held-out image pools. Upper panel: mean Brier risks in original units. Lower panel: paired contrasts in units of 10^{-3} Brier risk. Correction gain is weighting risk minus corrected risk within the same replication and predictor; positive gains favor correction. Differences in gains subtract the second listed predictor's gain from the first. The two primary contrasts share a Bonferroni-adjusted simultaneous 95% interval family; the four secondary contrasts form a separate family. Coverage is not joint across the two families. Estimated or true refers to the propensity used by both methods.

Predictor	Propensity	Weighting	Corrected
MLP, raw (g, V)	Estimated	0.429005	0.427588
PCA-feature MLP	Estimated	0.432864	0.433213
Four-parameter head	Estimated	0.430347	0.428629
MLP, raw (g, V)	True	0.426283	0.426299
PCA-feature MLP	True	0.437873	0.437861
Four-parameter head	True	0.426613	0.426641

Contrast	Mean	95% simultaneous CI	Positive runs
Primary family: estimated propensity			
Gain: MLP on (g, V)	1.417	[1.336, 1.498]	393/400
Gain difference: (g, V) MLP minus PCA MLP	1.765	[1.704, 1.827]	399/400
Secondary family			
Gain: MLP on (g, V), true propensity	-0.016	$[-0.072, 0.039]$	186/400
Gain difference: (g, V) MLP minus PCA MLP, true propensity	-0.028	$[-0.105, 0.050]$	188/400
Gain: four-parameter head, estimated propensity	1.718	[1.621, 1.815]	395/400
Gain: four-parameter head, true propensity	-0.028	$[-0.087, 0.030]$	182/400

Policy and arm-risk feature sensitivity. We use the same 400 draws and raw (g, V) inputs for two further comparisons. The first changes the policy to $\delta = 3$, retaining $\psi(C) = (1, S)^T$; the weighting, correction, penalty selection, and evaluation all use that policy. The second retains $\delta = 9$ and reduces the arm-risk features to an intercept. Its weighted heads and selected penalties are identical to those of the $\delta = 9$ comparison above, and only the corrected heads are refitted. Both settings use a ridge coefficient of $1/2048$ and penalize the intercept. Each environment-specific squared-error sum is divided by the total number of records in the regression fold. All combined per-record coefficients are positive in these three settings.

Table 6 reports each correction gain and its paired change from the $\delta = 9$, $(1, S)^T$ setting. The policy contrast subtracts two gains, each evaluated under its own policy, rather than subtracting raw risks

across different target laws. With estimated propensity, correction remains beneficial at $\delta = 3$ and with intercept-only arm regression, although both gains are smaller than the primary gain. With true propensity, the intercept-only correction has a small adverse effect; the other true-propensity intervals include zero. Reducing the arm-risk feature set is a sensitivity to regression specification. We do not measure r_{arm} here, so these results do not verify its increase or the remainder's product rate. All intervals describe repeated synthetic-record sampling conditional on both fixed image pools. These follow-up comparisons reuse the original draws and therefore do not constitute independent confirmations.

Table 6: Sensitivity of the MLP on raw (g, V), with 400 paired replications per setting conditional on the fixed training and held-out image pools. Upper panel: mean Brier risks in original units. Lower panel: contrasts in units of 10^{-3} Brier risk. Correction gain is weighting risk minus corrected risk. Change in gain subtracts the paired gain at the reference setting, $\delta = 9$ and $\psi(C) = (1, S)^T$, with the same propensity status. Within each setting, the two estimated-propensity contrasts and the two true-propensity contrasts have separate Bonferroni-adjusted simultaneous 95% interval families. Coverage is not joint across families or settings. The $\delta = 3$ change compares gains evaluated under their respective policies.

Setting	Propensity	Weighting	Corrected
$\delta = 3$, $\psi(C) = (1, S)^T$	Estimated	0.473109	0.472625
$\delta = 3$, $\psi(C) = (1, S)^T$	True	0.472115	0.472102
$\delta = 9$, $\psi(C) = 1$	Estimated	0.429005	0.427720
$\delta = 9$, $\psi(C) = 1$	True	0.426283	0.426341

Contrast	Mean	95% simultaneous CI	Positive runs
$\delta = 3$, $\psi(C) = (1, S)^T$			
Correction gain, estimated propensity	0.484	[0.438, 0.530]	359/400
Change in gain, estimated propensity	-0.933	$[-0.974, -0.892]$	2/400
Correction gain, true propensity	0.013	$[-0.012, 0.037]$	204/400
Change in gain, true propensity	0.029	$[-0.001, 0.059]$	228/400
$\delta = 9$, $\psi(C) = 1$			
Correction gain, estimated propensity	1.285	[1.230, 1.341]	398/400
Change in gain, estimated propensity	-0.132	$[-0.188, -0.077]$	174/400
Correction gain, true propensity	-0.058	$[-0.077, -0.038]$	163/400
Change in gain, true propensity	-0.041	$[-0.091, 0.009]$	195/400

B IDENTIFICATION WITH MISSING LABELS

Under the target intervention, draw $A^{q_\delta} \sim q_\delta(\cdot \mid C)$ independently of (Y, H^0, H^1) conditional on C, and set $H^{q_\delta} = H^{A^{q_\delta}}$. The joint counterfactual law is

$$P_\delta(dc, \{a\}, dh, dy) = P(dc) q_{\delta,a}(c) P(H^a \in dh, Y \in dy \mid C = c).$$

For finite positive δ, the intervention preserves assignment support: $q_{\delta,a}(c) > 0$ implies $P(A = a \mid C = c) > 0$. On these strata, conditional exchangeability and consistency give

$$\begin{aligned} P(H^a \in dh, Y \in dy \mid C = c) &= P(H^a \in dh, Y \in dy \mid A = a, C = c) \\ &= P(H \in dh, Y \in dy \mid A = a, C = c). \end{aligned}$$

It follows that

$$p_\delta(c, a, h, y) = p(c) q_{\delta,a}(c) p(h, y \mid a, c). \tag{16}$$

Conditional exchangeability also entails $Y \perp A \mid C$, so an equivalent factorization is

$$p_\delta(c, a, h, y) = p(c, y) q_{\delta,a}(c) p(h \mid a, c, y).$$

Consequently, $P_\delta(C, Y) = P(C, Y)$ and $P_\delta(Y) = P(Y)$, but the image marginal generally differs from $p(h)$:

$$p_\delta(h) = \int p(c) \sum_{a=0}^{1} q_{\delta,a}(c) p(h \mid a, c)\, dc.$$

MAR and label-observation positivity give $P(Y \in dy \mid A, C, H) = P(Y \in dy \mid A, C, H, R = 1)$ on the support of the post-intervention distribution. The observed data therefore identify the target law in Equation (16) and every integrable risk under this law. Assignment support and label-observation positivity play distinct roles: the intervention avoids unsupported values of A, but this property alone does not identify the label law in strata where labels are never observed. □

If A is observed at deployment but C is not, the same identification argument and fixed-predictor risk scores apply to $f(H, A)$ after replacing $f(H)$ throughout. In general, $P_\delta(Y \mid H, A)$ can still change: $P_\delta(C \mid H, A)$ depends on $q_{\delta,A}(C)$, even though $P_\delta(Y \mid H, A, C) = P(Y \mid H, A, C)$.

C Pseudo-outcome identity and the Brier Bayes rule

For candidate label-observation probabilities and class-probability regressions $(\bar{e}, \bar{\mu})$, with $\bar{e} > 0$, define

$$Z(\bar{e}, \bar{\mu}) = \bar{\mu}(A, C, H) + \frac{R}{\bar{e}(A, C, H)} \{\mathbf{Y} - \bar{\mu}(A, C, H)\}.$$

Conditional on (A, C, H), MAR gives

$$\begin{aligned} \mathbb{E}\{Z(\bar{e}, \bar{\mu}) \mid A, C, H\} &= \bar{\mu} + \frac{e_0}{\bar{e}}(\mu_0 - \bar{\mu}), \\ \mathbb{E}\{Z(\bar{e}, \bar{\mu}) \mid A, C, H\} - \mu_0 &= \frac{\bar{e} - e_0}{\bar{e}}(\bar{\mu} - \mu_0). \end{aligned} \tag{17}$$

At (e_0, μ_0),

$$\mathbb{E}(Z_0 \mid A, C, H) = \mu_0(A, C, H) = \mathbb{E}(\mathbf{Y} \mid A, C, H). \tag{18}$$

More generally, (17) gives $\mathbb{E}\{Z(\bar{e}, \bar{\mu}) \mid A, C, H\} = \mu_0$ if either $\bar{e} = e_0$ or $\bar{\mu} = \mu_0$.

On the historical support, P_δ is absolutely continuous with respect to P, with density ratio

$$\frac{dP_\delta}{dP}(C, A, H, Y) = \frac{q_{\delta,A}(C)}{P(A \mid C)} = w_\delta(C, A).$$

Combining this identity with (18), for every integrable scalar function $g(H)$,

$$\mathbb{E}\{w_\delta Z_0 g(H)\} = \mathbb{E}\{w_\delta \mathbf{Y} g(H)\} = \mathbb{E}_\delta\{\mathbf{Y} g(H)\}.$$

In particular, since $\|\mathbf{Y}\|_2^2 = 1$,

$$\mathcal{R}_{\delta,B}(f) = \mathbb{E}\left[w_\delta\{\|f(H)\|_2^2 - 2f(H)^T Z_0 + 1\}\right]. \tag{19}$$

Provided the relevant second moments are finite, $\mathbb{E}\{w_\delta \|f(H) - Z_0\|_2^2\}$ has the same minimizer as Equation (19), since their difference is the f-independent quantity $\mathbb{E}\left[w_\delta\{\|Z_0\|_2^2 - 1\}\right]$.

For any class $k \in \{0, \ldots, K-1\}$ and measurable set B,

$$\begin{aligned} P_\delta(Y = k, H \in B) &= \mathbb{E}[\mathbf{1}(H \in B)\mathbb{E}\{w_\delta Z_{0,k} \mid H\}], \\ P_\delta(H \in B) &= \mathbb{E}[\mathbf{1}(H \in B)\mathbb{E}(w_\delta \mid H)]. \end{aligned}$$

The conditional Radon–Nikodym identity therefore yields

$$P_\delta(Y = k \mid H = h) = \frac{\mathbb{E}\{w_\delta Z_{0,k} \mid H = h\}}{\mathbb{E}(w_\delta \mid H = h)}.$$

For finite positive δ, w_δ is bounded away from zero, so the denominator is positive almost surely under the historical image distribution. Equivalently,

$$\mathbb{E}(w_\delta Z_0 \mid H) = \mathbb{E}(w_\delta \mid H) P_\delta(Y = \cdot \mid H),$$

where $\mathbb{E}(w_\delta \mid H)$ need not equal one because the intervention can change $P(H)$. The weighted criterion in Equation (19) accounts for this change without separately estimating the marginal density ratio for H.

Let $f^*_\delta(H) = P_\delta(Y = \cdot \mid H)$. Expanding the conditional Brier loss gives

$$\mathbb{E}_\delta\big[\|f(H) - \mathbf{Y}\|_2^2 \mid H\big] = \|f(H) - f^*_\delta(H)\|_2^2 + \mathbb{E}_\delta\big[\|f^*_\delta(H) - \mathbf{Y}\|_2^2 \mid H\big].$$

Integrating over H gives the squared-$L_2(P^H_\delta)$ excess-risk identity. Since μ_0 is simplex-valued, the coordinates of Z_0 sum to one, although individual coordinates may lie outside $[0, 1]$. For candidate nuisances, coordinatewise clipping to $[0, 1]$ or Euclidean projection onto the simplex generally breaks (17). At the true nuisances, both preserve (18). □

D Efficient influence function for the risk of a fixed predictor

Treat the predictor f as fixed independently of the evaluation sample, and define

$$L_f = \|f(H) - \mathbf{Y}\|_2^2, \qquad \eta_f(a, c, h) = \mathbb{E}(L_f \mid A = a, C = c, H = h, R = 1).$$

By MAR, $\eta_f(A, C, H) = \mathbb{E}(L_f \mid A, C, H)$. Define

$$\begin{aligned} T_f &= \eta_f(A, C, H) + \frac{R}{e_0(A, C, H)}\{L_f - \eta_f(A, C, H)\}, \\ m_{a,f}(c) &= \mathbb{E}\{\eta_f(a, c, H) \mid A = a, C = c\}, \\ m_{q_\delta,f}(c) &= q_{\delta,1}(c)m_{1,f}(c) + q_{\delta,0}(c)m_{0,f}(c). \end{aligned} \tag{20}$$

The intervention-averaged conditional risk also satisfies $m_{q_\delta,f}(C) = \mathbb{E}(w_\delta T_f \mid C)$.

Consider a regular parametric submodel through the observed-data law in the unrestricted MAR model. Efficiency here is relative to this model, without imposing the restrictions implied by the causal assumptions. As in Proposition 2, assume that the influence function is square integrable. The submodel score decomposes as

$$S(O) = S_C(C) + S_A(A \mid C) + S_H(H \mid A, C) + S_R(R \mid A, C, H) + RS_Y(Y \mid A, C, H), \tag{21}$$

where every conditional score has conditional mean zero given its conditioning variables. The identified target-risk functional is $\mathcal{R}_{\delta,B}(f) = \mathbb{E}\{m_{q_\delta,f}(C)\}$. For a submodel derivative denoted by a dot,

$$\dot q_{\delta,1}(c) = \frac{\delta}{d_\delta(c)^2}\dot\pi_0(c), \qquad \dot\pi_0(c) = \mathbb{E}[\{A - \pi_0(c)\}S_A \mid C = c]. \tag{22}$$

The gradient components for the covariate, image, label, and assignment laws in (21) are

$$\begin{aligned} D_C(O) &= m_{q_\delta,f}(C) - \mathcal{R}_{\delta,B}(f), \\ D_H(O) &= w_\delta(C, A)\{\eta_f(A, C, H) - m_{A,f}(C)\}, \\ D_Y(O) &= w_\delta(C, A)\frac{R}{e_0(A, C, H)}\{L_f - \eta_f(A, C, H)\}, \\ D_A(O) &= \frac{\delta\{m_{1,f}(C) - m_{0,f}(C)\}}{d_\delta(C)^2}\{A - \pi_0(C)\}. \end{aligned} \tag{23}$$

Pairing D_A with S_A recovers the contribution from differentiating the intervention probabilities in $q_{\delta,1}m_{1,f} + q_{\delta,0}m_{0,f}$, using (22). The inner product of D_H and S_H gives the derivative with respect to the conditional image law. For the observed-label score, MAR and $\mathbb{E}(R \mid A, C, H) = e_0(A, C, H)$ give

$$\mathbb{E}\left[\frac{R}{e_0}(L_f - \eta_f)RS_Y \;\middle|\; A, C, H\right] = \mathbb{E}\{(L_f - \eta_f)S_Y \mid A, C, H\},$$

which represents the derivative with respect to the conditional label law. The target does not depend on the label-observation mechanism. Moreover, $\mathbb{E}(D_Y \mid A, C, H, R) = 0$, and the other components are functions of (A, C, H); their sum is therefore orthogonal to S_R.

Summing (23) gives the expanded canonical gradient

$$D_{\delta,f}(O) = m_{q_\delta,f}(C) + w_\delta\{T_f - m_{A,f}(C)\} - \mathcal{R}_{\delta,B}(f) + \frac{\delta\{m_{1,f}(C) - m_{0,f}(C)\}}{d_\delta(C)^2}\{A - \pi_0(C)\}. \tag{24}$$

The sum belongs to the observed-data tangent space and represents the pathwise derivative in every score direction, so it is the canonical gradient.

An incremental intervention expresses a relative change within each stratum of C; a fully known, fixed $q(A \mid C)$ instead defines an absolute policy with the usual fixed-policy score and no policy-derivative term.

For the incremental intervention, the expanded form simplifies pointwise. Put $\Delta_f(C) = m_{1,f}(C) - m_{0,f}(C)$. On supported arms,

$$m_{q_\delta,f}(C) - m_{A,f}(C) = -\frac{\delta^{1-A}}{d_\delta(C)}\{A - \pi_0(C)\}\Delta_f(C).$$

Multiplication by $w_\delta = \delta^A/d_\delta$ shows that

$$w_\delta\{m_{q_\delta,f} - m_{A,f}\} + \frac{\delta\Delta_f}{d_\delta^2}(A - \pi_0) = 0.$$

Substitution into (24) yields the compact form

$$\boxed{D_{\delta,f}(O) = m_{q_\delta,f}(C) + w_\delta(C,A)\{T_f(O) - m_{q_\delta,f}(C)\} - \mathcal{R}_{\delta,B}(f).} \tag{25}$$

The compact identity also holds P-almost surely when $\pi_0(C)$ equals zero or one. On such a stratum, $A - \pi_0(C) = 0$ almost surely, the intervention leaves assignment unchanged, and unsupported arm regressions do not affect the score. No inverse assignment probability occurs in (25). Label-observation positivity is still required for this representation, since T_f contains $1/e_0$.

The compact EIF yields the following special cases. If $\delta = 1$, then $w_\delta = 1$, and (25) becomes $D_{1,f} = T_f - \mathbb{E}(L_f)$, the MAR efficient influence function for the historical Brier risk. If $R \equiv 1$, then $e_0 \equiv 1$ and $T_f = L_f$, so (25) is the complete-data incremental-intervention EIF. If both $\delta = 1$ and $R \equiv 1$, it reduces further to $L_f - \mathbb{E}(L_f)$. □

E Exact remainder and cross-fitted one-step inference

For candidate nuisance functions $(\bar\pi, \bar e, \bar\mu, \bar m_{q_\delta,f})$, with $\bar\pi \in [0,1]$ and $\bar e > 0$, write

$$\begin{aligned} \bar d(C) &= 1 + (\delta - 1)\bar\pi(C), \qquad \bar w(C,A) = \frac{\delta^A}{\bar d(C)}, \\ \bar\eta_f(A,C,H) &= 1 + \|f(H)\|_2^2 - 2f(H)^T\bar\mu(A,C,H), \\ \bar T_f &= \bar\eta_f(A,C,H) + \frac{R}{\bar e(A,C,H)}\{L_f - \bar\eta_f(A,C,H)\}, \\ \bar\Gamma_{\delta,f}(O) &= \bar m_{q_\delta,f}(C) + \bar w(C,A)\{\bar T_f(O) - \bar m_{q_\delta,f}(C)\}. \end{aligned} \tag{26}$$

All barred functions are treated as fixed in the following calculation. First,

$$\mathbb{E}(\bar T_f \mid A,C,H) - \eta_f = \frac{\bar e - e_0}{\bar e}(\bar\eta_f - \eta_f). \tag{27}$$

Second, with $d_\delta = 1 + (\delta - 1)\pi_0$,

$$\mathbb{E}(\bar w \mid C) = \frac{d_\delta}{\bar d}, \qquad \mathbb{E}(\bar w\eta_f \mid C) = \frac{d_\delta}{\bar d}m_{q_\delta,f}(C).$$

Conditioning (26) on C, using these two identities, and subtracting $m_{q_\delta,f}(C)$ gives the exact remainder

$$\boxed{\begin{aligned} P\bar\Gamma_{\delta,f} - \mathcal{R}_{\delta,B}(f) = \mathbb{E}\left[\frac{(\delta-1)(\pi_0 - \bar\pi)}{\bar d}\{m_{q_\delta,f} - \bar m_{q_\delta,f}\}\right] \\ + \mathbb{E}\left[\bar w\frac{\bar e - e_0}{\bar e}(\bar\eta_f - \eta_f)\right]. \end{aligned}} \tag{28}$$

The two terms are products of errors in $(\pi, m_{q_\delta,f})$ and (e, η_f), respectively.

Fix f independently of the i.i.d. risk-evaluation sample. Partition this sample into a fixed number of data-independent folds, estimate all nuisance functions outside each evaluation fold, and average the resulting scores. Let $\widehat{\Gamma}_{\delta,f,-v(i)}$ denote the score fitted outside the fold containing observation i. The cross-fitted estimator is

$$\widehat{\mathcal{R}}^{\mathrm{EIF}}_{\delta,B}(f) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\Gamma}_{\delta,f,-v(i)}(O_i).$$

Let $P_n g = n^{-1}\sum_{i=1}^{n} g(O_i)$, and define the true uncentered score

$$\Gamma_{0,\delta,f}(O) = m_{q_\delta,f}(C) + w_\delta(C,A)\{T_f(O) - m_{q_\delta,f}(C)\} = D_{\delta,f}(O) + \mathcal{R}_{\delta,B}(f).$$

If, on every fold, candidate assignment probabilities lie in $[0,1]$, candidate label-observation probabilities are bounded below by a fixed $\epsilon > 0$, the estimated uncentered score converges to $\Gamma_{0,\delta,f}$ in $L_2(P)$, and

$$\|\widehat{\pi} - \pi_0\|_{L_2(P_C)}\|\widehat{m}_{q_\delta,f} - m_{q_\delta,f}\|_{L_2(P_C)} = o_p(n^{-1/2}),$$
$$\|\widehat{e} - e_0\|_{L_2(P_{ACH})}\|\widehat{\mu} - \mu_0\|_{L_2(P_{ACH};\ell_2)} = o_p(n^{-1/2}),$$

then (28) and the bound $\|\widehat{\eta}_f - \eta_f\|_{L_2(P_{ACH})} \le 2\|\widehat{\mu} - \mu_0\|_{L_2(P_{ACH};\ell_2)}$, together with conditional empirical-process control from cross-fitting, yield

$$\widehat{\mathcal{R}}^{\mathrm{EIF}}_{\delta,B}(f) - \mathcal{R}_{\delta,B}(f) = (P_n - P)D_{\delta,f} + o_p(n^{-1/2}).$$

For a fixed f with $P(D^2_{\delta,f}) < \infty$, the central limit theorem gives asymptotic normality of the root-n estimation error with variance $P(D^2_{\delta,f})$. The empirical variance of the cross-fitted uncentered scores consistently estimates this variance.

For an independently trained predictor, the argument is conditional on its training data. A sequence of fitted predictors that changes with sample size additionally requires a corresponding conditional central limit theorem. Fold-specific predictors target a fold-weighted average risk; inference for that target or evaluation on observations used to train the predictor requires additional conditions.

F Weighted baseline and its excess-risk bound

For comparison, consider a learner that uses the intervention weight and label correction without the conditional-risk augmentation in Equation (9). Denote this weighted predictor by $\widehat{f}^{\mathrm{w}}_\delta$.

Partition the sampling units into a fixed number K_{cf} of folds. For fold v, estimate π_0, e_0, μ_0 using the other folds and evaluate them on the held-out units. The subscript $-v$ denotes a fit excluding fold v. Write $\widehat{\pi}_i = \widehat{\pi}_{-v(i)}(C_i)$ and, for $\nu \in \{e, \mu\}$, $\widehat{\nu}_i = \widehat{\nu}_{-v(i)}(A_i, C_i, H_i)$. Then

$$\widehat{w}_{\delta,i} = \frac{\delta^{A_i}}{1 + (\delta - 1)\widehat{\pi}_i}, \qquad \widehat{Z}_i = \widehat{\mu}_i + \frac{R_i \mathbf{Y}_i - R_i\widehat{\mu}_i}{\widehat{e}_i}.$$

For Theorem 2, assume $\widehat{e}_{-v} \ge \epsilon > 0$. The baseline predictor approximately minimizes the quadratic criterion

$$\begin{aligned} \widehat{Q}_{n,\delta}(f) &= \frac{1}{n}\sum_{i=1}^{n} \widehat{w}_{\delta,i}\{\|f(H_i)\|_2^2 - 2f(H_i)^T\widehat{Z}_i\}, \\ \widehat{Q}_{n,\delta}(\widehat{f}^{\mathrm{w}}_\delta) &\le \inf_{f\in\mathcal{F}_n} \widehat{Q}_{n,\delta}(f) + \varepsilon^{\mathrm{opt}}_n. \end{aligned} \tag{29}$$

The weighted squared-error criterion $\sum_i \widehat{w}_{\delta,i}\|f(H_i) - \widehat{Z}_i\|_2^2$ has the same minimizer; its additional term does not depend on f. Only H is required at deployment, whereas (A, C, R) enter estimation.

Baseline guarantee. Let p_v be the proportion of observations in fold v, and let f°_n minimize target risk over $\mathcal{F}_n$. Set $\Lambda_\delta = |\delta - 1|W_\delta/\min(1,\delta)$. Let $\mathcal{E}_n$ denote the foldwise uniform empirical deviation defined in Appendix F.1.

Theorem 2. *Under the conditions stated in Appendix F.1, including $\widehat{e}_{-v} \geq \epsilon > 0$, every $\widehat{f}_\delta^{\mathrm{w}}$ satisfying Equation (29) obeys*

$$\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{\mathrm{w}}) - \mathcal{R}_{\delta,B}(f_n^\circ) \leq 2\mathcal{E}_n + 2\Lambda_\delta \sum_{v=1}^{K_{\mathrm{cf}}} p_v \|\widehat{\pi}_{-v} - \pi_0\|_{L_1(P_C)} + \frac{4W_\delta}{\epsilon} \sum_{v=1}^{K_{\mathrm{cf}}} p_v \|\widehat{e}_{-v} - e_0\|_{L_2(P_{ACH})} \|\widehat{\mu}_{-v} - \mu_0\|_{L_2(P_{ACH};\ell_2)} + \varepsilon_n^{\mathrm{opt}}. \tag{30}$$

All norms are under the historical distribution, using the Euclidean norm for the class-probability vector. This bound contains a first-order propensity-error term and a product of label-observation and class-probability regression errors. The first-order term does not imply that the learner's excess risk is necessarily first order; Proposition 4 gives a sharper bound under additional conditions on the prediction class. With complete labels, setting $\widehat{e} \equiv 1$ gives $\widehat{Z} = \mathbf{Y}$; if also $\delta = 1$, Equation (29) reduces to standard Brier ERM.

F.1 Proof of the weighted-baseline bound

Let I_v be fold v, $p_v = |I_v|/n$, and $P_{n,v}$ its empirical measure. Assume i.i.d. sampling, a fixed number of data-independent folds, and a class $\mathcal{F}_n$ of simplex-valued predictors that is deterministic or selected using independent data. Suppose the target is identified, $\widehat{\pi}_{-v} \in [0,1]$, $\widehat{\mu}_{-v} \in \Delta^{K-1}$ pointwise, and $\widehat{e}_{-v} \in [\epsilon, 1]$ for fixed $\epsilon > 0$. Assume that the target risk attains its minimum at $f_n^\circ \in \mathcal{F}_n$ and that $\widehat{f}_\delta^{\mathrm{w}}$ satisfies the approximate-ERM condition in Equation (29) with $\varepsilon_n^{\mathrm{opt}} \geq 0$. For a candidate assignment probability $\bar{\pi}$, define $w_{\delta,\bar{\pi}}(C,A) = \delta^A/\{1 + (\delta-1)\bar{\pi}(C)\}$. Define

$$\widehat{\ell}_{v,f}(O) = w_{\delta,\widehat{\pi}_{-v}}(C,A)\{\|f(H)\|_2^2 - 2f(H)^T Z(\widehat{e}_{-v}, \widehat{\mu}_{-v})\},$$
$$\ell_{0,f}(O) = w_\delta(C,A)\{\|f(H)\|_2^2 - 2f(H)^T Z_0\}.$$

Also let

$$\mathcal{E}_n = \sum_{v=1}^{K_{\mathrm{cf}}} p_v \sup_{f \in \mathcal{F}_n} |(P_{n,v} - P)\widehat{\ell}_{v,f}|,$$
$$\widehat{Q}_{n,\delta}(f) = \sum_{v=1}^{K_{\mathrm{cf}}} p_v P_{n,v} \widehat{\ell}_{v,f}, \qquad Q_0(f) = P\ell_{0,f}.$$

Also let $W_\delta = \max(\delta, \delta^{-1})$. The exact plug-in difference is

$$w_{\delta,\bar{\pi}} - w_{\delta,\pi_0} = -\frac{\delta^A(\delta-1)(\bar{\pi} - \pi_0)}{\{1 + (\delta-1)\bar{\pi}\}\{1 + (\delta-1)\pi_0\}}.$$

Since $\bar{\pi}$ and π_0 lie in $[0,1]$,

$$|w_{\delta,\bar{\pi}} - w_{\delta,\pi_0}| \leq \Lambda_\delta |\bar{\pi} - \pi_0|, \qquad \Lambda_\delta = \frac{|\delta - 1| \max(1,\delta)}{\min(1,\delta)^2}.$$

Fix a fold v and condition on the data used to fit its nuisance functions. Put

$$a_f(A,C,H) = \|f(H)\|_2^2 - 2f(H)^T \mu_0(A,C,H),$$
$$b_v(A,C,H) = \frac{\widehat{e}_{-v} - e_0}{\widehat{e}_{-v}}(\widehat{\mu}_{-v} - \mu_0).$$

By (17), taking conditional expectations of the foldwise criterion yields the exact decomposition

$$P\widehat{\ell}_{v,f} - P\ell_{0,f} = P[\{w_{\delta,\widehat{\pi}_{-v}} - w_\delta\}a_f] - 2P\{w_{\delta,\widehat{\pi}_{-v}} f(H)^T b_v\},$$

where $P\ell_{0,f} = \mathcal{R}_{\delta,B}(f) - 1$.

For $f, \mu_0 \in \Delta^{K-1}$, $|a_f| \le 1$. To see the lower bound, let $M = \max_k f_k$. Then $f^T\mu_0 \le M$, $\|f\|_2^2 \ge M^2$, and $a_f \ge M^2 - 2M \ge -1$; the upper bound follows from $\|f\|_2^2 \le 1$ and $f^T\mu_0 \ge 0$. Also, $\|f\|_2 \le 1$, $w_{\delta,\widehat{\pi}_{-v}} \le W_\delta$, and $\widehat{e}_{-v} \ge \epsilon$. Therefore

$$\begin{aligned}\sup_{f\in\mathcal{F}_n} |P\widehat{\ell}_{v,f} - P\ell_{0,f}| \le \Lambda_\delta \|\widehat{\pi}_{-v} - \pi_0\|_{L_1(P_C)} \\ + \frac{2W_\delta}{\epsilon}\|\widehat{e}_{-v} - e_0\|_{L_2(P_{ACH})}\|\widehat{\mu}_{-v} - \mu_0\|_{L_2(P_{ACH};\ell_2)}.\end{aligned} \tag{31}$$

The triangle inequality and (31) imply

$$\begin{aligned}\sup_{f\in\mathcal{F}_n} |\widehat{Q}_{n,\delta}(f) - Q_0(f)| \le \mathcal{E}_n + \Lambda_\delta \sum_{v=1}^{K_{\mathrm{cf}}} p_v \|\widehat{\pi}_{-v} - \pi_0\|_{L_1(P_C)} \\ + \frac{2W_\delta}{\epsilon}\sum_{v=1}^{K_{\mathrm{cf}}} p_v \|\widehat{e}_{-v} - e_0\|_{L_2(P_{ACH})}\|\widehat{\mu}_{-v} - \mu_0\|_{L_2(P_{ACH};\ell_2)}.\end{aligned} \tag{32}$$

The basic inequality for an approximate empirical risk minimizer gives

$$Q_0(\widehat{f}_\delta^{\mathrm{w}}) - Q_0(f_n^\circ) \le 2 \sup_{f\in\mathcal{F}_n} |\widehat{Q}_{n,\delta}(f) - Q_0(f)| + \varepsilon_n^{\mathrm{opt}}.$$

Because $Q_0(f) = \mathcal{R}_{\delta,B}(f) - 1$, substituting (32) gives Equation (30). □

G Calculation of the corrected learning objective

During classifier optimization, the fitted assignment and label nuisances, metadata features, and regularization parameter remain fixed. The arm-risk regressions vary with the candidate predictor. The matrix $\widehat{G}_a + \lambda I_J$ is symmetric positive definite because $\lambda > 0$. Substituting the regression coefficients in Equation (8) gives

$$P_\ell[(1-\widehat{w}_\delta)\widehat{m}_{q_\delta,f}] = \sum_{a=0}^{1} P_\ell[(1-\widehat{w}_\delta)\widehat{q}_{\delta,a}\psi]^T\widehat{\beta}_{a,f} = P_m[\omega\widehat{T}_f].$$

The last equality substitutes $\widehat{\beta}_{a,f}$, uses the symmetry of $(\widehat{G}_a + \lambda I_J)^{-1}$, and applies the definition of ω_j in Equation (10). Adding $P_\ell[\widehat{w}_\delta\widehat{T}_f]$ proves Equation (11).

Gradient calculation. For a differentiable class-probability predictor $f_\theta(H)$, let $J_\theta f_\theta(H)$ be its K-by-$\dim(\theta)$ Jacobian. Since $\nabla_\theta\widehat{T}_{f_\theta} = 2J_\theta f_\theta(H)^T\{f_\theta(H) - \widehat{Z}\}$, holding the nuisance fits and precomputed coefficients fixed gives

$$\begin{aligned}\nabla_\theta\widehat{\mathcal{J}}_\delta(f_\theta) = \frac{2}{n_\ell}\sum_{i\in S_\ell} \widehat{w}_{\delta,i} J_\theta f_\theta(H_i)^T\{f_\theta(H_i) - \widehat{Z}_i\} \\ + \frac{2}{n_m}\sum_{j\in S_m} \omega_j J_\theta f_\theta(H_j)^T\{f_\theta(H_j) - \widehat{Z}_j\}.\end{aligned} \tag{33}$$

The second sum accounts for the dependence of the conditional-risk regressions on f_θ. Treating $\widehat{m}_{q_\delta,f}$ as constant during differentiation would omit this term. Fresh uniform minibatches from S_ℓ and S_m, sampled independently of previous updates and normalized separately by their batch sizes, give an unbiased estimate of Equation (33) conditional on the data and current classifier parameters.

Image classifier and nuisance models. A standard image encoder with a softmax classification head can represent f_θ, for example a pretrained ResNet-18 (He et al., 2016). The assignment model uses pre-assignment metadata. The label-observation and class-probability models can combine image features with (A, C), using a binary head for e and a K-class head for μ. Their fitted parameters, including any image encoders, remain frozen during classifier optimization; the classifier may have its own trainable encoder. If nuisance models replace H with an embedding, MAR given (A, C, H)

need not hold given the reduced inputs. The stated criterion uses deterministic predictions for each image. Accordingly, batch-normalization statistics are fixed, and stochastic augmentation and dropout are disabled during classifier updates. The metadata features, regularization parameter, and predictor class are prespecified or selected using S_η, as in Section 3.1. Tuning and stopping use historical data only; test units are held out from fitting and model selection.

The bounds in Equation (34) ensure a finite criterion for fixed $\lambda > 0$, although it can take negative values.

H Uniform excess risk for the corrected learner

Throughout this section, $I_a = \mathbf{1}(A = a)$, $p_a(C) = P(A = a \mid C)$, and $q_a(C) = q_{\delta,a}(C)$. We use the fitted functions and predictor class defined in Section 3.1. Conditional expectations over a fresh observation are taken with all fitted functions held fixed. The predictor class $\mathcal{F}_n$, ψ, λ, and $\delta \in (0, \infty)$ are prespecified or selected using only S_η. We assume that all displayed suprema are measurable. A sufficient condition is that every member of each function class is a pointwise limit of a sequence from a countable subclass.

H.1 Uniform score bias and the excess-risk bound

We first bound the fitted losses and regressions, then control score bias and apply the approximate-minimization inequality.

Put $\widehat{\eta}_f = 1 + \|f(H)\|_2^2 - 2f(H)^T\widehat{\mu}$. Both $\widehat{\eta}_f$ and the observed Brier loss L_f lie in $[0, 2]$. If $R = 0$, $\widehat{T}_f = \widehat{\eta}_f$; if $R = 1$,

$$\widehat{T}_f = L_f/\widehat{e} + (1 - 1/\widehat{e})\widehat{\eta}_f \in [-2(1/\widehat{e} - 1),\ 2/\widehat{e}].$$

Thus $|\widehat{T}_f| \le M_T := 2/\epsilon$ uniformly. With $\widehat{p}_{ma} = P_m I_a$, the ridge normal equations give

$$\sup_f \|\widehat{\beta}_{a,f}\|_2 \le \frac{BM_T\widehat{p}_{ma}}{\lambda}, \qquad \sup_{f,c} |\widehat{m}_{a,f}(c)| \le \frac{B^2 M_T\widehat{p}_{ma}}{\lambda}. \tag{34}$$

Since $\widehat{p}_{m0} + \widehat{p}_{m1} = 1$, $M_\Delta \le B^2 M_T/\lambda$, including when an arm is empty. Also $|1 - \widehat{w}_\delta| \le W_\delta - 1$, which yields the score envelope M_Γ in Equation (43). These bounds require no inverse moments of the true observation probability e_0; its positivity is used for identification.

The conditional risks in Equation (20) are defined on the observed-arm support. Outside that support, bounded extensions leave the target risk unchanged. Put $u = \pi_0 - \widehat{\pi}$ and $d_\delta = 1 + (\delta - 1)\pi_0$. Since $q_{\delta,1} - \widehat{q}_{\delta,1} = \delta u/(d_\delta \widehat{d})$, the error in the fitted intervention risk satisfies

$$m_{q_\delta,f} - \widehat{m}_{q_\delta,f} = \sum_{a=0}^{1} q_{\delta,a}(m_{a,f} - \widehat{m}_{a,f}) + \frac{\delta u}{d_\delta \widehat{d}}(\widehat{m}_{1,f} - \widehat{m}_{0,f}). \tag{35}$$

Combining Equations (28) and (35) gives, for every candidate f,

$$\begin{aligned} P\widehat{\Gamma}_{\delta,f} - \mathcal{R}_{\delta,B}(f) = \mathbb{E}\left[\frac{(\delta-1)u}{\widehat{d}}\sum_a q_a(m_{a,f} - \widehat{m}_{a,f})\right] \\ + \mathbb{E}\left[\frac{\delta(\delta-1)u^2}{d_\delta \widehat{d}^2}(\widehat{m}_{1,f} - \widehat{m}_{0,f})\right] \\ + \mathbb{E}\left[\widehat{w}_\delta \frac{\widehat{e} - e_0}{\widehat{e}}(\widehat{\eta}_f - \eta_f)\right]. \end{aligned} \tag{36}$$

The first term is a product with the policy-weighted arm-regression error, and the second is quadratic in the propensity error. If the arm-regression error does not vanish, the first product can scale linearly with the propensity error. The decomposition remains valid when the fitted policy assigns mass outside historical assignment support; the target still depends only on supported arms. Both denominators $d_\delta, \widehat{d}$ are at least $\min(1, \delta)$. The absolute values of the first two terms are bounded by $B_\delta r_\pi r_{\text{arm}}$ and $C_\delta M_\Delta r_\pi^2$, respectively. For the last term, use

$$|\widehat{\eta}_f - \eta_f| = 2|f(H)^T(\widehat{\mu} - \mu_0)| \le 2\|\widehat{\mu} - \mu_0\|_2,$$

followed by Cauchy–Schwarz, $\widehat{e} \geq \epsilon$, and $\widehat{w}_\delta \leq W_\delta$. Consequently

$$\sup_f |P\widehat{\Gamma}_{\delta,f} - \mathcal{R}_{\delta,B}(f)| \leq B_\delta r_\pi r_{\text{arm}} + C_\delta M_\Delta r_\pi^2 + \frac{2W_\delta}{\epsilon} r_e r_\mu =: \mathcal{B}_\Gamma. \tag{37}$$

For any comparator $f \in \mathcal{F}_n$, approximate minimization, the empirical-process deviation, and the bias bound imply

$$\begin{aligned}\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{\text{aug}}) &\leq \widehat{\mathcal{J}}_\delta(\widehat{f}_\delta^{\text{aug}}) + \mathcal{E}_\Gamma + \mathcal{B}_\Gamma \\ &\leq \widehat{\mathcal{J}}_\delta(f) + \varepsilon_{\text{aug}}^{\text{opt}} + \mathcal{E}_\Gamma + \mathcal{B}_\Gamma \\ &\leq \mathcal{R}_{\delta,B}(f) + 2\mathcal{E}_\Gamma + 2\mathcal{B}_\Gamma + \varepsilon_{\text{aug}}^{\text{opt}}.\end{aligned}$$

Taking the infimum over comparators proves Theorem 1, without assuming that the infimum is attained.

Proof of Corollary 1. For fixed δ, ϵ, the coefficients of the three nuisance-error terms in Equation (13) are finite constants. The total nuisance-error contribution is therefore $o_p(\rho_n)$ under Equation (14). Substitution gives Equation (15). At $\delta = 1$, $B_\delta = C_\delta = 0$. □

H.2 Uniform error of the fitted arm regressions

Define the population ridge coefficients and the target-weighted Gram matrices by

$$\begin{aligned} G_a &= P[I_a \psi\psi^T], & H_{\delta,a} &= P[q_a \psi\psi^T], \\ \beta_{\lambda,a,f} &= (G_a + \lambda I_J)^{-1} P[I_a \psi \eta_f], & m_{\lambda,a,f} &= \psi^T \beta_{\lambda,a,f}. \\ A_{J,\lambda} &= \sup_{f\in\mathcal{F}_n} \left\{ \mathbb{E} \sum_{a=0}^{1} q_a (m_{\lambda,a,f} - m_{a,f})^2 \right\}^{1/2}. \end{aligned} \tag{38}$$

For this subsection set

$$\begin{aligned} \zeta_a &= \sup_f \|(P_m - P)[I_a \psi \widehat{T}_f]\|_2, & \nu_a &= \|\widehat{G}_a - G_a\|_{\text{op}}, \\ b_{\lambda,a} &= \sup_f \|\beta_{\lambda,a,f}\|_2, & L_{\lambda,a} &= \|H_{\delta,a}^{1/2} (\widehat{G}_a + \lambda I_J)^{-1}\|_{\text{op}}, \\ r_{e,a} &= \{P[I_a(\widehat{e} - e_0)^2]\}^{1/2}, & r_{\mu,a} &= \{P[I_a \|\widehat{\mu} - \mu_0\|_2^2]\}^{1/2}. \end{aligned}$$

The norms $\|\cdot\|_{\text{op}}$ are Euclidean operator norms; $H_{\delta,a}^{1/2}$ is the positive semidefinite square root. Singular population Gram matrices are allowed. The fitted inverse exists because $\lambda > 0$.

With $A_{J,\lambda}$ defined in Equation (38), the arm-regression error satisfies

$$r_{\text{arm}} \leq A_{J,\lambda} + \left[\sum_{a=0}^{1} L_{\lambda,a}^2 \left\{ \zeta_a + \nu_a b_{\lambda,a} + \frac{2B}{\epsilon} r_{e,a} r_{\mu,a} \right\}^2 \right]^{1/2}. \tag{39}$$

Here $A_{J,\lambda}$ measures the population ridge approximation error; the remaining terms arise from estimating its coefficients. To prove this, subtract the population and empirical normal equations:

$$\begin{aligned} (\widehat{G}_a + \lambda I_J)(\widehat{\beta}_{a,f} - \beta_{\lambda,a,f}) = (P_m - P)[I_a \psi \widehat{T}_f] \\ + P[I_a \psi(\widehat{T}_f - \eta_f)] - (\widehat{G}_a - G_a)\beta_{\lambda,a,f}. \end{aligned} \tag{40}$$

MAR implies

$$\mathbb{E}(\widehat{T}_f - \eta_f \mid A, C, H) = \frac{\widehat{e} - e_0}{\widehat{e}} (\widehat{\eta}_f - \eta_f).$$

Therefore the norm of the middle term in Equation (40), uniformly in f, is at most $2Br_{e,a}r_{\mu,a}/\epsilon$. Multiplying by $H_{\delta,a}^{1/2}(\widehat{G}_a + \lambda I_J)^{-1}$ bounds the target-weighted prediction error relative to population

ridge. Finally, conditional Jensen's inequality and the triangle inequality give

$$\begin{aligned}\left\|\sum_a q_a(m_{a,f}-\widehat{m}_{a,f})\right\|_{L_2(P_C)} &\le \left\{\mathbb{E}\sum_a q_a(m_{a,f}-\widehat{m}_{a,f})^2\right\}^{1/2}\\ &\le \left\{\mathbb{E}\sum_a q_a(m_{a,f}-m_{\lambda,a,f})^2\right\}^{1/2}\\ &\quad+\left\{\sum_a \|H_{\delta,a}^{1/2}(\widehat{\beta}_{a,f}-\beta_{\lambda,a,f})\|_2^2\right\}^{1/2}.\end{aligned}$$

Taking suprema proves Equation (39).

Support and approximation. Since $q_a \le W_\delta p_a$, $H_{\delta,a} \preceq W_\delta G_a$. The target-weighted arm-regression quantities do not depend on the extension of $m_{a,f}$ to strata where $p_a = 0$. The estimated propensity may nevertheless give $\widehat{q}_{\delta,a} > 0$ on such strata; this error is covered by Equation (36). Population ridge minimizes $P[I_a(m_{a,f}-\psi^T\beta)^2]+\lambda\|\beta\|_2^2$. It follows that

$$A_{J,\lambda}^2 \le W_\delta \sup_f \sum_a \inf_{\beta\in\mathbb{R}^J}\left\{P[I_a(m_{a,f}-\psi^T\beta)^2]+\lambda\|\beta\|_2^2\right\} \le 4W_\delta.$$

The last inequality uses $\beta = 0$; approximation consistency requires additional conditions on the dictionary and regularization. Also,

$$b_{\lambda,a} \le \frac{2BP(A=a)}{\lambda}, \qquad L_{\lambda,a} \le \frac{B\sqrt{Pq_a}}{\lambda}.$$

The quantities in Equation (39) are therefore well-defined without strict assignment positivity. These upper bounds can grow as the regularization parameter decreases.

H.3 High-probability bounds

For a class $\mathcal{A}$ fixed given a conditioning sample $\mathcal{S}$, define the expected absolute Rademacher complexity by

$$\mathfrak{R}_n(\mathcal{A}\mid\mathcal{S}) = \mathbb{E}_{O_{1:n},\sigma}\left[\sup_{g\in\mathcal{A}}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i g(O_i)\right| \,\middle|\, \mathcal{S}\right]. \tag{41}$$

Here the O_i are fresh independent observations from P, $P_n g = n^{-1}\sum_{i=1}^n g(O_i)$, and the independent signs σ_i equal ±1 with equal probability. We suppress $\mathcal{S}$ when the conditioning is stated. For a uniformly bounded class $|g| \le M$, symmetrization gives $\mathbb{E}\sup_g|(P_n-P)g| \le 2\mathfrak{R}_n(\mathcal{A})$. Replacing one observation changes the supremum by at most $2M/n$. The bounded-differences inequality therefore gives

$$\sup_{g\in\mathcal{A}}|(P_n-P)g| \le 2\mathfrak{R}_n(\mathcal{A}) + M\sqrt{2\log(1/\alpha)/n} \tag{42}$$

with conditional probability at least $1-\alpha$. Conditional on (S_η, S_m), let $\mathcal{G}_\delta = \{\widehat{\Gamma}_{\delta,f} : f \in \mathcal{F}_n\}$. Applying Equation (42) with the score envelope derived above gives, with conditional probability at least $1-\alpha$,

$$\mathcal{E}_\Gamma \le 2\mathfrak{R}_{n_\ell}(\mathcal{G}_\delta) + M_\Gamma\sqrt{\frac{2\log(1/\alpha)}{n_\ell}}, \quad M_\Gamma = \frac{2}{\epsilon}\left\{W_\delta + \frac{(W_\delta-1)B^2}{\lambda}\right\}. \tag{43}$$

We next bound the complexity of the correction term. Let $\mathcal{T}_\delta = \{\widehat{w}_\delta\widehat{T}_f : f \in \mathcal{F}_n\}$ and $h_a = (1-\widehat{w}_\delta)\widehat{q}_{\delta,a}$. By Equation (34) and Cauchy–Schwarz over the signs,

$$\begin{aligned}\mathbb{E}_\sigma\sup_f\left|\frac{1}{n_\ell}\sum_i\sigma_i\sum_a h_a(O_i)\psi(C_i)^T\widehat{\beta}_{a,f}\right| &\le \frac{BM_T}{\lambda n_\ell}\sum_a \widehat{p}_{ma}\mathbb{E}_\sigma\left\|\sum_i\sigma_i h_a(O_i)\psi(C_i)\right\|_2\\ &\le \frac{(W_\delta-1)B^2M_T}{\lambda\sqrt{n_\ell}}.\end{aligned}$$

Consequently

$$\mathfrak{R}_{n_\ell}(\mathcal{G}_\delta) \le \mathfrak{R}_{n_\ell}(\mathcal{T}_\delta) + \frac{(W_\delta - 1)B^2 M_T}{\lambda\sqrt{n_\ell}}. \tag{44}$$

The bound holds uniformly over $\widehat{\beta}_{a,f}$ as f varies.

For the regression sample, condition only on S_η. Define $\mathcal{H}_a = \{I_a(v^T\psi)\widehat{T}_f : f \in \mathcal{F}_n,\ \|v\|_2 \le 1\}$. Duality gives $\zeta_a = \sup_{h\in\mathcal{H}_a} |(P_m - P)h|$. The envelope is BM_T. For $0 < \alpha_m < 1$, a union bound over the two arms yields, simultaneously with probability at least $1 - \alpha_m$,

$$\zeta_a \le U_{\zeta,a} := 2\mathfrak{R}_{n_m}(\mathcal{H}_a \mid S_\eta) + BM_T\sqrt{\frac{2\log(2/\alpha_m)}{n_m}}, \quad a = 0, 1.$$

For the Gram matrices, $\|I_a\psi\psi^T\|_F \le B^2$, where $\|\cdot\|_F$ is the Frobenius norm. Independence gives

$$\mathbb{E}\|\widehat{G}_a - G_a\|_F \le \{\mathbb{E}\|\widehat{G}_a - G_a\|_F^2\}^{1/2} \le B^2/\sqrt{n_m}.$$

Replacing one observation changes this norm by at most $2B^2/n_m$. Applying bounded differences and $\|\cdot\|_{\mathrm{op}} \le \|\cdot\|_F$ gives, simultaneously with probability at least $1 - \alpha_G$,

$$\nu_a \le N_G := \frac{B^2}{\sqrt{n_m}} + B^2\sqrt{\frac{2\log(2/\alpha_G)}{n_m}}, \quad a = 0, 1.$$

For a growing dictionary, B may depend on its dimension.

Choose positive $\alpha_\ell, \alpha_m, \alpha_G$ whose sum is less than one, and define

$$U_\ell = 2\mathfrak{R}_{n_\ell}(\mathcal{G}_\delta \mid S_\eta, S_m) + M_\Gamma\sqrt{2\log(1/\alpha_\ell)/n_\ell},$$

$$U_{\mathrm{arm}} = A_{J,\lambda} + \left[\sum_a L_{\lambda,a}^2\{U_{\zeta,a} + N_G b_{\lambda,a} + 2Br_{e,a}r_{\mu,a}/\epsilon\}^2\right]^{1/2}.$$

The learning, response-moment, and Gram events jointly have probability at least $1 - \alpha_\ell - \alpha_m - \alpha_G$, conditional on S_η and hence also unconditionally. A union bound applies because the learning event has the stated conditional probability for every fixed S_m, without requiring independence of the three events. On their intersection,

$$\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{\mathrm{aug}}) - \inf_{f\in\mathcal{F}_n} \mathcal{R}_{\delta,B}(f) \le 2U_\ell + 2B_\delta r_\pi U_{\mathrm{arm}} + 2C_\delta\frac{B^2 M_T}{\lambda}r_\pi^2 + \frac{4W_\delta}{\epsilon}r_e r_\mu + \varepsilon_{\mathrm{aug}}^{\mathrm{opt}}.$$

The right-hand side separates approximation error, regression-sample error, classifier complexity, nuisance error, and optimization error. It involves unknown population quantities and therefore does not directly provide a computable confidence bound.

H.4 Rates for a partition dictionary

Let $\mathcal{P}_n = \{B_{n1}, \ldots, B_{nJ_n}\}$ be a prespecified partition of the metadata space, let $\psi_j(c) = \mathbf{1}(c \in B_{nj})$, and take $\lambda = 1/n_m$. Then the ridge fit in Equation (8) reduces to the bounded cell mean

$$\widehat{m}_{a,f}(c) = \frac{\sum_{i\in S_m}\mathbf{1}(A_i = a, C_i \in B_{nj})\widehat{T}_{f,i}}{N_{aj} + 1}, \qquad N_{aj} = \sum_{i\in S_m}\mathbf{1}(A_i = a, C_i \in B_{nj}), \quad c \in B_{nj}. \tag{45}$$

The Gram matrix $\widehat{G}_a$ is diagonal with entries N_{aj}/n_m, and $\lambda = 1/n_m$ gives the denominator $N_{aj}+1$. The shared coefficients are

$$v_{a,j} = \frac{n_m}{N_{aj} + 1}P_\ell[(1 - \widehat{w}_\delta)\widehat{q}_{\delta,a}\mathbf{1}(C \in B_{nj})], \qquad \omega_i = v_{A_i,j(C_i)}, \quad i \in S_m,$$

where $j(c)$ is the cell containing c. Given the cell indices, computing the counts and cell sums takes $O(n_m + n_\ell + J_n)$ time.

For $c \in B_{nj}$, define the population cell risk $m_{a,f}^{\mathcal{P}_n}(c) = \mathbb{E}[\eta_f \mid A = a, C \in B_{nj}]$, with value zero if the conditioning event has probability zero. Define the uniform approximation error and the L_2 norm of the label-error product by

$$a_{J_n} = \sup_{f \in \mathcal{F}_n} \left\{ \mathbb{E} \sum_a q_a (m_{a,f}^{\mathcal{P}_n} - m_{a,f})^2 \right\}^{1/2},$$

$$b_{e\mu} = \left\{ \mathbb{E}[(\widehat{e} - e_0)^2 \|\widehat{\mu} - \mu_0\|_2^2] \right\}^{1/2}.$$

Bounded probabilities and simplex-valued label regressions imply

$$b_{e\mu} \le \min(r_\mu, \sqrt{2} r_e), \qquad b_{e\mu} \le 2^{1/4} (r_e r_\mu)^{1/2}.$$

The quantity $b_{e\mu}$ controls label-error propagation into the arm regression. As an L_2 norm of a product, it need not be bounded by the product $r_e r_\mu$ appearing in Equation (13).

Proposition 3. *Under the conditions of Theorem 1, use Equation (45) with fixed δ, ϵ. Let $d_\infty(f, g) = \sup_h \|f(h) - g(h)\|_2$, and let $\mathcal{N}_n(\tau)$ be the size of a smallest τ-cover of $\mathcal{F}_n$ in this metric, with centers in $\mathcal{F}_n$. For a deterministic $\tau_n > 0$ with $\mathcal{N}_n(\tau_n) < \infty$, put*

$$\ell_n = \log\{2e\mathcal{N}_n(\tau_n)\}, \qquad \kappa_n = \tau_n + \sqrt{J_n \ell_n / n_m}.$$

Then $M_\Delta \le 4/\epsilon$, and

$$r_{\mathrm{arm}} \le a_{J_n} + \frac{2\sqrt{W_\delta}}{\epsilon} b_{e\mu} + O_p(\kappa_n), \qquad \mathcal{E}_\Gamma = O_p\left(\tau_n + \sqrt{\ell_n / n_\ell}\right). \tag{46}$$

The implicit constants depend only on δ, ϵ. No lower bound on an arm–cell probability is required.

Proof. Write $M_T = 2/\epsilon$. The cell average in Equation (45) has nonnegative weights summing to $N_{aj}/(N_{aj} + 1) \le 1$. Thus $|\widehat{m}_{a,f}| \le M_T$, $M_\Delta \le 2M_T$, and $|\widehat{\Gamma}_{\delta,f}| \le M_T(2W_\delta - 1)$. Also $\|\widehat{Z}\|_2 \le 2/\epsilon - 1$, so

$$|\widehat{T}_f - \widehat{T}_g| \le \{\|f + g\|_2 + 2\|\widehat{Z}\|_2\} \|f - g\|_2 \le L_T d_\infty(f, g), \qquad L_T = 4/\epsilon.$$

Cell averaging preserves this Lipschitz bound.

Condition on S_η, and set $p_{aj} = P(A = a, C \in B_{nj})$ and $t_{aj,f} = \mathbb{E}[\widehat{T}_f \mid A = a, C \in B_{nj}, S_\eta]$. Set $t_{aj,f} = 0$ if $p_{aj} = 0$. Write $\widehat{m}_{aj,f}$ for the fitted value on this cell. On a finite cover of size $N = \mathcal{N}_n(\tau_n)$, conditional Hoeffding bounds for the within-cell sample means, together with the shrinkage bias, give

$$\mathbb{E}\left[\max_{f \text{ in cover}} |\widehat{m}_{aj,f} - t_{aj,f}|^2 \,\middle|\, N_{aj} = k, S_\eta \right] \le \frac{8M_T^2 \log(2eN)}{k + 1}. \tag{47}$$

Indeed, for $k \ge 1$, a union bound shows that the maximum absolute centered sample mean over the cover exceeds $t > 0$ with conditional probability at most $2N \exp\{-kt^2/(2M_T^2)\}$. Integrating this tail bound and using $\widehat{m}_{aj,f} - t_{aj,f} = \{k/(k+1)\}(\overline{T}_f - t_{aj,f}) - t_{aj,f}/(k+1)$ proves the displayed bound. For $k = 0$, the error is at most M_T. Since N_{aj} is binomial,

$$p_{aj} \mathbb{E}[(N_{aj} + 1)^{-1}] = \frac{1 - (1 - p_{aj})^{n_m + 1}}{n_m + 1} \le \frac{1}{n_m + 1}. \tag{48}$$

Summing Equation (47) over the $2J_n$ cells with weights p_{aj} bounds the expected supremum of the squared $L_2(P_{AC})$ error over the cover by $16M_T^2 J_n \ell_n/(n_m + 1)$. Markov's inequality and a $2L_T\tau_n$ cover extension give $O_p(\kappa_n)$ uniformly over $\mathcal{F}_n$. Zero-probability cells contribute zero.

The MAR identity in Equation (27) yields

$$\left| \mathbb{E}[\widehat{T}_f - \eta_f \mid A, C, H, S_\eta] \right| \le \frac{2}{\epsilon} |\widehat{e} - e_0| \|\widehat{\mu} - \mu_0\|_2.$$

Conditional Jensen's inequality bounds the $L_2(P_{AC})$ difference between $t_{aj,f}$ and the population cell risk by $2b_{e\mu}/\epsilon$. Finally, $q_a \le W_\delta P(A = a \mid C)$, the triangle inequality, and the definition of a_{J_n} prove the first bound in Equation (46).

Conditional on S_η, S_m, the fitted score is $L_\Gamma = (2W_\delta - 1)L_T$-Lipschitz in d_∞, with envelope $M_\Gamma^{\mathcal{P}} = (2W_\delta - 1)M_T$. A cover and Hoeffding's inequality give, with probability at least $1 - \alpha$,

$$\mathcal{E}_\Gamma \le 2L_\Gamma \tau_n + M_\Gamma^{\mathcal{P}} \sqrt{2 \log\{2\mathcal{N}_n(\tau_n)/\alpha\}/n_\ell}.$$

This proves the second bound. Its constants do not depend on the cell counts, dictionary size, or $1/\lambda$. □

Sufficient rate conditions. By Proposition 3 and Corollary 1, the class-relative excess risk is $O_p(\rho_n)$ if $\tau_n + \sqrt{\ell_n/n_\ell} = O(\rho_n)$, $\varepsilon_{\text{aug}}^{\text{opt}} = O_p(\rho_n)$, and

$$r_\pi(a_{J_n} + b_{e\mu} + \kappa_n) + r_\pi^2 + r_e r_\mu = o_p(\rho_n).$$

The conditions depend on metadata approximation error and the number of cells relative to the regression sample size.

Approximation consistency follows without a smoothness assumption when $\mathcal{F}_n = \mathcal{F}$ is a fixed d_∞-totally-bounded class and the deterministic nested partitions generate $\sigma(C)$. For each fixed f, conditional-expectation convergence gives $m_{A,f}^{\mathcal{P}_n}(C) \to m_{A,f}(C)$ in $L_2(P_{AC})$. The conditional Brier risks and their cell projections are 4-Lipschitz in d_∞. A finite-cover argument, followed by $q_a \le W_\delta p_a$, therefore gives $a_{J_n} \to 0$. For expanding classes, a_{J_n} must instead be controlled uniformly over the chosen sequence.

H.5 Nonparametric nuisance rates

We give an example in which nonparametric propensity and arm-risk estimation jointly satisfy the rate conditions. The construction uses one-dimensional metadata and complete labels. We first consider a finite predictor class, then its convex hull.

Corollary 2. *Assume the identification conditions of Theorem 1, let $R = 1$ almost surely, and set $\widehat{e} \equiv 1$. Fix $\delta \in (0, \infty)$ and a nonempty prespecified finite class $\mathcal{F}$ of M simplex-valued predictors. Suppose $C \in [0,1]$, π_0 has a Lipschitz version with constant L_π, and each $m_{a,f}$ has a version that is L_m-Lipschitz on the support of $P(C \mid A = a)$, uniformly over $a \in \{0,1\}$ and $f \in \mathcal{F}$. No condition is imposed for an arm of probability zero.*

Write $n_\eta = |S_\eta|$. Partition $[0,1]$ into J equal intervals, $\mathcal{P} = \{B_1, \dots, B_J\}$. On S_η, estimate π_0 by the empirical mean of A in each interval, using $1/2$ in an empty interval. On the independent sample S_m, use the arm-regression fit in Equation (45) with the same partition and $\lambda = 1/n_m$. Then

$$\begin{aligned} \mathbb{E} r_\pi^2 &\le \frac{L_\pi^2}{J^2} + \frac{J}{2(n_\eta + 1)}, \\ \mathbb{E} r_{\text{arm}}^2 &\le \frac{2L_m^2}{J^2} + \frac{128 W_\delta J \log(2eM)}{n_m + 1}, \qquad M_\Delta \le 2. \end{aligned} \tag{49}$$

The expectations are over S_η, S_m. If all three split sizes are of order n and $J \asymp n^{1/3}$, then

$$\begin{aligned} r_\pi + r_{\text{arm}} &= O_p(n^{-1/3}), \\ \sup_{f \in \mathcal{F}} \left| P\widehat{\Gamma}_{\delta,f} - \mathcal{R}_{\delta,B}(f) \right| &= O_p(n^{-2/3}), \qquad \mathcal{E}_\Gamma = O_p(n^{-1/2}). \end{aligned} \tag{50}$$

Consequently, if $\varepsilon_{\text{aug}}^{\text{opt}} = O_p(n^{-1/2})$, then

$$\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{\text{aug}}) - \min_{f \in \mathcal{F}} \mathcal{R}_{\delta,B}(f) \le 2\mathcal{E}_\Gamma + \varepsilon_{\text{aug}}^{\text{opt}} + O_p(n^{-2/3}) = O_p(n^{-1/2}). \tag{51}$$

Neither a density bound for C nor a positive lower bound for the assignment probabilities is required.

Proof. Complete labels and $\widehat{e} = 1$ give $\widehat{Z} = \mathbf{Y}$ and $\widehat{T}_f = L_f \in [0,2]$, with $\mathbb{E}(L_f \mid A, C) = m_{A,f}(C)$. In particular, the arm fits take values in $[0,2]$, so $M_\Delta \le 2$.

For an interval B_j, write $p_j = P(C \in B_j)$, $\bar{\pi}_j = \mathbb{E}[A \mid C \in B_j]$, and $K_j = \sum_{i \in S_\eta} \mathbf{1}(C_i \in B_j)$. Cells of zero probability contribute nothing below. Conditional on $K_j = k \ge 1$, the variance of the empirical cell mean is at most $1/(4k) \le 1/\{2(k+1)\}$, and its conditional mean is $\bar{\pi}_j$. For $k = 0$, the squared error of the value $1/2$ is at most $1/4 \le 1/\{2(k+1)\}$. Let $\bar{\pi}(C) = \bar{\pi}_j$ on B_j. The cell-projection identity gives, for every fitted histogram,

$$r_\pi^2 = \|\pi_0 - \bar{\pi}\|_{L_2(P_C)}^2 + \sum_j p_j(\widehat{\pi}_j - \bar{\pi}_j)^2.$$

The first term is at most L_π^2/J^2. Applying the binomial identity in Equation (48) with sample size n_η to the second term proves the first bound in Equation (49).

The cell diameter and the uniform Lipschitz condition imply $a_J \le L_m/J$. To bound the stochastic arm error, use the finite-class argument in the proof of Proposition 3, with the class itself as a cover and $M_T = 2$. Writing $t_{aj,f} = m^{\mathcal{P}}_{a,f}|_{B_j}$, Equations (47) and (48) give

$$\mathbb{E}\sup_{f\in\mathcal{F}}\sum_{a,j} p_{aj}(\widehat{m}_{aj,f} - t_{aj,f})^2 \le \frac{64J\log(2eM)}{n_m+1}.$$

Here the supremum of the sum is bounded by the sum of the cellwise suprema. Because $q_a \le W_\delta P(A = a \mid C)$, Jensen's inequality for the arm mixture and $(x+y)^2 \le 2x^2 + 2y^2$ imply the second bound in Equation (49). Markov's inequality now gives the nuisance rates when $J \asymp n^{1/3}$.

The label-error terms vanish. Substitution into Equation (37), using $M_\Delta \le 2$, gives the bias rate. Conditional on S_η, S_m, the score class has M members and envelope $M^{\mathcal{P}}_\Gamma = 2(2W_\delta - 1)$. Thus, with conditional probability at least $1-\alpha$,

$$\mathcal{E}_\Gamma \le M^{\mathcal{P}}_\Gamma\sqrt{2\log(2M/\alpha)/n_\ell}.$$

Finally apply Theorem 1. □

Convex hull of a finite class. The same rates hold when the learner optimizes over convex combinations of fixed simplex-valued predictors $g_1, \dots, g_M$. Let $\mathcal{F} = \mathrm{conv}\{g_1, \dots, g_M\}$, and set

$$\mathcal{V} = \{g_i : 1 \le i \le M\} \cup \{(g_i + g_j)/2 : 1 \le i < j \le M\}.$$

Assume the Lipschitz condition in Corollary 2 for the arm risks of the finite set $\mathcal{V}$, with all other conditions unchanged. For $f_\alpha = \sum_i \alpha_i g_i$, where $\alpha_i \ge 0$ and $\sum_i \alpha_i = 1$, expansion of the Brier loss gives

$$L_{f_\alpha} = \sum_i \alpha_i(2\alpha_i - 1)L_{g_i} + 4\sum_{i<j}\alpha_i\alpha_j L_{(g_i+g_j)/2}. \tag{52}$$

The coefficients sum to one, and their absolute values sum to at most three: $\sum_i \alpha_i|2\alpha_i - 1| \le 1$ and $4\sum_{i<j}\alpha_i\alpha_j \le 2$. Conditional expectation, partition regression, and the fitted-score map are linear in the loss, with the nuisance fits held fixed. Thus Equation (52) holds for the corresponding arm risks, fitted arm risks, and scores as well. The triangle inequality consequently yields

$$r_{\mathrm{arm}}(\mathcal{F}) \le 3r_{\mathrm{arm}}(\mathcal{V}), \qquad \mathcal{E}_\Gamma(\mathcal{F}) \le 3\mathcal{E}_\Gamma(\mathcal{V}),$$

where the argument denotes the class over which the supremum is taken. Also $M_\Delta(\mathcal{F}) \le 2$, since convex combinations remain simplex-valued. Applying Corollary 2 to $\mathcal{V}$ therefore proves Equations (50) and (51) for $\mathcal{F}$, with constants depending on the fixed number of base predictors. Appendix I compares the two learners in this common nonparametric regime.

H.6 Consistency and rate conditions

For fixed δ, ϵ, B, with λ bounded away from zero, M_Δ and r_{arm} are bounded. Thus consistency of $\widehat{\pi}$, a vanishing label-error product, and vanishing learning and optimization errors suffice for class-relative consistency, even if the arm approximation error persists. In that case, $r_\pi r_{\mathrm{arm}}$ can be of the same order as r_π. Equation (39) separates the approximation and sampling contributions to this product.

For a general dictionary, the score envelopes and inverse-Gram factors may diverge as $\lambda \to 0$. A conservative sufficient condition for the squared assignment term to vanish is $B^2 r_\pi^2/(\epsilon\lambda) = o_p(1)$; a smaller fitted contrast can give a weaker requirement. For fixed remaining constants, sufficient conditions for the bound in Equation (44) to vanish are vanishing weighted-loss complexity and $\lambda\sqrt{n_\ell} \to \infty$. The partition construction in Proposition 3 instead has bounds independent of $1/\lambda$. Rates for other predictor classes require corresponding complexity and approximation bounds.

The squared propensity term can remain even with exact arm regressions because the target policy depends on π_0. If either $\widehat{e} = e_0$ or $\widehat{\mu} = \mu_0$ almost surely, the label-error products in Equations (37) and (39) vanish. At $\delta = 1$, both the assignment correction and its contribution to Equation (44) vanish. Selection of δ or models using S_ℓ, beyond minimization over the prespecified $\mathcal{F}_n$, requires uniform control over the enlarged choice set.

The fitted restriction $\widehat{e} \geq \epsilon$ also matters when interpreting nuisance consistency. In particular, $r_e^2 \geq \mathbb{E}[(\epsilon - e_0)_+^2]$. Thus a fixed floor is compatible with $r_e \to 0$ only if the right-hand side is zero. The product-rate conditions do not require individual consistency of both label nuisances: with a nonvanishing r_e, sufficiently accurate label regression can still make the label-error products small.

I BRIER GEOMETRY AND COMPARISON WITH WEIGHTED LEARNING

The linear propensity term in Theorem 2 bounds a uniform objective deviation and need not give a sharp bound on the learned predictor's excess risk. For complete labels, Brier curvature can yield a quadratic propensity-error contribution even for weighted learning. Throughout this section, set $R = \widehat{e} = 1$, use the same predictor class $\mathcal{F}_n$ and learning sample S_ℓ for both criteria, and condition on the fitted nuisance functions. The corrected criterion additionally uses arm regressions fitted on S_m. Write

$$\|f - g\|_\delta^2 = \mathbb{E}_\delta \|f(H) - g(H)\|_2^2, \qquad \mathcal{E}_w = \sup_{f \in \mathcal{F}_n} |(P_\ell - P)(\widehat{w}_\delta L_f)|.$$

Let $f_{\mathcal{F}}$ attain the minimum of $\mathcal{R}_{\delta,B}$ over $\mathcal{F}_n$. Assume either that $\mathcal{F}_n$ is convex as a set of prediction functions or that the Bayes predictor f_δ^* belongs to $\mathcal{F}_n$, in which case take $f_{\mathcal{F}} = f_\delta^*$. These conditions concern predictions, not convexity of a neural-network parameterization.

Proposition 4. *Under the preceding conditions and the identification assumptions, let $\widehat{\pi} \in [0,1]$, fix $0 < \delta < \infty$, and suppose that $\widehat{f}_w \in \mathcal{F}_n$ satisfies*

$$P_\ell(\widehat{w}_\delta L_{\widehat{f}_w}) \leq \inf_{f \in \mathcal{F}_n} P_\ell(\widehat{w}_\delta L_f) + \varepsilon_w^{\mathrm{opt}}.$$

Then

$$\mathcal{R}_{\delta,B}(\widehat{f}_w) - \mathcal{R}_{\delta,B}(f_{\mathcal{F}}) \leq 4\mathcal{E}_w + 2\varepsilon_w^{\mathrm{opt}} + 8B_\delta^2 r_\pi^2. \tag{53}$$

No lower bound on the assignment propensity is required.

Proof. Let $d_\delta = 1 + (\delta - 1)\pi_0$, $\widehat{d} = 1 + (\delta - 1)\widehat{\pi}$, and

$$h(C) = \frac{d_\delta(C)}{\widehat{d}(C)} - 1 = \frac{(\delta - 1)(\pi_0(C) - \widehat{\pi}(C))}{\widehat{d}(C)}.$$

The weight identity gives

$$P[\widehat{w}_\delta(L_f - L_g)] - \{\mathcal{R}_{\delta,B}(f) - \mathcal{R}_{\delta,B}(g)\} = \mathbb{E}_\delta[h(C)(L_f - L_g)].$$

For simplex-valued predictions and a one-hot label,

$$|L_f - L_g| = |(f - g)^T(f + g - 2\mathbf{Y})| \leq 2\sqrt{2}\|f - g\|_2.$$

Also $|h(C)| \leq B_\delta |\widehat{\pi}(C) - \pi_0(C)|$, and the intervention leaves the marginal distribution of C unchanged. Therefore Cauchy–Schwarz yields the contrast bound

$$|P[\widehat{w}_\delta(L_f - L_g)] - \{\mathcal{R}_{\delta,B}(f) - \mathcal{R}_{\delta,B}(g)\}| \leq 2\sqrt{2} B_\delta r_\pi \|f - g\|_\delta. \tag{54}$$

If $f_\delta^* \in \mathcal{F}_n$, Equation (5) directly gives

$$\|f - f_{\mathcal{F}}\|_\delta^2 \leq \mathcal{R}_{\delta,B}(f) - \mathcal{R}_{\delta,B}(f_{\mathcal{F}}).$$

The same inequality holds when $\mathcal{F}_n$ is convex and its minimum is attained: minimizing along the segment from $f_{\mathcal{F}}$ to f gives $\mathbb{E}_\delta[(f_{\mathcal{F}} - f_\delta^*)^T(f - f_{\mathcal{F}})] \geq 0$, and expansion of the Brier risk proves the inequality.

Write $x = \mathcal{R}_{\delta,B}(\widehat{f}_w) - \mathcal{R}_{\delta,B}(f_{\mathcal{F}})$. Approximate minimization and Equation (54) imply

$$x \leq 2\mathcal{E}_w + \varepsilon_w^{\mathrm{opt}} + 2\sqrt{2} B_\delta r_\pi \sqrt{x}.$$

Applying $ab \leq (a^2 + b^2)/2$ to the last term and rearranging gives Equation (53). □

This result sharpens Theorem 2 under prediction curvature. For the same predictor class and learning sample, the corrected learner's bound in Equation (13) has nuisance contribution $r_\pi r_{\mathrm{arm}} + M_\Delta r_\pi^2$ for complete labels. If r_π and r_{arm} are both $O_p(t_n)$, with $M_\Delta = O_p(1)$, both displayed bounds have an $O_p(t_n^2)$ nuisance contribution. Their sampling and optimization terms remain different. Consequently, the contrast between a linear baseline bound and a second-order corrected bound does not by itself establish a faster learning rate for the corrected learner.

A common nonparametric regime. Consider the convex-hull construction in Appendix H.5, with the same fitted histogram propensity, predictor class, and learning sample for both criteria. The minimum of the target risk is attained because the coefficient simplex is compact and Brier risk is continuous in the coefficients. The finite loss representation in Equation (52) also gives $\mathcal{E}_w(\mathcal{F}) \le 3\mathcal{E}_w(\mathcal{V})$. Since $0 \le \widehat{w}_\delta L_f \le 2W_\delta$, conditional concentration over the fixed finite set $\mathcal{V}$ yields $\mathcal{E}_w = O_p(n^{-1/2})$. Consequently,

$$\mathcal{R}_{\delta,B}(\widehat{f}_w) - \mathcal{R}_{\delta,B}(f_{\mathcal{F}}) \le 4\mathcal{E}_w + 2\varepsilon_w^{\mathrm{opt}} + O_p(n^{-2/3}),$$
$$\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{\mathrm{aug}}) - \mathcal{R}_{\delta,B}(f_{\mathcal{F}}) \le 2\mathcal{E}_\Gamma + \varepsilon_{\mathrm{aug}}^{\mathrm{opt}} + O_p(n^{-2/3}).$$

With optimization gaps of order $O_p(n^{-1/2})$, both upper bounds are $O_p(n^{-1/2})$; this comparison does not establish the exact rates of either learner. The corrected criterion has joint second-order population bias, whereas weighted learning can have a quadratic excess-risk contribution through Brier curvature despite first-order population bias.

J Cross-fitted learning

The regression and learning samples can exchange roles while sharing a single predictor. Let $S_1 = S_m, S_2 = S_\ell$ be independent samples of equal size n, independent of S_η, and write P_s for their empirical means, with $P_{1\cup 2} = (P_1 + P_2)/2$. The index $s \in \{1,2\}$ labels the sample folds. Conditional on S_η, fix the nuisance fits, predictor class $\mathcal{F}_n$, metadata features ψ, regularization parameter $\lambda > 0$, and intervention parameter δ, as in Section 3.1. Put $\bar{s} = 3 - s$.

Objective and precomputation. For $a \in \{0,1\}$ and $s \in \{1,2\}$, define

$$\widehat{G}_{a,s} = P_s[\mathbf{1}(A=a)\psi\psi^T],$$
$$\widehat{m}_{a,f}^{[s]}(c) = \psi(c)^T(\widehat{G}_{a,s} + \lambda I_J)^{-1} P_s[\mathbf{1}(A=a)\psi\widehat{T}_f],$$
$$\widehat{m}_{q_\delta,f}^{[s]} = \sum_a \widehat{q}_{\delta,a}\widehat{m}_{a,f}^{[s]},$$
$$\widehat{\Gamma}_f^{(-s)} = \widehat{w}_\delta\widehat{T}_f + (1-\widehat{w}_\delta)\widehat{m}_{q_\delta,f}^{[\bar{s}]}.$$

The averaged criterion and its approximate minimizer satisfy

$$\widehat{\mathcal{J}}_\delta^{\mathrm{sym}}(f) = \frac{1}{2}\sum_{s=1}^{2} P_s\widehat{\Gamma}_f^{(-s)},$$
$$\widehat{\mathcal{J}}_\delta^{\mathrm{sym}}(\widehat{f}_\delta^{\mathrm{sym}}) \le \inf_{f\in\mathcal{F}_n} \widehat{\mathcal{J}}_\delta^{\mathrm{sym}}(f) + \varepsilon_{\mathrm{sym}}^{\mathrm{opt}}. \tag{55}$$

Define

$$b_{a,s} = P_s[(1-\widehat{w}_\delta)\widehat{q}_{\delta,a}\psi], \qquad \omega_{s,i} = \psi(C_i)^T(\widehat{G}_{A_i,s} + \lambda I_J)^{-1} b_{A_i,\bar{s}}, \quad i \in S_s.$$

The criterion has the precomputed representation

$$\widehat{\mathcal{J}}_\delta^{\mathrm{sym}}(f) = \frac{1}{2}\sum_{s=1}^{2} P_s[(\widehat{w}_\delta + \omega_s)\widehat{T}_f] = \frac{1}{2n}\sum_{s=1}^{2}\sum_{i\in S_s}(\widehat{w}_{\delta,i} + \omega_{s,i})\widehat{T}_{f,i}. \tag{56}$$

Indeed, symmetry of each inverse gives

$$P_s[(1-\widehat{w}_\delta)\widehat{m}_{q_\delta,f}^{[\bar{s}]}] = \sum_a b_{a,s}^T(\widehat{G}_{a,\bar{s}} + \lambda I_J)^{-1} P_{\bar{s}}[\mathbf{1}(A=a)\psi\widehat{T}_f] = P_{\bar{s}}[\omega_{\bar{s}}\widehat{T}_f].$$

Summing over s proves Equation (56). Four J-dimensional linear-system solves, one for each arm and regression fold, suffice before optimization. The coefficients do not depend on the classifier parameters, may be negative, and are used without renormalization or clipping.

For a differentiable image-only predictor f_θ, let $J_\theta f_\theta(H)$ denote its K-by-$\dim(\theta)$ Jacobian. The exact gradient is

$$\nabla_\theta \widehat{\mathcal{J}}_\delta^{\mathrm{sym}}(f_\theta) = \frac{1}{n}\sum_{s=1}^{2}\sum_{i\in S_s}(\widehat{w}_{\delta,i} + \omega_{s,i}) J_\theta f_\theta(H_i)^T\{f_\theta(H_i) - \widehat{Z}_i\}.$$

This gradient includes the dependence of both sets of fitted arm regressions on f_θ. At $\delta = 1$, $\widehat{w}_1 = 1$ and $b_{a,s} = \omega_{s,i} = 0$, hence $\widehat{\mathcal{J}}_1^{\mathrm{sym}}(f) = P_{1\cup 2}\widehat{T}_f$. This is pooled MAR-corrected Brier learning on $S_1 \cup S_2$. With complete labels and $\widehat{e} \equiv 1$, it is pooled Brier ERM. For any δ, setting $\omega_s = 0$ in Equation (56) yields the intervention-weighted pseudo-loss on the same pooled sample.

For partition features $\psi_j(c) = \mathbf{1}(c \in B_{nj})$ and $\lambda = 1/n$, let $j(c)$ be the cell index and $N_{a,j,s} = \sum_{i\in S_s} \mathbf{1}(A_i = a, C_i \in B_{nj})$. The shared coefficient simplifies to

$$\omega_{s,i} = \frac{\sum_{r\in S_{\bar{s}}} \mathbf{1}\{j(C_r) = j(C_i)\}(1 - \widehat{w}_{\delta,r})\widehat{q}_{\delta,A_i}(C_r)}{N_{A_i,j(C_i),s} + 1}, \qquad i \in S_s.$$

The numerator uses $S_{\bar{s}}$, while the denominator counts observations in S_s. The denominator remains positive for empty arm–cell combinations.

Excess-risk guarantee. Use $r_\pi, r_e, r_\mu, B_\delta, C_\delta, W_\delta$ from Section 3.2, write $q_a = q_{\delta,a}$, and define

$$\begin{aligned}
\mathcal{E}_s &= \sup_{f\in\mathcal{F}_n} |(P_s - P)\widehat{\Gamma}_f^{(-s)}|, \\
r_{\mathrm{arm},s} &= \sup_{f\in\mathcal{F}_n} \left\| \sum_a q_a(m_{a,f} - \widehat{m}_{a,f}^{[s]}) \right\|_{L_2(P_C)}, \\
M_{\Delta,s} &= \sup_{f\in\mathcal{F}_n} \|\widehat{m}_{1,f}^{[s]} - \widehat{m}_{0,f}^{[s]}\|_{L_\infty(P_C)}.
\end{aligned}$$

As before, P integrates a fresh observation with the fitted functions held fixed.

Proposition 5. *Under the identification, fitted-function, bounded-feature, and measurability assumptions of Theorem 1, with the independent samples and common class specified above, every learner satisfying Equation (55) obeys*

$$\begin{aligned}
\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{\mathrm{sym}}) - \inf_{f\in\mathcal{F}_n} \mathcal{R}_{\delta,B}(f) \le \mathcal{E}_1 + \mathcal{E}_2 + B_\delta r_\pi(r_{\mathrm{arm},1} + r_{\mathrm{arm},2}) \\
+ C_\delta r_\pi^2(M_{\Delta,1} + M_{\Delta,2}) + \frac{4W_\delta}{\epsilon} r_e r_\mu + \varepsilon_{\mathrm{sym}}^{\mathrm{opt}}.
\end{aligned} \tag{57}$$

For both folds, $M_{\Delta,s} \le 2B^2/(\epsilon\lambda)$.

Proof. For the regression fitted on S_s, Equation (37) bounds the population score bias uniformly by

$$\mathcal{B}_s = B_\delta r_\pi r_{\mathrm{arm},s} + C_\delta M_{\Delta,s} r_\pi^2 + \frac{2W_\delta}{\epsilon} r_e r_\mu.$$

Thus $\sup_f |P\widehat{\Gamma}_f^{(-s)} - \mathcal{R}_{\delta,B}(f)| \le \mathcal{B}_{\bar{s}}$. The same envelope argument as Equation (34) gives the claimed bound on $M_{\Delta,s}$. By the triangle inequality,

$$\sup_{f\in\mathcal{F}_n} |\widehat{\mathcal{J}}_\delta^{\mathrm{sym}}(f) - \mathcal{R}_{\delta,B}(f)| \le \tfrac{1}{2}(\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{B}_1 + \mathcal{B}_2).$$

Applying this inequality once to $\widehat{f}_\delta^{\mathrm{sym}}$ and once to an arbitrary comparator, and using Equation (55), bounds excess risk by $\mathcal{E}_1 + \mathcal{E}_2 + \mathcal{B}_1 + \mathcal{B}_2 + \varepsilon_{\mathrm{sym}}^{\mathrm{opt}}$. Taking the infimum over comparators proves the result. □

Concentration and rates. For each s, condition on $(S_\eta, S_{\bar{s}})$. The score class $\mathcal{G}_s = \{\widehat{\Gamma}_f^{(-s)} : f \in \mathcal{F}_n\}$ is then fixed and independent of its evaluation sample S_s. Let $\mathfrak{R}_{n,s}(\mathcal{G}_s)$ be its conditional expected absolute Rademacher complexity, as in Equation (41). Both classes have the envelope M_Γ from Equation (43). For any $\alpha \in (0,1)$, with probability at least $1 - \alpha$ conditional on S_η, simultaneously for $s = 1, 2$,

$$\mathcal{E}_s \le 2\mathfrak{R}_{n,s}(\mathcal{G}_s) + M_\Gamma\sqrt{\frac{2\log(2/\alpha)}{n}}.$$

Apply the conditional concentration inequality with failure probability $\alpha/2$ for each direction, integrate over its regression sample, and take a union bound; the two criteria need not be independent.

For fixed δ, ϵ, if $\mathcal{E}_1 + \mathcal{E}_2 + \varepsilon_{\text{sym}}^{\text{opt}} = O_p(\rho_n)$ and

$$r_\pi(r_{\text{arm},1} + r_{\text{arm},2}) + r_\pi^2(M_{\Delta,1} + M_{\Delta,2}) + r_e r_\mu = o_p(\rho_n),$$

then Equation (57) gives class-relative excess risk $O_p(\rho_n)$, with negligible nuisance bias at that scale. The extension retains the product and squared-error structure of Theorem 1, with both folds contributing to the empirical-risk term.

J.1 Cross-fitting with rotating sample roles

The simulation uses a rotation of the same three roles. For $K_{\text{cf}} \geq 3$, partition the N observations into independent folds $I_1, \ldots, I_{K_{\text{cf}}}$, with $n_k = |I_k|$ and $a_k = n_k/N$. For rotation k, let $S_{\ell,k} = I_k$, let $S_{m,k} = I_{k+1}$ with cyclic indexing, and let $S_{\eta,k}$ contain the remaining $K_{\text{cf}} - 2$ folds. Write $P_{\ell,k} g = n_k^{-1} \sum_{i \in I_k} g(O_i)$ and $P_N g = N^{-1} \sum_{i=1}^N g(O_i)$ for the fold and full-sample means. Fit all nuisance functions on $S_{\eta,k}$, fit the conditional-risk regression on $S_{m,k}$, and evaluate the resulting score $\widehat{\Gamma}_{k,f}$ on $S_{\ell,k}$. The shared learner minimizes

$$\widehat{\mathcal{J}}_\delta^{(K_{\text{cf}})}(f) = \sum_{k=1}^{K_{\text{cf}}} a_k P_{\ell,k} \widehat{\Gamma}_{k,f}. \tag{58}$$

The weights a_k make this an observation-weighted average. In particular, at $\delta = 1$ with complete labels and $\widehat{e} \equiv 1$ in every rotation, $\widehat{\mathcal{J}}_1^{(K_{\text{cf}})} = P_N T_f$, including when the fold sizes are unequal. With selectively observed labels, the same nuisance fit from $S_{\eta,k}$ is used to form the pseudo-loss on both $S_{\ell,k}$ and $S_{m,k}$.

For each rotation, define $\mathcal{E}_k = \sup_{f \in \mathcal{F}_n} |(P_{\ell,k} - P)\widehat{\Gamma}_{k,f}|$ and let $\mathcal{B}_k$ be the uniform score-bias bound obtained from Equation (37) using the nuisance and arm-regression errors of rotation k. Conditional on $(S_{\eta,k}, S_{m,k})$, define $\mathcal{G}_k = \{\widehat{\Gamma}_{k,f} : f \in \mathcal{F}_n\}$, and let $\mathfrak{R}_{n_k,k}(\mathcal{G}_k)$ be its conditional expected absolute Rademacher complexity, as in Equation (41).

Proposition 6. *Under the assumptions of Theorem 1, suppose that the predictor class $\mathcal{F}_n$ and intervention parameter are common across rotations and fixed before observing the evaluation folds, and that every fitted function used by rotation k is measurable with respect to $(S_{\eta,k}, S_{m,k})$. If $\widehat{f}_\delta^{(K_{\text{cf}})}$ minimizes Equation (58) up to $\varepsilon_{K_{\text{cf}}}^{\text{opt}}$, then*

$$\mathcal{R}_{\delta,B}(\widehat{f}_\delta^{(K_{\text{cf}})}) - \inf_{f \in \mathcal{F}_n} \mathcal{R}_{\delta,B}(f) \leq 2 \sum_{k=1}^{K_{\text{cf}}} a_k(\mathcal{E}_k + \mathcal{B}_k) + \varepsilon_{K_{\text{cf}}}^{\text{opt}}. \tag{59}$$

Proof. For each k, condition on $(S_{\eta,k}, S_{m,k})$. The fitted score class is then fixed and independent of $S_{\ell,k}$, while the uniform score calculation gives $\sup_f |P\widehat{\Gamma}_{k,f} - \mathcal{R}_{\delta,B}(f)| \leq \mathcal{B}_k$. Hence

$$\sup_{f \in \mathcal{F}_n} |\widehat{\mathcal{J}}_\delta^{(K_{\text{cf}})}(f) - \mathcal{R}_{\delta,B}(f)| \leq \sum_{k=1}^{K_{\text{cf}}} a_k(\mathcal{E}_k + \mathcal{B}_k).$$

Applying this inequality to the learned predictor and to an arbitrary comparator, followed by the optimization inequality and an infimum over comparators, proves Equation (59). The events need not be independent across rotations; foldwise concentration followed by a union bound is sufficient. □

For a common envelope M_Γ, using failure probability α/K_{cf} in each rotation yields, simultaneously for all k,

$$\mathcal{E}_k \leq 2\mathfrak{R}_{n_k,k}(\mathcal{G}_k) + M_\Gamma \sqrt{\frac{2\log(K_{\text{cf}}/\alpha)}{n_k}}.$$

For fixed K_{cf}, this preserves the rate and the nuisance product structure of the two-fold result.