FIXED-CONFIDENCE IDENTIFICATION OF CAPACITY-CONSTRAINED TREATMENT ALLOCATIONS WITH PROGNOSTIC COVARIATES

Anonymous authors
Paper under double-blind review

ABSTRACT

Randomized experiments are often used to guide treatment allocation under limited capacity. Existing work has studied fixed-confidence identification under structured or constrained decision sets and covariate adjustment for improving treatment effect estimation. However, it remains unclear whether individual-level covariates that predict outcomes, which we call prognostic covariates, can reduce the sample complexity of identifying an optimal capacity-constrained group-level treatment allocation. To address this gap, we study fixed-confidence identification of capacity-constrained group-level treatment allocations with prognostic covariates. We characterize the first-order expected stopping-time benchmark when individual covariates are retained and show that it can be strictly smaller than the benchmark based on group-level information alone. We then develop an adaptive procedure that learns the unknown prognostic functions from accumulating data while controlling the allocation-selection error in finite samples, and prove that its expected stopping time asymptotically attains the same first-order benchmark as when the prognostic functions are known. Experimental results confirm the predicted information gap and show the learned procedure approaching the full-covariate benchmark.

1 INTRODUCTION

Randomized experiments are often used to guide treatment allocation when budget or capacity constraints prevent treating all units (Manski, 2004; Bhattacharya & Dupas, 2012; Kitagawa & Tetenov, 2018; Athey & Wager, 2021). Consider a service provider deciding which edge sites should receive low-latency computing capacity when only a subset can be supported. Before deployment, the provider can randomize incoming requests within each site between edge and remote-cloud processing while recording request-level baseline characteristics such as workload, computational demand, and network conditions. Some of these covariates may be predictive of outcomes, and we refer to them as *prognostic covariates*. Because both the experimental treatment probabilities and the final deployment decisions are defined at the site level, these individual-level covariates do not affect which site-level allocations are feasible. Their potential role is statistical: they may help learn the treatment effects that determine which feasible allocation is best (Tsiatis et al., 2008; Lin, 2013).

Understanding whether prognostic covariates can reduce the number of observations needed to identify the best feasible group-level allocation connects two lines of work. Fixed-confidence best-arm identification and its structured and constrained extensions study adaptive sampling for identifying the optimal decision with a prescribed error probability (Garivier & Kaufmann, 2016; Fiez et al., 2019; Degenne et al., 2020; Carlsson et al., 2024). These results characterize how the decision set and sampling constraints govern sample complexity. Covariate adjustment in randomized experiments shows that baseline covariates can improve the precision of treatment effect estimation (Tsiatis et al., 2008; Lin, 2013; Benkeser et al., 2021). Related sequential work studies subpopulation testing (Russac et al., 2021) and semiparametric reward models (Krishnamurthy et al., 2018; Kim et al., 2025; Kim, 2026). Contextual best-arm identification can use covariates observed before arm selection to identify the arm with the best marginal mean (Kato & Ariu, 2026), and post-action context can also improve pure exploration (Shahverdikondori et al., 2025). In our setting, treatment probabilities may depend on G but not on X beyond G, and the final decision is a capacity-constrained

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.

group-level allocation. Covariate-aware methods also provide asymptotic fixed-confidence guarantees (Cho & Kallus, 2026), anytime-valid off-policy inference (Waudby-Smith et al., 2024), and optimal policy identification from logged data (Molitor, 2026). Existing work therefore leaves open whether prognostic covariates that are excluded from both within-group randomization and deployment can reduce fixed-confidence sample complexity for constrained group-level allocation.

For a target error probability δ, fixed-confidence identification minimizes the expected number of experimental observations required to select the optimal allocation with probability at least $1-\delta$ (Garivier & Kaufmann, 2016). Its sample complexity is determined by how difficult it is to distinguish the true data-generating process from alternatives under which a competing feasible allocation is optimal (Kaufmann et al., 2016; Garivier & Kaufmann, 2016). Variance reduction for a particular estimator therefore does not automatically imply a smaller optimal stopping time. The prognostic relationship between baseline covariates and outcomes is also unknown and must be learned from the same sequential experiment while preserving finite-sample error control (Wasserman et al., 2020; Howard et al., 2021). These considerations lead to our central question:

> *Can individual-level prognostic covariates reduce the sample complexity of identifying the optimal capacity-constrained group-level treatment allocation?*

We answer this question by comparing the full-covariate and coarsened first-order characteristic times. We show that retaining these covariates can strictly reduce the first-order expected stopping-time coefficient, even though they affect neither the feasible group-level allocations nor the experimental treatment probabilities. We further characterize the full-covariate lower bound when the prognostic functions are unknown and show that the same first-order benchmark remains attainable. Finally, we develop an adaptive procedure that learns the prognostic functions online, updates the feasible treatment probabilities toward the information-optimal design, and ensures that the probability of selecting an incorrect allocation is at most δ. As the target error probability vanishes, the procedure attains the optimal full-covariate first-order expected stopping time.

Our **contributions** are threefold:

1. We compare the information available in the full-covariate and coarsened experiments, derive a coarsened-data lower bound, and show that discarding within-group prognostic covariates can strictly increase the first-order sample requirement even when the deployment capacity constraint is binding.
2. We derive an explicit full-covariate characteristic time. The least-favorable alternatives that change which allocation is optimal preserve the prognostic functions, so estimating these unknown functions does not change the first-order benchmark.
3. We develop an adaptive procedure that learns the prognostic functions from the sequential experiment, controls the allocation-selection error in finite samples, and asymptotically attains the full-covariate first-order benchmark.

2 Problem Setup

Deployment objective. Let $X \in \mathcal{X}$ denote baseline covariates with marginal distribution P_X, and let $G = g(X) \in [M]$, where $[M] = \{1, \dots, M\}$, denote a prespecified population group with probability $p_m = \mathbb{P}(G = m) > 0$. Treatments take values in a finite set $\mathcal{A}$, and $Y(a)$ denotes the potential outcome under treatment $a \in \mathcal{A}$ (Rubin, 1974). For each group-treatment cell (m, a), define the marginal mean

$$\theta_{m,a} = \mathbb{E}[Y(a) \mid G = m].$$

Let $\theta = (\theta_{m,a})_{m,a}$ collect these cell means. The deployment decision is chosen from a finite known allocation set $\mathcal{Z}$. Each allocation $z = (z_{m,a})_{m,a} \in \mathcal{Z}$ specifies the population mass assigned to each treatment within each group, with $z_{m,a} \geq 0$ and $\sum_a z_{m,a} = p_m$. Equivalently, $z_{m,a}/p_m$ is the fraction of group m assigned to treatment a. The set $\mathcal{Z}$ encodes the capacity constraints. The population value of z under model P is

$$V_P(z) = \sum_{m,a} z_{m,a}\theta_{m,a} = z^\top \theta, \qquad z^*(P) = \arg\max_{z \in \mathcal{Z}} V_P(z). \tag{1}$$

Assume that there exists a unique optimal allocation and write $z^* = z^*(P)$ when the model is clear.

Prognostic outcome model. Individual-level covariates may predict outcomes within a group-treatment cell even though they do not enter the deployment decision. We model this through

$$Y(a) \mid X = x \sim N\big(\theta_{m,a} + r_{m,a}(x), \sigma^2_{m,a}\big), \qquad g(x) = m, \tag{2}$$

where $r_{m,a}$ is an unknown prognostic function. For a cell $c = (m, a)$, write $\theta_c = \theta_{m,a}$, $r_c = r_{m,a}$, $\sigma^2_c = \sigma^2_{m,a}$, and $\mu_c(x) = \theta_c + r_c(x)$. Let ϕ_c be a feature map into a Hilbert space $\mathbb{V}_c$, and define the centered feature map

$$\widetilde{\phi}_c(x) = \phi_c(x) - \mathbb{E}[\phi_c(X) \mid G = m].$$

We assume $\sup_x \|\widetilde{\phi}_c(x)\|_{\mathbb{V}_c} \le \kappa_c$. Let $\widetilde{k}_c(x, x') = \langle \widetilde{\phi}_c(x), \widetilde{\phi}_c(x') \rangle_{\mathbb{V}_c}$ and let $\mathbb{H}_c$ denote the reproducing kernel Hilbert space (RKHS) induced by $\widetilde{k}_c$ (Aronszajn, 1950). We restrict

$$\mathcal{R}_c = \{r_c \in \mathbb{H}_c : \|r_c\|_{\mathbb{H}_c} \le B_c\}. \tag{3}$$

Every $r_c \in \mathcal{R}_c$ is centered within group, so $\mathbb{E}[r_c(X) \mid G = m] = 0$ and $\theta_{m,a}$ remains the marginal mean in Eq. (1). We assume $0 < \underline{\sigma}^2 \le \sigma^2_c \le \overline{\sigma}^2 < \infty$. Let $\mathcal{P}$ denote the resulting model class. We treat P_X, g, p_m, and σ^2_c as known, while θ and the prognostic functions are unknown.

Sequential experiment. Units arrive independently across rounds. Let $\mathcal{F}_{t-1}$ contain all observations through round $t-1$, and let $\ell = 1, \dots, L$ index the experimental resources. Before round t, the experiment chooses an $\mathcal{F}_{t-1}$-measurable vector of group-specific treatment probabilities $q_t = (q_{t,m,a})_{m,a}$ from

$$\mathcal{Q} = \left\{ q : q_{m,a} \ge 0, \quad \sum_a q_{m,a} = 1, \quad \sum_m p_m \sum_a c^{\exp}_\ell(m, a) q_{m,a} \le e_\ell \quad \forall \ell \right\}. \tag{4}$$

Here $c^{\exp}_\ell(m, a)$ is the per-unit consumption of experimental resource ℓ, and e_ℓ is its expected per-round budget. We assume that $\mathcal{Q}$ is nonempty, compact, and convex and contains a design with positive probability on every group-treatment cell whose allocation weight differs across two feasible allocations. For a cell $c = (m, a)$, we write $q_c = q_{m,a}$ whenever a design is indexed by cells.

At round t, $X_t \sim P_X$ is drawn independently of earlier rounds and $G_t = g(X_t)$ is observed. The experiment then draws $A_t \sim q_t(\cdot \mid G_t)$ and observes $Y_t = Y_t(A_t)$. Thus, treatment probabilities may adapt to past observations and the current group, but not to the remaining individual-level covariates of the current unit. Randomization and consistency imply (Rubin, 1974)

$$\mathbb{E}[Y_t \mid \mathcal{F}_{t-1}, G_t = m, A_t = a] = \theta_{m,a}. \tag{5}$$

The allocation set $\mathcal{Z}$ determines the feasible final allocations, while $\mathcal{Q}$ constrains how experimental observations can be collected.

Fixed-confidence objective. We compare two experiments under the same $\mathcal{Z}$ and $\mathcal{Q}$. The *full-covariate experiment* observes $O^{\mathrm{F}} = (X, G, A, Y)$, while the *coarsened experiment* observes only $O^{\mathrm{C}} = (G, A, Y)$. A sequential procedure specifies predictable designs $q_t \in \mathcal{Q}$, a stopping time τ_δ, and an output $\widehat{z}_{\tau_\delta} \in \mathcal{Z}$. It is uniformly δ-correct on $\mathcal{P}$ (Garivier & Kaufmann, 2016) if

$$\mathbb{P}_P(\widehat{z}_{\tau_\delta} = z^*(P)) \ge 1 - \delta \tag{6}$$

for every $P \in \mathcal{P}$ with a unique optimum. At a fixed instance P, the objective is to minimize $\mathbb{E}_P[\tau_\delta]$ subject to Eq. (6) and $q_t \in \mathcal{Q}$ at every round. We ask whether retaining the within-group covariate information in X beyond G can reduce this sample requirement and whether the resulting first-order benchmark remains attainable when the prognostic functions must be learned from the same experiment.

3 Fixed-Confidence Sample Complexity

3.1 Value of Retaining Prognostic Covariates

For a fixed design $q \in \mathcal{Q}$, let P^q_{F} and P^q_{C} denote the one-round laws of O^{F} and O^{C}. Let $D_{\mathrm{KL}}(P\|Q)$ denote the Kullback-Leibler (KL) divergence from P to Q. The statistical difficulty is governed

by alternatives under which another feasible allocation is optimal. Define the closed alternative set $\overline{\mathrm{Alt}}(P) = \{Q \in \mathcal{P} : \exists z \neq z^*(P) \text{ with } V_Q(z) \geq V_Q(z^*(P))\}$. This set contains models under which a different feasible allocation is optimal or ties with $z^*(P)$. Under design q, let $I_{\mathrm{F}}(q;P) = \inf_{Q\in\overline{\mathrm{Alt}}(P)} D_{\mathrm{KL}}(P_{\mathrm{F}}^q \| Q_{\mathrm{F}}^q)$ and $I_{\mathrm{C}}(q;P) = \inf_{Q\in\overline{\mathrm{Alt}}(P)} D_{\mathrm{KL}}(P_{\mathrm{C}}^q \| Q_{\mathrm{C}}^q)$. The corresponding characteristic times (Garivier & Kaufmann, 2016; Fiez et al., 2019) are

$$T_{\mathrm{F}}^*(P) = \left[\sup_{q\in\mathcal{Q}} I_{\mathrm{F}}(q;P)\right]^{-1}, \qquad T_{\mathrm{C}}^*(P) = \left[\sup_{q\in\mathcal{Q}} I_{\mathrm{C}}(q;P)\right]^{-1}.$$

We use the convention $1/0 = +\infty$. A smaller characteristic time means that the true model can be separated from every model in $\overline{\mathrm{Alt}}(P)$ with fewer observations at first order.

Proposition 1 (Value of retaining prognostic covariates)**.** *For every $P \in \mathcal{P}$, $T_{\mathrm{C}}^*(P) \geq T_{\mathrm{F}}^*(P)$. There are instances in the centered Gaussian RKHS model for which the inequality is strict, including instances with a binding deployment capacity constraint.*

Appendix A.1 gives the proof. The weak inequality follows from data processing because O^{C} retains G and discards the remaining within-group information in X from O^{F} (Cover & Thomas, 2006). Strict inequality already occurs with two equally sized groups and binary treatment. Let $X = (G, U)$, where $U \in \{-1, +1\}$ is uniform conditional on group, and let $r_c(X) = \nu_c U$ with $\nu_c \neq 0$ in every group-treatment cell. If U is retained, a cellwise mean shift h_c has Gaussian KL divergence $h_c^2/(2\sigma_c^2)$. After U is marginalized out, the cell outcome is a two-component Gaussian mixture, and the KL divergence under the same mean shift is strictly smaller whenever $h_c\nu_c \neq 0$. In the construction of Appendix A.1, the full-covariate least-favorable shift changes all four group-treatment cells and the feasible designs assign each such cell probability bounded away from zero. The KL contraction is therefore uniform over $q \in \mathcal{Q}$, which gives $T_{\mathrm{C}}^*(P) > T_{\mathrm{F}}^*(P)$ while the deployment constraint remains binding.

The information loss has a sequential consequence. Let τ_δ^{C} denote the stopping time of a procedure that observes only (G, A, Y).

Proposition 2 (Coarsened-data lower bound)**.** *Every uniformly δ-correct coarsened-data procedure satisfying $q_t \in \mathcal{Q}$ at every round obeys*

$$\liminf_{\delta\downarrow 0} \frac{\mathbb{E}_P[\tau_\delta^{\mathrm{C}}]}{\log(1/\delta)} \geq T_{\mathrm{C}}^*(P).$$

The proof is a stopped change-of-measure argument in the coarsened filtration (Kaufmann et al., 2016) and is given in Appendix A.2. Proposition 1 and Proposition 2 together show that discarding the within-group covariate information in X can strictly increase the first-order lower bound. It remains to determine whether the smaller full-covariate benchmark can be attained when the prognostic functions are unknown.

3.2 Full-Covariate Lower Bound

To characterize $T_{\mathrm{F}}^*(P)$, fix a competitor $z \neq z^*$ and define $d_z = z^* - z$ and $\Delta_z = d_z^\top \theta > 0$. Write $r = (r_c)_c$ and $r' = (r'_c)_c$. For a design $q \in \mathcal{Q}$, compare the true model $P = (\theta, r)$ with $Q = (\theta', r')$, and write $D_{\mathrm{KL}}^q(P\|Q) := D_{\mathrm{KL}}(P_{\mathrm{F}}^q \| Q_{\mathrm{F}}^q)$. In cell $c = (m, a)$, the conditional Gaussian KL contribution is $\{2\sigma_c^2\}^{-1}\mathbb{E}[\{(\theta_c - \theta'_c) + (r_c(X) - r'_c(X))\}^2 \mid G = m]$. Because r_c and r'_c are centered within group, this divergence is minimized at $r'_c = r_c$ for any fixed θ'_c. Thus, the least-favorable alternatives under which a competing allocation is optimal change only the group-treatment means.

For a comparison direction d, define

$$\Gamma_q(d) = \sum_{m,a} \frac{d_{m,a}^2 \sigma_{m,a}^2}{p_m q_{m,a}}, \tag{7}$$

where a summand is interpreted as zero when $d_{m,a} = 0$, including when $q_{m,a} = 0$, and $\Gamma_q(d) = +\infty$ if $d_{m,a} \neq 0$ and $q_{m,a} = 0$. Projecting the mean vector onto the boundary $d_z^\top \theta' \leq 0$ gives

$$\inf_{Q=(\theta',r')\in\mathcal{P}:\, d_z^\top\theta'\leq 0} D_{\mathrm{KL}}^q(P\|Q) = \frac{\Delta_z^2}{2\Gamma_q(d_z)}. \tag{8}$$

Hence $I_{\mathrm{F}}(q;P) = \mathcal{I}_P(q)$, with

$$\mathcal{I}_P(q) = \min_{z \neq z^*} \frac{\Delta_z^2}{2\Gamma_q(d_z)}, \quad q^* \in \arg\max_{q \in \mathcal{Q}} \mathcal{I}_P(q), \quad T_{\mathrm{F}}^*(P) = 2 \inf_{q \in \mathcal{Q}} \max_{z \neq z^*} \frac{\Gamma_q(d_z)}{\Delta_z^2}. \tag{9}$$

Corollary 1 (Characteristic time with unknown prognostic functions)**.** *For every $P \in \mathcal{P}$ with a unique optimal allocation, the full-covariate characteristic time $T_{\mathrm{F}}^*(P)$ is the same whether the prognostic functions are known in advance or are unknown over $\prod_c \mathcal{R}_c$.*

The corollary follows from the cellwise KL decomposition above. For any alternative cell means θ', the conditional Gaussian KL is minimized at $r' = r$, so enlarging the alternative class to unknown prognostic functions does not produce a closer allocation-changing alternative. Appendix B.1 gives the complete derivation. This equality concerns the first-order identification benchmark. In finite samples, r still has to be estimated to construct predictive likelihoods, and Section 4 quantifies the resulting remainder.

Eq. (9) also separates the roles of observed covariates and experimental design. The full-covariate characteristic time depends on the residual variances σ_c^2, the comparison directions d_z, the value gaps Δ_z, and the feasible design set $\mathcal{Q}$. The magnitude of $r_c(X)$ does not appear because the likelihood comparison conditions on X and the least-favorable alternative keeps r_c fixed. In contrast, the coarsened likelihood integrates over X and therefore depends on the marginal outcome distributions induced by $r_c(X)$. The difference between T_{F}^* and T_{C}^* is the KL information lost when the within-group covariates are discarded. Estimating r_c still affects finite-sample evidence, and Section 4 shows that this additional error is lower order.

Theorem 1 (Full-covariate lower bound)**.** *Let $P \in \mathcal{P}$ have a unique optimal allocation. Every sequential procedure that is uniformly δ-correct on $\mathcal{P}$ and satisfies $q_t \in \mathcal{Q}$ at every round obeys*

$$\liminf_{\delta \downarrow 0} \frac{\mathbb{E}_P[\tau_\delta]}{\log(1/\delta)} \geq T_{\mathrm{F}}^*(P).$$

Thus, every uniformly valid procedure has first-order expected stopping time at least $T_{\mathrm{F}}^*(P)\log(1/\delta)$. The least-favorable alternatives preserve the prognostic functions, so the benchmark is unchanged when those functions must be learned. Appendix B.2 gives the complete proof. Section 4 shows that the bound is asymptotically attainable.

4 Adaptive Procedure

We now construct a procedure that learns the prognostic functions online, controls the selection error at level δ, and asymptotically attains the full-covariate benchmark $T_{\mathrm{F}}^*(P)$. For $d = z - z'$, let $\mathcal{C}(d) = \{c : d_c \neq 0\}$ and $\mathcal{C}_{\mathrm{all}} = \bigcup_{z \neq z'} \mathcal{C}(z - z')$. Only cells in $\mathcal{C}_{\mathrm{all}}$ can affect the ordering of two feasible allocations. Write $C_t = (G_t, A_t)$. All set-valued optimization steps below use a fixed measurable tie-breaking rule.

4.1 Adaptive Design and Prognostic Learning

Adaptive design. For forced exploration (Garivier & Kaufmann, 2016), choose a baseline design $q^{(0)} \in \mathcal{Q}$ with $q_c^{(0)} > 0$ for every $c \in \mathcal{C}_{\mathrm{all}}$. For each such cell, let $N_c(t) = \sum_{s \leq t} \mathbf{1}\{C_s = c\}$ and, whenever $N_c(t) > 0$, $\widehat{\theta}_c(t) = N_c(t)^{-1} \sum_{s \leq t : C_s = c} Y_s$. Write $\widehat{\theta}_t = (\widehat{\theta}_c(t))_{c \in \mathcal{C}_{\mathrm{all}}}$. Once every cell in $\mathcal{C}_{\mathrm{all}}$ has been observed, set

$$\widetilde{z}_t \in \arg\max_{z \in \mathcal{Z}} \sum_{c \in \mathcal{C}_{\mathrm{all}}} z_c \widehat{\theta}_c(t), \qquad \widehat{\Delta}_{z,t} = \sum_{c \in \mathcal{C}(\widetilde{z}_t - z)} (\widetilde{z}_{t,c} - z_c)\widehat{\theta}_c(t).$$

If all estimated gaps are positive, update the estimated information-optimal design by

$$\widehat{q}_t \in \arg\min_{q \in \mathcal{Q}} \max_{z \neq \widetilde{z}_t} \frac{2\Gamma_q(\widetilde{z}_t - z)}{\widehat{\Delta}_{z,t}^2}. \tag{10}$$

Before this update is well defined, set $\widehat{q}_t = q^{(0)}$. With $\rho_t = t^{-\gamma}$ for $0 < \gamma < 1$, let $q_1 = q^{(0)}$ and

$$q_{t+1} = (1 - \rho_{t+1})\widehat{q}_t + \rho_{t+1} q^{(0)}. \tag{11}$$

The mixture keeps $q_t \in \mathcal{Q}$, maintains sampling of every cell in $\mathcal{C}_{\text{all}}$, and allows q_t to approach q^*.

Prognostic learning. Index observations in cell c by arrival order as $\{(X_{c,i}, Y_{c,i})\}_{i\geq 1}$. For $n \geq 1$, let $\bar{Y}_{c,n} = n^{-1}\sum_{i=1}^{n} Y_{c,i}$ and fit

$$\widehat{r}_{c,n} \in \operatorname*{arg\,min}_{r\in\mathcal{R}_c} \frac{1}{n}\sum_{i=1}^{n}\{Y_{c,i} - \bar{Y}_{c,n} - r(X_{c,i})\}^2.$$

Set $\widehat{\mu}_{c,n}(x) = \bar{Y}_{c,n} + \widehat{r}_{c,n}(x)$. We refit when the cell count reaches $1, 2, 4, \ldots$ and use the latest fit between updates. Let $\widehat{\mu}_c^{(t-1)}$ denote the predictor available before round t. With φ_σ the $N(0,\sigma^2)$ density, set $\widehat{p}_t(y \mid X_t, A_t) = \varphi_{\sigma_c}(y - \widehat{\mu}_c^{(t-1)}(X_t))$ for $c = C_t$. Before the first observation in a cell, set $\widehat{\mu}_c^{(t-1)} \equiv 0$. The predictor is therefore fixed before Y_t is observed, which is the property needed for sequential validity (Wasserman et al., 2020).

4.2 Pairwise Evidence and Stopping

Following Wasserman et al. (2020, Section 8), for an ordered pair (z, z'), write $d = z - z'$ and define the null model set $H_{0,d} := \{Q = (\theta', r') \in \mathcal{P} : d^\top\theta' \leq 0\}$. Let $p_Q(y \mid x, a)$ be the conditional Gaussian density under $Q = (\theta', r')$, and define

$$L_t^{(d)}(Q) = \prod_{s=1}^{t} p_Q(Y_s \mid X_s, A_s)^{\mathbf{1}\{C_s\in\mathcal{C}(d)\}}, \qquad \widehat{L}_t^{(d)} = \prod_{s=1}^{t} \widehat{p}_s(Y_s \mid X_s, A_s)^{\mathbf{1}\{C_s\in\mathcal{C}(d)\}}.$$

Because $d_c = 0$ outside $\mathcal{C}(d)$, $d^\top\theta'$ in the profile below means $\sum_{c\in\mathcal{C}(d)} d_c\theta'_c$. The null profile likelihood is

$$L_{0,t}^{\sup}(d) = \sup_{\substack{\{\theta'_c, r'_c : c\in\mathcal{C}(d)\} \\ d^\top\theta' \leq 0,\ r'_c\in\mathcal{R}_c}} L_t^{(d)}(\theta', r'), \tag{12}$$

and the ideal evidence is $E_t(d) = \widehat{L}_t^{(d)}/L_{0,t}^{\sup}(d)$. For any $P_0 \in H_{0,d}$, $E_t(d) \leq \widehat{L}_t^{(d)}/L_t^{(d)}(P_0)$. Let $\mathcal{G}_s = \sigma(\mathcal{F}_{s-1}, X_s, G_s, A_s)$. The fitted conditional density is determined before Y_s is observed, so each likelihood-ratio increment has conditional mean one under P_0 given $\mathcal{G}_s$. The right-hand side is therefore a nonnegative martingale, and Ville's inequality (Ville, 1939) gives

$$\mathbb{P}_{P_0}\left(\sup_t E_t(d) \geq 1/\alpha\right) \leq \alpha. \tag{13}$$

The profile in Eq. (12) is infinite-dimensional because $r_c \in \mathcal{R}_c$. The representer theorem (Schölkopf et al., 2001) reduces it to a finite-dimensional convex quadratically constrained quadratic program, given explicitly in Appendix E. Let $\mathcal{L}_{0,t}^*(d)$ denote its optimal Gaussian quadratic loss, with certified primal and dual bounds $\underline{\mathcal{L}}_{0,t}(d) \leq \mathcal{L}_{0,t}^*(d) \leq \overline{\mathcal{L}}_{0,t}(d)$ (Boyd & Vandenberghe, 2004, Chapter 5). Writing $\mathcal{N}_t(d) = \prod_{s\leq t: C_s\in\mathcal{C}(d)}(2\pi\sigma_{C_s}^2)^{-1/2}$, define $U_t(d) = \mathcal{N}_t(d)\exp\{-\underline{\mathcal{L}}_{0,t}(d)\} \geq L_{0,t}^{\sup}(d)$. The computable evidence therefore satisfies (Wasserman et al., 2020, Section 6)

$$\widetilde{E}_t(d) = \frac{\widehat{L}_t^{(d)}}{U_t(d)} \leq E_t(d). \tag{14}$$

We compute each profile until $0 \leq \overline{\mathcal{L}}_{0,t}(d) - \underline{\mathcal{L}}_{0,t}(d) \leq \zeta_t$, where $\zeta_t \leq C_{\text{opt}}\log(t+1)$ for a finite C_{opt}. Hence the numerical loss in log evidence is $o(t)$.

Let $J = |\mathcal{Z}| \geq 2$ and assign $\alpha_{z,z'} = \delta/[J(J-1)]$ to each ordered pair. We stop at the first time t for which some $z \in \mathcal{Z}$ satisfies $\widetilde{E}_t(z - z') \geq 1/\alpha_{z,z'}$ for every $z' \neq z$, and output $\widehat{z}_{\tau_\delta} = z$.

Complete procedure. At each round, the procedure uses observations available before the new outcome to form the predictive density and the next feasible design. After observing Y_t, it updates the cell mean and the prognostic fit when a refitting time is reached, recomputes the design in Eq. (10), and evaluates the certified pairwise evidence in Eq. (14). Stopping requires one allocation to cross its pairwise threshold against every competitor. The same data therefore serve three distinct

tasks: estimating the marginal cell means that determine the allocation, learning r_c for the predictive likelihood, and adapting q_t toward the design that maximizes the worst-case information rate. Predictability of the fitted density is used for finite-sample validity, while consistency of the fitted means and prognostic functions is used for the first-order stopping-time result.

The finite-sample validity result uses the model class and feasible predictable designs defined in Section 2. For the first-order stopping-time result, fix $P \in \mathcal{P}$ with a unique optimal allocation z^* and a unique information-optimal design q^* in Eq. (9). We also use the certified profile tolerance $\zeta_t \le C_{\mathrm{opt}} \log(t+1)$ specified above.

Theorem 2 (Validity and first-order optimality)**.** *Let $\rho_t = t^{-\gamma}$ with $0 < \gamma < 1$, and suppose the certified profile computation satisfies $\zeta_t \le C_{\mathrm{opt}} \log(t+1)$ for a finite C_{opt}. For every $\delta \in (0,1)$, the procedure is uniformly δ-correct on $\mathcal{P}$ in the sense of Eq.* (6)*, and $q_t \in \mathcal{Q}$ at every round before stopping. If $P \in \mathcal{P}$ has a unique optimal allocation z^* and Eq.* (9) *has a unique maximizer q^*, then*

$$\frac{\mathbb{E}_P[\tau_\delta]}{T^*_{\mathrm{F}}(P)\log(1/\delta)} \to 1 \qquad \text{as } \delta \downarrow 0.$$

The two conclusions use different parts of the construction. Finite-sample δ-correctness follows from the predictability of $\widehat{p}_t$, Ville's inequality, the conservative null-profile bound in Eq. (14), a union bound over ordered allocation pairs, and almost-sure stopping for every model with a unique optimal allocation. For first-order efficiency, forced exploration yields consistent cell means and prognostic fits, while uniqueness of q^* implies that the adaptive design tracks the information-optimal design. The predictive likelihood remainder from estimating r, the excess null-profile likelihood, and the certified numerical profile gap are all $o(t)$. Lemmas 4 and 5 show that the two statistical remainders are sublinear even though the same sequential observations are used to estimate r and test allocation comparisons. Consequently, the minimum log evidence against the true optimum grows as $t/T^*_{\mathrm{F}}(P) + o(t)$. Appendix C.2 gives the complete proof.

The expected-stopping-time conclusion requires a tail bound on this evidence growth, not only convergence of its normalized value. Lemma 6 proves that for every $\epsilon > 0$, there are finite constants C_ϵ and t_ϵ such that

$$\mathbb{P}_P\left(\min_{z \neq z^*} \log \widetilde{E}_t(z^* - z) < t\{[T^*_{\mathrm{F}}(P)]^{-1} - \epsilon\}\right) \le C_\epsilon t^{-3}, \qquad t \ge t_\epsilon.$$

Taking $t = (1+\eta)T^*_{\mathrm{F}}(P)\log(1/\delta)$ makes the evidence exceed every pairwise threshold with probability tending to one. Summing the t^{-3} tail controls the excess expected stopping time by $o\{\log(1/\delta)\}$. This is the step that upgrades first-order evidence growth to the expectation statement in Theorem 2.

Thus, learning the prognostic functions does not increase the first-order sample complexity. On every instance from Proposition 1 with $T^*_{\mathrm{C}}(P) > T^*_{\mathrm{F}}(P)$ that satisfies the theorem conditions,

$$\lim_{\delta \downarrow 0} \frac{\mathbb{E}_P[\tau_\delta]}{\log(1/\delta)} = T^*_{\mathrm{F}}(P) < T^*_{\mathrm{C}}(P) \le \liminf_{\delta \downarrow 0} \frac{\mathbb{E}_P[\tau^{\mathrm{C}}_\delta]}{\log(1/\delta)}.$$

Hence, retaining X yields a strict first-order reduction in expected sample size.

Deterministic checkpointed design refreshes and evidence evaluations preserve the validity and first-order coefficient under the conditions in Appendix C.3.

5 Experiments

In this section, we investigate three questions: (1) Does the experimental instance exhibit the information gap $T^*_{\mathrm{F}} < T^*_{\mathrm{C}}$ established by Proposition 1? (2) When the design is fixed at the information-optimal q^*, does learning the prognostic function add only finite-sample overhead while the stopping time approaches the full-covariate benchmark in Theorem 2? (3) Does the stopping-time ratio between the learned full-covariate procedure and the coarsened benchmark move toward $T^*_{\mathrm{F}}/\widehat{T}^*_{\mathrm{C}}$ as the confidence requirement increases? All experiments use the same two-group instance, deployment constraint, experimental constraint, and data-generating process. Appendix D gives the complete implementation and numerical results.

5.1 Experimental Design

Setup. We use the construction in Appendix A.1, with two equally sized groups, binary treatment, and a deployment constraint that treats exactly one group. In the cell order $((1,0),(1,1),(2,0),(2,1))$, the marginal means are $\theta = (0, 0.2, 0.2, 0)$. Conditional on group, the prognostic covariate $U \in \{-1, +1\}$ is uniform. Within each cell, $r_c(U) = \beta_c U$ with $|\beta_c| \leq B = 2$. The true instance has $\beta_c = \nu = 1$ in every cell, so $Y = \theta_c + \nu U + \varepsilon$ with $\varepsilon \sim N(0,1)$. The experimental treatment probabilities satisfy $q_{m,1} \in [0.1, 0.9]$. Let z^* denote the optimal allocation, let z denote its only competitor, and set $d = z^* - z$, so $d^\top \theta = 0.2$. With one competitor, the design update in Eq. (10) reduces to minimizing $\Gamma_q(d)$. The unique minimizer satisfies $q^*_{1,1} = q^*_{2,1} = 1/2$. We set $q^{(0)} = q^*$, which gives $q_t = q^*$ at every round. This choice isolates prognostic-function learning and evidence accumulation from adaptive design tracking. The full-covariate experiment observes U, while the coarsened experiment observes only (G, A, Y).

Comparators. We evaluate three full-covariate procedures. **Known** r uses the true prognostic function in both the predictive likelihood and the null profile. **Oracle predictor** uses the true prognostic function only in the predictive likelihood and treats r as unknown in the null profile. **Learned** r is the proposed procedure and estimates r predictably from past observations. Comparing Learned r with Oracle predictor isolates the cost of estimating r in the predictive likelihood. Comparing Oracle predictor with Known r isolates the additional cost of maximizing the null likelihood over unknown r and computing a certified profile bound. We also report a **Coarsened benchmark** that discards U. This benchmark uses the true marginal mixture density and the boundary alternative returned by the coarsened numerical optimization at $q_{1,1} = q_{2,1} = 1/2$. Its threshold-crossing time serves only as a numerical reference for the coarsened first-order information rate. Appendix F gives the rank-one profile calculation used by the full-covariate procedures.

Evaluation. Write $L_\delta = \log(1/\delta)$. We evaluate $L_\delta \in \{20, 50, 100, 200, 500, 1000, 2000\}$ using 100 paired trajectories at each level. Evidence is evaluated every 200 arrivals and the administrative horizon is 1.5×10^6. All three full-covariate procedures stop before the horizon, and the coarsened benchmark crosses its threshold before the horizon. Tables report means with Monte Carlo standard errors (MCSEs), while figures show mean ± 2 MCSEs. Within each replication, all four procedures are evaluated on the same simulated trajectory.

5.2 Results

Information gain from retaining the prognostic covariate. The full-covariate calculation gives $T^*_{\mathrm{F}} = 200$. Numerical optimization of the marginal Gaussian-mixture KL gives the estimate $\widehat{T}^*_{\mathrm{C}} = 363.953$, so $T^*_{\mathrm{F}}/\widehat{T}^*_{\mathrm{C}} \approx 0.5495$. The corresponding numerical boundary alternative has cell means approximately $(0.1, 0.1, 0.1, 0.1)$, which makes the two feasible allocations tie. Thus retaining U reduces the estimated first-order coefficient by about 45.0% on this instance. Table 1 reports the full-covariate benchmark and the numerical coarsened estimate, and Figure 1(a) shows them with the normalized mean stopping times. Appendix D.2 describes the coarsened optimization and its numerical stability checks. Since the deployment problem and experimental constraints are identical in the two experiments, the difference measures the KL information lost by discarding U.

Table 1: First-order information benchmarks for the experimental instance. The coarsened value is the numerical estimate from Appendix D.2.

ν	T^*_{F}	$\widehat{T}^*_{\mathrm{C}}$	$T^*_{\mathrm{F}}/\widehat{T}^*_{\mathrm{C}}$
1	200.00	363.95	0.5495

Adaptation to the full-covariate benchmark. Because $q_t = q^*$ at every round, Figure 1(b) isolates the prognostic-function estimation and evidence components of Theorem 2. The normalized mean stopping time $\mathbb{E}[\tau]/(T^*_{\mathrm{F}} L_\delta)$ for Learned r decreases from 5.251 at $L_\delta = 20$ to 1.463 at $L_\delta = 200$ and 1.055 at $L_\delta = 2000$. At $L_\delta = 2000$, the corresponding values are 1.036 for Known r and 1.037 for Oracle predictor. Table 3 in Appendix D.4 further shows Learned r/Oracle predictor decreasing to 1.018 and Oracle predictor/Known r decreasing to 1.001. These comparisons separate the finite-sample cost of estimating r in the predictive likelihood from the cost of treating r

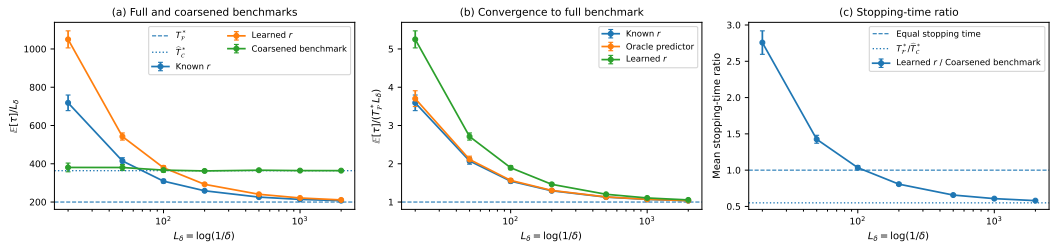

Figure 1: Information and stopping-time behavior on the experimental instance at $\nu = 1$. Panel (a) plots $\mathbb{E}[\tau]/L_\delta$ for Known r, Learned r, and the coarsened benchmark, with horizontal lines at $T_{\mathrm{F}}^* = 200$ and $\widehat{T}_{\mathrm{C}}^* = 363.953$. Panel (b) plots $\mathbb{E}[\tau]/(T_{\mathrm{F}}^* L_\delta)$ for the three full-covariate procedures. Panel (c) plots the Learned r/coarsened benchmark ratio and $T_{\mathrm{F}}^*/\widehat{T}_{\mathrm{C}}^* \approx 0.5495$. For the coarsened benchmark, τ denotes its threshold-crossing time under the numerical boundary alternative. Error bars are ± 2 MCSEs over 100 paired trajectories.

as unknown under the null. At $L_\delta = 2000$, the larger Learned r/Oracle predictor gap shows that predictive estimation is the larger of the two measured overheads.

Stopping-time separation from the coarsened benchmark. Table 2 in Appendix D.4 reports all stopping times over the confidence grid, and Figure 1(c) plots the Learned r/coarsened benchmark ratio. The coarsened benchmark closely tracks $\widehat{T}_{\mathrm{C}}^* L_\delta$, with normalized mean threshold-crossing time 1.046 at $L_\delta = 20$ and 1.0004 at $L_\delta = 2000$. Learned r has a larger finite-sample overhead at small L_δ, but its normalized mean approaches the smaller coefficient $T_{\mathrm{F}}^* = 200$.

The stopping-time ratio decreases from 2.758 at $L_\delta = 20$ to 1.034 at $L_\delta = 100$, falls below one at $L_\delta = 200$, and reaches 0.580 at $L_\delta = 2000$. The numerical first-order ratio is $T_{\mathrm{F}}^*/\widehat{T}_{\mathrm{C}}^* \approx 0.5495$. At $L_\delta = 2000$, the normalized means are 1.0554 for Learned r and 1.0004 for the coarsened benchmark, which imply a stopping-time ratio of 0.5797. The remaining difference from 0.5495 is consistent with the finite-sample overhead of Learned r.

6 Conclusion

In this paper, we study fixed-confidence identification of capacity-constrained group-level treatment allocations when individual-level prognostic covariates are observed during experimentation but excluded from randomization and deployment. We show that retaining these covariates can strictly reduce the optimal first-order expected sample requirement and derive the corresponding full-covariate characteristic time. We characterize the least-favorable allocation-changing alternatives and show that they preserve the prognostic functions, so learning these functions does not change the first-order benchmark. Our adaptive procedure estimates the prognostic functions from the same sequential experiment, controls incorrect allocation selection in finite samples, and asymptotically attains the lower bound. The results establish when within-group prognostic covariates can reduce the experimental sample requirement for constrained treatment allocation. The gain occurs even when the covariates are excluded from both deployment and within-group randomization.

Limitations and future work. The analysis assumes a finite allocation set, a prespecified centered Gaussian RKHS outcome model, a known covariate distribution for centering, known residual variances, and a unique information-optimal design. Estimating the centering distribution and residual variances, or extending beyond Gaussian conditional outcomes, while retaining finite-sample validity and first-order optimality is an interesting next step. Nonunique information-optimal designs require set-valued design tracking. Larger constrained allocation problems also motivate scalable methods for repeated composite-null profile optimization and direct evaluation of adaptive design tracking.

AI USE STATEMENT

Generative AI tools were used for language polishing. The authors independently derived and checked the theoretical derivations, experimental procedures, and reported results.

REFERENCES

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1): 133–161, 2021.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of machine learning research*, 3(Nov):463–482, 2002.

David Benkeser, Iván Díaz, Alex Luedtke, Jodi Segal, Daniel Scharfstein, and Michael Rosenblum. Improving precision and power in randomized trials for covid-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes. *Biometrics*, 77(4):1467–1481, 2021.

Debopam Bhattacharya and Pascaline Dupas. Inferring welfare maximizing treatment assignment under budget constraints. *Journal of Econometrics*, 167(1):168–196, 2012.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. ISBN 9780199535255. doi: 10.1093/acprof:oso/9780199535255.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Richard P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, 1973.

Emil Carlsson, Debabrota Basu, Fredrik Johansson, and Devdatt Dubhashi. Pure exploration in bandits with linear constraints. In *International Conference on Artificial Intelligence and Statistics*, pp. 334–342. PMLR, 2024.

Brian M Cho and Nathan Kallus. Exploration in the limit. *arXiv preprint arXiv:2601.00084*, 2026.

Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.

Rémy Degenne, Pierre Ménard, Xuedong Shang, and Michal Valko. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pp. 2432–2442. PMLR, 2020.

Tanner Fiez, Lalit Jain, Kevin Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pp. 100–118, 1975.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pp. 998–1027. PMLR, 2016.

Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with numpy. *nature*, 585(7825):357–362, 2020.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

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.

Masahiro Kato and Kaito Ariu. The role of contextual information in best arm identification. *Journal of Machine Learning Research*, 27(51):1–61, 2026.

Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. *The Journal of Machine Learning Research*, 17(1):1–42, 2016.

Seok-Jin Kim. Nearly optimal best arm identification for semiparametric bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 4402–4410. PMLR, 2026.

Seok-Jin Kim, Gi-Soo Kim, and Min-hwan Oh. Experimental design for semiparametric bandits. In *The Thirty Eighth Annual Conference on Learning Theory*, pp. 3215–3252. PMLR, 2025.

Toru Kitagawa and Aleksey Tetenov. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Akshay Krishnamurthy, Zhiwei Steven Wu, and Vasilis Syrgkanis. Semiparametric contextual bandits. In *International Conference on Machine Learning*, pp. 2776–2785. PMLR, 2018.

Winston Lin. Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, pp. 295–318, 2013.

Charles F Manski. Statistical treatment rules for heterogeneous populations. *Econometrica*, 72(4): 1221–1246, 2004.

Daniel Molitor. Anytime-valid optimal policy identification. *arXiv preprint arXiv:2606.17515*, 2026.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Yoan Russac, Christina Katsimerou, Dennis Bohle, Olivier Cappé, Aurélien Garivier, and Wouter M Koolen. A/b/n testing with control in the presence of subpopulations. *Advances in Neural Information Processing Systems*, 34:25100–25110, 2021.

Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A generalized representer theorem. In *International conference on computational learning theory*, pp. 416–426. Springer, 2001.

Mohammad Shahverdikondori, Amir Mohammad Abouei, Alireza Rezaeimoghadam, and Negar Kiyavash. Pure exploration beyond reward feedback: The role of post-action context. *arXiv preprint arXiv:2502.03061*, 2025.

Anastasios A Tsiatis, Marie Davidian, Min Zhang, and Xiaomin Lu. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine*, 27(23):4658–4677, 2008.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Jean Ville. *Etude critique de la notion de collectif*, volume 3 of *Monographies des probabilites*. Gauthier-Villars, Paris, 1939.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Larry Wasserman, Aaditya Ramdas, and Sivaraman Balakrishnan. Universal inference. *Proceedings of the National Academy of Sciences*, 117(29):16880–16890, 2020.

Ian Waudby-Smith, Lili Wu, Aaditya Ramdas, Nikos Karampatziakis, and Paul Mineiro. Anytime-valid off-policy inference for contextual bandits. *ACM/IMS Journal of Data Science*, 1(3):1–42, 2024.

A Full-Covariate versus Coarsened Sample Complexity

A.1 Proof of Proposition 1

Proof. Fix $q \in \mathcal{Q}$. Since $O^{\mathrm{C}} = (G, A, Y)$ is a measurable function of $O^{\mathrm{F}} = (X, G, A, Y)$, the data-processing inequality gives

$$D_{\mathrm{KL}}(P^q_{\mathrm{C}} \| Q^q_{\mathrm{C}}) \le D_{\mathrm{KL}}(P^q_{\mathrm{F}} \| Q^q_{\mathrm{F}})$$

for every $P, Q \in \mathcal{P}$. Taking the infimum over $Q \in \overline{\mathrm{Alt}}(P)$ and then the supremum over $q \in \mathcal{Q}$ yields $T^*_{\mathrm{C}}(P) \ge T^*_{\mathrm{F}}(P)$.

For strict inequality, consider two equally sized groups with $p_1 = p_2 = 1/2$, binary treatment, and covariates $X = (G, U)$, where $U \in \{-1, +1\}$ is defined below. In the coordinate order $((1,0),(1,1),(2,0),(2,1))$, let

$$z^{(1)} = (0, p_1, p_2, 0), \qquad z^{(2)} = (p_1, 0, 0, p_2), \qquad \mathcal{Z} = \{z^{(1)}, z^{(2)}\}.$$

Thus the deployment requirement assigns treatment 1 to exactly one of the two equally sized groups, so the active-treatment capacity is fixed at its binding value $1/2$. Let $d = z^{(1)} - z^{(2)}$ and assume $\Delta = d^\top \theta > 0$. Conditional on $G = m$, let $U \in \{-1, +1\}$ be uniform and set $r_c(X) = \nu_c U$ with $\nu_c \neq 0$ for every cell c. This is a centered rank-one RKHS model with kernel $\widetilde{k}_c(u, u') = uu'$. Choose $B_c \ge |\nu_c|$ so that $r_c \in \mathcal{R}_c$. Fix $0 < \eta < 1/2$ and take

$$\mathcal{Q} = \{q : q_{m,0} + q_{m,1} = 1,\ \eta \le q_{m,a} \le 1 - \eta\}.$$

For binary treatment, these box constraints are a special case of Eq. (4): an upper bound on each arm probability also gives the corresponding lower bound on the other arm because $q_{m,0} + q_{m,1} = 1$.

For a fixed $q \in \mathcal{Q}$, the least-favorable full-covariate alternative keeps r fixed and shifts θ by

$$\min_h \frac{1}{2} \sum_c \frac{p_m q_c}{\sigma_c^2} h_c^2 \quad \text{subject to} \quad d^\top h \le -\Delta.$$

The constraint is active, and the solution is

$$h^*_c(q) = -\frac{\Delta d_c \sigma_c^2}{p_m q_c \Gamma_q(d)}, \qquad I_{\mathrm{F}}(q; P) = \frac{\Delta^2}{2\Gamma_q(d)}. \tag{15}$$

Every coordinate of $h^*(q)$ is nonzero because $d_c \neq 0$ and $q_c \ge \eta$.

After marginalizing U, the outcome density in cell c is

$$f_{c,\theta_c}(y) = \tfrac{1}{2}\varphi_{\sigma_c}(y - \theta_c - \nu_c) + \tfrac{1}{2}\varphi_{\sigma_c}(y - \theta_c + \nu_c).$$

By translation invariance, define $D_c(h) = D_{\mathrm{KL}}(f_{c,\theta_c} \| f_{c,\theta_c + h}) = D_{\mathrm{KL}}(f_{c,0} \| f_{c,h})$. If U is retained, the same location shift h has KL divergence $h^2/(2\sigma_c^2)$. Marginalizing U gives

$$D_c(h) \le \frac{h^2}{2\sigma_c^2}.$$

The inequality is strict whenever $h\nu_c \neq 0$. Equality in data processing would require the likelihood ratio for (U, Y) to be measurable with respect to Y (Cover & Thomas, 2006, Chapter 2), which fails because both values of U have positive posterior probability for every finite Y under Gaussian noise. Evaluating the coarsened divergence at the full-covariate least-favorable shift gives

$$\begin{aligned} I_{\mathrm{C}}(q; P) &\le \sum_c p_m q_c D_c(h^*_c(q)) \\ &< \frac{1}{2} \sum_c \frac{p_m q_c}{\sigma_c^2} \{h^*_c(q)\}^2 = I_{\mathrm{F}}(q; P). \end{aligned}$$

More explicitly, the evaluated gap

$$G(q) := \frac{1}{2} \sum_c \frac{p_m q_c}{\sigma_c^2} \{h^*_c(q)\}^2 - \sum_c p_m q_c D_c(h^*_c(q))$$

is continuous and strictly positive on $\mathcal{Q}$. Because $q_c \in [\eta, 1-\eta]$ and every d_c and ν_c is nonzero, the shifts $|h_c^*(q)|$ are uniformly bounded away from zero on this compact set. Hence $\epsilon_0 := \inf_{q\in\mathcal{Q}} G(q) > 0$, and for every $q \in \mathcal{Q}$,

$$I_{\mathrm{C}}(q; P) \le I_{\mathrm{F}}(q; P) - \epsilon_0.$$

Taking suprema yields

$$\sup_{q\in\mathcal{Q}} I_{\mathrm{C}}(q; P) < \sup_{q\in\mathcal{Q}} I_{\mathrm{F}}(q; P),$$

which gives $T_{\mathrm{C}}^*(P) > T_{\mathrm{F}}^*(P)$. □

A.2 Proof of Proposition 2

Here $\mathrm{kl}(u, v) = u\log(u/v) + (1-u)\log\{(1-u)/(1-v)\}$ denotes the binary KL divergence used below.

Proof. If $\mathbb{E}_P[\tau_\delta^{\mathrm{C}}] = +\infty$, the claim is immediate. Otherwise, fix a strict alternative Q whose unique optimal allocation differs from $z^*(P)$, and define $E = \{\widehat{z}_{\tau_\delta^{\mathrm{C}}} = z^*(P)\}$. Uniform δ-correctness gives $\mathbb{P}_P(E) \ge 1-\delta$ and $\mathbb{P}_Q(E) \le \delta$. Let $\mathbb{P}_{P,\mathrm{C}}^{\tau}$ denote the law of the coarsened experiment stopped at time τ under model P. The stopped change-of-measure inequality in the coarsened filtration (Kaufmann et al., 2016, Lemma 1) yields

$$\mathrm{kl}(1-\delta, \delta) \le \mathbb{E}_P\left[\log \frac{d\mathbb{P}_{P,\mathrm{C}}^{\tau_\delta^{\mathrm{C}}}}{d\mathbb{P}_{Q,\mathrm{C}}^{\tau_\delta^{\mathrm{C}}}}\right].$$

For fixed P and Q, the one-round coarsened KL is linear in the design,

$$D_{\mathrm{KL}}(P_{\mathrm{C}}^q \| Q_{\mathrm{C}}^q) = \sum_{m,a} p_m q_{m,a} D_{\mathrm{KL}}\Big(P^{Y|G=m,A=a} \| Q^{Y|G=m,A=a}\Big).$$

Define the expected average design

$$\bar{q}_{m,a} = \frac{\mathbb{E}_P[\sum_{t\le\tau_\delta^{\mathrm{C}}} q_{t,m,a}]}{\mathbb{E}_P[\tau_\delta^{\mathrm{C}}]}.$$

Convexity of $\mathcal{Q}$ gives $\bar{q} \in \mathcal{Q}$. Therefore

$$\mathrm{kl}(1-\delta, \delta) \le \mathbb{E}_P[\tau_\delta^{\mathrm{C}}] D_{\mathrm{KL}}(P_{\mathrm{C}}^{\bar{q}} \| Q_{\mathrm{C}}^{\bar{q}}).$$

Taking the infimum over strict alternatives and approaching the decision boundary gives

$$\mathrm{kl}(1-\delta, \delta) \le \mathbb{E}_P[\tau_\delta^{\mathrm{C}}] I_{\mathrm{C}}(\bar{q}; P) \le \frac{\mathbb{E}_P[\tau_\delta^{\mathrm{C}}]}{T_{\mathrm{C}}^*(P)}.$$

Since $\mathrm{kl}(1-\delta, \delta) = (1+o(1))\log(1/\delta)$, the result follows. □

Together with Theorem 2, Proposition 2 shows that these instances admit an attainable first-order gap between retaining and discarding the within-group covariate information in X.

B Proof of the Full-Covariate Lower Bound

B.1 Least-favorable alternative for a fixed competitor

Fix $q \in \mathcal{Q}$ and models $P = (\theta, r)$ and $Q = (\theta', r')$ with the same known residual variances. For cell $c = (m, a)$, the conditional Gaussian KL is

$$\frac{1}{2\sigma_c^2}\mathbb{E}\big[\{(\theta_c - \theta_c') + (r_c(X) - r_c'(X))\}^2 \mid G = m\big].$$

Centering of r_c and r_c' removes the cross term, so the numerator equals

$$(\theta_c - \theta_c')^2 + \mathbb{E}[(r_c(X) - r_c'(X))^2 \mid G = m].$$

For every fixed θ', the divergence is minimized by $r' = r$.

Fix $z \neq z^*$, write $d = d_z$ and $\Delta = d^\top \theta > 0$, and set $h = \theta' - \theta$. The competitor reaches the decision boundary when $d^\top h \leq -\Delta$. If $q_c = 0$ for some cell with $d_c \neq 0$, the KL infimum is zero and Eq. (16) below holds because $\Gamma_q(d) = +\infty$. Otherwise, changes in coordinates with $d_c = 0$ can only increase the KL objective, so set $h_c = 0$ there and restrict to the coordinates $c = (m, a)$ with $d_c \neq 0$. On these coordinates define

$$W_{q,d} = \mathrm{diag}_{c=(m,a)\in\mathcal{C}(d)} \left(\frac{p_m q_c}{\sigma_c^2} \right),$$

which is positive definite. The least-favorable mean shift solves

$$\min_h \frac{1}{2} h^\top W_{q,d} h \quad \text{subject to} \quad d^\top h \leq -\Delta,$$

where h and d in this display are restricted to the same coordinates. The Lagrange equations give $h = -\lambda W_{q,d}^{-1} d$ with $\lambda = \Delta / (d^\top W_{q,d}^{-1} d)$. Since

$$d^\top W_{q,d}^{-1} d = \Gamma_q(d) = \sum_{m,a} \frac{d_{m,a}^2 \sigma_{m,a}^2}{p_m q_{m,a}},$$

with the convention from Eq. (7), we obtain

$$\inf_{Q=(\theta',r')\in\mathcal{P}:\, d^\top\theta' \leq 0} D_{\mathrm{KL}}^q(P\|Q) = \frac{\Delta^2}{2\Gamma_q(d)}. \tag{16}$$

Minimizing over competitors and maximizing over $q \in \mathcal{Q}$ gives Eq. (9).

B.2 Proof of Theorem 1

Proof. If $\mathbb{E}_P[\tau_\delta] = +\infty$, the claim is immediate. Otherwise, fix a strict alternative Q under which a competitor is uniquely optimal, and let $E = \{\widehat{z}_{\tau_\delta} = z^*\}$. Uniform δ-correctness implies $\mathbb{P}_P(E) \geq 1 - \delta$ and $\mathbb{P}_Q(E) \leq \delta$. Let $\mathbb{P}_{P,\mathrm{F}}^\tau$ denote the law of the full-covariate experiment stopped at time τ under model P. The stopped change-of-measure inequality (Kaufmann et al., 2016, Lemma 1) gives

$$\mathrm{kl}(1-\delta, \delta) \leq \mathbb{E}_P \left[\log \frac{d\mathbb{P}_{P,\mathrm{F}}^{\tau_\delta}}{d\mathbb{P}_{Q,\mathrm{F}}^{\tau_\delta}} \right]. \tag{17}$$

The treatment-randomization kernel is predictable and identical under P and Q, so it cancels from the likelihood ratio. For $c = (m, a)$, let $D_c(P, Q)$ denote the conditional KL divergence of (X, Y) given $G = m$, $A = a$. Because P_X is the same under P and Q, this is equivalently the $P_X(\cdot \mid G = m)$-average of the conditional Gaussian outcome KL. Then the right-hand side of Eq. (17) equals

$$\sum_c \mathbb{E}_P[N_c(\tau_\delta)] D_c(P, Q).$$

Predictability and the group probabilities give

$$\mathbb{E}_P[N_c(\tau_\delta)] = p_m \mathbb{E}_P \left[\sum_{t \leq \tau_\delta} q_{t,c} \right], \qquad c = (m, a).$$

Define

$$\bar{q}_{m,a} = \frac{\mathbb{E}_P[\sum_{t\leq\tau_\delta} q_{t,m,a}]}{\mathbb{E}_P[\tau_\delta]}.$$

Since $q_t \in \mathcal{Q}$ at every round and $\mathcal{Q}$ is convex, $\bar{q} \in \mathcal{Q}$. Hence

$$\mathrm{kl}(1-\delta, \delta) \leq \mathbb{E}_P[\tau_\delta] D_{\mathrm{KL}}^{\bar{q}}(P\|Q).$$

For a fixed competitor z, take the infimum over strict alternatives under which z beats z^* and approach the tie boundary. Eq. (16) yields

$$\mathrm{kl}(1-\delta, \delta) \leq \mathbb{E}_P[\tau_\delta] \frac{\Delta_z^2}{2\Gamma_{\bar{q}}(d_z)}.$$

Taking the minimum over competitors gives

$$\mathrm{kl}(1-\delta, \delta) \leq \mathbb{E}_P[\tau_\delta] \mathcal{I}_P(\bar{q}) \leq \frac{\mathbb{E}_P[\tau_\delta]}{T_{\mathrm{F}}^*(P)}.$$

Using $\mathrm{kl}(1-\delta, \delta) = (1 + o(1)) \log(1/\delta)$ proves the theorem. □

C Proof of the Adaptation Theorem

Let

$$\Delta_{\min} = \min_{z \neq z^*} \Delta_z > 0, \qquad \mathcal{D}_* = \{z^* - z : z \neq z^*\}.$$

For $c = (m, a)$, let $S_{c,i} = \inf\{t : N_c(t) = i\}$ and $(X_{c,i}, Y_{c,i}) = (X_{S_{c,i}}, Y_{S_{c,i}})$. Let P_c denote the joint cell law of (X, Y) and let $P_{X,c} = P_X(\cdot \mid G = m)$ denote its covariate marginal. Since treatment assignment depends on the current covariates only through G_t,

$$\mathrm{Law}(X_t, Y_t \mid \mathcal{F}_{t-1}, C_t = c) = P_c, \qquad X \sim P_{X,c}, \quad Y \mid X \sim N(\theta_c + r_c(X), \sigma_c^2) \text{ under } P_c. \tag{18}$$

Thus, observations indexed by arrival order within a cell are independent and identically distributed from the joint law P_c.

For the design analysis, define

$$\widehat{\mathcal{I}}_t(q) = \min_{z \neq \widetilde{z}_t} \frac{\widehat{\Delta}_{z,t}^2}{2\Gamma_q(\widetilde{z}_t - z)}$$

whenever all estimated gaps are positive. The optimizer in Eq. (10) maximizes $\widehat{\mathcal{I}}_t(q)$.

For $d \in \mathcal{D}_*$, let Q_d^* denote the least-favorable null model under q^* and define

$$\mathcal{R}_t^{\mathrm{pred}}(d) = \log L_t^{(d)}(P) - \log \widehat{L}_t^{(d)}, \qquad \mathcal{R}_t^0(d) = \log L_{0,t}^{\sup}(d) - \log L_t^{(d)}(Q_d^*),$$

$$\mathcal{R}_t^{\mathrm{comp}}(d) = \log U_t(d) - \log L_{0,t}^{\sup}(d).$$

Then

$$\log \widetilde{E}_t(d) = \log \frac{L_t^{(d)}(P)}{L_t^{(d)}(Q_d^*)} - \mathcal{R}_t^{\mathrm{pred}}(d) - \mathcal{R}_t^0(d) - \mathcal{R}_t^{\mathrm{comp}}(d). \tag{19}$$

The proof shows that the average design converges to q^* and that all three remainder terms are $o(t)$. Consequently, for every $\epsilon > 0$ and all sufficiently large t,

$$\mathbb{P}_P\left(\min_{d \in \mathcal{D}_*} \log \widetilde{E}_t(d) < t\{[T_{\mathrm{F}}^*(P)]^{-1} - \epsilon\}\right) \leq C_\epsilon t^{-3}. \tag{20}$$

C.1 Auxiliary lemmas

Lemma 1 (Concentration of cell means and tracked designs)**.** *For a vector v indexed by cells and a cell set S, write $\|v\|_{\infty,S} = \max_{c \in S} |v_c|$. For every $\eta > 0$, there exist $C_\eta, c_\eta > 0$ such that, for all sufficiently large t,*

$$\mathbb{P}_P\Big(\|\widehat{\theta}_t - \theta\|_{\infty,\mathcal{C}_{\mathrm{all}}} > \eta\Big) \leq C_\eta e^{-c_\eta t^{1-\gamma}}.$$

Consequently, $\widetilde{z}_t = z^*$ except on an event with a summable probability. For every neighborhood U of q^*,

$$\mathbb{P}_P(\widehat{q}_t \notin U) \leq C_U e^{-c_U t^{1-\gamma}}.$$

Let

$$\bar{q}_t = \frac{1}{t}\sum_{s=1}^{t} q_s, \qquad \widehat{w}_{c,t} = \frac{N_c(t)}{t}, \qquad w_c^* = p_m q_c^*, \quad c = (m, a).$$

Define the corresponding vectors over $\mathcal{C}_{\mathrm{all}}$ by

$$\widehat{w}_t = (\widehat{w}_{c,t})_{c \in \mathcal{C}_{\mathrm{all}}}, \qquad w^* = (w_c^*)_{c \in \mathcal{C}_{\mathrm{all}}}.$$

Then, for every $\eta > 0$,

$$\mathbb{P}_P(\|\bar{q}_t - q^*\|_1 > \eta) \leq C_\eta e^{-c_\eta t^{1-\gamma}}, \qquad \mathbb{P}_P(\|\widehat{w}_t - w^*\|_1 > \eta) \leq C_\eta e^{-c_\eta t^{1-\gamma}}.$$

Proof. Fix $c = (m, a) \in \mathcal{C}_{\text{all}}$ and define its predictable cumulative exposure

$$\Lambda_c(t) = \sum_{s=1}^{t} p_m q_{s,c}.$$

Eq. (11) gives $q_{s,c} \geq \rho_s q_c^{(0)}$. Since $\rho_s = s^{-\gamma}$ with $0 < \gamma < 1$, there is $a_c > 0$ such that

$$\Lambda_c(t) \geq a_c t^{1-\gamma}$$

for all sufficiently large t. The process $N_c(t) - \Lambda_c(t)$ is a bounded-increment martingale, so a Bernstein or Freedman inequality (Freedman, 1975; Howard et al., 2020) gives

$$\mathbb{P}_P\Big(N_c(t) < \frac{a_c}{2} t^{1-\gamma}\Big) \leq e^{-b_c t^{1-\gamma}} \tag{21}$$

for some $b_c > 0$.

At cell arrivals, define $\varepsilon_{c,i} := Y_{c,i} - \theta_c - r_c(X_{c,i}) \sim N(0, \sigma_c^2)$. Then $Y_{c,i} - \theta_c = r_c(X_{c,i}) + \varepsilon_{c,i}$ is mean zero and sub-Gaussian (Boucheron et al., 2013). The RKHS bound gives $|r_c(X)| \leq B_c \kappa_c$, so there is a finite variance proxy v_c^2 such that, for every deterministic n,

$$\mathbb{P}_P\left(\left|\sum_{i=1}^{n}(Y_{c,i} - \theta_c)\right| \geq \eta n\right) \leq 2\exp\left(-\frac{\eta^2 n}{2v_c^2}\right).$$

Summing this bound over $n \geq n_0$ yields

$$\mathbb{P}_P\left(\exists n \geq n_0 : \left|\frac{1}{n}\sum_{i=1}^{n} Y_{c,i} - \theta_c\right| > \eta\right) \leq C_{c,\eta} e^{-c_{c,\eta} n_0}.$$

Choose $n_0 = \lfloor a_c t^{1-\gamma}/2 \rfloor$ and combine the last display with Eq. (21). Then

$$\mathbb{P}_P(|\widehat{\theta}_c(t) - \theta_c| > \eta) \leq C_{c,\eta} e^{-c_{c,\eta} t^{1-\gamma}}.$$

A finite union bound over $\mathcal{C}_{\text{all}}$ gives the first claim.

Since $\mathcal{Z}$ is finite and $\Delta_{\min} > 0$, any

$$\eta < \frac{\Delta_{\min}}{2\max_{z \neq z^*} \|d_z\|_1}$$

is small enough that $\|\widehat{\theta}_t - \theta\|_{\infty, \mathcal{C}_{\text{all}}} \leq \eta$ implies $\widetilde{z}_t = z^*$. On this event, all estimated gaps converge to their population values. The map $1/\Gamma_q(d)$ extends continuously to zero when a required exposure vanishes. Since the design set is compact and the number of competitors is finite,

$$\sup_{q \in \mathcal{Q}} |\widehat{\mathcal{I}}_t(q) - \mathcal{I}_P(q)| \to 0$$

with the same stretched-exponential control. For any neighborhood U of the unique maximizer q^*, compactness and uniqueness imply

$$\mathcal{I}_P(q^*) - \sup_{q \notin U} \mathcal{I}_P(q) > 0.$$

If the uniform error is smaller than one third of this positive separation, every maximizer $\widehat{q}_t$ lies in U, which proves the stated tail for $\widehat{q}_t$.

To control the average design, fix $\eta > 0$ and choose a neighborhood U of q^* small enough for the desired ℓ_1 error. Choose $\lambda \in (0, 1)$ so that the first λt rounds contribute at most $\eta/3$ to the average. If $\widehat{q}_{s-1} \in U$ for every $s \in [\lambda t, t]$, then Eq. (11) and $t^{-1}\sum_{s \leq t} \rho_s \to 0$ imply $\|\bar{q}_t - q^*\|_1 \leq \eta$ for all sufficiently large t. Therefore

$$\mathbb{P}_P(\|\bar{q}_t - q^*\|_1 > \eta) \leq \sum_{s=\lceil \lambda t \rceil}^{t} \mathbb{P}_P(\widehat{q}_{s-1} \notin U) \leq C_\eta e^{-c_\eta t^{1-\gamma}}.$$

Finally, for $c = (m, a)$,

$$D_{c,t} = N_c(t) - \sum_{s=1}^{t} p_m q_{s,c}$$

is a bounded-increment martingale. Standard martingale concentration (Freedman, 1975) gives

$$\mathbb{P}_P(|D_{c,t}| > \eta t) \le 2e^{-c\eta^2 t}.$$

Since

$$\widehat{w}_{c,t} - w_c^* = \frac{D_{c,t}}{t} + p_m(\bar{q}_{t,c} - q_c^*),$$

the claim for $\widehat{w}_t$ follows by a finite union bound. □

Lemma 2 (Positive limiting exposure on comparison cells)**.** *For every $d \in \mathcal{D}_*$ and every $c \in \mathcal{C}(d)$,*

$$q_c^* > 0.$$

Hence

$$\underline{q}_* := \min_{d \in \mathcal{D}_*} \min_{c \in \mathcal{C}(d)} q_c^* > 0.$$

Moreover, for all sufficiently large t,

$$\mathbb{P}_P\left(\exists d \in \mathcal{D}_* : \min_{c=(m,a)\in\mathcal{C}(d)} N_c(t) < \frac{1}{4} \min_{c=(m,a)\in\mathcal{C}(d)} p_m q_c^* \, t\right) \le Ce^{-ct^{1-\gamma}}.$$

Proof. The baseline design $q^{(0)}$ is positive on $\mathcal{C}_{\text{all}}$, so $\mathcal{I}_P(q^{(0)}) > 0$. Fix $d \in \mathcal{D}_*$ and $c \in \mathcal{C}(d)$. If $q_c = 0$, then $\Gamma_q(d) = +\infty$, which makes the information rate for this comparison equal to zero and hence $\mathcal{I}_P(q) = 0$. Since $q^{(0)}$ has positive information, every maximizer must assign positive exposure to each cell in a comparison set $\mathcal{C}(d)$. Thus $q_c^* > 0$ on every such cell. Finiteness of $\mathcal{D}_*$ gives $\underline{q}_* > 0$.

Lemma 1 implies that late designs remain in a sufficiently small neighborhood of q^* except on a stretched-exponentially small event. On the complementary event, $q_{t,c} \ge q_c^*/2$ for every c in every comparison set $\mathcal{C}(d)$. Predictable Bernoulli concentration (Howard et al., 2020) then gives the stated linear lower bound for the corresponding cell counts. □

Lemma 3 (Cellwise RKHS prediction error)**.** *Consider n independent observations from one cell,*

$$Y_i = \theta + r(X_i) + \varepsilon_i, \qquad X_i \sim P_{X,c}, \qquad r \in \mathcal{R}_c, \qquad \varepsilon_i \sim N(0, \sigma_c^2),$$

where the pairs (X_i, ε_i) are identically distributed across i. Let

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad \widehat{r}_n \in \arg\min_{r' \in \mathcal{R}_c} \frac{1}{n}\sum_{i=1}^{n} \{Y_i - \bar{Y}_n - r'(X_i)\}^2,$$

and define $\widehat{\mu}_n(x) = \bar{Y}_n + \widehat{r}_n(x)$ and $\mu(x) = \theta + r(x)$. There is a cell-dependent constant $C_c < \infty$ such that, for every $\beta \in (0, e^{-1})$ and all sufficiently large n,

$$\mathbb{P}\left(\|\widehat{\mu}_n - \mu\|_{L_2(P_{X,c})}^2 > C_c \left\{\sqrt{\frac{\log(1/\beta)}{n}} + \frac{\log(1/\beta)}{n}\right\}\right) \le \beta.$$

Proof. Let Π_c denote expectation under independent draws $X \sim P_{X,c}$ and $\varepsilon \sim N(0, \sigma_c^2)$, and let $\Pi_{c,n}$ denote the corresponding empirical average. Thus,

$$\Pi_c f = \mathbb{E}[f(X, \varepsilon)], \qquad \Pi_{c,n} f = \frac{1}{n}\sum_{i=1}^{n} f(X_i, \varepsilon_i).$$

Write $\delta_\theta = \bar{Y}_n - \theta$ and $h = \widehat{r}_n - r$. Since $r(X)$ is centered and bounded and ε is Gaussian,

$$\mathbb{P}(|\delta_\theta| > u) \le 2e^{-cnu^2}$$

for a cell-dependent constant $c > 0$ in the required range. Both r and $\widehat{r}_n$ belong to $\mathcal{R}_c$, so $\|h\|_{\mathbb{H}_c} \le 2B_c$, $|h(x)| \le 2B_c\kappa_c$, and $\Pi_c h = 0$.

The basic inequality for constrained least squares is

$$\Pi_{c,n} h^2 \le 2\Pi_{c,n}\{(\varepsilon - \delta_\theta)h\}. \tag{22}$$

The RKHS Rademacher bound (Bartlett & Mendelson, 2002, Lemma 22) gives, for Rademacher variables $\xi_1, \dots, \xi_n$,

$$\mathbb{E}_\xi \sup_{\|f\|_{\mathbb{H}_c} \le 2B_c} \frac{1}{n}\sum_{i=1}^n \xi_i f(X_i) \le \frac{2B_c\kappa_c}{\sqrt{n}}.$$

The class has envelope $2B_c\kappa_c$, and the multiplier ε is Gaussian with variance σ_c^2. These bounds provide the uniform envelope and sub-Gaussian multiplier conditions required for the empirical-process inequalities below. Symmetrization and contraction (Bartlett & Mendelson, 2002), together with Gaussian multiplier concentration (Boucheron et al., 2013), therefore give, with probability at least $1 - \beta$ after adjusting constants,

$$\sup_{\|f\|_{\mathbb{H}_c} \le 2B_c} |(\Pi_{c,n} - \Pi_c)(\varepsilon f)| \le C\sqrt{\frac{\log(1/\beta)}{n}},$$

$$\sup_{\|f\|_{\mathbb{H}_c} \le 2B_c} |(\Pi_{c,n} - \Pi_c) f| \le C\sqrt{\frac{\log(1/\beta)}{n}},$$

and

$$\sup_{\|f\|_{\mathbb{H}_c} \le 2B_c} |(\Pi_{c,n} - \Pi_c) f^2| \le C\sqrt{\frac{\log(1/\beta)}{n}}.$$

Using $\Pi_c(\varepsilon h) = 0$ and $\Pi_c h = 0$ in Eq. (22) gives

$$\Pi_c h^2 \le 2|(\Pi_{c,n} - \Pi_c)(\varepsilon h)| + 2|\delta_\theta|\,|(\Pi_{c,n} - \Pi_c)h| + |(\Pi_{c,n} - \Pi_c)h^2|.$$

Combining the three empirical-process bounds with $|\delta_\theta|^2 = O\{\log(1/\beta)/n\}$ yields

$$\Pi_c h^2 \le C_c \left\{ \sqrt{\frac{\log(1/\beta)}{n}} + \frac{\log(1/\beta)}{n} \right\}.$$

Finally, $\Pi_c h = 0$ implies

$$\|\widehat{\mu}_n - \mu\|^2_{L_2(P_{X,c})} = \delta_\theta^2 + \Pi_c h^2,$$

which proves the claim. □

Lemma 4 (Predictive-remainder tail)**.** *For $d \in \mathcal{D}_*$, recall $\mathcal{R}_t^{\mathrm{pred}}(d)$ from Eq.* (19)*. For every $\epsilon > 0$, there exist $C_\epsilon, t_\epsilon < \infty$ such that*

$$\mathbb{P}_P\left(\max_{d \in \mathcal{D}_*} \mathcal{R}_t^{\mathrm{pred}}(d) > \epsilon t \right) \le C_\epsilon t^{-4}, \qquad t \ge t_\epsilon.$$

More precisely, outside an event of probability $O(t^{-6})$,

$$\max_{d \in \mathcal{D}_*} \mathcal{R}_t^{\mathrm{pred}}(d) \le C_P \left\{ \sqrt{t}(\log t)^{3/2} + t^{1/4}(\log t)^{5/4} + (\log t)^2 \right\}$$

for a finite constant C_P.

Proof. Fix a deterministic horizon $T \ge 3$. For cell $c \in \mathcal{C}_{\mathrm{all}}$, use dyadic training sizes $n_j = 2^j \le T$. At size n_j, define

$$e_{c,j}(x) = \widehat{\mu}_{c,n_j}(x) - \mu_c(x), \qquad R_{c,j} = \|e_{c,j}\|^2_{L_2(P_{X,c})}.$$

Apply Lemma 3 to every eligible dyadic fit with confidence level T^{-12}. A union bound over the finite cells and $O(\log T)$ fits gives an event $\mathcal{E}_T^{\mathrm{fit}}$ with

$$\mathbb{P}_P\{(\mathcal{E}_T^{\mathrm{fit}})^c\} \le CT^{-10}$$

on which

$$R_{c,j} \le C_c \left\{ \sqrt{\frac{\log T}{n_j}} + \frac{\log T}{n_j} \right\}$$

for every fit. The same event can be chosen so that $|e_{c,j}(x)| \le B_T := C_P\sqrt{\log T}$ for all relevant c, j, x.

During the arrival block $\{n_j + 1, \dots, 2n_j\}$, the fitted predictor is fixed. Let $\mathscr{H}_{c,n_j}$ denote the sigma-field generated by observations through the n_jth arrival in cell c. Conditional on $\mathscr{H}_{c,n_j}$, future within-cell covariates are fresh draws from $P_{X,c}$ by Eq. (18). For

$$U_{c,j,k} = e_{c,j}(X_{c,n_j+k})^2,$$

we therefore have

$$\mathbb{E}[U_{c,j,k} \mid \mathscr{H}_{c,n_j}] = R_{c,j}, \qquad 0 \le U_{c,j,k} \le B_T^2.$$

Hoeffding's inequality (Hoeffding, 1963) applied to every deterministic prefix of every dyadic block, followed by a union bound, gives an event $\mathcal{E}_T^{\mathrm{prefix}}$ with failure probability $O(T^{-9})$ on which all such prefixes satisfy the corresponding concentration bound. Substituting the realized cell counts only after this simultaneous event has been constructed avoids conditioning on the random terminal count.

Let

$$\delta_s = \widehat{\mu}_{C_s}^{(s-1)}(X_s) - \{\theta_{C_s} + r_{C_s}(X_s)\}.$$

On $\mathcal{E}_T^{\mathrm{fit}} \cap \mathcal{E}_T^{\mathrm{prefix}}$, each complete or partial dyadic block contributes at most

$$C_P\{\sqrt{n_j}(\log T)^{3/2} + \log T\}$$

to the cumulative squared prediction error. Since $\sum_{j:2^j \le N} \sqrt{2^j} \le C\sqrt{N}$,

$$\sum_{s=1}^{T} \delta_s^2 \le C_P\{\sqrt{T}(\log T)^{3/2} + (\log T)^2\}. \tag{23}$$

outside an event of probability $O(T^{-9})$. The finitely many pre-initialization observations are absorbed into the lower-order term on a simultaneous Gaussian-tail event.

Write $\varepsilon_s = Y_s - \theta_{C_s} - r_{C_s}(X_s)$. For $d \in \mathcal{D}_*$, the Gaussian log-loss difference is

$$\mathcal{R}_T^{\mathrm{pred}}(d) = \frac{1}{2} \sum_{s \le T} \mathbf{1}\{C_s \in \mathcal{C}(d)\} \frac{\delta_s^2}{\sigma_{C_s}^2} - M_T(d),$$

where

$$M_T(d) = \sum_{s \le T} \mathbf{1}\{C_s \in \mathcal{C}(d)\} \frac{\varepsilon_s \delta_s}{\sigma_{C_s}^2}.$$

With $\mathcal{G}_s = \sigma(\mathcal{F}_{s-1}, X_s, G_s, A_s)$, the multiplier is $\mathcal{G}_s$-measurable and $\varepsilon_s \mid \mathcal{G}_s \sim N(0, \sigma_{C_s}^2)$. Thus $M_T(d)$ is a Gaussian martingale transform with predictable quadratic variation bounded by a constant multiple of the right-hand side of Eq. (23). The exponential martingale inequality (Howard et al., 2020) and a finite union bound over $\mathcal{D}_*$ give

$$\max_{d \in \mathcal{D}_*} |M_T(d)| \le C_P \left\{ T^{1/4}(\log T)^{5/4} + \log T \right\}$$

outside an event of probability $O(T^{-6})$. Combining the two terms proves the explicit bound. Its right-hand side is $o(T)$, which yields the stated T^{-4} tail for all sufficiently large T. □

Lemma 5 (Null-profile remainder tail)**.** *For $d = d_z \in \mathcal{D}_*$, recall $\mathcal{R}_t^0(d)$ from Eq.* (19)*. For every $\epsilon > 0$, there exist $C_\epsilon, t_\epsilon < \infty$ such that*

$$\mathbb{P}_P\left(\max_{d \in \mathcal{D}_*} \mathcal{R}_t^0(d) > \epsilon t \right) \le C_\epsilon t^{-4}, \qquad t \ge t_\epsilon.$$

Proof. Fix a deterministic horizon T and $d \in \mathcal{D}_*$. For nonnegative weights $w = (w_c)_{c\in\mathcal{C}(d)}$, define

$$\ell_{w,d}(Q) = \sum_{c\in\mathcal{C}(d)} w_c \mathbb{E}_{P_c}[\log p_Q(Y \mid X, c)], \qquad w_c^* = p_m q_c^*.$$

For $Q = (\theta', r')$, the Q-dependent part of $\ell_{w^*,d}(Q)$ is

$$-\frac{1}{2}\sum_{c\in\mathcal{C}(d)} \frac{w_c^*}{\sigma_c^2}\mathbb{E}\big[\{(\theta_c - \theta_c') + (r_c - r_c')\}^2 \mid G = m\big].$$

Centering separates the mean and prognostic terms. The population null likelihood is therefore maximized by $r' = r$ together with the weighted projection from Appendix B.1. Hence

$$Q_d^* \in \arg\max_{Q\in H_{0,d}} \ell_{w^*,d}(Q). \tag{24}$$

Lemma 2 gives a linear number of observations in every cell of $\mathcal{C}(d)$ except on an event of probability $Ce^{-cT^{1-\gamma}}$. Cell-arrival outcomes have sub-Gaussian tails, so their squared values are sub-exponential. Uniform concentration over deterministic cell-prefix lengths between a fixed linear fraction of T and T gives constants $a_{P,d}, K_{P,d} > 0$ such that

$$N_c(T) \geq a_{P,d}T, \qquad \frac{1}{N_c(T)}\sum_{s\leq T: C_s = c} Y_s^2 \leq K_{P,d}$$

for every $c \in \mathcal{C}(d)$ except on an event of probability at most $Ce^{-cT^{1-\gamma}}$.

For $Q = (\theta', r')$, write the negative pairwise Gaussian log likelihood up to a Q-independent constant as

$$\mathcal{S}_T(Q) = \sum_{c\in\mathcal{C}(d)} \frac{1}{2\sigma_c^2}\sum_{s\leq T: C_s = c} \{Y_s - \theta_c' - r_c'(X_s)\}^2.$$

Let $\widehat{Q}_T = (\widehat{\theta}_T', \widehat{r}_T')$ be an empirical maximizer of the null likelihood. Then $\mathcal{S}_T(\widehat{Q}_T) \leq \mathcal{S}_T(Q_d^*) \leq C_{P,d}T$. Since $|r_c'(x)| \leq B_c\kappa_c$ and $(u-v)^2 \geq u^2/2 - v^2$,

$$\mathcal{S}_T(\widehat{Q}_T) \geq \frac{N_c(T)}{2\overline{\sigma}^2}\left\{\frac{(\widehat{\theta}_{T,c}')^2}{2} - 2K_{P,d} - 2B_c^2\kappa_c^2\right\}.$$

The linear count lower bound therefore implies $|\widehat{\theta}_{T,c}'| \leq R_{P,d}$ for a deterministic constant $R_{P,d}$. Enlarge $R_{P,d}$ so that Q_d^* also satisfies this bound and define

$$H_{0,d}^R = \left\{Q = (\theta', r') \in H_{0,d} : \max_{c\in\mathcal{C}(d)} |\theta_c'| \leq R_{P,d}\right\}. \tag{25}$$

There is an event $\mathcal{E}_{\mathrm{loc}}(T, d)$ with

$$\mathbb{P}_P\{\mathcal{E}_{\mathrm{loc}}(T,d)^c\} \leq Ce^{-cT^{1-\gamma}} \tag{26}$$

such that

$$Q_d^* \in H_{0,d}^R, \qquad L_{0,T}^{\sup}(d) = \sup_{Q\in H_{0,d}^R} L_T^{(d)}(Q). \tag{27}$$

For $c \in \mathcal{C}(d)$, let $P_{c,n}$ denote the empirical measure of the first n joint cell observations $(X_{c,i}, Y_{c,i})$, and let

$$\mathcal{M}_{c,P,d} = \{x \mapsto \theta_c' + r_c'(x) : |\theta_c'| \leq R_{P,d},\ r_c' \in \mathcal{R}_c\}.$$

This class is uniformly bounded because $|\theta_c'| \leq R_{P,d}$ and $|r_c'(x)| \leq B_c\kappa_c$. After removing the common Y^2 term, the Gaussian criterion is proportional to $\psi_f(X,Y) = 2Yf(X) - f(X)^2$. The term $f(X)^2$ has a bounded envelope, while $Yf(X)$ is a uniformly bounded multiple of the sub-Gaussian outcome Y. For deterministic cell-prefix length n, the bounded-RKHS Rademacher bound and contraction (Bartlett & Mendelson, 2002), together with sub-Gaussian multiplier concentration (Boucheron et al., 2013), give

$$\mathbb{P}_P\left(\sup_{f\in\mathcal{M}_{c,P,d}} |(P_{c,n} - P_c)\psi_f| > C_{c,P,d}\left\{\sqrt{\frac{\log(1/\beta)}{n}} + \frac{\log(1/\beta)}{n}\right\}\right) \leq \beta. \tag{28}$$

The common Y^2 term satisfies the same form of deviation bound by sub-exponential concentration (Boucheron et al., 2013). Set $\beta_T = T^{-11}$ and apply Eq. (28) for every $n \le T$, every relevant cell, and every $d \in \mathcal{D}_*$. A finite union bound gives total failure probability $O(T^{-9})$. On the complementary event, the realized cell counts can be substituted into the simultaneous deterministic-prefix bounds. With $\widehat{w}_{c,T} = N_c(T)/T$,

$$\sup_{Q \in H^R_{0,d}} \left| T^{-1} \log L_T^{(d)}(Q) - \ell_{\widehat{w}_T,d}(Q) \right| \le C_{P,d} \left\{ \sqrt{\frac{\log T}{T}} + \frac{\log T}{T} \right\}. \tag{29}$$

The population criterion is uniformly bounded on $H^R_{0,d}$, so

$$|\ell_{w,d}(Q) - \ell_{w',d}(Q)| \le C_{P,d} \|w - w'\|_1.$$

Lemma 1 gives $\widehat{w}_T \to w^*$ with stretched-exponential tails. Therefore, for every fixed $\epsilon > 0$ and all sufficiently large T,

$$\sup_{Q \in H^R_{0,d}} \left| T^{-1} \log L_T^{(d)}(Q) - \ell_{w^*,d}(Q) \right| \le \epsilon/2 \tag{30}$$

outside an event of probability $O(T^{-9}) + Ce^{-cT^{1-\gamma}}$.

On $\mathcal{E}_{\mathrm{loc}}(T,d)$, Eqs. (27) and (24) give

$$\begin{aligned} \frac{\mathcal{R}_T^0(d)}{T} &= \sup_{Q \in H^R_{0,d}} T^{-1} \log L_T^{(d)}(Q) - T^{-1} \log L_T^{(d)}(Q_d^*) \\ &\le 2 \sup_{Q \in H^R_{0,d}} \left| T^{-1} \log L_T^{(d)}(Q) - \ell_{w^*,d}(Q) \right|. \end{aligned}$$

Eq. (30) implies $\mathcal{R}_T^0(d) \le \epsilon T$. A finite union bound over $d \in \mathcal{D}_*$ gives the stated T^{-4} tail after adjusting constants. □

Lemma 6 (Uniform pairwise evidence growth)**.** *For $d = d_z \in \mathcal{D}_*$, define*

$$I_d^* = \frac{\Delta_z^2}{2\Gamma_{q^*}(d_z)}.$$

Then $\min_{d \in \mathcal{D}_*} I_d^* = [T_{\mathrm{F}}^*(P)]^{-1}$. For every $\epsilon > 0$, there exist $C_\epsilon, t_\epsilon < \infty$ such that

$$\mathbb{P}_P\left(\min_{d \in \mathcal{D}_*} \log \widetilde{E}_t(d) < t\{[T_{\mathrm{F}}^*(P)]^{-1} - \epsilon\} \right) \le C_\epsilon t^{-3}, \qquad t \ge t_\epsilon.$$

Proof. Fix $d \in \mathcal{D}_*$ and let Q_d^* be the least-favorable null under q^*. Define

$$\mathcal{R}_t^{\mathrm{comp}}(d) = \log U_t(d) - \log L_{0,t}^{\sup}(d), \qquad 0 \le \mathcal{R}_t^{\mathrm{comp}}(d) \le \zeta_t \le C_{\mathrm{opt}} \log(t+1).$$

The implemented evidence satisfies

$$\log \widetilde{E}_t(d) = \log \frac{L_t^{(d)}(P)}{L_t^{(d)}(Q_d^*)} - \mathcal{R}_t^{\mathrm{pred}}(d) - \mathcal{R}_t^0(d) - \mathcal{R}_t^{\mathrm{comp}}(d). \tag{31}$$

Lemmas 4 and 5 control the first two remainder terms with $O(t^{-4})$ tails. The computational term is deterministically $o(t)$.

By the least-favorable construction in Appendix B.1, the mean shift from P to Q_d^* is supported only on $\mathcal{C}(d)$, so P and Q_d^* agree outside $\mathcal{C}(d)$. The pairwise likelihood ratio in Eq. (31) is therefore the full likelihood ratio for this alternative. For any design q, define the pair-restricted one-round KL divergence by

$$D_{\mathrm{KL}}^{q,(d)}(P\|Q) := \sum_{c=(m,a) \in \mathcal{C}(d)} p_m q_c \, \mathbb{E}_{P_c}\left[\log \frac{p_P(Y \mid X, c)}{p_Q(Y \mid X, c)} \right].$$

Its predictable drift is

$$\sum_{s \le t} D_{\mathrm{KL}}^{q_s,(d)}(P\|Q_d^*),$$

and linearity in the design gives

$$\frac{1}{t}\sum_{s\le t} D_{\mathrm{KL}}^{q_s,(d)}(P\|Q_d^*) = D_{\mathrm{KL}}^{\bar{q}_t,(d)}(P\|Q_d^*).$$

Lemma 1 gives $\bar{q}_t \to q^*$ with stretched-exponential tails, so the drift is at least $t(I_d^* - \epsilon)$ outside such an event for all sufficiently large t. The centered log-likelihood ratio is a martingale with uniformly bounded conditional sub-exponential moments and predictable quadratic variation of order t. A martingale Bernstein inequality (Howard et al., 2020) yields

$$\mathbb{P}_P\left(\log\frac{L_t^{(d)}(P)}{L_t^{(d)}(Q_d^*)} < t(I_d^* - 2\epsilon)\right) \le C_\epsilon e^{-c_\epsilon t} + C_\epsilon e^{-c_\epsilon t^{1-\gamma}}.$$

A finite union bound over $d \in \mathcal{D}_*$ and the two remainder lemmas proves the result after absorbing fixed constants into ϵ. The identity $\min_d I_d^* = \mathcal{I}_P(q^*) = [T_{\mathrm{F}}^*(P)]^{-1}$ follows from Eq. (9). □

Lemma 7 (Almost-sure stopping under a unique allocation optimum)**.** *For every $P \in \mathcal{P}$ with a unique optimal allocation z^* and every fixed $\delta \in (0,1)$, the stopping rule in Section 4.2 satisfies*

$$\mathbb{P}_P(\tau_\delta < \infty) = 1.$$

Proof. Let $\mathcal{I}_P^* = \max_{q\in\mathcal{Q}} \mathcal{I}_P(q)$. The baseline design is positive on every cell in $\mathcal{C}_{\mathrm{all}}$ and $\mathcal{Z}$ is finite, so uniqueness of z^* gives $\mathcal{I}_P^* > 0$. The cell-count and cell-mean argument in Lemma 1 does not use uniqueness of q^*. By Borel-Cantelli, $\widehat{\theta}_t \to \theta$ and $\widetilde{z}_t = z^*$ eventually almost surely. The same uniform-convergence argument used in that proof gives

$$\sup_{q\in\mathcal{Q}} |\widehat{\mathcal{I}}_t(q) - \mathcal{I}_P(q)| \to 0 \qquad \text{almost surely.}$$

Hence $\mathcal{I}_P(\widehat{q}_t) \ge \mathcal{I}_P^*/2$ eventually almost surely. For $d = z^* - z$ and $c = (m,a) \in \mathcal{C}(d)$,

$$\frac{\Delta_z^2}{2\Gamma_{\widehat{q}_t}(d)} \ge \mathcal{I}_P(\widehat{q}_t) \ge \frac{\mathcal{I}_P^*}{2}$$

eventually. Therefore

$$\widehat{q}_{t,c} \ge \frac{\mathcal{I}_P^* d_c^2 \sigma_c^2}{p_m \Delta_z^2} > 0.$$

Since $\rho_t \to 0$, every cell in every comparison set $\mathcal{C}(d)$ receives a probability bounded away from zero for all sufficiently large t. A martingale strong law for the realized cell indicators then gives constants $a_{P,d,c} > 0$ such that

$$\liminf_{t\to\infty} \frac{N_c(t)}{t} \ge a_{P,d,c} \qquad \text{almost surely.}$$

Fix $d = z^* - z$. Write $w_{c,t} = N_c(t)/t$ on $\mathcal{C}(d)$. For nonnegative weights w, define $\ell_{w,d}$ as in the proof of Lemma 5. Centering of the prognostic functions gives

$$\ell_{w,d}(P) - \sup_{Q\in H_{0,d}} \ell_{w,d}(Q) = \frac{\Delta_z^2}{2\sum_{c\in\mathcal{C}(d)} d_c^2\sigma_c^2/w_c}.$$

The linear count bound places w_t eventually in a compact set on which every component is bounded away from zero. The last display is therefore bounded below by a constant $\kappa_{P,d} > 0$. Using any fixed null model with $r' = r$ and $d^\top\theta' = 0$ gives the localization step in the proof of Lemma 5. The uniform empirical-process argument leading to Eq. (29), with the localization radius enlarged to contain the true mean vector, then yields

$$\frac{1}{t}\left\{\log L_t^{(d)}(P) - \log L_{0,t}^{\sup}(d)\right\} \ge \kappa_{P,d} - o(1) \qquad \text{almost surely.}$$

The tail bound in Lemma 4 and Borel-Cantelli give $\mathcal{R}_t^{\mathrm{pred}}(d) = o(t)$ from above almost surely. The certified profile gap satisfies $0 \le \mathcal{R}_t^{\mathrm{comp}}(d) \le C_{\mathrm{opt}}\log(t+1) = o(t)$. Consequently,

$$\log\widetilde{E}_t(d) = \log L_t^{(d)}(P) - \log L_{0,t}^{\sup}(d) - \mathcal{R}_t^{\mathrm{pred}}(d) - \mathcal{R}_t^{\mathrm{comp}}(d) \to +\infty$$

almost surely. There are finitely many competitors, so the evidence for z^* crosses every fixed pairwise threshold in finite time almost surely. Hence $\tau_\delta < \infty$ almost surely. □

C.2 Proof of Theorem 2

Proof. For finite-sample correctness, fix any $P \in \mathcal{P}$ with a unique optimal allocation z^*. Lemma 7 gives $\mathbb{P}_P(\tau_\delta < \infty) = 1$. Suppose the procedure outputs $z \neq z^*$. The stopping rule then requires

$$\widetilde{E}_t(z - z^*) \geq 1/\alpha_{z,z^*}$$

at some time t. Since $(z - z^*)^\top \theta < 0$, the true model belongs to $H_{0,z-z^*}$. Eq. (13) and the conservative profile computation imply

$$\mathbb{P}_P\left(\sup_t \widetilde{E}_t(z - z^*) \geq 1/\alpha_{z,z^*}\right) \leq \alpha_{z,z^*}.$$

Summing over $z \neq z^*$ gives

$$\mathbb{P}_P(\widehat{z}_{\tau_\delta} \neq z^*) \leq \sum_{z \neq z^*} \alpha_{z,z^*} \leq \delta.$$

Combining this bound with almost-sure stopping gives $\mathbb{P}_P(\widehat{z}_{\tau_\delta} = z^*) \geq 1 - \delta$. Because P was arbitrary, the procedure is uniformly δ-correct on $\mathcal{P}$. The design constraint $q_t \in \mathcal{Q}$ holds pathwise because Eq. (11) is a convex combination of feasible designs.

For the expected stopping time, fix $P \in \mathcal{P}$ with a unique optimal allocation z^* and a unique maximizer q^* in Eq. (9). Let $L_\delta = \log(1/\delta)$ and fix $\eta > 0$. Set

$$t_\delta = (1 + \eta) T^*_{\mathrm{F}}(P) L_\delta.$$

Since J is fixed,

$$\log(1/\alpha_{z,z'}) = L_\delta + O(1)$$

for every ordered pair. Choose $\epsilon > 0$ such that

$$(1 + \eta) T^*_{\mathrm{F}}(P) \{[T^*_{\mathrm{F}}(P)]^{-1} - \epsilon\} > 1.$$

For all sufficiently small δ, Lemma 6 gives

$$\mathbb{P}_P(\tau_\delta > t) \leq C_\epsilon t^{-3}, \qquad t \geq t_\delta.$$

Therefore

$$\mathbb{E}_P[(\tau_\delta - t_\delta)_+] = \sum_{s \geq t_\delta} \mathbb{P}_P(\tau_\delta > s) \leq C_\epsilon \sum_{s \geq t_\delta} s^{-3} = o(L_\delta).$$

It follows that

$$\limsup_{\delta \downarrow 0} \frac{\mathbb{E}_P[\tau_\delta]}{T^*_{\mathrm{F}}(P) \log(1/\delta)} \leq 1 + \eta.$$

Letting $\eta \downarrow 0$ gives the upper bound 1. Theorem 1 supplies the matching lower bound, which proves

$$\frac{\mathbb{E}_P[\tau_\delta]}{T^*_{\mathrm{F}}(P) \log(1/\delta)} \to 1.$$

□

C.3 Deterministic implementation checkpoints

The experiments update the design at deterministic refresh times. Let $s(t) \leq t$ be the latest refresh time and assume $s(t) \geq c_0 t$ eventually for some $c_0 > 0$. Replacing $\widehat{q}_t$ in Eq. (11) by the optimizer computed from observations through $s(t)$ preserves predictability and feasibility. Lemma 1 applied at $s(t)$ gives

$$\mathbb{P}_P(\widehat{q}_{s(t)} \notin U) \leq C_U e^{-c_U s(t)^{1-\gamma}} \leq C_U e^{-c'_U t^{1-\gamma}}.$$

The average-design, cell-count, and likelihood arguments therefore retain the same rates. A deterministic multiplicative refresh schedule with a fixed factor satisfies the required lower bound on $s(t)$.

Suppose evidence is evaluated at deterministic times t_k with gaps bounded by a fixed integer b, and let $\tau^{\mathrm{chk}}_\delta$ denote the resulting stopping time. Restricting the stopping rule to these times preserves Eq. (13). The evidence-growth bound applies at every checkpoint. For large t, the preceding checkpoint is at least $t - b$, so

$$\mathbb{P}_P(\tau^{\mathrm{chk}}_\delta > t) \leq C(t - b)^{-3}$$

after the checkpoint exceeds the asymptotic threshold used above. Summing this tail adds only $O(b)$ to the expected-time bound. Hence the checkpointed implementation has the same first-order expected stopping time as the every-round procedure.

D Experimental Details and Additional Results

This appendix gives the data-generating process, the numerical computation of the coarsened characteristic time, the sequential implementation, and the complete results for Section 5.

D.1 Experimental Data-Generating Process

There are two groups with $p_1 = p_2 = 1/2$ and two treatments. The deployment decision treats exactly one group. In the cell order $((1,0),(1,1),(2,0),(2,1))$, we use

$$\theta = (0, \Delta, \Delta, 0), \qquad \Delta = 0.2.$$

Conditional on group, U is uniform on $\{-1, +1\}$ and

$$Y = \theta_c + \nu U + \varepsilon, \qquad \varepsilon \sim N(0,1), \qquad \nu = 1.$$

This is the rank-one centered RKHS model used in the two-group construction of Appendix A.1, with $\widetilde{k}_c(u,u') = uu'$ and $B_c = B = 2$ in every cell. The experimental treatment probabilities satisfy $q_{m,1} \in [0.1, 0.9]$. With $z^{(1)} = (0, 1/2, 1/2, 0)$ and $z^{(2)} = (1/2, 0, 0, 1/2)$, let $d = z^{(1)} - z^{(2)}$. Then $d^\top \theta = \Delta = 0.2$ and

$$\Gamma_q(d) = \frac{1}{2} \sum_{m=1}^{2} \left\{ \frac{1}{q_{m,1}} + \frac{1}{1 - q_{m,1}} \right\}.$$

The minimum is attained at $q_{1,1} = q_{2,1} = 1/2$, which gives

$$T_{\mathrm{F}}^* = \frac{8}{\Delta^2} = 200.$$

D.2 Coarsened Characteristic Time

After the within-group covariate U is marginalized out, the outcome distribution in one cell is the symmetric Gaussian mixture

$$f_{\theta,\beta}(y) = \frac{1}{2}\varphi_1(y - \theta - \beta) + \frac{1}{2}\varphi_1(y - \theta + \beta),$$

where φ_1 is the standard normal density and $|\beta| \le B$. For a fixed design q, we compute the coarsened information rate by minimizing the marginal KL divergence over cell means satisfying the competing-allocation constraint and all admissible rank-one prognostic coefficients. The optimization problem is

$$I_{\mathrm{C}}(q; P) = \inf_{\substack{d^\top \theta' \le 0 \\ |\beta'_c| \le B}} \sum_c p_m q_c D_{\mathrm{KL}}\big(f_{\theta_c,\nu} \,\big\|\, f_{\theta'_c,\beta'_c}\big).$$

The one-dimensional mixture KL terms are evaluated by Gauss-Hermite quadrature using NumPy's `hermgauss` (Harris et al., 2020). We evaluate the optimized information rate on a 9×9 grid over $(q_{1,1}, q_{2,1}) \in [0.1, 0.9]^2$ and refine the best design with SciPy's sequential least-squares programming routine (Virtanen et al., 2020). At each grid point the inner constrained problem is initialized from three feasible points. At the selected design, we repeat the inner optimization from 64 dispersed feasible initializations using seed 20260925. All 64 runs agree within 2.0×10^{-12} in information rate, and the largest grid value occurs at $q_{1,1} = q_{2,1} = 1/2$.

The refined numerical solution is $q_{1,1} = q_{2,1} = 1/2$. Writing $\widehat{I}_{\mathrm{C}}^*$ for the optimized numerical information rate gives

$$\widehat{I}_{\mathrm{C}}^* \approx 0.0027476075, \qquad \widehat{T}_{\mathrm{C}}^* = (\widehat{I}_{\mathrm{C}}^*)^{-1} \approx 363.9530.$$

The returned boundary solution has cell means approximately $(0.1, 0.1, 0.1, 0.1)$ and prognostic coefficient magnitudes approximately 1.00362 in all four cells.

D.3 Sequential Procedures and Monte Carlo Protocol

For the full-covariate procedures, observations in cell c are indexed by arrival order. The prognostic predictor is updated when the local cell count reaches $1, 2, 4, \ldots$. At each update, the cell intercept is the current sample mean and the rank-one coefficient is the least-squares coefficient clipped to $[-B, B]$. Each new outcome is scored before it enters the next fit, so the predictive likelihood remains predictable.

For the unknown-r null profile, the finite-sample optimization in this rank-one model is

$$\min_{\{\theta'_c, \beta'_c\}} \frac{1}{2} \sum_c \sum_{i: C_i = c} \{Y_i - \theta'_c - \beta'_c U_i\}^2$$

subject to $d^\top \theta' \le 0$ and $|\beta'_c| \le B$. The implementation solves the Lagrange multiplier for the active half-space constraint by scalar root finding and evaluates the stopping statistic with the resulting dual lower bound on the null loss. A feasible projected point gives the corresponding primal upper bound. Appendix F gives the resulting rank-one calculation. The Known r benchmark fixes $\beta'_c = \beta_c = \nu$ in both the predictive likelihood and the null profile. The Oracle predictor uses the true $\beta_c = \nu$ only in the predictive likelihood and retains the unknown-r null profile.

The coarsened benchmark uses only (G, A, Y) in its evidence. It accumulates the log likelihood ratio between the true marginal mixture law and the boundary alternative returned by the optimization in Appendix D.2. It uses the selected design $q_{1,1} = q_{2,1} = 1/2$. The benchmark uses the true coarsened law and the numerically selected boundary alternative, so we use its threshold-crossing time only as a numerical reference for the coarsened information rate.

All four procedures use $q_{1,1} = q_{2,1} = 1/2$. For the full-covariate procedures this equals the design path generated by Eq. (10) when $q^{(0)} = q^*$, as shown in Section 5.1. Pairwise evidence is evaluated every 200 arrivals. With two feasible allocations, the threshold is

$$L_\delta + \log 2.$$

We use 100 replications with base seed 20260915 and administrative horizon 1.5×10^6. All full-covariate trajectories stop at every reported confidence level, and every coarsened benchmark trajectory crosses its threshold before the horizon. For paired ratios of sample means, the MCSE is computed by the paired delta method (van der Vaart, 1998, Chapter 3).

D.4 Complete Results

Table 2: Complete results for the experimental instance. The three full-covariate columns report stopping times. The coarsened benchmark column reports threshold-crossing times under the numerical boundary alternative. Times and MCSEs are in thousands of arrivals. The normalized columns use T^*_{F} and $\widehat{T}^*_{\mathrm{C}}$, respectively. The final column is the Learned r/coarsened benchmark ratio with paired MCSE in parentheses. Normalized quantities and ratios are computed from the unrounded replication means.

L_δ	Known r	Oracle predictor	Learned r	Coarsened benchmark	Learned r: $\frac{\mathbb{E}[\tau]}{T^*_{\mathrm{F}} L_\delta}$	Coarsened: $\frac{\mathbb{E}[\tau]}{\widehat{T}^*_{\mathrm{C}} L_\delta}$	$\frac{\text{Learned } r}{\text{Coarsened}}$
20	14.37 (0.41)	14.81 (0.41)	21.00 (0.45)	7.62 (0.23)	5.251	1.046	2.758 (0.081)
50	20.74 (0.42)	21.13 (0.43)	27.12 (0.47)	19.03 (0.39)	2.712	1.046	1.426 (0.027)
100	30.92 (0.46)	31.32 (0.47)	37.93 (0.52)	36.68 (0.52)	1.897	1.008	1.034 (0.015)
200	51.77 (0.54)	52.13 (0.54)	58.51 (0.59)	72.51 (0.70)	1.463	0.996	0.807 (0.008)
500	112.89 (0.68)	113.35 (0.67)	120.28 (0.68)	183.13 (1.15)	1.203	1.006	0.657 (0.004)
1000	213.84 (0.88)	214.24 (0.88)	221.31 (0.91)	364.28 (1.56)	1.107	1.001	0.608 (0.003)
2000	414.43 (1.16)	414.88 (1.16)	422.17 (1.16)	728.22 (1.98)	1.055	1.000	0.580 (0.002)

Table 3: Finite-sample adaptation-cost decomposition. Entries are ratios of mean stopping times computed from paired replications, with paired delta-method MCSEs in parentheses.

L_δ	Learned r/Known r	Learned r/Oracle predictor	Oracle predictor/Known r
20	1.4615 (0.0203)	1.4183 (0.0192)	1.0305 (0.0036)
50	1.3078 (0.0118)	1.2837 (0.0114)	1.0188 (0.0021)
100	1.2267 (0.0092)	1.2110 (0.0090)	1.0129 (0.0015)
200	1.1302 (0.0043)	1.1223 (0.0043)	1.0070 (0.0007)
500	1.0655 (0.0023)	1.0612 (0.0022)	1.0041 (0.0005)
1000	1.0349 (0.0010)	1.0330 (0.0010)	1.0019 (0.0003)
2000	1.0187 (0.0005)	1.0176 (0.0006)	1.0011 (0.0001)

The Learned r/coarsened benchmark ratios in the final column are plotted in Figure 1(c). Across the seven confidence levels, the ratio decreases toward the numerical first-order value $T^*_{\mathrm{F}}/\widehat{T}^*_{\mathrm{C}} = 0.5495$ reported in the main text.

E Computational Details for Pairwise Profiles

Fix an allocation comparison d and recall the centered kernel from Section 2,

$$\widetilde{k}_c(x, x') = \langle \widetilde{\phi}_c(x), \widetilde{\phi}_c(x') \rangle_{\mathbb{V}_c}.$$

For cell $c \in \mathcal{C}(d)$, let $X_c = (X_{c,1}, \ldots, X_{c,N_c(t)})$, let $y_c \in \mathbb{R}^{N_c(t)}$ be the corresponding outcome vector, let $\mathbf{1}_c \in \mathbb{R}^{N_c(t)}$ be the all-ones vector, and let $K_{c,t}$ be the centered Gram matrix. The Gaussian null profile minimizes

$$\sum_{c \in \mathcal{C}(d)} \frac{1}{2\sigma_c^2} \|y_c - \theta'_c \mathbf{1}_c - r'_c(X_c)\|_2^2$$

subject to $d^\top \theta' \le 0$ and $\|r'_c\|_{\mathbb{H}_c} \le B_c$. By the representer theorem (Schölkopf et al., 2001), an optimizer has the form

$$r'_c(\cdot) = \sum_{i=1}^{N_c(t)} a_{c,i} \widetilde{k}_c(X_{c,i}, \cdot).$$

The exact profile is therefore

$$\begin{aligned} \min_{\{\theta'_c, a_c : c \in \mathcal{C}(d)\}} \quad & \sum_{c \in \mathcal{C}(d)} \frac{1}{2\sigma_c^2} \|y_c - \theta'_c \mathbf{1}_c - K_{c,t} a_c\|_2^2 \\ \text{s.t.} \quad & d^\top \theta' \le 0, \qquad a_c^\top K_{c,t} a_c \le B_c^2 \quad \forall c \in \mathcal{C}(d). \end{aligned} \tag{32}$$

The objective is convex quadratic and the RKHS constraints are convex because each $K_{c,t}$ is positive semidefinite. Its optimal objective value is $\mathcal{L}^*_{0,t}(d)$ from Section 4.2, and

$$L^{\sup}_{0,t}(d) = \mathcal{N}_t(d) \exp\{-\mathcal{L}^*_{0,t}(d)\}.$$

A dual-feasible value lower-bounds the optimal loss, which gives an upper bound on the composite-null likelihood after exponentiation. A primal-feasible value upper-bounds the optimal loss (Boyd & Vandenberghe, 2004). Their difference gives an observable profile-optimization gap. Substituting the dual lower bound into the likelihood denominator yields Eq. (14). The condition $\zeta_t \le C_{\mathrm{opt}} \log(t+1)$ makes the resulting loss in log evidence $o(t)$.

F Rank-One Profile Computation for the Experiment

The main experiments use a rank-one specialization of the profile problem in Appendix E. Pairwise evidence is evaluated only after every cell in $\mathcal{C}_{\mathrm{all}}$ has at least one observation. For cell c, let n_c be its observation count, $\bar{Y}_c$ its sample mean, and $S_{U,c} = \sum_{i:C_i=c} U_i$. Write θ'_c and β'_c for the null cell

mean and null prognostic coefficient. For a Lagrange multiplier $\lambda \geq 0$ associated with $d^\top \theta' \leq 0$, the coefficient $\beta'_c(\lambda)$ minimizes the cellwise Lagrangian over $[-B, B]$, and

$$\theta'_c(\lambda) = \bar{Y}_c - \frac{S_{U,c}}{n_c}\beta'_c(\lambda) - \frac{\lambda d_c}{n_c}.$$

The map $\lambda \mapsto d^\top \theta'(\lambda)$ is nonincreasing. If the unconstrained solution violates the null, the implementation solves $d^\top \theta'(\lambda) = 0$ by Brent's method (Brent, 1973), implemented with SciPy's `brentq` (Virtanen et al., 2020). For the returned λ, let $\mathcal{L}(\lambda)$ be the quadratic loss. The dual bound is

$$\underline{\mathcal{L}}_{0,t}(d) = \mathcal{L}(\lambda) + \lambda d^\top \theta'(\lambda).$$

A feasible projection of $\theta'(\lambda)$ onto $d^\top \theta' \leq 0$ gives the primal upper bound $\overline{\mathcal{L}}_{0,t}(d)$. The stopping statistic uses $\underline{\mathcal{L}}_{0,t}(d)$, so the null-likelihood denominator is conservative as required by Eq. (14).

For Known r, fixing $\beta'_c = \beta_c = \nu$ reduces the null profile to the weighted Euclidean projection of the adjusted cell means onto $d^\top \theta' \leq 0$. These formulas give the rank-one implementation used in the experiment. The general RKHS procedure is defined in Section 4 and Appendix E.