WHEN COMPARISON IS EASIER THAN EVALUATION: LEARNING BETTER RANDOMIZATION DESIGNS

Anonymous authors
Paper under double-blind review

ABSTRACT

Randomization identifies treatment effects, but its dependence structure controls precision. Across repeated cohorts, practitioners may choose among independent assignment, matching, rerandomization, and balancing schemes that all preserve treatment probability one half. Learning the most precise law from observed outcomes seems impossible: quadratic risk depends on both potential outcomes, yet only one is observed. We show that comparison escapes this obstruction. Centered sign designs share a unit covariance diagonal, so subtracting two risks removes the unobserved within-unit products. Under independent fair reference assignment, no absolute risk is universally unbiasedly estimable from one cohort without outcome restrictions, whereas every pairwise risk difference is. This identity yields Prediction-Augmented Comparative Randomization Learning (PA-CRL): occasional reference rounds and simulated ghost assignments provide sampler-only unbiased contrasts, while an inverse-reference residual corrects any clipped predictable contrast model. Optimistic mirror descent gives finite-horizon expected regret against a fixed finite contextual-policy class, and cohort-size weighting converts it into expected excess-MSE control for unequal cohorts. Fresh centered deployment preserves deterministic-horizon unbiasedness and yields an exact MSE identity; bounded equal-size cohorts additionally admit anytime-valid confidence sequences. Polynomial-hint PA-CRL improves on vanilla CRL in all 17 controlled synthetic cells (1.83–10.28%; median 5.66%). Against a raw linear plug-in from the same misspecified predictor class, it improves by 3.4–42.3% throughout a contiguous 16/36-cell region at $\beta \geq .625$, with all 16 pointwise seed-level 95% confidence intervals excluding zero, and remains below Bernoulli risk across the full surface.

1 INTRODUCTION

Randomized experiments identify causal effects, but their assignment law determines how much precision a fixed cohort provides. Independent assignment, blocking, matching, rerandomization, minimax allocation, and discrepancy balancing preserve randomization while inducing different cross-unit dependence (Morgan & Rubin, 2012; Kallus, 2018; Harshaw et al., 2024; Arbour et al., 2022). Consider a hospital network repeating the same two-arm study across sites. It may assign independently, pair patients on different biomarkers, or rerandomize for multivariate balance. Every candidate gives each patient treatment probability one half and targets the same design-based estimand, yet matching on a prognostic direction can sharply reduce variance while matching on an irrelevant direction cannot. Because that direction may change across sites, the network must learn from completed cohorts which dependence law is precise before assigning the next cohort.

Most design methods commit in advance to one target geometry: rerandomization fixes a balance region, minimax and kernel designs optimize over a declared outcome class, and Gram–Schmidt Walk or online balancing controls a specified discrepancy (Morgan & Rubin, 2012; Kallus, 2018; Harshaw et al., 2024; Arbour et al., 2022). Prior outcomes can tune covariate weights within a sequential-matching procedure (Kapelner & Krieger, 2023). We study the complementary problem of selecting among an arbitrary prespecified library of centered samplers according to outcome-relevant randomization risk. The policy class contains fixed designs and allows the preferred sampler to vary with site context. Unlike adaptive Neyman allocation (Dai et al., 2023; Noarov et al., 2025;

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.

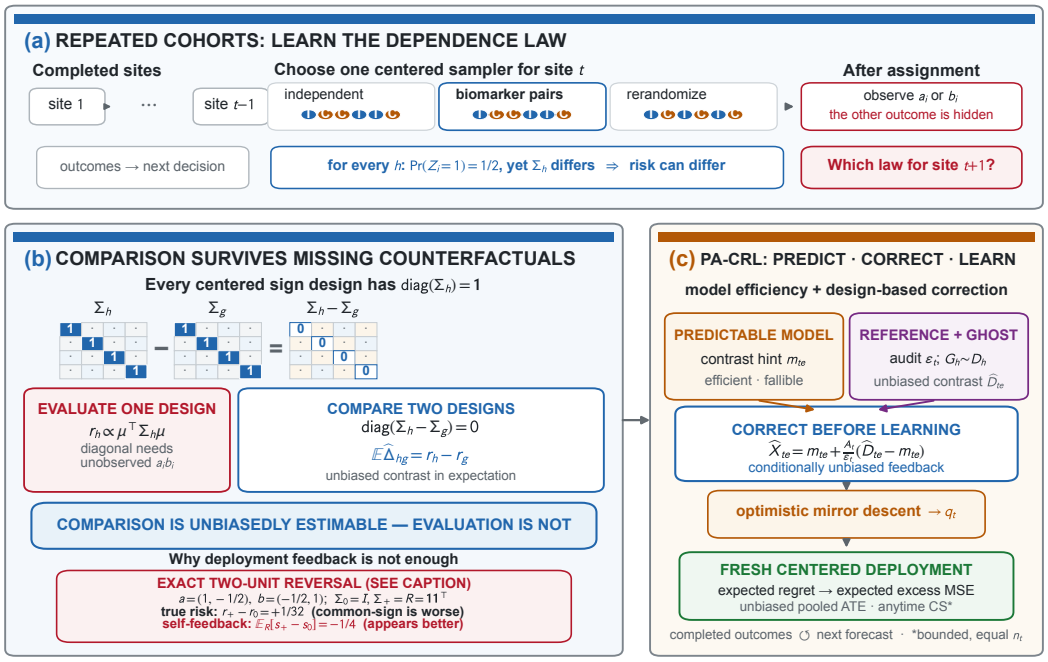

Figure 1: **Pairwise design risk remains learnable when absolute risk does not.** (a) Repeated sites choose among centered laws with identical treatment marginals but different covariance and precision; one potential outcome per patient remains hidden. (b) Absolute risk contains inaccessible within-unit products, whereas $\mathrm{diag}(\Sigma_h - \Sigma_g) = 0$ cancels them and makes $r_h - r_g$ unbiasedly estimable from an independent-reference cohort. The lower inset is the exact instance of Proposition 3.3: for $a = (1, -1/2)^\top$, $b = (-1/2, 1)^\top$, $\Sigma_0 = I$, and $\Sigma_+ = R = \mathbf{1}\mathbf{1}^\top$, the true contrast is $+1/32$, but common-sign self-feedback gives $-1/4$ and reverses the decision. (c) PA-CRL predicts design contrasts, corrects them with occasional reference and ghost feedback, updates an optimistic learner, and deploys a fresh centered law, linking unbiased comparison, adaptive precision, and valid inference (Theorems 3.1, 4.1, 5.1, 6.1 and 6.2 and Proposition 4.2).

Chen et al., 2025), this adaptation changes only cross-unit dependence: every marginal treatment probability remains one half.

Observed outcomes do not make this selection routine. If a_i and b_i are unit i's potential outcomes and $\mu_i = a_i + b_i$, sampler h has risk proportional to $\mu^\top \Sigma_h \mu$, whose diagonal contains the unobserved product $a_i b_i$. A direct plug-in succeeds only when its outcome model ranks the samplers correctly; an omitted interaction or a site shift can reverse that ranking. More strikingly, deployment feedback itself can mislead systematically: the two-unit construction in Proposition 3.3 has true common-sign-minus-independent risk contrast $+1/32$, while common-sign self-feedback reports $-1/4$ and selects the worse design. The practical and theoretical question is therefore the same: *can the network learn the more precise dependence law, correct prediction errors, and retain design-based inference without independently randomizing every cohort?*

The answer rests on a separation between evaluation and decision. Every centered sign design has $\mathrm{diag}(\Sigma_h) = \mathbf{1}$, so $\Sigma_h - \Sigma_g$ has zero diagonal and removes exactly the unobserved within-unit products. If an independently randomized reference cohort reveals $P = 2Y$, then

$$\mathrm{diag}(\Sigma_h - \Sigma_g) = 0 \quad \Longrightarrow \quad \mathbb{E}\big[P^\top (\Sigma_h - \Sigma_g) P\big] = \mu^\top (\Sigma_h - \Sigma_g)\mu. \tag{1}$$

Consequently, no absolute candidate risk is universally unbiasedly estimable from one cohort without outcome restrictions, whereas every pairwise risk difference is (Theorem 3.1). Treatment-effect heterogeneity contributes the same offset to every candidate score, so the ordering remains valid: the hospital can compare matching, rerandomization, and independence without estimating any design's absolute risk.

Prediction-Augmented Comparative Randomization Learning (PA-CRL) turns this cancellation into a predict–correct–learn procedure. On occasional independent-reference rounds, fresh simulated ghost assignments score the entire sampler library without constructing dense covariance matrices (Theorem 4.1). Between those rounds, a clipped predictable contrast model supplies an efficient ranking; an inverse-reference residual restores conditional unbiasedness under misspecification (Proposition 4.2). Entropic optimistic mirror descent converts the corrected contrasts into a contextual policy (Rakhlin & Sridharan, 2013), with expected regret that separates reference opportunity cost, policy complexity, and squared reference residual; cohort-size weighting yields expected excess-MSE control for unequal cohorts (Theorem 5.1). Each deployed assignment remains fresh and centered, so adaptation changes precision while preserving deterministic-horizon unbiasedness and an exact MSE identity; bounded equal-size cohorts further admit anytime-valid confidence sequences, and standard variance regularity yields asymptotic normality (Theorems 6.1 to 6.3).

The experiments follow this theorem chain. Exact enumeration reproduces the $+1/32 \mapsto -1/4$ reversal, ghost error contracts at the predicted $M^{-1/2}$ rate, and controlled schedules, a fixed misspecification surface, and public-data covariate replays test prediction correction and adaptive precision. Polynomial-hint PA-CRL improves on vanilla CRL in all 17 controlled cells (1.83–10.28%; median 5.66%). Under a matched misspecified linear predictor class, it improves on the raw plug-in by 3.4–42.3% throughout a contiguous 16/36-cell region at $\beta \geq .625$, with every pointwise seed-level 95% confidence interval excluding zero, while remaining below Bernoulli risk across the full surface.

Our contributions are threefold. **(i) Comparative estimability:** We characterize universal one-cohort unbiased quadratic-risk estimation, prove every centered-design contrast estimable under independent fair reference assignment, and give an exact correlated-feedback reversal (Theorem 3.1, Corollary 3.2, and Proposition 3.3). **(ii) Prediction-corrected learning:** We combine a sampler-only ghost oracle, conditionally unbiased prediction correction, and contextual optimistic mirror descent to obtain finite-horizon regret and expected excess-MSE control for unequal cohorts (Theorems 4.1 and 5.1, Proposition 4.2, and Algorithm 1). **(iii) Adaptive validity and evidence:** We establish deterministic-horizon unbiasedness, an exact MSE identity, anytime-valid inference for bounded equal-size cohorts, and asymptotic normality, then test each mechanism in synthetic and real-covariate experiments (Theorems 6.1 to 6.3 and Section 7).

Following Figure 1, Section 2 defines the decision problem; Sections 3 to 6 move from comparative identification to computable feedback, learning, and inference; Sections 7 and 8 present the evidence and implications; and the appendix contains related work, complete proofs, implementation and provenance details, and the machine-checked Lean subset.

2 Finite-population decision problem

Across repeated cohorts, the theory connects a design-specific precision target, a context-dependent rule for choosing among candidate samplers, and an effect estimator whose validity survives that adaptation. We first derive each sampler's finite-population risk, then lift candidate samplers to contextual policies, and finally separate the learning, estimation, and inference targets used in Steps 1–4 (Equations (2) and (3), Figure 2, and Sections 3 to 6).

For cohort t of size n_t, let $a_t, b_t \in [-1,1]^{n_t}$ be fixed treated and control potential outcomes, with $\mu_t = a_t + b_t$, $\delta_t = a_t - b_t$, and $\tau_t = n_t^{-1}\mathbf{1}^\top \delta_t$. A sign vector $Z_t \in \{-1,1\}^{n_t}$ reveals

$$Y_t = \tfrac{1}{2}(\mu_t + Z_t \odot \delta_t), \qquad \widehat{\tau}_t = \frac{2}{n_t} Z_t^\top Y_t = \tau_t + \frac{1}{n_t}\mu_t^\top Z_t. \tag{2}$$

Before assignment, let $\mathcal{F}_{t-1}$ be the sigma-field generated by completed cohorts and prior algorithmic randomness, and set $\mathcal{G}_t = \sigma(\mathcal{F}_{t-1}, X_t)$ after observing current covariates X_t. Candidate samplers and all quantities called predictable below are $\mathcal{G}_t$-measurable; the current reference decision, deployed assignment, and ghost draws use fresh random bits. A candidate $h \in [H]$ is a centered assignment sampler D_{th}, possibly a prespecified function of X_t, with actual covariance $\Sigma_{th} = \mathbb{E}_{D_{th}}[Z_t Z_t^\top]$. Necessarily $\mathrm{diag}(\Sigma_{th}) = \mathbf{1}$. The conditional randomization risk and its normalized form are

$$R_{th} = \mathbb{E}_{D_{th}}(\widehat{\tau}_t - \tau_t)^2 = n_t^{-2}\mu_t^\top \Sigma_{th}\mu_t, \qquad r_{th} = R_{th}/4 \in [0,1]. \tag{3}$$

Independent fair assignment is indexed by 0, with $\Sigma_{t0} = I$ and $r_{t0} \leq 1/n_t$.

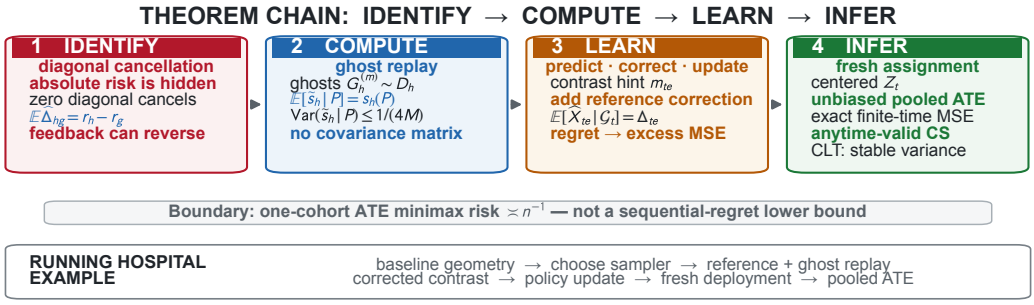

Figure 2: **The theorem chain converts hidden absolute risk into computable comparison, adaptive precision, and valid inference.** Zero-diagonal cancellation leads to sampler-only correction, contextual regret and excess MSE, then finite-horizon, anytime, and asymptotic inference (Theorems 3.1, 3.4, 4.1, 5.1 and 6.1 to 6.3 and Propositions 4.2 and 3.3). The lower ribbon follows Example 2.1.

Example 2.1 (Running hospital network)**.** *At each site, the network chooses among independent assignment, prespecified baseline-score matching, and rerandomization. The site context X_t records the cohort's announced baseline geometry; all candidates preserve treatment probability one half but induce different off-diagonal covariances. An independent-reference site provides unbiased relative-risk comparisons (Step 1), ghost assignments compute all candidate scores without new participants (Step 2), a model predicts the next site's contrasts and reference feedback corrects its residual (Step 3), and the freshly randomized site enters the pooled ATE with valid adaptive inference (Step 4).*

Declare a finite nonempty policy class $\mathcal{E}$, where $e \in \mathcal{E}$ maps X_t to $h_t(e) = e(X_t)$ and incurs $r_{te} = r_{t,h_t(e)}$; fixed labels, thresholds, and finite trees are special cases. Learning compares r_{te}, estimation targets $\tau_{1:T} = N_T^{-1} \sum_t n_t \tau_t$ with $\widehat{\tau}_{1:T} = N_T^{-1} \sum_t n_t \widehat{\tau}_t$, and inference covers that same average after predictable sampler choices, where $N_T = \sum_t n_t$. Potential outcomes may vary arbitrarily within the declared bounds; fresh centered assignment, no interference, a predictable clipped contrast predictor, and a predeclared finite policy class are the operative restrictions (Algorithm 1 and Table 1).

Equal treatment marginals therefore do not make the candidates equivalent: precision resides in off-diagonal dependence. Step 1 asks whether one observed reference cohort can compare that dependence even when it cannot reveal any candidate's absolute risk (Theorem 3.1).

3 Step 1: When comparison is estimable

Suppress t and let $P = 2Y = \mu + Z \odot \delta$ on an independent fair reference cohort. Step 1 moves from absolute-risk obstruction to zero-diagonal cancellation, a common-offset score, correlated-feedback reversal, pair-support validity, and the absolute minimax boundary (Theorems 3.1 and 3.4, Corollary 3.2, and Propositions C.1 and 3.3).

Theorem 3.1 (Absolute obstruction, comparative cancellation)**.** *For a fixed symmetric matrix A, an integrable statistic of (Z, Y), possibly using independent auxiliary randomness, is unbiased for $\mu^\top A \mu$ for every $(a, b) \in [-1, 1]^{2n}$ under independent fair assignment if and only if $\mathrm{diag}(A) = 0$. When this holds, $P^\top A P$ is unbiased. Hence no candidate's absolute risk has a universally unbiased estimator, whereas for every h, g,*

$$\widehat{\Delta}_{hg} = \frac{P^\top (\Sigma_h - \Sigma_g) P}{4n^2}, \qquad \mathbb{E}\widehat{\Delta}_{hg} = r_h - r_g. \tag{4}$$

The theorem concerns universal one-cohort unbiased estimability. Each diagonal term μ_i^2 contains the unobserved product $2a_i b_i$, and necessity rules out recovering it with any integrable, possibly randomized statistic without additional outcome structure. Covariance differences delete this term, so the reference site in Example 2.1 can unbiasedly compare matching with independence without

estimating either absolute risk. The resulting common-offset score is the algorithmic interface used below.

Corollary 3.2 (Common-offset feedback)**.** *For $s_h(P) = P^\top \Sigma_h P/(4n^2)$ and $d = \|\delta\|_2^2/(4n^2)$, we have $0 \le s_h \le \min\{1, \lambda_{\max}(\Sigma_h)/n\}$ and $\mathbb{E}s_h = r_h + d$. Thus $\arg\min_h \mathbb{E}s_h = \arg\min_h r_h$, including under arbitrary treatment-effect heterogeneity.*

All scores share the nuisance d, so treatment-effect heterogeneity changes the offset without changing their ordering in expectation.

Proposition 3.3 (Correlated-feedback bias)**.** *If feedback is generated by a centered assignment with covariance R, then the bias of the uncorrected h-minus-g score is*

$$\mathbb{E}_R[s_h(P) - s_g(P)] - (r_h - r_g) = \frac{\delta^\top\{(\Sigma_h - \Sigma_g) \circ R\}\delta}{4n^2}. \tag{5}$$

Why deployment feedback can reverse the decision. Take $n = 2$, $a = (1, -1/2)^\top$, $b = (-1/2, 1)^\top$, $\Sigma_0 = I$, and common-sign $\Sigma_+ = R = \mathbf{1}\mathbf{1}^\top$. Then $r_0 = 1/32$ and $r_+ = 1/16$, but $\mathbb{E}_R[s_+(P) - s_0(P)] = -1/4 < 0 < r_+ - r_0 = 1/32$: common-sign assignment is truly worse yet appears better under its own feedback. Equation (5) vanishes for $R = I$ but not for every marginally fair correlated design. Thus unbiased treatment-effect estimation can coexist with systematically reversed design-learning feedback.

How much reference randomization is needed? Independent reference is sufficient; Proposition C.1 permits any law with positive support on every pair cell used by the zero-diagonal quadratic form, whereas marginal support alone is insufficient.

Theorem 3.4 (Unrestricted minimax rate)**.** *Over every outcome-independent assignment distribution and every estimator based on (Z, Y) and independent auxiliary randomness,*

$$\frac{1}{n} \le \inf_{\mathcal{D}, \widetilde{\tau}} \sup_{a,b \in [-1,1]^n} \mathbb{E}_{\mathcal{D}}(\widetilde{\tau} - \tau)^2 \le \frac{4}{n}. \tag{6}$$

The sandwich in (6) concerns absolute one-cohort ATE estimation; together with Proposition C.2, it gives order n^{-1} and shows that no centered covariance uniformly dominates on every bounded finite population. The comparison task is therefore context-specific: PA-CRL learns which declared law fits the realized sequence. Step 1 identifies valid feedback, but direct evaluation of $P^\top \Sigma_h P$ appears to require every dense covariance; Step 2 removes that bottleneck using only sampler calls (Theorem 4.1).

4 Step 2: Computing comparisons from samplers

Step 2 turns the valid contrast into a sampler-only ghost oracle, then combines it with a predictable model whose residual is corrected on reference rounds; these objects feed the sequential learner (Theorems 4.1 and 5.1, Proposition 4.2, and Algorithm 1).

Theorem 4.1 (Sampler-based comparison score)**.** *For $M \ge 1$, after an independent-reference cohort reveals P, draw fresh simulated assignments $G_h^{(1)}, \dots, G_h^{(M)} \overset{\text{iid}}{\sim} D_h$, henceforth* ghost draws*, using random bits independent of the experiment and set*

$$\widetilde{s}_h^{(M)} = \frac{1}{M} \sum_{m=1}^{M} \frac{(P^\top G_h^{(m)})^2}{4n^2}. \tag{7}$$

Then $0 \le \widetilde{s}_h^{(M)} \le 1$, $\mathbb{E}[\widetilde{s}_h^{(M)} \mid P] = s_h(P)$, $\mathbb{E}\widetilde{s}_h^{(M)} = r_h + d$, and $\mathrm{Var}(\widetilde{s}_h^{(M)} \mid P) \le 1/(4M)$. The scores for H candidates cost $O(MHn)$ arithmetic plus sampler calls and require no covariance storage.

One reference vector now scores an entire library: the network in Example 2.1 can replay independence, matching, rerandomization, and GSW samplers without assigning another patient. A ghost draw is a simulated assignment, not an imputed outcome; increasing M removes conditional Monte Carlo error at rate $M^{-1/2}$ through the variance bound in Theorem 4.1, not randomness in P or the need for valid reference feedback. The next result uses prediction to reduce how often that feedback is requested.

Algorithm 1 Prediction-Augmented Comparative Randomization Learning (PA-CRL)

Require: centered samplers; policies $\mathcal{E}$; η, M; predictable ϵ_t and contrast predictor; fixed nonnegative ω_t

initialize $\pi_{1e} = 1/|\mathcal{E}|$ for every e
for $t = 1, \dots, T$ **do**
  observe X_t; from completed cohorts compute $h_t(e)$ and clipped m_{te}, setting $m_{te} = 0$ when $h_t(e) = 0$; form $q_{te} \propto \pi_{te} e^{-\eta \omega_t m_{te}}$
  draw $A_t \sim \text{Bernoulli}(\epsilon_t)$
  if $A_t = 1$ **then**
    draw $Z_t \sim D_{t0}$; observe Y_t; simulate fresh assignments and compute every $\widehat{D}_{te}$ (set $\widehat{D}_{te} = 0$ if $h_t(e) = 0$); set $\widehat{X}_{te} = m_{te} + (\widehat{D}_{te} - m_{te})/\epsilon_t$
  else
    draw $E_t \sim q_t$ and then $Z_t \sim D_{t,h_t(E_t)}$; observe Y_t; set $\widehat{X}_{te} = m_{te}$ for every e
  end if
  update the pooled effect, set $\pi_{t+1,e} \propto \pi_{te} e^{-\eta \omega_t \widehat{X}_{te}}$, and train the next predictor
end for

4.1 The correction primitive

Reference randomization identifies the contrast, while prediction fills intervening rounds. Proposition 4.2 gives one-round unbiasedness and exact variance before Algorithm 1 and Theorem 5.1 accumulate the corrections. Write $\Delta_{te} = r_{te} - r_{t0}$, let $\mathcal{G}_t$ be pre-decision information, and freeze any clipped prediction $m_{te} \in [-1, 1]$ before drawing A_t; on a reference round define

$$\widehat{D}_{te} = \widetilde{s}^{(M)}_{t,h_t(e)} - \frac{\|P_t\|_2^2}{4n_t^2}, \qquad \mathbb{E}[\widehat{D}_{te} \mid \mathcal{G}_t, A_t = 1] = \Delta_{te}, \tag{8}$$

$$\widehat{X}_{te} = \begin{cases} m_{te}, & A_t = 0, \\ m_{te} + \epsilon_t^{-1}(\widehat{D}_{te} - m_{te}), & A_t = 1. \end{cases} \tag{9}$$

This construction yields the exact correction guarantee.

Proposition 4.2 (Unbiased augmentation and exact variance)**.** *If $m_{te} \in [-1, 1]$ and $\epsilon_t \in (0, 1]$ are $\mathcal{G}_t$-measurable, $A_t \mid \mathcal{G}_t \sim \text{Bernoulli}(\epsilon_t)$, and the reference assignment and ghost bits are fresh and satisfy* (8)*, then $\mathbb{E}[\widehat{X}_{te} \mid \mathcal{G}_t] = \Delta_{te}$. With $v_{te} = \text{Var}(\widehat{D}_{te} \mid \mathcal{G}_t, A_t = 1)$,*

$$\text{Var}(\widehat{X}_{te} \mid \mathcal{G}_t) = \frac{v_{te}}{\epsilon_t} + \left(\frac{1}{\epsilon_t} - 1\right)(\Delta_{te} - m_{te})^2. \tag{10}$$

Thus calibration removes the missing-reference component of variance, not the intrinsic noise of the reference contrast.

Equation (10) separates identification from efficiency: positive-probability reference randomization identifies every predictable contrast, accurate prediction shrinks only the missing-reference term, and larger M shrinks only ghost Monte Carlo noise, leaving dense reference-contrast noise. In Example 2.1, a reference round corrects an omitted severity interaction that reverses the predicted ranking; the resulting unbiased loss vectors feed Algorithm 1 and Theorem 5.1.

5 Step 3: Prediction-corrected contextual learning

Step 3 converts the one-round correction into predict–correct–update learning: Algorithm 1 gives the loop, Theorem 5.1 separates reference cost, policy complexity, and reference residual, and (12) yields excess MSE for unequal cohorts.

The base π_t summarizes corrected past contrasts, while $q_{te} \propto \pi_{te} \exp(-\eta \omega_t m_{te})$ tilts it by the current hint. After a predictable reference decision and correction, PA-CRL updates $\pi_{t+1,e} \propto \pi_{te} \exp(-\eta \omega_t \widehat{X}_{te})$; Algorithm 1 states the complete loop.

The next theorem accumulates these conditionally unbiased innovations and exposes the three sources of regret.

Theorem 5.1 (Prediction-augmented contextual-policy regret)**.** *Let the finite populations, contexts, candidate samplers, deterministic nonnegative weights, and finite policy class $\mathcal{E}$ be fixed before the experiment. Let $\epsilon_t \in (0,1]$ and $\boldsymbol{m}_t \in [-1,1]^{|\mathcal{E}|}$ be predictable; conditionally on $\mathcal{G}_t$, draw $A_t \sim \text{Bernoulli}(\epsilon_t)$, and use fresh reference-assignment and ghost randomness so that* (8) *holds. Define $\bar{r}_t = \epsilon_t r_{t0} + (1-\epsilon_t)\sum_{e\in\mathcal{E}} q_{te} r_{te}$ and $V_T = \mathbb{E}\sum_t \frac{\omega_t^2}{\epsilon_t}\mathbb{E}[\|\widehat{\boldsymbol{D}}_t - \boldsymbol{m}_t\|_\infty^2 \mid \mathcal{G}_t, A_t = 1]$. For every $\eta > 0$,*

$$\mathbb{E}\sum_{t=1}^{T} \omega_t \bar{r}_t - \min_{e\in\mathcal{E}} \sum_{t=1}^{T} \omega_t r_{te} \le \mathbb{E}\sum_{t=1}^{T} \frac{\omega_t \epsilon_t}{n_t} + \frac{\log|\mathcal{E}|}{\eta} + \frac{\eta}{2} V_T. \tag{11}$$

The three terms in (11) charge reference opportunity cost, finite-policy complexity, and squared reference residual; $V_T \le 4\mathbb{E}\sum_t \omega_t^2/\epsilon_t$ without model correctness, and accurate contrast prediction can tighten it. In Example 2.1, the comparator is the best predeclared context-to-sampler rule, not merely the best fixed sampler; dynamic comparators and outcome-adaptive libraries require additional complexity control.

For $n_t \le n_{\max}$ and $\omega_t = n_t^2/n_{\max}^2$, the same bound has an exact statistical interpretation:

$$\mathbb{E}(\widehat{\tau}_{1:T} - \tau_{1:T})^2 - \min_{e\in\mathcal{E}} \frac{4}{N_T^2}\sum_{t=1}^{T} n_t^2 r_{te} \le \frac{4n_{\max}^2}{N_T^2} B_T, \tag{12}$$

where B_T is the right-hand side of (11). The weighting matches each cohort's contribution to the pooled estimator, so online loss control becomes expected excess MSE rather than a surrogate guarantee. Step 4 now separates that precision guarantee from inferential validity (Theorems 6.1 to 6.3).

6 Step 4: Inference survives adaptive selection

Adaptive sampler choice changes conditional covariance, not the estimand: with $\Sigma_t^{\text{sel}} = \mathbb{E}[Z_t Z_t^\top \mid \mathcal{G}_t]$, fresh centered assignment makes $\mu_t^\top Z_t$ a martingale difference even when past outcomes determine the current design. Step 4 therefore proceeds from exact deterministic-horizon risk to anytime coverage and asymptotic normality (Theorems 6.1 to 6.3).

Theorem 6.1 (Exact risk after predictable selection)**.** *Under no interference and the standing predictability and freshness conditions above, so that $\mathbb{E}[Z_t \mid \mathcal{G}_t] = 0$, for every deterministic T,*

$$\mathbb{E}\widehat{\tau}_{1:T} = \tau_{1:T}, \qquad \mathbb{E}(\widehat{\tau}_{1:T} - \tau_{1:T})^2 = N_T^{-2}\sum_{t=1}^{T} \mathbb{E}[\mu_t^\top \Sigma_t^{\text{sel}} \mu_t]. \tag{13}$$

The expression remains bounded above after replacing each selected covariance by an almost-sure positive-semidefinite upper certificate.

Hospital interpretation. The sampler may depend on all earlier sites but is fixed before current outcomes; centering then gives the unbiasedness and exact MSE in (13), while Σ_t^{sel} changes precision. Current-outcome assignment or cross-cohort interference lies outside this martingale argument.

For equal cohort sizes, inference can also be time-uniform using observed data. Put $X_t^{\text{obs}} = Z_t^\top Y_t/n$ and $Q_t = \sum_{b\le t}(X_b^{\text{obs}})^2$. Let $\psi(\lambda) = -\log(1-\lambda) - \lambda$, $\lambda_j = 2^{-(j+1)}$, $w_j = 6/\{\pi^2(j+1)^2\}$, and

$$u_\alpha(q) = \inf_{j\ge 0} \frac{\log\{2/(\alpha w_j)\} + \psi(\lambda_j) q}{\lambda_j}. \tag{14}$$

Theorem 6.2 (Observable confidence sequence)**.** *Under the conditions of Theorem 6.1, if $n_t \equiv n$, then for every $\alpha \in (0,1)$, with probability at least $1-\alpha$, simultaneously for all $t \ge 1$,*

$$\tau_{1:t} \in \left[\widehat{\tau}_{1:t} - \frac{2u_\alpha(Q_t)}{t}, \widehat{\tau}_{1:t} + \frac{2u_\alpha(Q_t)}{t}\right] \cap [-2,2]. \tag{15}$$

Any fixed nonempty finite subset of this dyadic grid remains valid with its displayed weights, and the infinite-grid boundary has the empirical-Bernstein iterated-logarithm order established in Appendix G.

For the hospital network, (15) remains valid under continuous monitoring and data-dependent stopping, without waiting for a fixed terminal site. This finite-sample guarantee may be conservative; the next result gives a Gaussian approximation when no cohort dominates and conditional variance stabilizes.

Theorem 6.3 (Martingale central limit condition)**.** *Under the conditions of Theorem 6.1, let* $D_t = \mu_t^\top Z_t$ *and* $s_T^2 = \sum_{t\le T} \mathbb{E}D_t^2$*. If* $s_T^2 \to \infty$*,* $\max_{t\le T} n_t/s_T \to 0$*, and* $s_T^{-2}\sum_{t\le T}\mathbb{E}[D_t^2 \mid \mathcal{G}_t] \to 1$ *in probability, then*

$$\frac{N_T(\widehat{\tau}_{1:T} - \tau_{1:T})}{s_T} = \frac{\sum_{t\le T} D_t}{s_T} \Rightarrow \mathcal{N}(0,1). \tag{16}$$

The scopes differ: Theorem 6.1 allows unequal cohorts, Theorem 6.2 uses bounded outcomes and equal sizes, and the limit in (16) adds diverging variance, no dominance, and variance stabilization; feasible Gaussian intervals also need an estimator of the oracle scale s_T. Prediction affects selected precision, not validity. The observable clock Q_t drives absolute uncertainty and includes risk plus τ_t^2 in conditional expectation (Proposition C.9); Section 7 tests the completed chain.

7 Experiments: evidence for the theorem chain

We probe observable consequences along the theorem chain rather than treating experiments as a leaderboard. Independent-reference feedback should preserve risk contrasts when deployment-correlated feedback does not; ghost scores should converge at the Monte Carlo rate; prediction should help precisely when it reduces the reference residual; and fresh centered deployment should preserve pooled inference. The candidate library contains Bernoulli assignment, three prognostic-pairing rules, and threshold rerandomization. We compare vanilla CRL with zero hint, linear and polynomial ridge plug-ins, and PA-CRL with matched hints. The primary metric is exact conditional design risk divided by Bernoulli risk. Potential outcomes are visible only to the evaluator, and best-fixed, context-map, and dynamic oracles are hindsight comparators.

Protocols. The scaling study uses $n = 12$, $p = 8$, three contexts, $T \in \{64, 128, 256, 512, 1024\}$, three frozen finite-population paths, and 64 randomizations per path. The $p \times T$ grid uses two paths and 48 repetitions per cell; each real-covariate replay uses 128. The fixed misspecification surface crosses nine interaction weights with four horizons, eight populations, and 128 histories per population–cell–method, retaining all 73,728 method-path rows. A separate 4×4 grid crosses reference rates $\{.05, .1, .2, .4\}$ with learning rates $\{1, 2, 4, 8\}$ on the same three $T = 512$ paths. Appendix K gives every construction, seed, baseline, and information boundary.

Identification and sampler-only computation. In the exact two-unit construction, the true common-sign-minus-independent contrast is $1/32$, whereas deployment-correlated feedback reports $-1/4$. Across 60 scenario–round–candidate cells, the largest discrepancy between the exact quadratic score and the ghost expectation is 1.39×10^{-17}, while empirical RMSE contracts as $M^{-1/2}$. With $M = 16$, ghost scoring changes end-to-end risk by only 0.02–0.64% across seven schedules; in the archived single-host run with identical rejection-sampler calls, dense/ghost wall-clock time rises from 1.12 at $n = 64$ to 1.98 at $n = 512$ (Figure 3f). These checks validate the constructive identities in Proposition 3.3 and Theorem 4.1; Theorems 3.1 and 3.4 establish the impossibility and minimax statements analytically.

Prediction correction, advantage region, and sensitivity. Polynomial-hint PA-CRL improves on vanilla CRL in all 17 controlled synthetic cells: risk falls by 1.8–10.3% (median 5.7%), including 4.5% at $T = 1024$ (Figure 3a). The sharper matched-model-class comparison exposes a contiguous robustness region. For every $\beta \ge .625$ and every displayed horizon, PA-CRL beats the raw linear plug-in in 16 of 36 complete-grid cells, by 3.4–42.3%; all 16 pointwise seed-level 95% intervals exclude zero, 15 win on all eight populations, and the gain increases with T in every row (Figure 3b,e). Here the linear predictor misorders candidates on 69.6–90.2% of rounds; on reference rounds, correction reverses 77.5–94.0% of those misorderings. Across the full surface, PA-CRL risk remains 0.171–0.714 of Bernoulli; the signed direct-model boundary, including the strong polynomial negative control, is reported in Appendix K. The 4×4 tuning grid likewise stays below Bernoulli and spans 0.171–0.632 (declared cell 0.253; Figure 3d). The retained residual-fit ablation worsens reference-residual MSE at every horizon (Figure 9), as predicted by Theorem 5.1.

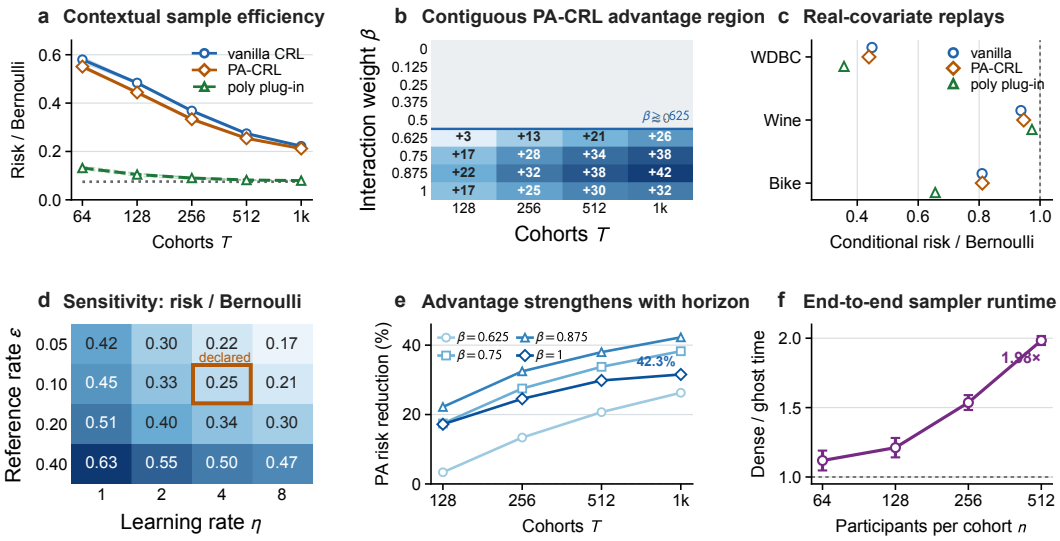

Figure 3: **The main evidence centers the confirmed advantage region.** (a) Conditional risk across horizon; faint curves show all three finite-population paths and bold curves their mean. (b) Full fixed interaction-weight–horizon surface comparing PA-CRL with the raw plug-in under a matched misspecified linear predictor specification; blue cells form the contiguous region whose pointwise seed-level 95% intervals exclude zero, and entries are percentage risk reductions, while gray cells retain the complementary grid without foregrounding its signed values (Table 8). (c) Exact risk on three zero-effect UCI covariate replays; intervals are 1.96 randomization Monte Carlo standard errors. (d) Complete reference-rate–learning-rate sensitivity grid at $T = 512$; orange marks the declared setting. (e) All 16 highlighted cells, showing that the advantage increases with horizon for every $\beta \geq .625$. (f) End-to-end dense/ghost timing with 1.96 standard errors over 20 runs.

Real-covariate replays and inference. For PA-CRL/vanilla CRL/polynomial plug-in, risk ratios are $0.438/0.448/0.356$ on WDBC, $0.946/0.937/0.973$ on Wine, and $0.810/0.810/0.656$ on Bike (Figure 3c) (Wolberg et al., 1993; Cortez et al., 2009; Fanaee-T, 2013). The mixed ordering shows that empirical geometry determines whether correction or direct modeling is more efficient. Across seven original synthetic and RAND schedules, the largest standardized pooled-effect mean error is 2.16 Monte Carlo standard errors; empirical all-time coverage is 99.8–100%, with terminal width reductions of 0–45.5% relative to Bernoulli (Appendix I). These zero-effect replays test design precision and inference under observed covariate geometry; they are not causal analyses of the source datasets.

8 Discussion

Comparison makes adaptive design learning possible even when absolute design risk cannot be evaluated: shared covariance diagonals cancel, ghost assignments compute the resulting contrasts, prediction reduces reference use, and inverse-reference correction restores the contrast target. This chain yields finite-horizon policy regret, expected excess-MSE control, unbiased pooled effects, and equal-cohort anytime coverage. In the executed study, PA-CRL improves on vanilla CRL in all 17 controlled cells and on the raw plug-in throughout the contiguous $\beta \geq .625$ region (3.4–42.3%); all 16 pointwise seed-level 95% confidence intervals exclude zero, while risk stays below Bernoulli in all 36 cells. Appendix K reports the complete signed boundary, including settings where a strong direct model is more efficient. Equation (10) turns that boundary into an operating rule: choose the reference rate by balancing held-out contrast error plus v_{te}/ϵ_t against the cost of reference cohorts. The present scope is a fixed finite policy library, bounded outcomes, fresh cohorts, and no interference. Extending the comparison principle to adaptive libraries, dependent cohorts, data-driven reference rates, and joint randomization–adjustment selection would support experiments under distribution shift, delayed outcomes, and network interference.

AI USE STATEMENT

Generative AI tools were used as assistive tools to brainstorm and refine the research framework, formulate candidate mathematical claims and proof arguments, design experiments, implement and test code, generate synthetic data through the documented simulation programs, clean and reformat the cited public datasets, interpret numerical results, design figures, and draft and edit text. The authors selected the research questions and methodology, re-derived and checked every theoretical claim and proof, and independently inspected every reported result; the formally encoded subset was additionally verified by the Lean kernel without placeholders. All numerical results were regenerated by executable code from declared inputs and checked against saved machine-readable outputs, and citations and closest-work statements were verified against primary sources. The authors reviewed every AI-assisted artifact and take responsibility for the paper's final content. The same disclosure will be entered in the OpenReview submission form.

ETHICS STATEMENT

This work studies statistical design rather than a deployed intervention. The empirical evaluation is confined to synthetic populations and publicly available datasets used as fixed-covariate, zero-effect design replays; it entails no new participant recruitment, data collection, or prospective treatment assignment and supports no causal conclusions about the source datasets. In practice, adaptive randomization changes the dependence structure of treatment assignments and can therefore affect subgroup balance and exposure even when every individual retains marginal treatment probability one half. Responsible deployment requires prospective specification of admissible designs and safety constraints, domain and ethics review, monitoring of subgroup assignments and outcomes, and independent oversight. Statistical precision is one design objective and must be considered jointly with participant welfare, fairness, privacy, scientific validity, and applicable legal and institutional requirements.

REPRODUCIBILITY STATEMENT

The supplement provides complete proofs, assumption-to-diagnostic mappings, pseudocode, data provenance, seeds, and the exact preprocessing and evaluation protocol. The accompanying archive contains documented reproduction entry points, raw repetition-level outputs, figure sources, exact finite-design enumerations, sampler-scaling benchmarks, and Lean source and build logs. No reported table entry is transcribed from an interactive analysis.

REFERENCES

David Arbour, Drew Dimmery, Tung Mai, and Anup Rao. Online balanced experimental design. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 844–864, 2022. URL https://proceedings.mlr.press/v162/arbour22a.html.

Peter M. Aronow and Cyrus Samii. Conservative variance estimation for sampling designs with zero pairwise inclusion probabilities. *Survey Methodology*, 39(1):231–241, 2013. URL https://www150.statcan.gc.ca/n1/pub/12-001-x/2013001/article/11831-eng.pdf.

Laura B. Balzer, Erica Cai, Lucas Godoy Garraza, and Pracheta Amaranath. Adaptive selection of the optimal strategy to improve precision and power in randomized trials. *Biometrics*, 80(1):ujad034, 2024. doi: 10.1093/biomtc/ujad034.

Nicolò Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006. doi: 10.1017/CBO9780511546921.

Nicolò Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. Minimizing regret with label efficient prediction. *IEEE Transactions on Information Theory*, 51(6):2152–2162, 2005. doi: 10.1109/TIT.2005.847729.

Fangyi Chen, Shu Ge, Jian Qian, and Christopher Harshaw. Sigmoid-FTRL: Design-based adaptive Neyman allocation for AIPW estimators, 2025. URL https://arxiv.org/abs/2511.19905v2. Version 2, revised 9 March 2026.

Paulo Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. Wine quality [dataset], 2009.

Jessica Dai, Paula Gradu, and Christopher Harshaw. CLIP-OGD: An experimental design for adaptive neyman allocation in sequential experiments. In *Advances in Neural Information Processing Systems*, volume 36, pp. 32235–32269, 2023. doi: 10.52202/075280-1400. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/661d4fda173120a2f119e0319e6bcf97-Abstract-Conference.html.

Miroslav Dudik, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 1097–1104. ACM, 2011. URL https://icml.cc/2011/papers/554_icmlpaper.pdf.

Hadi Fanaee-T. Bike sharing [dataset], 2013.

Peter Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980.

Christopher Harshaw, Fredrik Sävje, Daniel A. Spielman, and Peng Zhang. Balancing covariates in randomized experiments with the Gram–Schmidt Walk design. *Journal of the American Statistical Association*, 119(548):2934–2946, 2024. doi: 10.1080/01621459.2023.2285474.

Christopher Harshaw, Joel A. Middleton, and Fredrik Sävje. Optimized variance estimation under interference and complex experimental designs. *Journal of the American Statistical Association*, 2026. doi: 10.1080/01621459.2026.2627027. URL https://doi.org/10.1080/01621459.2026.2627027. Published online 13 May 2026.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020. doi: 10.1214/18-PS321.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021. doi: 10.1214/20-AOS1991.

Per Johansson, Donald B. Rubin, and Mårten Schultzberg. On optimal rerandomization designs. *Journal of the Royal Statistical Society: Series B*, 83(2):395–403, 2021. doi: 10.1111/rssb.12417.

Nathan Kallus. Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society: Series B*, 80(1):85–112, 2018. doi: 10.1111/rssb.12240.

Nathan Kallus. On the optimality of randomization in experimental design: How to randomize for minimax variance and design-based inference. *Journal of the Royal Statistical Society: Series B*, 83(2):404–409, 2021. doi: 10.1111/rssb.12412. Response.

Adam Kapelner and Abba Krieger. A matching procedure for sequential experiments that iteratively learns which covariates improve power. *Biometrics*, 79(1):216–229, 2023. doi: 10.1111/biom.13561. URL https://academic.oup.com/biometrics/article/79/1/216/7478046.

Xinran Li, Peng Ding, and Donald B. Rubin. Asymptotic theory of rerandomization in treatment-control experiments. *Proceedings of the National Academy of Sciences*, 115(37):9157–9162, 2018. doi: 10.1073/pnas.1808191115.

Joel A. Middleton. Unifying design-based inference: On bounding and estimating the variance of any linear estimator in any experimental design, 2021. URL https://arxiv.org/abs/2109.09220.

Kari Lock Morgan and Donald B. Rubin. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2):1263–1282, 2012. doi: 10.1214/12-AOS1008.

Georgy Noarov, Riccardo Fogliato, Martin Andres Bertran, and Aaron Roth. Stronger neyman regret guarantees for adaptive experimental design. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 46735–46761, 2025. URL https://proceedings.mlr.press/v267/noarov25a.html.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pp. 993–1019. PMLR, 2013. URL https://proceedings.mlr.press/v30/Rakhlin13.html.

Alejandro Schuler, David Walsh, Diana Hall, Jon Walsh, and Charles Fisher. Increasing the efficiency of randomized trial estimates via linear adjustment for a prognostic score. *The International Journal of Biostatistics*, 18(2):329–356, 2022. doi: 10.1515/ijb-2021-0072.

statsmodels developers. RAND health insurance experiment data, n.d. URL https://www.statsmodels.org/stable/datasets/generated/randhie.html. Dataset documentation and distributed source; accessed 8 September 2026.

William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer wisconsin (diagnostic) [dataset], 1993.

Khurram Yamin, Edward Kennedy, and Bryan Wilder. Dependent randomized rounding for budget constrained experimental design. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 4681–4700, 2025. URL https://proceedings.mlr.press/v286/yamin25a.html.

APPENDIX CONTENTS

A Relation to existing paradigms **14**

B New oracle and policy proofs **14**

C Deferred statements supporting the main text **16**

C.1 Geometry and covariance approximation 16

C.2 Fixed-label learner retained as a special case 17

C.3 Auxiliary inference identity 17

D Proofs for observable comparisons **17**

E Proofs for covariance geometry **20**

F Proofs for learning and covariance approximation **21**

G Proofs for inference **23**

H Implementation and verification boundaries **24**

I Experimental implementation and complete uncertainty reporting **25**

I.1 Executable specification and random-number control 25

I.2 Finite assignment distributions 26

I.3 Exact GSW endpoint enumeration 26

I.4 Potential outcomes and the information firewall 27

I.5 Real-data provenance and preprocessing 27

I.6 Normalization, Monte Carlo summaries, and coverage 27

I.7 Scope of sensitivity analyses and figure generation 28

I.8 Additional theory diagnostics: variance, pair support, and limits 28

J Additional executed experiments and stress tests **29**

J.1 Sampler-only ghost feedback 29

J.2 Strong outcome-model selectors and reference schedules 31

J.3 Frozen contextual-policy campaign 31

J.4 High-dimensional stress test 33

J.5 RAND cohort-order robustness 34

K Prediction-augmented comparative learning: additional experiments **34**

L Direct finite checks of correction, adaptive inference, and regret **36**

M Scope of machine verification **41**

Table 1: Assumptions are tied to an operational diagnostic and a defined fallback.

Condition	Why it enters	What is checked	If it fails
Centered fresh sampler; no interference	treatment-effect error is a martingale difference	enumerate/test marginals; assess scientific interference	redefine the randomization unit or estimand
Known bounded endpoint	bounded feedback and anytime boundary	declare range before outcomes	rescale a natural bound; clipping defines a new estimand
Independent reference assignments and fresh reference ghosts	reference contrasts have their conditional target	separated experimental, predictive-ghost, and reference-ghost seeds	pair-IPW only with known supported pair probabilities
Predictable clipped contrast predictor	supplies an efficient hint without identifying the target	freeze m_{te} before A_t and current outcomes	set $m_{te} = 0$ or retain correction-only CRL
Exogenous finite policy class	deterministic best-policy comparator	freeze policies and context features	per-policy bounds; dynamic comparators need new analysis

A Relation to existing paradigms

Constructing and adapting a design. Rerandomization restricts assignments to an announced balance region (Morgan & Rubin, 2012; Li et al., 2018); minimax design optimizes against a stated function class, with an explicit debate over precision and randomization support (Kallus, 2018; Johansson et al., 2021; Kallus, 2021); GSW and online balancing control declared covariate discrepancy (Harshaw et al., 2024; Arbour et al., 2022). Dependent randomized rounding constructs one budget-constrained law that preserves fixed marginals while inducing useful dependence (Yamin et al., 2025); we select sequentially among arbitrary centered laws from outcome feedback. Kapelner & Krieger (2023) adapt later pair matching using earlier responses. We instead compare arbitrary prespecified centered assignment laws through unbiased covariance-risk contrasts; the learned object is the dependence law, not a metric inside one matching family.

Estimability under a design. Zero pairwise inclusion probabilities obstruct unbiased variance estimation (Aronow & Samii, 2013). General matrix characterizations identify estimable or conservatively bounded quadratic forms under complex designs (Middleton, 2021; Harshaw et al., 2026). We use that theory at a decision boundary: the shared unit diagonal makes every difference of candidate covariances compatible with independent-reference data. The ghost construction then converts this structural fact into a loss oracle with linear arithmetic cost conditional on sampler calls. It is the contrast, rather than an absolute variance proxy, that drives selection.

Prediction, correction, and inference. Predictable-sequence online learning exploits useful hints while retaining worst-case guarantees (Rakhlin & Sridharan, 2013); label-efficient prediction studies the price of requested feedback (Cesa-Bianchi et al., 2005; Cesa-Bianchi & Lugosi, 2006). Our hint is an outcome-model risk contrast, and our identifying correction is the independent-reference contrast. This resembles the model-plus-inverse-propensity residual pattern of offline contextual-bandit evaluation (Dudik et al., 2011); here identification comes solely from the reference law, while the model controls efficiency. Prognostic adjustment and data-adaptive analysis selection use outcome models to improve estimator precision (Schuler et al., 2022; Balzer et al., 2024); our model chooses the randomization law for a future cohort. Adaptive Neyman methods instead learn marginal treatment probabilities (Dai et al., 2023; Noarov et al., 2025; Chen et al., 2025), whereas every marginal here remains one half. Time-uniform inference follows separately from predictable fresh assignments (Howard et al., 2020; 2021).

B New oracle and policy proofs

Proof of Theorem 4.1. Fix P. Since $G_h^{(m)}$ is centered with second moment Σ_h,

$$\mathbb{E}\Big[(P^\top G_h^{(m)})^2 \mid P\Big] = P^\top \mathbb{E}[G_h^{(m)} G_h^{(m)\top}] P = P^\top \Sigma_h P.$$

Averaging over m proves the conditional identity. Taking expectation over the independent reference assignment and applying Corollary 3.2 gives $\mathbb{E}\widetilde{s}_h^{(M)} = r_h + d$. Moreover $|P_i| = 2|Y_i| \le 2$ and $|G_{hi}^{(m)}| = 1$, so $|P^\top G_h^{(m)}| \le 2n$ and every summand in (7) belongs to $[0,1]$. Conditional independence across m and the elementary bound $\mathrm{Var}(V) \le 1/4$ for $V \in [0,1]$ yield $\mathrm{Var}(\widetilde{s}_h^{(M)} \mid P) \le 1/(4M)$. Each inner product takes $O(n)$ operations, proving the stated complexity. □

Proof of Proposition 4.2. Condition on $\mathcal{G}_t$. The predictor and reference probability are fixed, while A_t is Bernoulli and, on $\{A_t = 1\}$, the reference contrast has conditional mean Δ_{te}. The piecewise definition in (9) therefore gives

$$\mathbb{E}[\widehat{X}_{te} \mid \mathcal{G}_t] = (1-\epsilon_t)m_{te} + \epsilon_t \mathbb{E}\Big[m_{te} + \epsilon_t^{-1}(\widehat{D}_{te} - m_{te})\Big|\mathcal{G}_t, A_t = 1\Big] = \Delta_{te}.$$

Write $R_{te} = \widehat{D}_{te} - m_{te}$ on $\{A_t = 1\}$. Conditioning on the two values of A_t yields

$$\begin{aligned}\mathbb{E}[\widehat{X}_{te}^2 \mid \mathcal{G}_t] &= (1-\epsilon_t)m_{te}^2 + \epsilon_t \mathbb{E}[(m_{te} + \epsilon_t^{-1}R_{te})^2 \mid \mathcal{G}_t, A_t = 1] \\ &= m_{te}^2 + 2m_{te}(\Delta_{te} - m_{te}) + \epsilon_t^{-1}\mathbb{E}[R_{te}^2 \mid \mathcal{G}_t, A_t = 1] \\ &= \Delta_{te}^2 + \frac{v_{te}}{\epsilon_t} + \left(\frac{1}{\epsilon_t} - 1\right)(\Delta_{te} - m_{te})^2.\end{aligned}$$

Subtracting the squared conditional mean proves (10). The same calculation gives the innovation identity

$$\mathbb{E}[(\widehat{X}_{te} - m_{te})^2 \mid \mathcal{G}_t] = \epsilon_t^{-1}\mathbb{E}[(\widehat{D}_{te} - m_{te})^2 \mid \mathcal{G}_t, A_t = 1]. \tag{17}$$

For finite M, the law of total variance further separates v_{te} into the variance of the exact dense reference contrast and the conditional ghost variance; the latter is at most $1/(4M)$ when only the candidate score is ghost-approximated. Thus increasing M cannot remove the former component. □

Proof of Theorem 5.1 and (12). Let $R(p) = \sum_e p_e \log p_e$, so $D_R(u,p) = \mathrm{KL}(u\|p)$. For a base distribution p define

$$\mathrm{prox}_p(g) = \arg\min_{q \in \Delta(\mathcal{E})}\{\eta\langle q, g\rangle + \mathrm{KL}(q\|p)\}.$$

Its solution is proportional to $p_e e^{-\eta g_e}$. Put $M_t = \omega_t \boldsymbol{m}_t$, $G_t = \omega_t \widehat{\boldsymbol{X}}_t$, $q_t = \mathrm{prox}_{\pi_t}(M_t)$, and $\pi_{t+1} = \mathrm{prox}_{\pi_t}(G_t)$.

The entropic prox optimality condition and the Bregman three-point identity give the two inequalities

$$\begin{aligned}\eta\langle \pi_{t+1} - u, G_t\rangle &\le \mathrm{KL}(u\|\pi_t) - \mathrm{KL}(u\|\pi_{t+1}) - \mathrm{KL}(\pi_{t+1}\|\pi_t), \\ \eta\langle q_t - \pi_{t+1}, M_t\rangle &\le \mathrm{KL}(\pi_{t+1}\|\pi_t) - \mathrm{KL}(\pi_{t+1}\|q_t) - \mathrm{KL}(q_t\|\pi_t).\end{aligned}$$

Adding them to the identity $\langle q_t - u, G_t\rangle = \langle \pi_{t+1} - u, G_t\rangle + \langle q_t - \pi_{t+1}, M_t\rangle + \langle q_t - \pi_{t+1}, G_t - M_t\rangle$ yields, for every distribution u,

$$\begin{aligned}\eta\langle q_t - u, G_t\rangle \le \mathrm{KL}(u\|\pi_t) - \mathrm{KL}(u\|\pi_{t+1}) + \eta\langle q_t - \pi_{t+1}, G_t - M_t\rangle \\ - \mathrm{KL}(\pi_{t+1}\|q_t) - \mathrm{KL}(q_t\|\pi_t).\end{aligned}$$

Pinsker's inequality gives $\mathrm{KL}(\pi_{t+1}\|q_t) \ge \|\pi_{t+1} - q_t\|_1^2/2$. Hölder's inequality and $ab - a^2/2 \le b^2/2$, with $a = \|q_t - \pi_{t+1}\|_1$ and $b = \eta\|G_t - M_t\|_\infty$, then bound the last display by

$$\mathrm{KL}(u\|\pi_t) - \mathrm{KL}(u\|\pi_{t+1}) + \frac{\eta^2}{2}\|G_t - M_t\|_\infty^2.$$

Summing, taking π_1 uniform, and choosing a point mass at fixed policy e yields the pathwise optimistic-mirror-descent inequality

$$\sum_t \langle q_t - e_e, G_t\rangle \le \frac{\log|\mathcal{E}|}{\eta} + \frac{\eta}{2}\sum_t \|G_t - M_t\|_\infty^2. \tag{18}$$

Conditional unbiasedness and predictability of q_t give

$$\mathbb{E}\langle q_t - e_e, G_t\rangle = \mathbb{E}\left[\omega_t\left(\sum_f q_{tf} r_{tf} - r_{te}\right)\right].$$

Moreover, $G_t - M_t = (\omega_t A_t/\epsilon_t)(\widehat{\boldsymbol{D}}_t - \boldsymbol{m}_t)$, so the expectation of the final sum in (18) is exactly V_T. Because each component of $\widehat{\boldsymbol{D}}_t$ and $\boldsymbol{m}_t$ lies in $[-1,1]$, $\|\widehat{\boldsymbol{D}}_t - \boldsymbol{m}_t\|_\infty^2 \le 4$ and hence $V_T \le 4\mathbb{E}\sum_t \omega_t^2/\epsilon_t$. Finally,

$$\bar{r}_t - r_{te} \le \sum_f q_{tf} r_{tf} - r_{te} + \frac{\epsilon_t}{n_t},$$

because candidate risks are nonnegative and $r_{t0} = \|\mu_t\|_2^2/(4n_t^2) \le 1/n_t$. Multiplying by ω_t, taking expectations, and selecting the deterministic hindsight minimizer proves (11).

For the pooled estimator, (2) and conditional centering make $\mu_t^\top Z_t$ martingale differences. Conditional on $\mathcal{G}_t$, the reference/exploitation mixture gives

$$\mathbb{E}[(\mu_t^\top Z_t)^2 \mid \mathcal{G}_t] = 4n_t^2 \left\{ \epsilon_t r_{t0} + (1-\epsilon_t) \sum_e q_{te} r_{te} \right\} = 4n_t^2 \bar{r}_t.$$

All cross-cohort products have expectation zero by the tower property. Therefore

$$\mathbb{E}(\widehat{\tau}_{1:T} - \tau_{1:T})^2 = \frac{4}{N_T^2} \sum_t n_t^2 \mathbb{E}\bar{r}_t.$$

With $\omega_t = n_t^2/n_{\max}^2$, multiply (11) by $4n_{\max}^2/N_T^2$ to obtain (12). □

C Deferred statements supporting the main text

This section records secondary results whose proofs follow in subsequent appendices. They are separated from the headline theorem chain to keep the main argument within the ICLR page limit.

Proposition C.1 (Pair support and all-cohort feedback)**.** *Let a centered sign design have covariance R, and fix a symmetric matrix A. A universally unbiased estimator of $\mu^\top A\mu$ exists exactly when $A_{ii} = 0$ and all four pair cells have positive probability whenever $A_{ij} \neq 0$. Under that condition,*

$$U_A = \sum_{i \neq j} A_{ij} \frac{Y_i Y_j}{p_{ij}(Z_i, Z_j)}, \quad p_{ij}(s,u) = \frac{1 + suR_{ij}}{4} \tag{19}$$

is unbiased. If the design mixes independent signs with probability $\gamma > 0$, then

$$\mathbb{E}U_A^2 \le \frac{16}{\gamma} \left(\sum_{i \neq j} |A_{ij}| \right)^2. \tag{20}$$

Proposition C.2 (Restricted robustness benchmark)**.** *Among centered sign designs used with the unadjusted Horvitz–Thompson estimator,*

$$\inf_{\mathcal{D}: \mathbb{E}_{\mathcal{D}} Z = 0} \sup_{a,b \in [-1,1]^n} \tfrac{1}{4}\mathbb{E}_{\mathcal{D}}(\widehat{\tau} - \tau)^2 = 1/n. \tag{21}$$

Every covariance-I design attains the value. Any two distinct centered covariance matrices are each favored by some bounded finite population.

Under coordinatewise independent fair reference signs, the exact variance of the reference contrast used in diagnostics is

$$\mathrm{Var}(\widehat{\Delta}_{hg}) = \frac{4\|D_\delta A\mu\|_2^2 + 2\|D_\delta A D_\delta\|_F^2}{16n^4}, \quad A = \Sigma_h - \Sigma_g. \tag{22}$$

C.1 Geometry and covariance approximation

Let B have full column rank and columns of norm at most one, $G = B^\top B$, and $C = G^{-1}$. If $P = \mu + Z \odot \delta$ is observed under an independent-sign reference assignment, then

$$\mathbb{E}[P^\top C P] = \mu^\top C \mu + \sum_i C_{ii} \delta_i^2. \tag{23}$$

For the GSW sampler analyzed by Harshaw et al. (2024), the same C is a valid risk certificate:

Proposition C.3 (GSW certificate). *For all u and λ, $\mathbb{E}e^{\lambda u^\top Z} \le e^{\lambda^2 u^\top C u/2}$ and $\Sigma \preceq C$. The inequalities remain conditional when B is selected before fresh walk randomness.*

Proposition C.4 (Ellipsoidal certificate). *For $K, C \succeq 0$ and $\mathcal{U} = \{K^{1/2}v : \|v\|_2 \le r\}$, $\sup_{u\in\mathcal{U}} u^\top C u = r^2\lambda_{\max}(K^{1/2}CK^{1/2})$.*

Proposition C.5 (Uniform covariance simulation). *For H fixed sign samplers on n units and m iid assignments per sampler, let $\widehat{\Sigma}_h = m^{-1}\sum_j Z_h^{(j)} Z_h^{(j)\top}$ and $e_m = \sqrt{2\log(2Hn^2/\beta)/m}$. With probability at least $1-\beta$,*

$$\max_{h,i,j} |(\widehat{\Sigma}_h - \Sigma_h)_{ij}| \le e_m, \qquad \sup_{\mu\in[-2,2]^n} \max_h \frac{|\mu^\top(\widehat{\Sigma}_h - \Sigma_h)\mu|}{4n^2} \le e_m.$$

An estimated-risk minimizer is consequently within $2e_m$ of the minimum true normalized risk.

C.2 Fixed-label learner retained as a special case

For compatibility with the original finite-library analysis, define

$$\widehat{\ell}_{bh} = \frac{A_b}{\epsilon}\frac{P_b^\top \Sigma_{bh} P_b}{4n_b^2}, \qquad w_{b+1,h} = w_{bh}e^{-\eta\widehat{\ell}_{bh}}. \tag{24}$$

Write the cumulative actual-risk regret against the best fixed label as

$$\mathcal{R}_T := \mathbb{E}\left[\sum_{b=1}^T \bar{r}_b\right] - \min_{h\in[H]} \sum_{b=1}^T r_{bh}. \tag{25}$$

Theorem C.6 (Fixed-label actual-risk regret). *If $c_b \ge \max_h \min\{1, \lambda_{\max}(\Sigma_{bh})/n_b\}$, then*

$$\mathcal{R}_T \le \epsilon \sum_b \frac{1}{n_b} + \frac{\log H}{\eta} + \frac{\eta}{2\epsilon}\sum_b c_b^2. \tag{26}$$

In particular,

$$\mathcal{R}_T \le \epsilon T + \frac{\log H}{\eta} + \frac{\eta T}{2\epsilon}, \tag{27}$$

$$\mathcal{R}_T \le (1+\sqrt{2})T^{2/3}(\log H)^{1/3} \tag{28}$$

under the usual joint choice when $T \ge \log H$.

Corollary C.7 (Predictable-library boundary). *For a library frozen before each fresh assignment, the same right-hand side bounds*

$$\mathbb{E}\sum_b (\bar{r}_b - r_{bh}) \tag{29}$$

for every label h fixed before the experiment; it does not automatically permit a random pathwise hindsight minimizer.

Proposition C.8 (Covariance-approximation stability). *If positive semidefinite unit-diagonal matrices $\widetilde{\Sigma}_{bh}$ satisfy $\|\widetilde{\Sigma}_{bh} - \Sigma_{bh}\|_{\mathrm{op}} \le \xi_b$, run (24) with $\widetilde{\Sigma}_{bh}$ in place of Σ_{bh} and choose $c_b \ge \max_h \min\{1, \lambda_{\max}(\widetilde{\Sigma}_{bh})/n_b\}$. Then (26) gains at most $2\sum_b \xi_b/n_b$. A joint simulation failure probability β contributes at most an additional βT in unconditional expected regret.*

C.3 Auxiliary inference identity

Proposition C.9 (Observable-clock decomposition). $\mathbb{E}[(X_b^{\mathrm{obs}})^2 \mid \mathcal{F}_{b-1}, X_b] = \{R_b^{\mathrm{sel}} + \tau_b^2\}/4$.

D Proofs for observable comparisons

Proof of Theorem 3.1. Write $P = \mu + D_\delta Z$, where $D_\delta = \mathrm{diag}(\delta)$. Independent fair assignment gives $\mathbb{E}Z = 0$ and $\mathbb{E}ZZ^\top = I$. Therefore

$$\mathbb{E}[P^\top A P] = \mu^\top A\mu + \mathbb{E}[Z^\top D_\delta A D_\delta Z]$$

$$= \mu^\top A\mu + \sum_i A_{ii}\delta_i^2.$$

If $\mathrm{diag}(A) = 0$, this proves sufficiency.

For necessity, fix i and set all other potential outcomes to zero. Integrate any auxiliary randomization into the statistic. If its expectation exists, averaging over assignments with $Z_i = 1$ produces a function $F(a_i)$, and averaging over assignments with $Z_i = -1$ produces a function $G(b_i)$. The full expectation is $F(a_i) + G(b_i)$; the factors $1/2$ have been absorbed into these functions. This argument allows the statistic to depend on the complete assignment vector. The four-point rectangular difference of an additive function is zero. The corresponding difference of its proposed target, evaluated at $(a_i, b_i) \in \{0, 1\}^2$, is

$$A_{ii}\{(1+1)^2 - (1+0)^2 - (0+1)^2 + (0+0)^2\} = 2A_{ii}.$$

All four populations are allowed. Unbiasedness at all four therefore forces $A_{ii} = 0$. Repeating this argument for every i proves necessity.

For a centered sign design, $\Sigma_{ii} = \mathbb{E}Z_i^2 = 1$. Hence its absolute quadratic form has a nonzero diagonal and falls under the impossibility result. For two such designs, $\mathrm{diag}(\Sigma_h - \Sigma_g) = 0$; sufficiency followed by division by $4n^2$ gives Equation (4). □

Proof of Theorem 3.4. Fix any design distribution $\mathcal{D}$ and any estimator. Place the uniform prior on the 2^{2n} finite populations with all a_i, b_i independently taking values -1 and 1. This prior is supported inside the stated parameter space and is independent of the chosen design. For each realized assignment vector z of positive design probability, the observed outcome selects one of the two independent prior signs at each unit. Conditional on this z and any observed outcome vector $y \in \{-1, 1\}^n$, the n missing potential outcomes W_i remain independent fair signs. To see this directly, each specified vector of missing signs completes (z, y) to exactly one of the 2^{2n} equally weighted prior populations. The conditional mass of each of the 2^n completions is therefore 2^{-n}.

The treatment effect decomposes as

$$\tau = \frac{1}{n}\sum_{i=1}^n z_i y_i - \frac{1}{n}\sum_{i=1}^n z_i W_i.$$

Indeed, for $z_i = 1$ the corresponding term is $a_i - b_i = y_i - W_i$, whereas for $z_i = -1$ it is $a_i - b_i = W_i - y_i$. Conditional independence and $\mathrm{Var}(W_i) = 1$ now give

$$\mathbb{E}[\tau \mid Z, Y] = \frac{1}{n}\sum_i Z_i Y_i, \qquad \mathrm{Var}(\tau \mid Z, Y) = \frac{1}{n^2}\sum_i Z_i^2 = \frac{1}{n}.$$

Conditioning additionally on independent estimator randomization leaves this conditional variance unchanged. The conditional squared-error decomposition consequently implies

$$\mathbb{E}[(\widetilde{\tau} - \tau)^2 \mid Z, Y, \text{auxiliary randomness}] = (\widetilde{\tau} - \mathbb{E}[\tau \mid Z, Y])^2 + \frac{1}{n} \geq \frac{1}{n}.$$

If the estimator has infinite risk the inequality is immediate; otherwise every displayed conditional expectation is finite. Taking expectations proves a Bayes risk of at least $1/n$. A supremum over finite populations is at least their finite prior average. Since the design and estimator were arbitrary, taking their infimum proves the lower bound.

For the upper bound, use independent fair assignment and the Horvitz–Thompson estimator. It is unbiased and its mean squared error is

$$\frac{1}{n^2}\sum_i (a_i + b_i)^2 \leq \frac{4n}{n^2} = \frac{4}{n}.$$

This upper bound holds for every bounded finite population and proves the other inequality. □

Proof of Corollary 3.2 and the exact contrast-variance formula (22). The covariance Σ_h is positive semidefinite and has trace n. Its largest eigenvalue is at most n. Since $|P_i| = 2|Y_i| \leq 2$, $\|P\|_2^2 \leq 4n$. Consequently

$$0 \leq \frac{P^\top \Sigma_h P}{4n^2} \leq \frac{\lambda_{\max}(\Sigma_h)\|P\|_2^2}{4n^2} \leq \frac{\lambda_{\max}(\Sigma_h)}{n} \leq 1.$$

The expectation calculation in the preceding proof, now with $A = \Sigma_h$ and unit diagonal, gives $\mathbb{E}s_h = r_h + \|\delta\|_2^2/(4n^2)$.

For the variance calculation, put $A = \Sigma_h - \Sigma_g$, $v = D_\delta A\mu$, and $H = D_\delta A D_\delta$. The matrix H is symmetric with zero diagonal. Expanding the centered numerator yields

$$P^\top AP - \mu^\top A\mu = 2v^\top Z + Z^\top HZ.$$

The first summand has mean zero and second moment $4\|v\|_2^2$. The second has mean zero because H has zero diagonal. Their covariance is zero: every term in its expansion is proportional to $\mathbb{E}[Z_k Z_i Z_j]$ with $i \neq j$, and this expectation vanishes by independence and symmetry, including when $k = i$ or $k = j$. Finally,

$$Z^\top HZ = 2\sum_{i<j} H_{ij} Z_i Z_j.$$

Products belonging to two distinct unordered pairs are orthogonal in L^2, since their product contains at least one sign to an odd power. Hence

$$\mathbb{E}[(Z^\top HZ)^2] = 4\sum_{i<j} H_{ij}^2 = 2\|H\|_F^2.$$

Adding the two orthogonal second moments and dividing by $(4n^2)^2$ proves Equation (22). □

Proof of Proposition 3.3. For any centered assignment with covariance R, expansion gives

$$\mathbb{E}[P^\top AP] = \mu^\top A\mu + \operatorname{tr}(D_\delta A D_\delta R) = \mu^\top A\mu + \delta^\top (A \circ R)\delta.$$

Set $A = \Sigma_h - \Sigma_g$ and divide by $4n^2$.

For the stated example, $\mu = (1/2, 1/2)^\top$ and $\delta = (3/2, -3/2)^\top$. Direct evaluation gives $r_0 = 1/32$ and $r_+ = 1/16$. Under common-sign feedback, $Z = (S, S)^\top$ with S a fair sign, and

$$P = (1/2 + 3S/2,\ 1/2 - 3S/2)^\top.$$

Its coordinate sum equals one and its squared Euclidean norm equals five. Therefore $s_+ = 1/16$, $s_0 = 5/16$, and $\mathbb{E}(s_+ - s_0) = -1/4$, whereas $r_+ - r_0 = 1/32 > 0$. All four potential outcomes lie in $[-1, 1]$. □

Proof of Proposition C.1. For fair marginal signs, solve the four linear equations specifying total probability one, $\mathbb{E}Z_i = \mathbb{E}Z_j = 0$, and $\mathbb{E}Z_iZ_j = R_{ij}$. The unique solution is $p_{ij}(s,t) = (1 + stR_{ij})/4$.

If all four probabilities are positive, summing over the four cells gives

$$\begin{aligned}\mathbb{E}\left[\frac{Y_iY_j}{p_{ij}(Z_i, Z_j)}\right] &= a_ia_j + a_ib_j + b_ia_j + b_ib_j \\ &= (a_i + b_i)(a_j + b_j) = \mu_i\mu_j.\end{aligned}$$

Linearity proves unbiasedness of Equation (19) when the diagonal is zero.

The diagonal necessity argument from Theorem 3.1 uses only positive marginal probabilities and therefore still applies. Now suppose $A_{ij} \neq 0$ and a cell (s,t) has probability zero. Let u be unit i's potential outcome for sign s, and let v be unit j's potential outcome for sign t. Set every other potential outcome to zero. Because u and v are never jointly observed, the expectation of any statistic is of the form $F(u) + G(v) + c$. Its rectangular difference at $u, v \in \{0, 1\}$ is zero. In the target $\mu^\top A\mu$, the only term with a nonzero such mixed difference is $2A_{ij}uv$, whose difference is $2A_{ij} \neq 0$. This contradiction proves pair-support necessity.

In a γ mixture with independent fair signs, every pair cell has probability at least $\gamma/4$. Since each potential outcome has absolute value at most one,

$$\mathbb{E}\left[\left\{\frac{Y_iY_j}{p_{ij}(Z_i, Z_j)}\right\}^2\right] = \sum_{s,t} \frac{Y_i(s)^2 Y_j(t)^2}{p_{ij}(s,t)} \leq \frac{16}{\gamma}.$$

Write $L_A = \sum_{i \neq j} |A_{ij}|$. If $L_A = 0$ the claimed bound is immediate. Otherwise weighted Cauchy–Schwarz, applied pointwise, implies

$$U_A^2 \leq L_A \sum_{i \neq j} |A_{ij}| \left\{ \frac{Y_i Y_j}{p_{ij}(Z_i, Z_j)} \right\}^2.$$

Taking expectations proves Equation (20). No independence between different pairs is used. □

Proof of Proposition C.2. Every $\mu \in [-2, 2]^n$ is realized by the bounded potential outcomes $a = b = \mu/2$. Let S be an auxiliary vector of independent fair signs, used only to average over allowable finite populations. For any centered design covariance Σ,

$$\mathbb{E}_S \left[\frac{(2S)^\top \Sigma (2S)}{4n^2} \right] = \frac{\operatorname{tr}(\Sigma)}{n^2} = \frac{1}{n}.$$

The supremum over allowable μ is at least its average, so the minimax value is at least $1/n$. If $\Sigma = I$, then $\mu^\top \Sigma \mu / (4n^2) \leq 4n/(4n^2) = 1/n$ for every allowable μ, proving attainment and equality.

For the final statement, put $A = \Sigma_h - \Sigma_g$. It is nonzero, symmetric, and has trace zero. A nonzero symmetric matrix with trace zero has at least one positive and one negative eigenvalue: otherwise the sum of its eigenvalues could not be zero. Select corresponding nonzero eigenvectors v_+ and v_-. Multiply each by a sufficiently small positive scalar so that every coordinate has absolute value at most two. These scaled vectors are allowable μ vectors, and their quadratic forms with A have opposite signs. Setting $a = b = \mu/2$ realizes the two required bounded finite populations. □

E PROOFS FOR COVARIANCE GEOMETRY

Proof of Proposition C.3. Theorem 6.6 of Harshaw et al. (2024) gives, for the fixed-pivot Gram–Schmidt Walk and every vector θ, the centered bound $\mathbb{E} \exp\{\theta^\top (BZ - \mathbb{E}[BZ])\} \leq \exp(\|\theta\|_2^2/2)$. The walk starts from the zero vector, for which its stated mean-preservation property gives $\mathbb{E}[Z] = 0$; hence the same inequality holds with BZ in place of $BZ - \mathbb{E}[BZ]$. Fix u, λ and put $\theta = \lambda BCu$. Then $B^\top \theta = \lambda u$, $\theta \in \operatorname{col}(B)$, and

$$\|\theta\|_2^2 = \lambda^2 u^\top C B^\top B C u = \lambda^2 u^\top C u.$$

The imported inequality applied to θ proves the displayed mgf bound. Since the law is centered and has finite support, its mgf is analytic. Comparing the second derivatives at zero gives $u^\top \Sigma u \leq u^\top C u$ for every u, which is exactly the Loewner inequality. Equivalently one can divide the sum of the two inequalities for λ and $-\lambda$ minus two by λ^2 and let $\lambda \to 0$.

For the conditional statement, condition on the selected matrix and pre-assignment history. The walk then has exactly the deterministic-input law to which the imported theorem applies. The same algebra holds for each conditioning value. This conditional substitution requires fresh randomization and the declared pivot rule; marginal fairness alone is insufficient. □

Proof of Proposition C.4. Put $A = K^{1/2} C K^{1/2} \succeq 0$. The finite-dimensional spectral theorem gives an orthonormal eigenbasis with eigenvalues $\lambda_i \geq 0$. For $\|v\|_2 \leq r$, $v^\top A v \leq \lambda_{\max}(A) \|v\|_2^2 \leq r^2 \lambda_{\max}(A)$. Equality is attained by r times a unit eigenvector for the largest eigenvalue, including the case $A = 0$; if $r = 0$ both sides are zero. Finally, $\gamma I - A \succeq 0$ holds if and only if $\gamma \geq \lambda_{\max}(A)$. This proves both formulas without requiring K to be invertible. □

Derivation of (23)**.** Under independent reference signs, $P = \mu + Z \odot \delta$. For $i \neq j$, $\mathbb{E}[P_i P_j] = \mu_i \mu_j$, whereas $\mathbb{E}[P_i^2] = \mu_i^2 + \delta_i^2$. Expand the quadratic form entry by entry. The extra diagonal term is precisely the one stated in (23).

Proof of Proposition C.5. For each fixed triple (h, i, j) the summands $Z_i^{(k)} Z_j^{(k)}$ are independent and lie in $[-1, 1]$. The bounded-variable Hoeffding inequality gives

$$\mathbb{P}\{|(\widehat{\Sigma}_h - \Sigma_h)_{ij}| > e\} \leq 2 \exp(-me^2/2).$$

A union bound over at most Hn^2 triples at $e = e_m$ proves the first assertion. On that event, for every μ in the stated cube,

$$|\mu^\top(\widehat{\Sigma}_h - \Sigma_h)\mu| \le e_m(\sum_i |\mu_i|)^2 \le 4n^2 e_m.$$

The event is independent of which μ is later inserted, so this is already uniform and requires no additional net or union bound over outcomes. If $\widehat{h}$ minimizes estimated normalized risk and h^* minimizes true normalized risk, insert and subtract the two estimated risks to obtain

$$r_{\widehat{h}} - r_{h^*} \le |r_{\widehat{h}} - \widehat{r}_{\widehat{h}}| + (\widehat{r}_{\widehat{h}} - \widehat{r}_{h^*}) + |\widehat{r}_{h^*} - r_{h^*}| \le 2e_m.$$

All minima exist because the library is finite. The claim about the empirical law follows by direct averaging of $(\pm Z^{(k)})(\pm Z^{(k)})^\top$; the extra global sign forces the mean to zero without changing the second moment. □

F Proofs for learning and covariance approximation

Lemma F.1 (Exponential weights with nonnegative estimated losses)**.** *Let $\widehat{\ell}_{bh} \ge 0$ be arbitrary finite numbers, $w_{1h} = 1$, and update weights by Equation* (24)*. For every h and every $\eta > 0$,*

$$\sum_{b=1}^{T}\sum_{g} q_{bg}\widehat{\ell}_{bg} - \sum_{b=1}^{T}\widehat{\ell}_{bh} \le \frac{\log H}{\eta} + \frac{\eta}{2}\sum_{b=1}^{T}\sum_{g} q_{bg}\widehat{\ell}^2_{bg}. \tag{30}$$

Proof. For $x \ge 0$, define $f(x) = 1 - x + x^2/2 - e^{-x}$. Then $f(0) = f'(0) = 0$ and $f''(x) = 1 - e^{-x} \ge 0$, so $f(x) \ge 0$. Thus $e^{-x} \le 1 - x + x^2/2$ for every nonnegative x, with no upper restriction on x.

Let $W_b = \sum_h w_{bh}$. The update gives

$$\begin{aligned}\frac{W_{b+1}}{W_b} &= \sum_h q_{bh} e^{-\eta\widehat{\ell}_{bh}} \\ &\le 1 - \eta\sum_h q_{bh}\widehat{\ell}_{bh} + \frac{\eta^2}{2}\sum_h q_{bh}\widehat{\ell}^2_{bh}.\end{aligned}$$

The expression on the last line is positive because it bounds the positive expression on the preceding line. Using $\log(1 + x) \le x$ for $x > -1$ and summing over b yields

$$\log\frac{W_{T+1}}{W_1} \le -\eta\sum_{b,h} q_{bh}\widehat{\ell}_{bh} + \frac{\eta^2}{2}\sum_{b,h} q_{bh}\widehat{\ell}^2_{bh}.$$

On the other hand, $W_1 = H$ and $W_{T+1} \ge w_{T+1,h} = \exp(-\eta\sum_b \widehat{\ell}_{bh})$ for each h. Therefore

$$\log\frac{W_{T+1}}{W_1} \ge -\eta\sum_b \widehat{\ell}_{bh} - \log H.$$

Combining the two inequalities and dividing by η proves the lemma. □

Proof of Theorem C.6. Condition on $\mathcal{H}_{b-1}$. The distribution q_b is fixed under this conditioning, and the reference indicator has conditional probability ϵ. On a reference round, the assignment is independent fair Bernoulli and the candidate covariance matrices were fixed before the assignment. By Corollary 3.2,

$$\mathbb{E}[\widehat{\ell}_{bh} \mid \mathcal{H}_{b-1}] = r_{bh} + d_b, \qquad d_b = \frac{\|\delta_b\|_2^2}{4n_b^2}. \tag{31}$$

The same conditioning and $s_{bh} \le c_b$ give

$$\mathbb{E}\left[\sum_h q_{bh}\widehat{\ell}^2_{bh}\,\middle|\,\mathcal{H}_{b-1}\right] = \frac{1}{\epsilon}\sum_h q_{bh}\mathbb{E}[s^2_{bh} \mid A_b = 1, \mathcal{H}_{b-1}] \le \frac{c_b^2}{\epsilon}. \tag{32}$$

All expectations are finite because the losses are bounded by c_b/ϵ. Take expectations in Equation (30). The offset d_b cancels because $\sum_h q_{bh} = 1$, producing, for each fixed h,

$$\mathbb{E}\sum_{b=1}^{T}\left(\sum_g q_{bg} r_{bg} - r_{bh}\right) \le \frac{\log H}{\eta} + \frac{\eta}{2\epsilon}\sum_{b=1}^{T} c_b^2. \tag{33}$$

For the deployed design,

$$\begin{aligned}\bar{r}_b - r_{bh} &= \sum_g q_{bg} r_{bg} - r_{bh} + \epsilon\left(r_{b0} - \sum_g q_{bg} r_{bg}\right)\\ &\le \sum_g q_{bg} r_{bg} - r_{bh} + \frac{\epsilon}{n_b}.\end{aligned}$$

The final inequality uses $r_{bg} \ge 0$ and $r_{b0} = \|\mu_b\|_2^2/(4n_b^2) \le 1/n_b$. Sum and use Equation (33). Since the population and library sequence are fixed in advance, a minimizer of $\sum_b r_{bh}$ is a deterministic label. Choosing that label gives Equation (26).

Setting $c_b = 1$ and $n_b \ge 1$ proves Equation (27). For a fixed ϵ, the two learning terms there are minimized at $\eta = \sqrt{2\epsilon \log H/T}$, with sum $\sqrt{2T \log H/\epsilon}$. The proposed $\epsilon = (\log H/T)^{1/3}$ belongs to $(0, 1]$ when $T \ge \log H$. Substitution gives a reference-round term $T^{2/3}(\log H)^{1/3}$ and a learning term $\sqrt{2}T^{2/3}(\log H)^{1/3}$, proving Equation (28). □

Proof of Corollary C.7. Enlarge the pre-assignment sigma-field to contain the current finite population and covariates, but require that it contain neither the reference indicator nor the fresh assignment randomness. Freeze the candidate library at this stage. Conditional versions of Equations (31) and (32) then hold unchanged. Their tower expectations give Equation (33) for every label h fixed before the experiment. The reference-round cost argument is pointwise and still applies. This proves Equation (29).

The step that chooses a deterministic minimizer in the previous proof is unavailable when the losses themselves depend on prior random observations. Conditional unbiasedness for every fixed label does not justify substituting a label selected from the complete random loss path. This is why the corollary asserts only the displayed fixed-label comparison. □

Proof of Proposition C.8. Define $\widetilde{r}_{bh} = \mu_b^\top \widetilde{\Sigma}_{bh}\mu_b/(4n_b^2)$. The unit diagonals imply that each approximate score has conditional expectation $\widetilde{r}_{bh} + d_b$, with the same offset as before. Positive semidefiniteness gives nonnegative scores and the spectral bounds used to define c_b. Therefore the proof of Equation (33) gives the same bound with all r replaced by $\widetilde{r}$.

The operator-norm inequality and bounded outcomes imply, for every b, h,

$$|\widetilde{r}_{bh} - r_{bh}| \le \frac{\xi_b\|\mu_b\|_2^2}{4n_b^2} \le \frac{\xi_b}{n_b}.$$

It follows pointwise that

$$\sum_g q_{bg} r_{bg} - r_{bh} \le \sum_g q_{bg}\widetilde{r}_{bg} - \widetilde{r}_{bh} + \frac{2\xi_b}{n_b}.$$

Summation followed by the unchanged actual-risk reference-round cost bound proves the deterministic approximation claim.

For the probability statement, condition first on all approximation randomness. On the joint good event the just-proved conditional expected bound applies. On its complement the actual regret is at most T, since each normalized actual risk lies in $[0, 1]$ and each comparator risk is nonnegative. The independent sources of simulation and experimental random bits ensure that conditioning on the simulation matrices preserves the conditional Bernoulli reference-round law. The observed outcomes themselves need not be independent of these matrices, because the matrices influence design selection. The deterministic envelopes c_b and ξ_b make the good-event upper bound nonrandom. Combining the two events adds at most βT. This step would require a different conditional argument if approximation randomness were selected using the current experimental outcomes. □

G PROOFS FOR INFERENCE

Proof of Theorem 6.1. For each cohort put

$$D_b = 2Z_b^\top Y_b - \mathbf{1}^\top \delta_b, \qquad M_t = \sum_{b \le t} D_b.$$

The observation equation gives

$$2Z_b^\top Y_b - \mathbf{1}^\top \delta_b = Z_b^\top \mu_b + \sum_i (Z_{bi}^2 - 1)\delta_{bi} = Z_b^\top \mu_b,$$

because $Z_{bi}^2 = 1$. Thus $D_b = Z_b^\top \mu_b$ is $\mathcal{F}_b$-measurable and $\mathbb{E}[D_b \mid \mathcal{G}_b] = 0$, because the selected law, including the reference and candidate mixture, is centered. Since $\mathcal{F}_{b-1} \subseteq \mathcal{G}_b$, the tower property gives $\mathbb{E}[D_b \mid \mathcal{F}_{b-1}] = 0$. Bounded outcomes ensure square integrability for each finite cohort, so $(D_b, \mathcal{F}_b)$ is a martingale-difference sequence and, for $a < b$, $\mathbb{E}[D_a D_b] = 0$. Hence

$$\mathbb{E}M_t = 0, \qquad \mathbb{E}M_t^2 = \sum_{b \le t} \mathbb{E}D_b^2 = \sum_{b \le t} \mathbb{E}[\mu_b^\top \Sigma_b^{\mathrm{sel}} \mu_b].$$

The sample size N_t is deterministic. Dividing the first identity by N_t gives unbiasedness and dividing the second by N_t^2 gives the exact MSE. The bound follows pointwise from positive semidefinite ordering. □

Lemma G.1 (Bounded empirical-Bernstein factor)**.** *If $|X| \le 1$ almost surely, $\mathcal{F}$ is a sigma-field, $m = \mathbb{E}[X \mid \mathcal{F}]$, and $0 < \lambda < 1$, then*

$$\mathbb{E}[\exp\{\lambda(X - m) - \psi(\lambda)X^2\} \mid \mathcal{F}] \le 1.$$

Proof. For $x \in [-1, 1]$, the power series for $\log(1 + \lambda x)$ converges absolutely. Its remainder satisfies

$$\lambda x - \log(1 + \lambda x) = \sum_{k \ge 2} \frac{(-1)^k (\lambda x)^k}{k} \le x^2 \sum_{k \ge 2} \frac{\lambda^k}{k} = \psi(\lambda) x^2.$$

Every inequality is valid also for $x = \pm 1$, since $\lambda < 1$. Exponentiating gives $e^{\lambda x - \psi(\lambda)x^2} \le 1 + \lambda x$; the right side is positive. Taking conditional expectation and multiplying by $e^{-\lambda m}$ yields a bound of $(1 + \lambda m)e^{-\lambda m} \le 1$. The last inequality follows from $1 + y \le e^y$ for real y. All conditional expectations exist by boundedness. □

Proof of Theorem 6.2. Set $S_t = \sum_{b \le t} X_b^{\mathrm{obs}}$, $\theta_b = \tau_b/2$, and $W_t = S_t - \sum_{b \le t} \theta_b$. We have $|X_b^{\mathrm{obs}}| \le 1$, and conditional centering together with (2) gives $\mathbb{E}[X_b^{\mathrm{obs}} \mid \mathcal{G}_b] = \theta_b$. Lemma G.1 therefore yields, for every fixed $\lambda \in (0, 1)$,

$$\mathbb{E}\big[\exp\big\{\lambda(X_b^{\mathrm{obs}} - \theta_b) - \psi(\lambda)(X_b^{\mathrm{obs}})^2\big\}\big|\mathcal{G}_b\big] \le 1.$$

Because $\mathcal{F}_{b-1} \subseteq \mathcal{G}_b$ and X_b^{obs} is $\mathcal{F}_b$-measurable, another application of the tower property shows that

$$L_t^+(\lambda) = \exp\{\lambda W_t - \psi(\lambda) Q_t\}$$

is a nonnegative $(\mathcal{F}_t)$-supermartingale starting at one. Applying the same argument to $-X_b^{\mathrm{obs}}$ gives $L_t^-(\lambda) = \exp\{-\lambda W_t - \psi(\lambda) Q_t\}$.

For completeness, the maximal inequality used here follows by stopping a nonnegative supermartingale L at the first crossing of $c > 0$, truncated at T. Repeated conditional expectation gives $\mathbb{E}L_{\sigma \wedge T} \le 1$, and nonnegativity implies $c\mathbb{P}(\sigma \le T) \le 1$. Letting $T \to \infty$ proves $\mathbb{P}(\exists t : L_t \ge c) \le 1/c$. This is Ville's inequality.

Apply it to each sign at λ_j with threshold $2/(\alpha w_j)$. A union bound over both signs and all j costs at most $\sum_j \alpha w_j = \alpha$. On the complementary event, simultaneously for all t, j,

$$|W_t| < \frac{\log\{2/(\alpha w_j)\} + \psi(\lambda_j) Q_t}{\lambda_j}.$$

Taking the infimum gives $|W_t| \le u_\alpha(Q_t)$. The bound is finite, and the expression tends to infinity as $j \to \infty$ for each fixed q, so the infimum is attained. Since $2W_t/t = \widehat{\tau}_{1:t} - \tau_{1:t}$, the closed interval covers. Intersecting with $[-2, 2]$ preserves coverage, because the target is in that interval. Restricting to a fixed nonempty finite subset of the declared dyadic grid removes crossing events from the union and preserves the same error bound while retaining the corresponding weights w_j without renormalization.

We give details of the rate. The series for ψ gives $\psi(\lambda) \le \lambda^2$ for $0 < \lambda \le 1/2$. For $q \ge 1$, let

$$L(q) = \log\left[\frac{\pi^2\{2 + \log_2(q+1)\}^2}{3\alpha}\right], \quad a(q) = \min\{1/2, \sqrt{L(q)/q}\}.$$

Choose a dyadic grid value λ_j with $a(q)/2 < \lambda_j \le a(q)$. Because $L(q) > 1$, such an index satisfies $j + 1 \le 2 + \log_2(q+1)$. Therefore the logarithmic term in (14) is at most $L(q)$ and

$$u_\alpha(q) \le \frac{2L(q)}{a(q)} + a(q)q.$$

If $a(q) = \sqrt{L(q)/q}$ this is at most $3\sqrt{qL(q)}$. Otherwise $q \le 4L(q)$ and it is at most $6L(q)$. On $0 \le q < 1$, the fixed choice $j = 0$ bounds $u_\alpha(q)$ by a constant multiple of $1 + \log(1/\alpha)$. These cases prove the displayed asymptotic order. This order statement concerns the infinite grid; a finite implementation remains valid but eventually has linear growth in q at its smallest grid point. □

Proof of Proposition C.9. Since $2X_b^{\mathrm{obs}} = \widehat{\tau}_b$ and $\mathbb{E}\widehat{\tau}_b = \tau_b$ conditionally on the pre-assignment history,

$$4\mathbb{E}[(X_b^{\mathrm{obs}})^2 \mid \mathcal{G}_b] = \mathrm{Var}(\widehat{\tau}_b \mid \mathcal{G}_b) + \tau_b^2.$$

The conditional variance equals R_b^{sel} by (3). This proves the claim, including history-dependent candidate probabilities. □

Proof of Theorem 6.3. To match the conditioning in the variance assumption, use the interlaced filtration

$$\mathcal{H}_{t,2b-2} = \mathcal{F}_{b-1}, \qquad \mathcal{H}_{t,2b-1} = \mathcal{G}_b, \qquad \mathcal{H}_{t,2b} = \mathcal{F}_b, qquad1 \le b \le t,$$

and the triangular-array increments $\xi_{t,2b-1} = 0$ and $\xi_{t,2b} = D_b/s_t$. The inclusions $\mathcal{F}_{b-1} \subseteq \mathcal{G}_b \subseteq \mathcal{F}_b$ and $\mathbb{E}[D_b \mid \mathcal{G}_b] = 0$ make this a martingale-difference array. Its predictable quadratic variation is exactly

$$\sum_{k=1}^{2t} \mathbb{E}[\xi_{t,k}^2 \mid \mathcal{H}_{t,k-1}] = s_t^{-2} \sum_{b \le t} \mathbb{E}[D_b^2 \mid \mathcal{G}_b],$$

which converges in probability to one by assumption. Moreover $|D_b| \le \sum_i |\mu_{bi}| \le 2n_b$. For every fixed $\varepsilon > 0$, the condition $\max_{b \le t} n_b/s_t \to 0$ implies that $|D_b| \le \varepsilon s_t$ simultaneously for all $b \le t$, for all sufficiently large t. Thus the conditional Lindeberg sum

$$s_t^{-2} \sum_{b \le t} \mathbb{E}[D_b^2 \mathbf{1}\{|D_b| > \varepsilon s_t\} \mid \mathcal{G}_b]$$

is eventually identically zero. The martingale-array central limit theorem with predictable quadratic-variation convergence and conditional Lindeberg condition (Hall & Heyde, 1980) now gives $\sum_{b \le t} D_b/s_t \Rightarrow \mathcal{N}(0, 1)$. This proof invokes that established limit theorem; it does not assume independence of the adaptive cohort increments. □

H IMPLEMENTATION AND VERIFICATION BOUNDARIES

The executable implementation is `code/reproduce.py`, and `code/plot_results.py` creates the figures from generated data. The main candidate library consists of independent Bernoulli assignment, three explicit matching rules, and a threshold rerandomization law with enumerated support. Every retained assignment is paired with its sign reversal. This guarantees centering and permits exact covariance computation. It does not imply independent signs, which is why the reference indicator determines whether the simple common-offset feedback can be used.

For a matching partition, each pair receives opposite fair signs, independently across pairs. Its covariance has diagonal one, covariance minus one within each pair, and zero across pairs. Its eigenvalues are zero and two. For an enumerated symmetric assignment set $\mathcal{Z}$, the uniform law has covariance $|\mathcal{Z}|^{-1}\sum_{z\in\mathcal{Z}} zz^\top$. These identities are direct calculations of the implemented law, not approximations to a nominal sampler.

The separate GSW diagnostic recursively enumerates both legal endpoint updates and their exact mean-preserving transition probabilities. The pivot persists within a phase until it reaches an endpoint. Floating-point least-squares solutions and numerical tolerances mean that this diagnostic is a numerical implementation check. It evaluates covariance bounds, selected mgf directions, and feedback substitutions, but cannot replace the all-direction algorithmic theorem.

All methods are evaluated on the same fixed population schedule within a scenario. Their learning inputs contain only the allowed observed outcomes. Potential-outcome vectors stored in environment files are used by evaluation code to compute risk and coverage; they are not used in the weight update. Repetition-level outputs allow every reported mean and Monte Carlo standard error to be recomputed. The real-data provenance file records source, row selection, normalization, and checksum. Seeds and hyperparameters are fixed in the configuration file.

The prediction-augmented campaign under `experiments/calibrated_crl/` is a separate frozen entry point. It implements the predict–correct–update loop in Equation (9) and Algorithm 1, crossed $p \times T$ experiments, an archived-schedule model-failure stress test, three disjoint calibration/replay splits from pinned UCI archives, and an end-to-end threshold-rerandomization benchmark. The runner verifies the pinned archive SHA-256 hashes before preprocessing. The separate validator checks cardinalities, exact-comparator consistency, recorded provenance, phase-grid completeness, stress outputs, and that the three generated figure PDFs contain vector text with no embedded raster images.

Theoretical proofs, numerical experiments, and Lean kernel checks address different propositions. Numerical enumeration can find a contradiction to an implementation claim but does not prove the general theorem. The handwritten arguments explicitly invoke established GSW and martingale limit results where indicated. Appendix M states the exact extent of machine verification; unformalized probability arguments are not counted as kernel-checked merely because their algebraic identities compile.

I Experimental implementation and complete uncertainty reporting

I.1 Executable specification and random-number control

The complete entry point for the original low-dimensional and RAND replay campaign is `code/reproduce.py`. With Python, NumPy, pandas, SciPy, and Matplotlib available, run

```
python code/reproduce.py --repetitions 512 --horizon 1024
```

The script regenerates the data summaries, figure PDFs/SVGs/PNGs, LaTeX tables, and numerical macros. It reads the included public-domain `data/randhie_source.csv` and does not download data. `code/plot_results.py` regenerates figures from saved outputs; `code/report_tables.py` regenerates manuscript numbers. The executed runtime used Python 3.12.14, NumPy 2.5.3, pandas 2.2.3, SciPy 1.18.1, and Matplotlib 3.11.1. Library versions are recorded because pseudorandom streams and numerical linear algebra should not be assumed bitwise invariant across environments.

For synthetic populations the generator is `numpy.random.default_rng(20260908)`. Each scenario first generates the same $T \times 12 \times 2$ covariate array and the same independent $T \times 12$ uniform-noise array. Scenario definitions then deterministically form μ and δ. Population generation is completed before randomization repetitions start. The randomized health-data row permutation uses seed 20260909. The exact GSW diagnostic uses seed 998 for its six-by-two covariate matrix. Indexing scenarios from zero in the order listed in the main text, fixed-assignment simulations use seed $7100 + s$ and each adaptive ablation uses seed $8100 + s$. Sensitivity cells use seed 9200 and

256 repetitions. All streams are NumPy generators; no global random state or current outcome enters a design decision.

Each experiment vectorizes the independent randomization repetitions. At each cohort it first normalizes the weights, draws an independent reference indicator, and samples a candidate on deployment repetitions. Fresh assignment randomness samples uniformly from the selected support. The learner's score is computed after assignment only on reference rounds. We subtract the largest log weight before exponentiating, which stabilizes normalization without changing the probabilities.

I.2 Finite assignment distributions

Write $\mathcal{Z}_n = \{-1, 1\}^n$ and list all vectors in binary-index order. Bernoulli assignment is uniform on $\mathcal{Z}_{12}$. In Pair 1 and Pair 2, participants are stably sorted by the specified covariate, with original order breaking ties. Consecutive participants form pairs. Pair all takes the first remaining participant and matches it to the remaining participant of smallest squared Euclidean distance, resolving distance ties by participant index, and repeats. There are six pairs and 64 equally likely independent pair orientations. Pair all is a completely specified greedy design; it is not asserted to solve an optimal matching problem.

For rerandomization, compute $d(z) = \|X_b^\top z\|^2$ for all 4096 sign vectors. Set c_b to NumPy's empirical 0.10 quantile with the `higher` convention and retain every z with $d(z) \le c_b + 10^{-10}$. The numerical tolerance only resolves floating-point ties. Since $d(-z) = d(z)$, this produces a sign-symmetric support and exact marginal centering. No treatment-count restriction is imposed. It is a threshold design rather than the original draft's best-of-64 candidate heuristic.

For each support $\mathcal{S}_{bh}$ the covariance used by both sampling and feedback is

$$\Sigma_{bh} = \frac{1}{|\mathcal{S}_{bh}|} \sum_{z \in \mathcal{S}_{bh}} z z^\top.$$

The script verifies zero column means to tolerance 10^{-12} and a unit covariance diagonal to the same tolerance. Covariances are not substituted by inverse-Gram bounds. Support enumeration has exponential cost and is appropriate here as a reference calculation; it is not a proposed large-cohort implementation.

I.3 Exact GSW endpoint enumeration

The separate diagnostic constructs

$$B_\phi = \begin{pmatrix} \sqrt{\phi} I_6 \\ \sqrt{1-\phi}\, X^\top / \xi \end{pmatrix}, \qquad \xi = \max_i \|x_i\|.$$

Starting at zero, the largest live coordinate is the pivot. Since it remains the largest live coordinate until it is frozen, this is the fixed-order classical GSW pivot rule. The other live coordinates minimize the Euclidean norm of $B_\phi u$ subject to $u_p = 1$. The least-squares system is solved using NumPy. Both maximal endpoint steps are followed recursively with their martingale probabilities. Each step freezes at least one coordinate; the recursion therefore has depth at most six. For each of the four displayed ϕ values, the implemented tree has 64 endpoint leaves. The covariance is the probability-weighted sum of their outer products.

The diagnostic outcome schedule is $\mu_i = 1.5 \tanh(2x_{i1})$ and $\delta_i = 0.4 \sin(3x_{i2})$. Expected feedback scores are averaged over all 64 independent Bernoulli reference vectors, regardless of which candidate covariance is scored. The common offset is 0.00455409121501031. The maximum centering residual is 6.3×10^{-17}, the maximum diagonal residual is 3.4×10^{-16}, and the maximum common-bias-identity residual is 5.3×10^{-18}. The smallest computed eigenvalue of $(B_\phi^\top B_\phi)^{-1} - \Sigma_\phi$ is nonnegative up to floating-point precision in every case. These checks validate the implementation at the displayed inputs; the general GSW inequality is an imported theorem, and its mathematical status does not depend on this numerical check.

I.4 Potential outcomes and the information firewall

For every synthetic scenario, define $a_{bi} = (\mu_{bi} + \delta_{bi})/2$ and $b_{bi} = (\mu_{bi} - \delta_{bi})/2$. The code verifies $|\mu_{bi}| + |\delta_{bi}| < 2$. The six outcome-sum definitions are

Signal 1:	$1.5\tanh(2.5x_{bi1})$,
Signal 2:	$1.5\tanh(2.5x_{bi2})$,
Interaction:	$1.5\tanh(4x_{bi1}x_{bi2})$,
No signal:	$1.5u_{bi}, \quad u_{bi} \sim \mathrm{Unif}[-1, 1]$,
Drift:	Signal 1 for $b \leq 512,$ then Signal 2,
Heterogeneous effects:	Signal 1.

The effect vector is $\delta_{bi} = 0.2$ except in the heterogeneous-effect scenario, where it is $0.45\sin(4x_{bi2})$. The true effect is consequently allowed to vary across participants and cohorts in that scenario. The design engine receives the current baseline covariates and historical reference scores. The simulation engine uses both potential outcomes to generate the one observed outcome and to evaluate risk, just as a standard potential-outcome simulation must; it does not pass the unobserved sum vector to the weight update. In particular, on a reference round the code forms $P_b = \mu_b + Z_b \odot \delta_b = 2Y_b$ and computes the quadratic score from that observed vector.

The no-update ablation has the same reference probability as the reference-only learner and therefore isolates the effect of learning. The uniform mixture removes both updating and the extra Bernoulli-reference mixture. Both are needed: comparing only with a uniform mixture would confound the update's benefit with the precision cost of the reference policy.

I.5 Real-data provenance and preprocessing

The source file is the unchanged `statsmodels` RAND HIE CSV, whose SHA-256 digest is

`9f6c87d05aef087a82cc4465310c8cd3f38327be6eafa43bd81fb98c4f3d088c`.

Its official documentation identifies `disea` as chronic disease burden and the four used indicators as physical limitation and self-rated health; it also notes that the table is a subset of the original study data (statsmodels developers, n.d.). The data are public domain. The study does not use historical insurance assignment as the replay treatment and does not infer counterfactual insurance outcomes.

The first 2019 indices in the fixed permutation are the calibration set; the next 12,288 indices form the replay, with no overlap or replacement. Their zero-based source row indices are stored separately, with each replay record's cohort and within-cohort slot. The calibration disease mean and population-standard-deviation convention yield $m = 11.144205477959387$ and $s = 6.743732680770911$. Every covariate is centered and scaled using its calibration mean and population standard deviation, with a lower standard-deviation floor of 0.1. The target is the bounded transformation $g_i = \tanh\{(\texttt{disea}_i - m)/s\}$. This transformation is part of the estimand definition, rather than an assertion that the original endpoint was bounded. The replay fixes $a_i = b_i = g_i$ and hence $\mu_i = 2g_i$, $\delta_i = 0$. Pair 1 sorts physical limitation; Pair 2 sorts the good-health indicator; Pair all and rerandomization use all four standardized indicators.

All revealed outcomes can enter the inference clock. Only designated reference-round outcomes enter the weight update. Although the negative-control construction makes both potential outcomes identical, the learner is not given that identity and does not receive deployment-round score updates. The evaluation exploits the known identity to compute the true randomization risk.

I.6 Normalization, Monte Carlo summaries, and coverage

The output files distinguish three quantities. First, `risk` is mean normalized cohort risk $T^{-1}\sum_b \bar{r}_b$, where $r_{bh} = \mu_b^\top \Sigma_{bh}\mu_b/(4n^2)$. Second, `final_error` is half of the pooled raw Horvitz–Thompson error:

$$e_T = \frac{1}{T}\sum_{b=1}^{T}\frac{Z_b^\top \mu_b}{2n} = \frac{\widehat{\tau}_{1:T} - \tau_{1:T}}{2}.$$

Third, the summary's `mse_final` is the empirical mean of e_T^2. The main text's raw MSE table multiplies this by four. Therefore the theoretical cross-check is $\mathbb{E}e_T^2 = T^{-1}\mathbb{E}\bar{r}_T$. No additional

Table 2: Complete inference evaluation for the principal learner and Bernoulli. Coverage is empirical all-time coverage in percent; widths are on the original bounded-outcome treatment-effect scale. The final column is percentage width reduction.

Scenario	CRL cov.	Bern. cov.	CRL width	Bern. width	Reduction (%)
Signal 1	100.00	100.00	0.0558	0.0920	39.4
Signal 2	100.00	100.00	0.0558	0.0921	39.4
Interaction	100.00	100.00	0.0661	0.0808	18.2
No signal	100.00	100.00	0.0721	0.0721	0.0
Drift	100.00	100.00	0.0576	0.0924	37.7
Heterogeneous	100.00	99.80	0.0433	0.0795	45.5
RAND HIE	99.80	100.00	0.0707	0.0729	3.1

factor of 12 is needed, because both the cohort estimator and the quadratic risk already contain their respective n and n^2 denominators.

Monte Carlo standard errors are the sample standard deviations divided by $\sqrt{512}$, with the usual $R-1$ denominator. Risk contrasts between adaptive ablations use paired repetitions and the sample variance of the within-repetition difference. Fixed-design conditional risks have zero Monte Carlo error because the support is enumerated, but their realized final squared errors do not. The largest absolute standardized mean-error discrepancy among the 63 method–scenario combinations is 2.16 standard errors. The corresponding maximum discrepancy between empirical final MSE and its theoretical expectation is 3.94 estimated standard errors. These are descriptive multiple comparisons, not acceptance thresholds. A separately seeded 10,000-repetition rerun removes the finite-Monte-Carlo RAND uniform-mixture reversal and agrees with the exact conditional-risk ordering; both runs remain archived. The exact risk calculation is the primary criterion and avoids selecting a favorable Monte Carlo realization.

For inference, the code accumulates

$$X_b^{\mathrm{obs}} = \frac{Z_b^\top \mu_b}{2n} + \frac{\tau_b}{2} = \frac{Z_b^\top Y_b}{n}, \qquad Q_t = \sum_{b \le t} (X_b^{\mathrm{obs}})^2,$$

and evaluates the minimum in (14) at the 20 indices $j = 0, \dots, 19$. The w_j are not renormalized. The equality to the observed score is a simulation algebra identity; the implemented interval requires no access to μ_b or τ_b. The truth appears only in the all-time coverage check $|\sum_{b \le t}(X_b^{\mathrm{obs}} - \tau_b/2)| \le u_\alpha(Q_t)$. Reported full interval width is $4u_\alpha(Q_t)/t$. At the final horizon the intervals lie inside $[-2, 2]$, so the target-range intersection does not change the reported widths. At early horizons intersecting with that range preserves coverage.

I.7 Scope of sensitivity analyses and figure generation

The complete sensitivity table contains all 60 combinations of three populations, five reference probabilities, and four learning rates. Every cell is retained. The chosen main setting $(\epsilon, \eta) = (0.1, 4)$ is fixed in the entry point and is not replaced by the best cell. The learning rate called theoretical is optimized for the coarse, universally bounded loss calculation; a sharper covariance envelope can improve that tuning. The study makes no claim that this rate is empirically optimal.

Figures use only saved arrays or deterministic functions of those arrays. Heatmaps display actual risk ratios, expected biases, or average candidate probabilities. The synthetic density plot uses SciPy's Gaussian-kernel density estimator for display and overlays raw empirical quantiles; KDE smoothing changes no reported statistic. The RAND cumulative-distribution panel plots quantiles from 0.02 through 0.98; its caption identifies the display rather than claiming exact stochastic ordering. Error bars are Monte Carlo intervals, not patient-level confidence intervals. All figures are exported as editable PDF/SVG together with PNG previews, and none is manually altered after generation.

I.8 Additional theory diagnostics: variance, pair support, and limits

The independent script `code/theory_diagnostics.py`, invoked by the complete entry point, uses seed 38472. For six covariate vectors uniform on $[-1, 1]^2$, set $\mu_i = 1.2 \tanh(3x_{i1})$ and

$\delta_i = 0.6\sin(2x_{i2})$. Evaluate all ten covariance contrasts from the five-rule library over the complete 64-assignment Bernoulli support. The largest absolute errors in the variance formula (22) and expected contrast are 2.2×10^{-19} and 7.0×10^{-18}, respectively. Figure 4a includes every contrast.

For Proposition C.1, fix $A = \Sigma_{\text{Pair 1}} - I$ and sample from the mixture with covariance $R_\gamma = \gamma I + (1-\gamma)\Sigma_{\text{Pair 1}}$. Enumerate the mean and variance of $U_A/(4n^2)$ at $\gamma \in \{0.01, 0.05, 0.2, 0.5, 1\}$ using exact assignment probabilities. The largest mean error is 2.1×10^{-17}. Variance rises from 0.0002308 at $\gamma = 1$ to 0.0442094 at $\gamma = 0.01$. At $\gamma = 0$, missing within-pair same-sign cells preclude universal unbiased estimation; this cell is recorded as unavailable. Weak pair support has a precision cost, without implying optimality of the reference learner.

For the Gaussian-limit diagnostic, $10{,}000$ repetitions use independent common-sign cohorts with $a_i = b_i = 1$, giving a fair-sign observed score. At $T = 16, 64, 256$, standardized sums have empirical Kolmogorov distances 0.1149, 0.0510, and 0.0303 from a standard Gaussian, and fourth moments 2.722, 2.749, and 2.978. These illustrate a submodel of Theorem 6.3.

The counterexample keeps these fixed outcomes and $n = 4$, but selects either Bernoulli or common-sign assignment with equal probability at time zero and retains that label. Conditional on it, cohorts use fresh randomization. Score variances are $1/4$ and 1; their unconditional mean is $5/8$, to which conditional variance does not stabilize. The $T = 256$ sum normalized by $\sqrt{5T/8}$ has empirical fourth moment 4.225. Its limiting variance mixture has fourth moment 4.08, versus the Gaussian value 3. Centering and fixed outcomes persist, as does finite-sample confidence-sequence validity.

Finally, Bernoulli assignment with $a_i = b_i = 1$ has exact raw HT MSE $4/n$. At each $n \in \{1, 2, 4, 8, 16, 32, 64, 128\}$, $10{,}000$ independent assignments estimate this MSE and its Monte Carlo standard error. This checks the classical restricted benchmark in Proposition C.2.

J Additional executed experiments and stress tests

This section reports every campaign added after the original archive review. Each configuration, seed list, and evaluation rule is fixed in the executable archive; no population seed, dimension, repetition, or adverse comparison is omitted. Exact conditional risks use the finite assignment support whenever it can be enumerated. For a history-dependent method, "MCSE" is the standard error over independent learning histories conditional on the finite population. Variation across independently generated finite populations is reported separately.

J.1 Sampler-only ghost feedback

For three representative archived schedules (Signal 1, heterogeneous effects, and RAND HIE), we selected cohorts $t \in \{0, 256, 512, 1023\}$ and all five candidate laws. At each of these 60 cells we enumerated the candidate support to compute $P^\top \Sigma_h P/(4n^2)$ and the exact mean of $(P^\top G_h)^2/(4n^2)$. The largest absolute discrepancy over all cells was 1.39×10^{-17}. We then generated $2{,}000$ independent ghost batches for $M \in \{1, 4, 16, 64\}$. The saved file `data/supplemental/ghost_mc_tradeoff.csv` contains every empirical bias, RMSE, and conditional standard deviation. The RMSE contracts at the predicted $M^{-1/2}$ scale.

The end-to-end comparison reuses the same reference indicators and candidate-selection uniforms for the dense and ghost learners, so their difference is paired. Across 512 histories, $M = 16$ increases mean risk relative to exact covariance scoring by 0.02–0.64% over the seven schedules. At $M = 4$ the range is 0.09–1.59%; at $M = 1$ it is 0.10–6.43%. Thus the oracle identity is exact, whereas a finite ghost batch introduces an explicitly controlled learning noise rather than a hidden covariance approximation.

The scaling microbenchmark uses five antithetic empirical sign laws, each with 64 support vectors, and $n \in \{16, 32, 64, 128, 256, 512, 1024\}$. At $n = 1024$, evaluating five dense quadratic forms takes 6.62 ms after 214 ms of covariance construction and stores 40 MiB of covariance matrices. A 16-ghost batch takes 0.049 ms and the explicit support used only for this benchmark occupies 2.5 MiB. A procedural sampler need not store that support. These are scoring microbenchmarks on one host, not wall-clock guarantees for arbitrary design samplers.

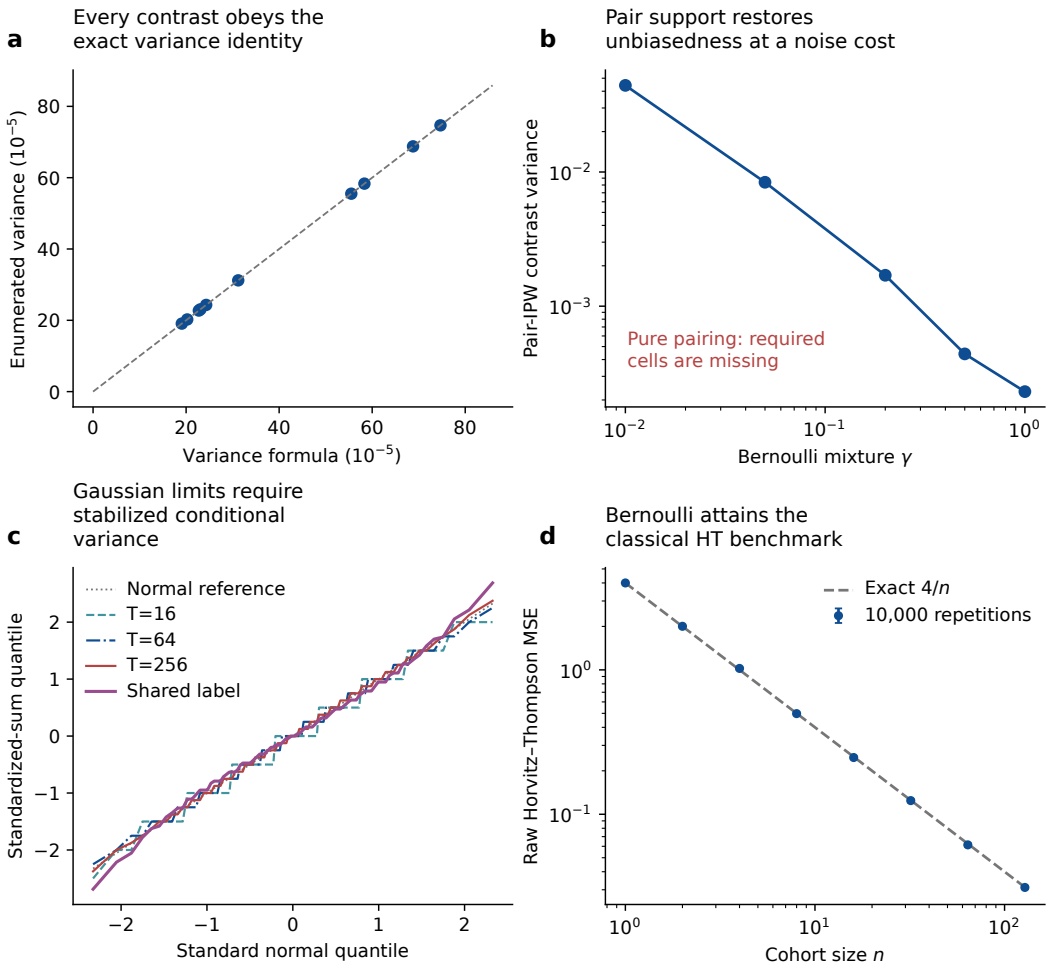

Figure 4: **Additional diagnostics verify finite identities and expose the assumptions behind limiting claims.** (a) All ten covariance contrasts satisfy the exact reference-variance formula under complete enumeration. (b) Normalized pair-IPW variance under a Bernoulli–pair mixture; every positive mixture weight gives unbiased feedback, but pure pairing lacks required cells. (c) Standardized common-sign sums approach Gaussian quantiles as T grows; retaining one randomly selected design label across all cohorts produces a non-Gaussian variance mixture. There are $10{,}000$ independent repetitions per displayed construction. (d) The empirical raw Bernoulli MSE at the worst-case outcome vector follows the exact $4/n$ benchmark; bars are 1.96 Monte Carlo standard errors. All plotted numbers are generated by the supplemental script.

Table 3: Theorem–experiment map. The last column states the observation that would have contradicted the claimed mechanism rather than merely weakened a performance comparison.

Claim	Intervention	Recorded evidence	Falsifying diagnostic
Common-offset identity	enumerate all reference signs and candidate supports	exact score, risk, and offset residual	candidate-specific residual above numerical tolerance
Ghost oracle	vary $M \in \{1, 4, 16, 64\}$ and $n \leq 1024$	conditional bias, RMSE, end-to-end risk, time, storage	persistent bias or RMSE failing to contract with M
Policy learning	remove updates while matching reference indicators	exact risk and deployment path	learned and no-update paths indistinguishable under recurring signal
Contextual comparator	switch the observed regime every 64 cohorts	policy risk and context-specific deployment	no separation between context-conditioned actions
Dimension stress	compare fixed nuisance dimensions at common T, n	risk by method, seed, and p	omitted dimensions/seeds or hidden access to the outcome sum
Predictable reference rate	replace the fixed reference probability by a frozen decay schedule	reference fraction and paired risk	current-outcome-dependent reference decisions
Real-covariate stability	permute frozen RAND cohorts 20 times	exact conditional risk and MCSE	gain driven by one favorable cohort order

J.2 Strong outcome-model selectors and reference schedules

The linear plug-in regresses all previously observed $2Y$ values on an intercept and the declared covariates. The polynomial plug-in adds cubic terms and pairwise interactions. Both use ridge penalty one, every completed cohort, and the same current candidate library; neither is charged a reference-round mixture. In the original finite population, polynomial ridge has lower mean risk than CRL in every archived schedule. Eight additional independently generated populations retain the same ordering for Signal 1, interaction, and drift; the polynomial map also wins on seven of eight no-signal populations. Linear ridge loses to CRL on all eight interaction populations, while the declared polynomial map wins on all eight. This isolates response-surface specification as the relevant difference: direct contrast feedback does not dominate a richer polynomial-feature outcome model, but its targeting identity does not require one.

The predictable schedule $\epsilon_t = \max\{0.025, 0.4t^{-1/3}\}$ is fixed before the campaign. It lowers the mean reference-round fraction from 10.0% to 5.9% and, relative to fixed $\epsilon = 0.1$, lowers risk by 22.6% in both single-signal schedules, 4.1% under interaction, and 15.2% under drift. The change is negligible in the no-signal and RAND schedules. The entropy-triggered diagnostic is also retained in the machine-readable results but is not used to make the main comparison.

J.3 Frozen contextual-policy campaign

The binary context is observed before assignment and switches every 64 cohorts. For context zero, $\mu = 1.5\tanh(2.5x_1)$; for context one, $\mu = 1.5\tanh(2.5x_2)$; $\delta = 0.2$ throughout. The policy class is the complete set of 25 maps from the two contexts to the five candidate designs. We use $T = 1024$, $n = 12$, $\epsilon = 0.1$, $\eta = 4$, eight population seeds 20261200–20261207, and 512 randomization histories for every stochastic method and population. The archive records these settings alongside every output.

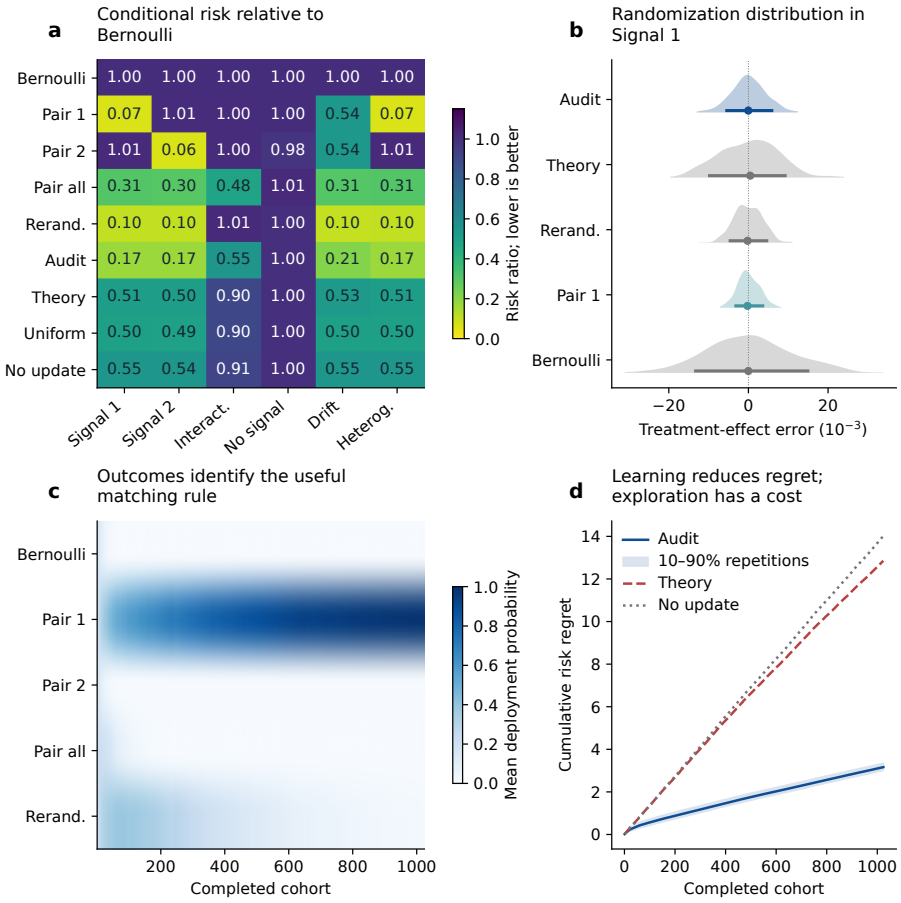

Figure 5: **Complete low-dimensional campaign.** (a) Exact conditional risk relative to Bernoulli for every fixed design and learning ablation. (b) Full randomization distributions of the pooled treatment-effect error in Signal 1. (c) Mean non-reference deployment probabilities. (d) Cumulative risk regret against the best fixed label; shading is the 10–90% range over 512 histories.

Table 4: Contextual campaign. Risk is exact conditional MSE averaged over cohorts. The parenthesized MCSE is over learning histories; seed SD measures variation over the eight finite populations. The best 25-policy map is restricted to one design per binary context and is not an unrestricted cohort-wise oracle.

Method	Risk $\times 10^3$ (MCSE)	Seed SD $\times 10^3$	Reference % (MCSE)	Risk/Bern.
Bernoulli	28.3508 (0.0000)	0.1124	0.00 (0.000)	1.000
Best fixed	2.9350 (0.0000)	0.0072	0.00 (0.000)	0.104
Oracle 25-policy	1.8512 (0.0000)	0.0652	0.00 (0.000)	0.065
Contextual exact-score CRL	5.3204 (0.0031)	0.0470	10.00 (0.015)	0.188
Static exact-score CRL	5.9888 (0.0032)	0.0113	10.00 (0.015)	0.211
No update	15.4609 (0.0000)	0.0912	10.00 (0.015)	0.545
Linear plug-in	2.0781 (0.0001)	0.0288	0.00 (0.000)	0.073
Polynomial plug-in	1.6333 (0.0000)	0.0343	0.00 (0.000)	0.058

Table 5: High-dimensional stress test. Entries are mean exact risk divided by Bernoulli risk, with between-population standard errors over all eight seeds.

Method	$p = 8$	$p = 64$	$p = 512$
Bernoulli	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000
Best fixed	0.038 ± 0.001	0.038 ± 0.001	0.038 ± 0.001
Exact-score CRL	0.203 ± 0.001	0.203 ± 0.001	0.206 ± 0.002
Linear plug-in	0.041 ± 0.001	0.045 ± 0.001	0.080 ± 0.003
Polynomial/RF plug-in	0.042 ± 0.001	0.046 ± 0.001	0.081 ± 0.002

All eight restricted oracles map context zero to Pair 1 and context one to Pair 2. On non-reference rounds exact-score contextual CRL deploys those actions with probabilities 0.697 and 0.702, respectively, reducing risk by 11.16% relative to the static five-label reference learner. Its reference-round and learning overhead remains visible: it is 81.28% above the best single fixed design and 187.40% above the best restricted map. Polynomial ridge is 69.30% below exact-score contextual CRL and 11.77% below the restricted map because it can choose a design separately for each realized cohort after observing the full covariate matrix. The latter result therefore does not violate the restricted comparator definition.

J.4 High-dimensional stress test

We fix $T = 256$, $n = 16$, $p \in \{8, 64, 512\}$, eight population seeds 20261100–20261107, and 64 randomization histories per seed. Each seed generates one $T \times n \times 512$ covariate tensor; lower dimensions use its first 8 or 64 coordinates, making comparisons across p paired. Only x_1 is prognostic: $\mu = 1.5 \tanh(2.5x_1)$ and $\delta = 0.2$. The four design laws are Bernoulli, pairing by x_1, pairing by x_2, and all-feature rerandomization. The latter begins with 256 random signs and their antithetic partners and retains the 10% with smallest all-feature imbalance. At $p = 8$ the polynomial plug-in uses the original cubic-and-interaction map. Before seeing outcomes, the otherwise infeasible expansions at $p = 64$ and 512 were replaced by 32 fixed Rademacher projections with squared and cubic features. Every learner sees current covariates and only its own completed outcomes; the outcome sum and exact risks are confined to the generator and evaluator.

Pairing on x_1 is the best fixed law in all 24 finite populations. The exact-score CRL mean varies only from 0.203 to 0.206, whereas both outcome-model selectors have larger mean ratios at $p = 512$ than at $p = 8$; nevertheless, each plug-in remains better than CRL in every finite population at this horizon. Because dimensions are nested within seed, the plotted cross-dimension trajectories are paired. The all-feature rerandomization means are 0.376, 0.751, and 0.915 for $p = 8, 64, 512$. This demonstrates the cost of irrelevant omnibus balance in the declared sparse-signal distribution, not a claim about dense-signal problems.

Table 6: Conditional risk divided by Bernoulli risk. Parentheses are randomization Monte Carlo standard errors; exact hindsight comparators omit them.

Method	Context	WDBC	Wine	Bike
Bernoulli	1.000	1.000	1.000	1.000
Best fixed	0.374	0.296	0.780	0.658
Vanilla CRL	0.222 (0.001)	0.448 (0.003)	0.937 (0.001)	0.810 (0.001)
Linear plug-in	0.079 (0.000)	0.338 (0.000)	0.964 (0.000)	0.653 (0.000)
Polynomial plug-in	0.079 (0.000)	0.356 (0.000)	0.973 (0.000)	0.656 (0.000)
PA-CRL (poly hint)	0.212 (0.001)	0.438 (0.001)	0.946 (0.002)	0.810 (0.002)
PA-CRL + residual fit	0.212 (0.001)	0.438 (0.002)	0.943 (0.002)	0.810 (0.002)

J.5 RAND COHORT-ORDER ROBUSTNESS

The same 1,024 frozen RAND replay cohorts were independently permuted 20 times; records and cohort membership were unchanged. Each ordering used 256 learning histories. CRL's reduction relative to Bernoulli ranges from 5.43% to 5.82%, while fixed rerandomization remains 3.54–3.97% better. A separate 10,000-repetition check removes the finite-Monte-Carlo reversal previously observed for the uniform-mixture pooled MSE and restores agreement with the exact conditional risk ordering. All order-specific risks, MCSEs, seeds, and high-repetition values are stored in `data/supplemental/`.

K PREDICTION-AUGMENTED COMPARATIVE LEARNING: ADDITIONAL EXPERIMENTS

Protocol and information firewall. For a candidate-minus-Bernoulli risk forecast m_{th} available before cohort-t outcomes, prediction-augmented CRL uses

$$\widehat{X}_{th} = m_{th} + \frac{A_t}{\epsilon}(\widehat{\Delta}_{th} - m_{th}), \qquad q_t(h) \propto g_t(h)e^{-\eta m_{th}}, \qquad g_{t+1}(h) \propto g_t(h)e^{-\eta \widehat{X}_{th}}.$$

The outcome models see current covariates and only their own completed $P = 2Y$ history. The current potential outcomes and exact risks are evaluator-only; reference indicators are independent Bernoulli(0.1) draws. The optional *PA-CRL + residual fit* ablation starts from the polynomial forecast and fits a regularized candidate-specific affine innovation model from completed reference rounds only; it is not the theorem-facing method. The CRL and PA-CRL learners use $\eta = 4$ and the same five-design library; the deterministic plug-in comparators select the smallest forecast and therefore have no learning rate.

The contextual study uses three frozen $p = 8$ finite-population paths, three observed contexts, $n = 12$, $T \in \{64, 128, 256, 512, 1024\}$, and 64 randomization repetitions per path. The phase diagram uses two paths per cell, $p \in \{8, 32, 128\}$, $T \in \{64, 128, 256, 512\}$, and 48 repetitions per path. All cells and seeds are retained.

Context, dimension, and real-covariate results. At $T = 1024$, PA-CRL with the polynomial hint has risk/Bernoulli 0.212 and is 4.5% below vanilla CRL, but 167.1% above the polynomial plug-in and 180.9% above the hindsight context-map oracle. It beats the polynomial plug-in in 0/12 predeclared phase cells. The three real-data analyses use disjoint calibration and replay rows from UCI WDBC (DOI 10.24432/C5DW2B), white Wine Quality (10.24432/C56S3T), and hourly Bike Sharing (10.24432/C5W894), all under CC BY 4.0. Calibration means, scales, row identifiers, selected prognostic coordinates, download URLs, versions, and SHA-256 hashes appear in `experiments/calibrated_crl/real_data_provenance.json` and `experiments/calibrated_crl/real_row_manifest.csv`. Both potential outcomes equal the same bounded transformation of the observed held-out label; these are prospective zero-effect design replays, not causal analyses.

Reference-rate and learning-rate sensitivity. To separate the declared tuning choice from retrospective optimization, we cross $\epsilon \in \{0.05, 0.10, 0.20, 0.40\}$ with $\eta \in \{1, 2, 4, 8\}$ on the same three

Table 7: Conditional risk divided by Bernoulli risk in the complete 4×4 sensitivity grid. Each entry averages three frozen population paths with 64 randomization histories per path; the predeclared value is $\epsilon = 0.10$, $\eta = 4$.

Reference rate ϵ	$\eta = 1$	$\eta = 2$	$\eta = 4$	$\eta = 8$
0.05	0.419	0.297	0.216	0.171
0.10	0.448	0.330	0.253	0.212
0.20	0.509	0.405	0.336	0.298
0.40	0.632	0.554	0.501	0.473

Table 8: Complete fixed-factorial advantage surface. Entries are relative risk reduction $100(1 - R_{\mathrm{PA}}/R_{\mathrm{plugin}})$ in percent, so positive values favor PA-CRL. Bold cells have pointwise paired seed-level 95% confidence intervals for $R_{\mathrm{PA}} - R_{\mathrm{plugin}}$ entirely below zero. All 36 cells use eight populations and 128 histories per population.

Interaction weight β	$T = 128$	$T = 256$	$T = 512$	$T = 1024$
0	-190.8	-164.5	-148.8	-137.4
0.125	-177.7	-147.2	-130.3	-117.8
0.250	-115.1	-92.4	-79.2	-69.7
0.375	-57.1	-45.2	-39.2	-34.3
0.500	-19.9	-11.3	-6.7	-1.5
0.625	**3.4**	**13.4**	**20.7**	**26.3**
0.750	**17.4**	**27.5**	**33.8**	**38.2**
0.875	**22.2**	**32.5**	**38.0**	**42.3**
1.000	**17.2**	**24.5**	**29.8**	**31.5**

$T = 512$, $p = 8$ finite-population paths, using 64 randomization histories per path and retaining all 3,072 rows. Every cell improves on Bernoulli, but the risk ratio ranges from 0.171 to 0.632; the declared $(0.10, 4)$ cell is 0.253, whereas the best cell is $(0.05, 8)$ at 0.171. All 16 cells remain above the polynomial plug-in value 0.082, so the grid documents genuine tuning sensitivity rather than manufacturing a favorable winner.

Fixed factorial interaction-misspecification surface. We fix a complete 9×4 grid crossing interaction weight $\beta \in \{0, 0.125, \ldots, 1\}$ with $T \in \{128, 256, 512, 1024\}$, then retain eight independently generated finite populations and 128 randomization histories per seed and cell. Both methods use the same ridge-linear predictor specification $[1, x_1, x_2]$, fitted to each method's own observed history; it excludes the x_1x_2 interaction from the response surface $(1-\beta)\, 1.5 \tanh(2.5x_1) + \beta\, 1.5 \tanh(4x_1x_2)$. The two methods share candidate-specific assignment draws. The raw plug-in selects by its predictor; PA-CRL adds only the declared independent reference contrasts. Exact risks remain evaluator-only. The five fixed candidate laws are Bernoulli assignment, pairing on x_1, pairing on x_2, greedy all-feature pairing, and fully enumerated 10.06% threshold rerandomization over the 2^{12} sign vectors. The matched factorial comprises $9 \times 4 \times 8 \times 128$ cases and 73,728 method-path rows. This design isolates the value of reference correction as the matched predictor class moves from an adequate ranking toward systematic misspecification.

PA-CRL improves on the raw linear plug-in throughout the contiguous $\beta \geq 0.625$ region: all 16 horizon cells favor PA-CRL, every paired pointwise seed-level 95% confidence interval for the risk difference excludes zero, and the relative risk reduction ranges from 3.4% to 42.3%. The maximum reduction occurs at $(\beta, T) = (0.875, 1024)$. For $\beta \leq 0.375$, the raw linear plug-in is better in all 16 cells; at $\beta = 0.5$, the first three horizons favor it and $T = 1024$ is the transition cell (risk ratio 1.015, with the clustered interval crossing zero). The complementary polynomial-hint phase grid places the efficiency boundary: the richer polynomial-feature plug-in retains lower risk in all 12 $p \times T$ cells. Together, the two surfaces identify the operating regime: reference correction buys robustness when the matched predictor class persistently misorders candidate laws, while an adequate hint preserves the plug-in's efficiency advantage.

Archived model-failure stress. We next freeze the original $T = 1024$, $n = 12$, seed-20260908 single-signal, interaction, no-signal, and drift/switch paths and reconstruct their exact saved design

Table 9: Complete archived-schedule risk and residual diagnostics. Repair conditions on a reference-round hint ordering being wrong; harm conditions on it being correct.

Method	Single	Interaction	No signal	Switch
Best fixed	0.066	0.479	0.976	0.103
Vanilla CRL	0.175 (0.000)	0.552 (0.001)	0.995 (0.000)	0.213 (0.001)
Linear plug-in	0.072 (0.000)	0.901 (0.001)	1.003 (0.000)	0.127 (0.000)
PA-CRL (linear hint)	0.172 (0.000)	0.551 (0.000)	0.996 (0.000)	0.211 (0.001)
Polynomial plug-in	0.057 (0.000)	0.435 (0.000)	0.985 (0.000)	0.165 (0.000)
PA-CRL (polynomial hint)	0.172 (0.000)	0.549 (0.000)	0.995 (0.000)	0.211 (0.001)

Diagnostic	Single	Interaction	No signal	Switch
Linear: innovation MSE	2.34e-05	1.03e-04	3.61e-05	1.76e-04
Linear: repair %	46.0	88.7	80.2	79.3
Linear: harm %	68.8	12.3	19.1	75.8
Polynomial: innovation MSE	9.71e-06	2.49e-05	3.68e-05	1.66e-04
Polynomial: repair %	39.7	61.6	79.7	85.4
Polynomial: harm %	67.0	46.8	18.5	70.8

covariances to numerical error below 10^{-12}. There are 256 repetitions per method and schedule. Raw linear and polynomial plug-ins are compared with PA-CRL under matched predictor specifications, and both PA learners share random seeds with vanilla CRL where the action paths permit pairing. Linear-hint PA-CRL lowers risk by 38.8% relative to its badly misspecified raw linear selector on the interaction schedule, but the raw plug-in is better in the single-signal and drift schedules. Polynomial PA-CRL does not beat the raw polynomial plug-in on any of the four schedules. Thus the correction supplies robustness value under a concrete wrong ordering, not blanket dominance.

Optional residual fit and computation. The candidate-specific two-parameter residual fit does not improve the raw innovation in this finite campaign: at $T = 1024$ its reference-residual-MSE ratio is 1.075. Each context supplies only about $0.1T/3 \approx 34$ reference rounds per repetition to each online fit, and heterogeneous effects make innovations noisy; the observed degradation is consistent with calibration-estimation variance, although the experiment does not isolate a unique cause. This does not affect the reference-contrast unbiasedness identity, which holds for every predictable m_{th}. The retained residual plot reports the full path rather than suppressing this failure.

On the archived host, the procedural benchmark gives dense/ghost wall-clock ratios 1.12 at $n = 64$ and 1.98 at $n = 512$. Both branches make the same calls to a threshold-rerandomization rejection sampler; unlike a covariance-free arithmetic benchmark, the dense branch includes construction of every $n \times n$ Monte Carlo covariance. All repetition rows, configurations, source hashes, timing rows, vector figures, and validation reports are archived under `experiments/calibrated_crl/`.

Evidence boundary. The results condition on the declared finite populations and design library. Hindsight comparators are not deployable, the phase grid is not an asymptotic phase transition, and the real-data replays do not estimate historical treatment effects. Most importantly, the polynomial plug-in remains the strongest method in every $p \times T$ cell and in most archived schedules; the role of PA-CRL is to retain the correct contrast target when the hint is wrong.

L Direct finite checks of correction, adaptive inference, and regret

This section targets the identities that connect the one-round reference score to sequential learning and inference. A fixed $n = 8$ population is used for the score and regret calculations. Complete enumeration of all 2^8 independent-reference assignments and all 2^4 paired ghost orientations gives the exact target $\Delta = -0.0087109375$ and exact score moments; Monte Carlo is then used only to check the executed estimator against those quantities.

Correction and finite-M variance. The augmented-score grid crosses five hint errors $\{-0.25, -0.10, 0, 0.10, 0.25\}$, five reference probabilities $\{0.05, 0.10, 0.25, 0.50, 1\}$, and $M \in \{1, 8, 64\}$, for 75 cells with 200,000 repetitions each. The largest absolute discrepancies for the augmented mean, exact variance in (10), and innovation identity are 3.259, 3.236, and 3.236 Monte

Carlo standard errors. At full reference probability, exact contrast variance falls from 3.4512×10^{-6} at $M = 1$ to 2.5894×10^{-6} at $M = 64$ because only the ghost term shrinks; the reference component remains 2.5758×10^{-6} (Figure 10a–b).

Unequal-cohort adaptation and Gaussian scaling. We next use periodic cohort sizes $4, 6, 8, 10, 12$ and a predictable rule that starts with independent assignment and thereafter selects paired assignment exactly when the completed-cohort pooled-effect numerator is nonnegative. Across $60{,}000$ paths at each $T \in \{16, 32, 64, 128, 256\}$, the largest standardized bias and exact-MSE discrepancies are 1.381 and 1.415 Monte Carlo standard errors, respectively, testing Theorem 6.1. Oracle-studentized 95% coverage ranges from 0.9493 to 0.9513; at $T = 256$ the standardized statistic has mean 0.0570, standard deviation 1.0032, Kolmogorov distance 0.0470 from a standard Gaussian, and mean largest conditional-variance share 0.0129, as required by the no-dominance mechanism in Theorem 6.3. This studentization uses the pathwise predictable variance and is a diagnostic of the limit theorem rather than a feasible interval construction (Figure 10c).

Regret decomposition. Finally, a two-policy instance uses reference probability 0.05, $M = 8$, and learning rate 12. Each horizon has $30{,}000$ paths. At $T = 256$, expected risk regret is 0.165703 with Monte Carlo standard error 0.000053, decomposing into 0.111500 reference excess and 0.054203 exploitation excess, while the three terms in (11) sum to 1.740251. The 99% Monte Carlo upper endpoint lies below the theorem bound at every tested horizon (Figure 10d). This verifies the executed decomposition and inequality on the declared finite instance; it does not assert that the upper bound is tight.

The finite diagnostics test unbiased correction, exact variance, predictable-selection MSE, an oracle-studentized Gaussian approximation, and one fixed-policy regret instance. Time-uniform confidence-sequence coverage, feasible estimation of the oracle CLT scale, uniformity outside the declared grids, and dynamic or infinite policy classes remain separate questions governed by their stated assumptions.

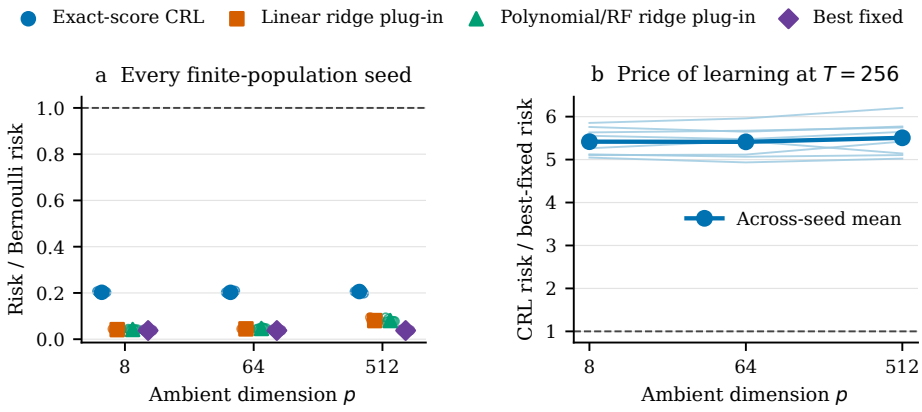

Figure 6: **Nuisance dimension changes the costs of modeling and reference feedback.** (a) Risk relative to Bernoulli for every nested finite-population seed; jitter is only for visibility. (b) Exact-score CRL risk relative to the best fixed design at $T = 256$. Thin lines pair the same seed across nested dimensions and the thick line is their mean. The fixed reference rate makes learning cost visible even though its mean risk ratio varies little across dimensions.

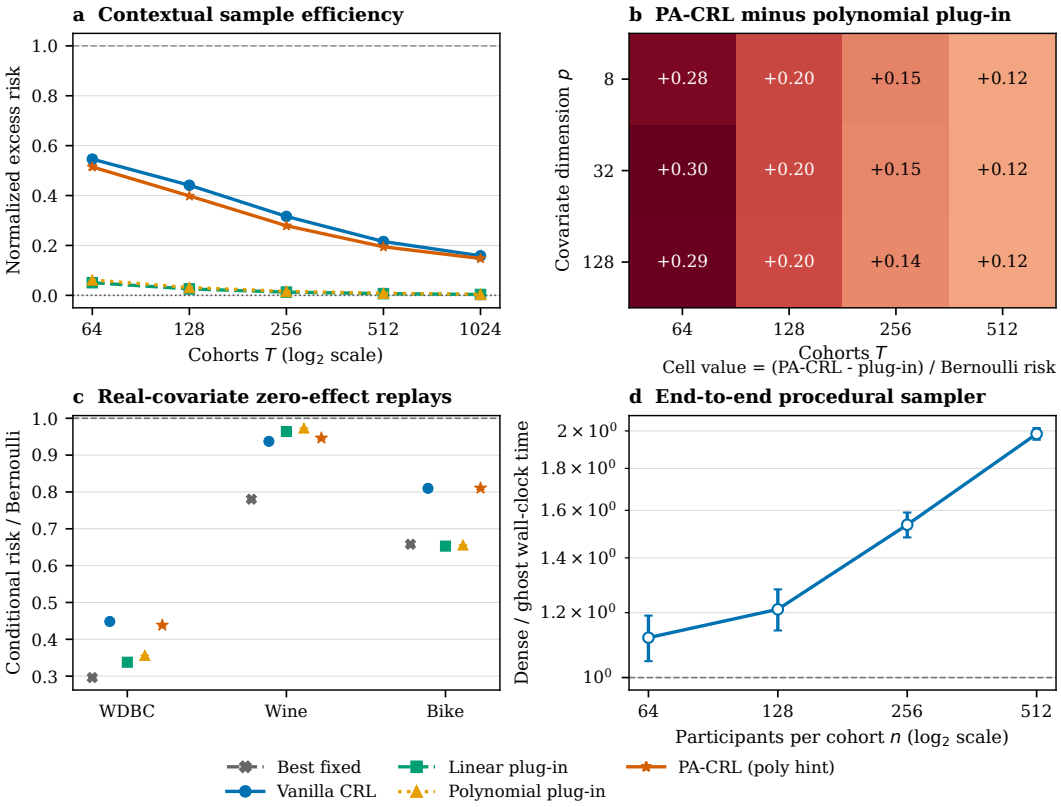

Figure 7: **Prediction augmentation improves vanilla CRL but does not erase the plug-in advantage.** (a) Contextual excess risk normalized between the hindsight context map (zero) and Bernoulli (one); bars are 1.96 between-path standard errors. (b) PA-CRL (poly hint) minus the polynomial plug-in, normalized by Bernoulli risk; positive cells favor the plug-in. (c) Three UCI zero-effect replays, with 1.96 randomization Monte Carlo standard errors. (d) End-to-end rejection-sampler timing, including covariance construction, over twenty single-threaded runs.

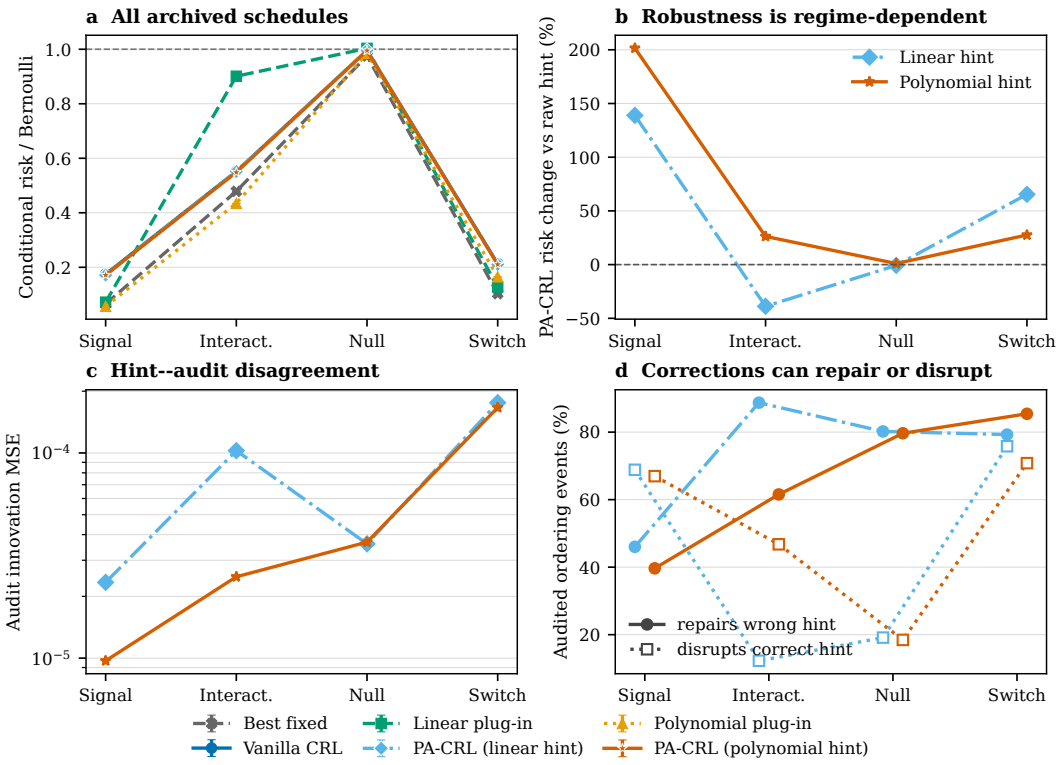

Figure 8: **Reference correction repairs some wrong model orderings and also introduces finite-reference noise.** (a) All predeclared schedules and adverse baselines. (b) PA-CRL risk change relative to the raw selector using the matched predictor class; negative is better. (c) Mean squared reference residual. (d) Conditional fractions of wrong orderings repaired and correct orderings disrupted on reference rounds.

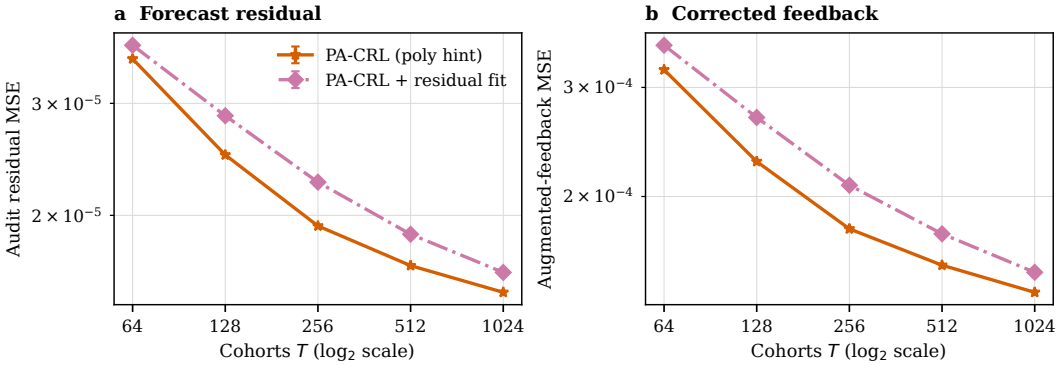

Figure 9: **The optional residual fit worsens, rather than reduces, the observed innovation.** Reference-residual MSE and augmented-feedback MSE are shown over every contextual horizon for theorem-facing PA-CRL (poly hint) and PA-CRL + residual fit; bars are 1.96 between-path standard errors and both axes are logarithmic.

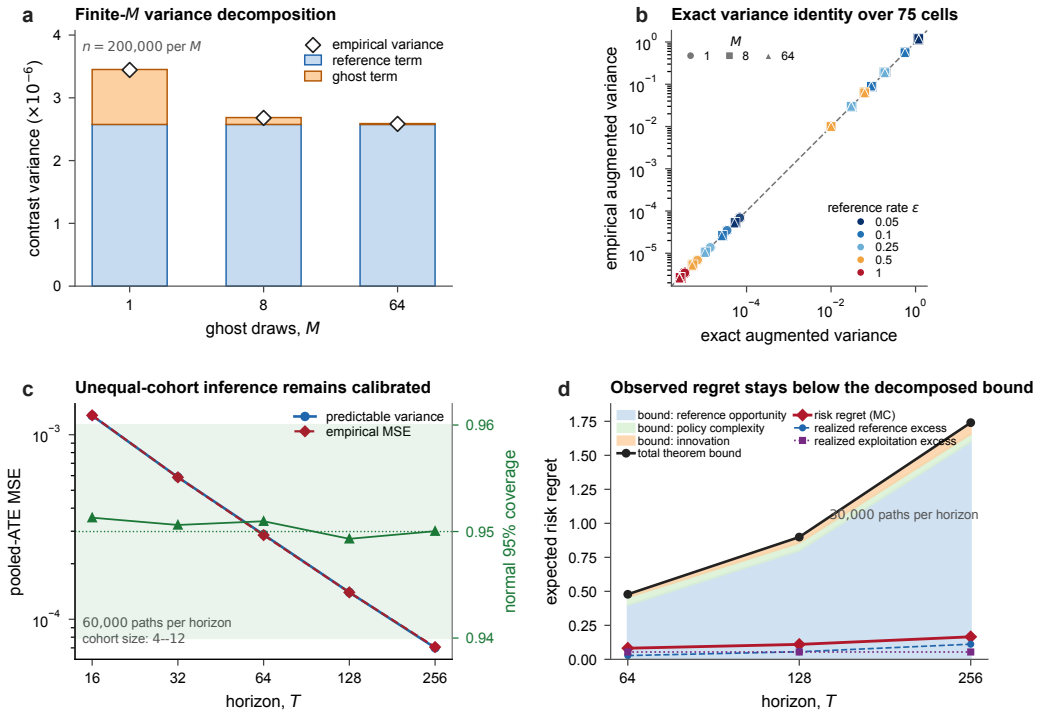

Figure 10: **Executed checks align the empirical mechanisms with the stated theory.** (a) Exact reference and ghost components of contrast variance, with empirical variance overlaid; each M uses 200,000 draws. (b) Empirical versus exact augmented variance in all 75 hint-error–reference-rate–ghost-budget cells on logarithmic axes. (c) Empirical pooled-ATE MSE matches the predictable-variance identity under unequal cohorts and outcome-adaptive sampler selection; the secondary axis reports oracle-studentized 95% coverage over 60,000 paths per horizon. (d) Realized risk regret and its reference/exploitation decomposition remain below the three-term upper bound over 30,000 paths per horizon. Scripts, raw cells, seeds, and validation thresholds are archived in `experiments/main_evidence/`.

M SCOPE OF MACHINE VERIFICATION

The accompanying `LeanProofs` project contains 77 proved declarations (75 theorems and two private auxiliary lemmas) in seven modules checked with Lean 4 and mathlib, both pinned to version `v4.34.0-rc1`. The required verification script completed with exit code zero after running the source checker with warnings treated as errors and a complete `lake build -wfail`. The final compiler and build output is included in `LeanProofs/logs/check-prediction-calibration-final.log`. This verifies the finite-population observability and exact contrast-variance core, the finite-space empirical-Bernstein factor, the finite-law ghost-oracle expectation, range, and finite-average variance contraction, deterministic exponential-weights algebra, the finite-law prediction-augmentation mean, variance, and innovation identities, and the observable-clock mean-square algebra; it does not verify the entire manuscript.

The formal probability model is explicit. For every finite index type I, the assignment space is $\Omega = \{\text{false}, \text{true}\}^I$, and the expectation of f is defined as $|\Omega|^{-1} \sum_{\omega \in \Omega} f(\omega)$. Coordinate and global sign flips are proved to be permutations of Ω. Reindexing finite sums establishes first, second, third, and fourth moments, rather than assuming the moment identities needed by the conclusion. The observed-outcome map, the A, μ, δ decomposition, linear–quadratic orthogonality, exact centering, and the full $16n^4$ variance denominator are all verified. A separate finite weighted-law module covers correlated assignments and the single-cohort Horvitz–Thompson risk identity.

The ghost module treats the observed vector P as fixed, the finite conditional representation of drawing fresh ghost randomness independently of P. It proves the second-moment quadratic identity, preservation of that expectation under a nonempty finite Monte Carlo average, a pointwise range bound for sign-bounded draws, and the $1/(4M)$ variance contraction under pairwise-uncorrelated one-draw scores. The Hedge module is deterministic: for a nonempty finite expert class, strictly positive weights, arbitrary nonnegative losses, and every $\eta > 0$, it verifies the actual one-step exponential-update ratio and log-potential bounds, the finite-horizon telescoping implication, and the reference/deployment mixture algebra. The prediction-calibration module uses an explicit finite weighted law and proves the target mean, central-second-moment/variance decomposition, and innovation second moment for prediction plus an inverse-probability reference residual once the requisite indicator-weighted moments are supplied. The same module verifies the finite-law algebra behind the observable-clock mean-square decomposition.

Table 10: Exact scope of the principal Lean declarations. Homogeneous expectation identities omit deterministic normalization; the contrast-variance declaration includes it.

Lean declaration	Verified mathematical statement
`avg_sign`, `avg_pair`	The complete finite independent-sign law has zero first moments and identity second moments.
`common_feedback_observed`	$\mathbb{E}[(2Y)^\top A(2Y)] = \mu^\top A\mu + \sum_i A_{ii}\delta_i^2$ for every finite real matrix and finite real potential-outcome population.
`contrast_unbiased`	Differences of unit-diagonal score matrices remove the common offset exactly.
`bounded_unbiased_quadratic_iff`	An observable statistic unbiased for all $a, b \in [-1, 1]^I$ exists exactly when the coefficient diagonal is zero.
`observable_rectangle`	An arbitrary statistic of the actual observed data has zero rectangular difference when one unit's two potential outcomes are varied.
`pair_probabilities`	Centered pair marginals and correlation ρ determine the four cell probabilities $(1 \pm \rho)/4$.
`correlated_contrast_bias`	The complete correlated-assignment feedback bias in Proposition 3.3.
`reversal_observed_feedback`	The actual finite common-sign feedback contrast in the counterexample is $-1/4$; separate declarations verify its true matrix risks and bounded outcomes.
`ht_finite_risk`	The exact single-cohort MSE numerator equals its quadratic form in the actual assignment pair moments.

Lean declaration	Verified mathematical statement
`finite_clock_decomposition`	The observable-clock mean-square decomposition once the finite-law Horvitz–Thompson mean and error second moment are supplied.
`normalized_observed_contrast_variance`	The full observed-data variance formula in (22), including actual finite expectation, fourth moments, potential-outcome decomposition, and normalization.
`empiricalBernstein_factor`	The exact pointwise exponential inequality for $0 < \lambda < 1$ and every $x \in [-1, 1]$, including both endpoints.
`empiricalBernstein_finite_factor`	The centered exponential factor has expectation at most one under every finite probability law; its mean is the actual weighted mean. This verifies the finite-law factor underlying Lemma G.1.
`ghost_square_mean_eq_quad`	Conditional on fixed P, any finite weighted ghost law with second moments Σ satisfies $\mathbb{E}(P^\top G)^2 = P^\top \Sigma P$.
`ghost_empirical_mean_eq_quad`, `ghost_empirical_range`	Every nonempty finite average retains that expectation when each draw has the stated marginal second moment; sign-bounded draws give the squared-ℓ_1 range bound.
`ghost_average_variance_le`	Pairwise-uncorrelated scores with one-draw variance at most $1/4$ have finite-average variance at most $1/(4M)$.
`hedge_one_step_ratio`, `hedge_one_step_log`	The actual finite-expert exponential update obeys the quadratic normalized-potential and log-potential bounds for nonnegative losses and every $\eta > 0$.
`hedge_pathwise_of_log_potential`	The standard one-step log and comparator lower-bound premises telescope to the displayed finite-horizon Hedge inequality with constants $\log \lvert\mathcal{E}\rvert/\eta$ and $\eta/2$.
`deployment_risk_identity`, `deployment_risk_le`	The exact reference/exploitation mixture decomposition and its additive ϵ/n charge under nonnegative candidate risk and reference risk at most $1/n$.
`prediction_augmented_unbiased`	Reference mass ϵ and indicator-weighted reference-score mean $\epsilon\Delta$ imply exact unbiasedness of $m + A(\widehat{D} - m)/\epsilon$ for every m and every nonzero ϵ.
`prediction_augmented_central_second_moment`	The exact central-second-moment, hence variance, decomposition for prediction plus inverse-reference residual correction under the stated finite-law first and second moments.
`prediction_augmented_innovation_second_moment`	The correction innovation has second moment equal to its indicator-weighted residual second moment divided by the reference mass.

The estimability theorem quantifies over arbitrary deterministic real-valued observable statistics. The written proof's extension to independent auxiliary randomization integrates that randomization into the statistic; an auxiliary probability space is not separately formalized. The ghost declarations verify the $1/(4M)$ finite-average contraction once pairwise-uncorrelated one-draw moments are supplied; they do not derive those moments from product-law ghost independence, encode the outer expectation over the reference assignment, or verify computational complexity. The prediction-calibration declarations verify the augmentation mean, exact central-second-moment/variance, and innovation identities after their finite-law reference moments are supplied; they do not derive those moments from a Bernoulli history filtration or formalize optimistic mirror descent. The finite-clock declaration verifies the mean-square decomposition once the required finite-law mean and error moments are supplied; conditional history packaging remains unformalized. The full regret theorem is not encoded. PSD score bounds, the general pair-support characterization, the minimax lower bound, covariance-approximation bounds, infinite-horizon confidence sequences, the CLT, and the imported GSW guarantee likewise remain outside the machine-checked scope. They must be assessed from their written proofs and cited foundations.

For each central declaration, `#print axioms` reports only `propext`, `Classical.choice`, and `Quot.sound`. These are standard classical foundations of mathlib. The project introduces no additional axioms or unproved placeholders. The companion `VERIFICATION.md` records exact propositions, commands, limitations, and reproduction instructions.