CERTIPLAN: DECISION-CLOSED SIGN RELEASE IN PRICED, FAILURE-PRONE INTERVENTION LIBRARIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Budgeted causal sign release has three non-interchangeable outcomes: capped cost to a posterior-risk threshold on a common shadow trajectory, sign error at a fixed budget, and campaign-level incidence of a wrong voluntary release. A policy can improve one while missing another when graph information guides acquisition, entropy governs stopping, and interventions have unequal prices and observable failures. CertiPlan aligns the three through a decision-closed posterior. Predictive ancestral closure preserves the coordinates needed for every future library outcome; finite-tree dynamic programming supplies non-myopic cost–risk targets; and an amortized action policy and ESS-adjusted release score read the same posterior labels. Across linear-Gaussian, sparse-additive, and Sachs libraries, CertiPlan reduces capped shadow cost-to-ϵ by 23.8–27.8%, fixed-budget sign error by 31.3–43.2%, and harmful voluntary-release incidence by 40.8–43.5% relative to mutual-information BOED. These comparisons use identical architectures, training exposure, tuning budgets, compute, and 2,000 held-out episodes per family. Component swaps separate the mechanisms: mutual-information acquisition raises cost by 30.8–37.8%, whereas entropy stopping approximately doubles harmful-release incidence while preserving acquisition cost within uncertainty. Cost reductions reach 33.1–38.9% for local queries with dispersed prices; target-disjoint Perturb-seq replay retains 21.7% lower cost and 26.0% lower sign error. The results show why acquisition efficiency and release behavior must be assessed through the terminal causal-decision posterior rather than a graph-wide surrogate.

1 INTRODUCTION

A causal sign-release campaign ultimately reports one bit: whether $\mathbb{E}[Y \mid \mathrm{do}(A)] - \mathbb{E}[Y \mid \mathrm{do}(\emptyset)]$ is positive beyond an indifference margin. The cost of reaching a posterior-risk threshold, error after a fixed budget, and the event of voluntarily releasing the wrong bit answer different scientific questions. Reporting them separately is essential: an acquisition policy determines how quickly evidence accumulates, while a stopping functional determines which histories terminate.

Prices and observable assay failures make this separation causal and sequential. After choosing I, a failed assay records $(F = 1, D = \varnothing)$ and spends c_I. Its likelihood reweights the reliability coordinate through ρ_I; under the independent failure prior evaluated here, it leaves the exact sign posterior unchanged. A successful batch instead reweights the causal-effect coordinates. Both branches still change continuation value because the next action is chosen with less budget and updated knowledge of assay reliability. CertiPlan retains exactly the coordinates needed to carry this transition into both future acquisition and release.

The key gap extends beyond non-myopic planning. Mutual-information BOED can prefer graph distinctions that leave the sign unchanged; active Bayesian causal inference and GO-CBED target causal-query information rather than 0–1 release loss (Toth et al., 2022; Zhang et al., 2025); Action-BED optimizes downstream loss without an explicit causal release state (Rossa et al., 2026); and optimal-stopping BED learns a generic continuation-value boundary (Cheng & Huan, 2025). The missing interface is a posterior closed under every priced library outcome whose terminal labels directly govern both decision-risk acquisition and release.

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.

CertiPlan makes the terminal decision the state interface. Every particle retains the intervention-effect vector and the predictive ancestral topology, mechanisms, noise variables, and failure propensity needed by the library observation map. This decision closure makes future outcomes, terminal Bayes risk, action value, and the released label measurable from one posterior. On tractable libraries, dynamic programming over a sampled finite outcome tree provides non-myopic teachers; imitation and PPO amortize their cost–risk policy, while release remains an explicit posterior-label computation. This construction is technically different from replacing mutual information by a one-step value-of-information score: it specifies the retained causal state, its two-branch transition, the multi-step objective, and the release functional together.

The technical contribution is a hybrid finite-budget decision process: paid continuation, the finite-tree teacher, PPO rewards, and budget-exhaustion loss share Eq. 1, while Eq. 3 imposes a transparent release rule from the same decision labels rather than a learned stop logit. Its predictively closed particle state supports both observation branches and the terminal library decision; Proposition 1 establishes exact posterior sufficiency for graph-SCM particles, and Lemma 1 bounds one-step risk and cost-normalized score perturbations under retained-posterior error.

The empirical contribution tests two mechanism predictions. Decision targeting should reduce cost most when the sign depends on a small causal subspace and prices are heterogeneous; reading the release score from the decision posterior should alter harmful-release incidence without changing the acquisition path. Component swaps and a locality–price interaction map test those predictions under matched capacity and compute, while finite-tree planning, categorical best-action transfer, misspecification, and target-disjoint Perturb-seq replay probe their range.

2 Related Work and Distinction

Bayesian sequential design combines experimental cost with posterior utility (Lindley, 1956; Chaloner & Verdinelli, 1995; Ryan et al., 2016), and DAD amortizes non-myopic design over simulated histories (Foster et al., 2021; Rainforth et al., 2024). Recent task-driven BED directly minimizes downstream expected loss (Rossa et al., 2026), while optimal-stopping BED learns a continuation-value boundary jointly with a design policy (Cheng & Huan, 2025). These works establish the decision-theoretic and amortized foundations. CertiPlan contributes the causal state interface for a different operating constraint: a priced, failure-prone intervention library must support a designated causal decision, its future observation laws, and an explicit posterior-risk release rule from the same retained posterior.

Within causal design, active Bayesian causal inference targets a causal query myopically (Toth et al., 2022), and GO-CBED targets query information with an amortized non-myopic policy (Zhang et al., 2025). Both move beyond full-graph discovery. CertiPlan targets the Bayes risk of a discrete terminal action rather than query entropy, represents prices, batch sizes, and observed failures in the transition, and uses the same posterior labels for acquisition and release. Causal bandits and causal Bayesian optimization instead optimize intervention rewards (Lattimore et al., 2016; Lee & Bareinboim, 2018; Lu et al., 2021; Aglietti et al., 2020; Sussex et al., 2023); active and budgeted causal discovery target graph structure (Spirtes et al., 2000; Chickering, 2002; Hauser & Bühlmann, 2012; Ghassami et al., 2018; Brouillard et al., 2020; Scherrer et al., 2021).

Fixed-confidence best-arm methods coordinate sampling and stopping when arm evidence is observed directly (Garivier & Kaufmann, 2016). Here, one intervention updates coupled library effects through an uncertain SCM, so the state required for a valid next-outcome distribution is larger than a vector of arm means. Predictive ancestral closure supplies that state without retaining causal coordinates irrelevant to every library outcome.

Table 1 isolates the technical delta instead of assigning novelty to non-myopia alone. Query-targeted BED supplies the closest acquisition precedent, loss-driven BED supplies the closest objective precedent, and optimal-stopping BED supplies the closest stopping precedent. CertiPlan's evaluated unit is the state-and-release interface specific to causal sign decisions: predictive closure of a known intervention library, discrete terminal Bayes risk, explicit paid failure transitions, and one posterior-label representation for action and release. Myopic VoI is retained as a direct baseline for decision-risk acquisition without multi-step planning.

Table 1: Nearest-neighbor comparison by valued object, planning horizon, and termination interface.

Family	Acquisition object	Planning	Termination
MI-BOED / DAD	model information	amortized sequence	fixed horizon or tuned entropy
Active Bayesian causal inference	causal-query information	myopic	fixed budget
GO-CBED	causal-query information	amortized non-myopic	fixed horizon
Action-BED	downstream expected loss	amortized task-driven	fixed horizon
Optimal-stopping BED	terminal reward	joint value-based	learned value boundary
CertiPlan	terminal causal-decision risk	finite-tree DP and amortization	shared posterior-label score

3 Problem Formulation

Let $\mathcal{M} = (\mathcal{G}, \mathcal{F}, \mathcal{P}_\varepsilon)$ be an SCM over d variables (Pearl, 2009; Peters et al., 2017). A known library $\mathcal{I} = \{I_1, \ldots, I_m\}$ stores each intervention's target assignment $\mathrm{do}(X_{T_i} = x_i)$ (or named biological perturbation), observation map O_i, price $c_i > 0$, and successful-batch size n_i; graph structure and mechanisms remain uncertain. If action I_t is chosen, the observed outcome is $U_t = (F_t, D_t)$, where $F_t \in \{0, 1\}$ is the assay-failure flag. If $F_t = 0$, then $D_t = \{X^{(q)}_{O_{I_t}}\}_{q=1}^{n_{I_t}}$ is the protocol-defined multivariate batch; if $F_t = 1$, then $D_t = \varnothing$. Both outcomes consume c_{I_t}. Particle k has failure probability $\rho^{(k)}_{I_t}$, giving $\bar{\rho}_{I_t,t} = \sum_k w^{(k)}_t \rho^{(k)}_{I_t}$; a failure updates weights through $\rho^{(k)}_{I_t}$, and a success through $(1 - \rho^{(k)}_{I_t}) p(D_t \mid I_t, \Theta^{(k)}_{\mathrm{ret}})$.

The primary decision is the treatment–control sign. With $\Delta_A = \mathbb{E}[Y \mid \mathrm{do}(A)] - \mathbb{E}[Y \mid \mathrm{do}(\emptyset)]$ and indifference margin $\delta = 0.05$, evaluated episodes satisfy $|\Delta_A| > \delta$ and use $Z_{\mathrm{sign}} = \mathbf{1}\{\Delta_A > \delta\} \in \{0, 1\}$; boundary cases are assigned to the indifference set before episode sampling. The transfer task uses $Z_{\mathrm{best}} = \arg\max_{I \in \mathcal{I}} \mathbb{E}[Y \mid \mathrm{do}(I)]$, with the lowest library index breaking exact ties. We write Z for either latent terminal decision.

Given history $\mathcal{H}_t = \{(I_s, F_s, D_s)\}_{s<t}$, the Bayes action and risk are

$$\hat{Z}(\mathcal{H}) = \arg\min_a \mathbb{E}[\ell(a, Z) \mid \mathcal{H}], \qquad R_Z(\mathcal{H}) = \mathbb{E}[\ell(\hat{Z}(\mathcal{H}), Z) \mid \mathcal{H}].$$

Let $C_t = \sum_{s<t} c_{I_s}$ and let τ_{budget} be the first history at which no library action is affordable. CertiPlan's release time τ_π is the first crossing of the posterior-label score in Eq. 3, and $\tau_B = \min\{\tau_\pi, \tau_{\mathrm{budget}}\}$. The action policy, finite-tree teacher, and PPO learner share the operational objective

$$\min_\pi \quad J_\lambda(\pi) = \mathbb{E}\Big[C_{\tau_B} + \lambda\, \mathbf{1}\{\tau_\pi \geq \tau_{\mathrm{budget}}\} \ell\Big(\hat{Z}(\mathcal{H}_{\tau_B}), Z\Big)\Big], \qquad C_{\tau_B} \leq B. \tag{1}$$

The indicator makes the terminal penalty explicit: a threshold crossing releases immediately, whereas budget exhaustion returns the current Bayes action and charges its residual loss in cost units. The multiplier λ is validation-selected and fixed before testing. Section 5 evaluates three consequences of this objective separately: capped acquisition cost on a common shadow trajectory, fixed-budget loss, and harmful voluntary-release incidence. Both tasks use 0–1 decision loss. For best-intervention selection, simple regret is $\max_I \mu_I - \mu_{\hat{Z}}$.

4 CertiPlan

4.1 Decision-closed posterior state

Let $\mathcal{I}^+$ contain all library actions, the designated treatment, and control. Particle k stores

$$S^{(k)}_t = (\Theta^{(k)}_{\mathrm{ret}}, \mu^{(k)}_{\mathcal{I}^+}, w^{(k)}_t), \qquad \mu^{(k)}_{\mathcal{I}^+} = \{\mathbb{E}[Y \mid \mathrm{do}(I)]\}_{I \in \mathcal{I}^+}.$$

The policy state is the weighted ensemble $\mathcal{S}_t = \{S_t^{(k)}\}_{k=1}^K$. For graph particle $G^{(k)}$, define the predictive ancestral closure

$$\mathcal{C}^{(k)} = \bigcup_{I \in \mathcal{I}^+} \mathrm{An}_{G^{(k)}_{\overline{T_I}}}(O_I \cup \{Y\}) \cup T_I,$$

where incoming edges to an intervened target are removed. For graph-SCM families, $\Theta_{\mathrm{ret}}^{(k)}$ contains the induced topology, local mechanisms, noise laws, and failure propensities on $\mathcal{C}^{(k)}$. Thus it determines $p(U \mid I, S_t^{(k)})$, every entry of $\mu_{\mathcal{I}^+}^{(k)}$, and the next retained state; the supported union is rebuilt after resampling. This is an exact posterior projection at the model level. Finite particles, clipped likelihood ratios, resampling, and Monte Carlo effect refresh define a numerical approximation to that projection, analyzed separately in Section 4.3. Empirical families use the analogous protocol closure: the target-indexed response density, control response, and failure likelihood needed for every library outcome. Appendix A gives both constructions.

The state inventory preserves a common Bayes update with family-specific likelihood and effect operators. Linear-Gaussian particles retain closure coefficients and noise variances, giving Gaussian batch likelihoods and analytic interventional means. Sparse-additive particles retain the g_{ij} modules and noise scales needed to factor the observed batch, and refresh effects by a topological Monte Carlo sweep. Sachs particles retain intervention-indexed 11-protein response densities. Replogle particles retain target-indexed pseudobulk response densities and guide-efficiency likelihoods. Each refreshed effect vector induces either a binary sign label or a categorical best-action label.

4.2 Failure-aware value, planning, and release

For $H_t^{I,f,D} = \mathcal{H}_t \cup \{(I, f, D)\}$, the exact one-step decision value includes both observed outcome branches:

$$\begin{aligned} \Delta_Z(I \mid \mathcal{H}_t) = R_Z(\mathcal{H}_t) &- \bar{\rho}_{I,t} R_Z(H_t^{I,1,\varnothing}) \\ &- (1 - \bar{\rho}_{I,t}) \mathbb{E}_{D|I,F=0,\mathcal{H}_t} R_Z(H_t^{I,0,D}), \\ I^*_{\mathrm{VoI}} = \arg &\max_{I: c_I \le B - C_t} \Delta_Z(I \mid \mathcal{H}_t)/c_I. \end{aligned} \tag{2}$$

Equation 2 permits label-correlated failure models, but the evaluated synthetic prior factorizes failure propensity from graph mechanisms. In that regime a failure updates assay reliability and paid budget, not the exact Z posterior, and the value reduces to success probability times success-conditioned decision value. Estimation draws $L = 32$ pseudo-outcomes per candidate, including Bernoulli failure flags, with common random numbers across actions. Log likelihood ratios are clipped to $[-30, 30]$, weights use log-sum-exp normalization, and systematic resampling occurs at $N_{\mathrm{eff}} < K/2$. With $K = 500$, mean batch size $\bar{n}$, and retained refresh cost r, one all-action estimate costs $O(mLK(\bar{n} + r))$ time and $O(K(m + r))$ memory.

Equation 2 defines the Myopic VoI baseline and supplies rollout values, but deployed CertiPlan is non-myopic. Let $\mathcal{A}_b = \{I : c_I \le b\}$. On tractable SCMs, $L = 32$ common-random-number pseudo-outcomes per action induce a sampled finite predictive tree. The teacher memoizes each posterior-and-budget node and solves that tree exactly on the fixed pseudo-outcome support:

$$V_b(\mathcal{H}) = \begin{cases} 0, & \tilde{p}_{\mathrm{err}}(\mathcal{H}) \le \alpha, \\ \lambda R_Z(\mathcal{H}), & \mathcal{A}_b = \varnothing, \\ \min_{I \in \mathcal{A}_b} \{c_I + \mathbb{E}_{U|I,\mathcal{H}} V_{b-c_I}(\mathcal{H} \cup \{I, U\})\}, & \text{otherwise.} \end{cases}$$

The terminal multiplier λ is selected on validation and then fixed. At an exhausted leaf, the expected penalty $\lambda\ell$ equals λR_Z under the current posterior; a threshold release has zero continuation value, matching Eq. 1. Tree-DP state–action values generate imitation targets over all reachable teacher nodes. A four-layer transformer with $d_{\mathrm{model}} = 256$ forms one action token from its posterior predictive and effect features, price, batch size, failure probability, and remaining budget, then maps the token set to masked action logits. PPO (Schulman et al., 2017) uses reward $-c_I$, fallback terminal reward $-\lambda\ell$ only at budget exhaustion, and unit discount. Stopping is not a policy logit. At deployment, CertiPlan checks Eq. 3, takes the highest-logit affordable action if continuation is required, observes U_t, and updates $\mathcal{S}_t$. Appendix I isolates tree approximation, teacher initialization, and PPO effects.

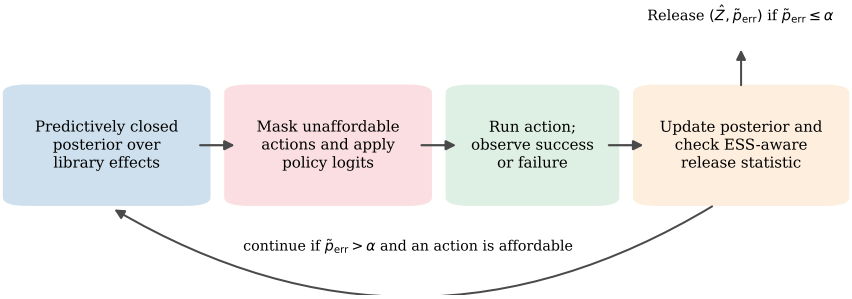

Figure 1: CertiPlan deployment loop. A predictively closed library-effect posterior is checked for release, mapped to affordable action logits when continuation is required, and updated by either a failed or successful intervention outcome.

For the current Bayes action, particle labels $Z^{(k)}$ yield wrong-decision mass

$$\hat{p}_{\text{err},t} = \sum_{k=1}^{K} w_t^{(k)} \mathbf{1}\{Z^{(k)} \neq \hat{Z}(\mathcal{H}_t)\}, \qquad N_{\text{eff},t} = \left(\sum_k (w_t^{(k)})^2 \right)^{-1}.$$

The stopping statistic and time are

$$\tilde{p}_{\text{err},t} = \hat{p}_{\text{err},t} + \sqrt{\frac{\hat{p}_{\text{err},t}(1 - \hat{p}_{\text{err},t})}{\max\{N_{\text{eff},t}, 1\}}}, \qquad \tau_\pi = \inf\{t : \tilde{p}_{\text{err},t} \le \alpha\}. \tag{3}$$

For categorical Z, each particle contributes its best-action label, $\hat{Z}$ is the posterior Bayes action, and the indicator measures posterior top-1 error. The additive term increases with weight degeneracy when $0 < \hat{p}_{\text{err},t} < 1$. The released posterior summary is $(\hat{Z}, \tilde{p}_{\text{err},t})$. Prediction, action features, stopping, and release all read the same closed state.

4.3 Stability under retained-posterior error

Let P_t and Q_t be the full and retained posteriors over Z, and let P_t^I and Q_t^I be their joint predictive laws over (Z, U), including success and failure. If both total-variation distances are at most η_t, then 0–1 terminal risks differ by at most η_t, one-step risk reductions differ by at most $2\eta_t$, and cost-normalized scores differ by at most $2\eta_t/c_I$. Appendix B proves the statement. If actions I and J have a full-posterior score gap greater than $2\eta_t/c_I + 2\eta_t/c_J$, their order is preserved by the retained state. For particle-independent failure, the tighter factor $2(1 - \bar{\rho}_{I,t})\eta_t/c_I$ additionally requires the same η_t bound for the success-conditioned joint laws. Proposition 1 gives exact model-level closure; Lemma 1 transfers any numerical retained-posterior error summarized by η_t to terminal risk and one-step action scores.

5 Experiments

The experiments resolve three questions in sequence. First, does decision-closed design improve acquisition and release under a fully matched training interface? Second, do acquisition and stopping swaps recover the two proposed mechanisms? Third, do query locality, price heterogeneity, decision cardinality, model shift, and observed perturbation data move performance in the directions predicted by the method?

5.1 Settings, protocols, and baselines

The core evaluation uses linear-Gaussian random DAGs, sparse-additive SCMs, and the 11-node Sachs protein-signaling setting (Sachs et al., 2005). Synthetic training draws $d \in \{5, 10, 20, 50\}$, edge probability 0.3, library size $m \in [12, 32]$, intervention arity 1–3, and batch size $n_i \in \{16, 32, 64\}$.

Costs are log-uniform on $[1, 50]$ and failures follow Beta$(2, 10)$ independently of graph and mechanism parameters. Treatment and readout are sampled from connected pairs with $|\Delta_A| > 0.05$. We generate 120,000/10,000/10,000 train/validation/test graphs from disjoint seeds; the principal synthetic tables use held-out $d = 20$ graphs. Sachs uses protocol-derived relative costs and failure annotations for method ranking. In its held-out-perturbation test, response densities from the held-out perturbations are excluded from initialization and become observable only when the corresponding replay action is selected; terminal labels remain evaluator-only.

The external evaluation uses the public CC-BY-4.0 Replogle K562 and RPE1 Perturb-seq data (Replogle et al., 2022) through CausalBench preprocessing (Chevalley et al., 2025). A replay action returns one target-indexed pseudobulk response and an observed guide-success flag; non-targeting guides define control. Ribosomal-stress, cell-cycle, and mitochondrial gene modules supply three scalar readouts, and their treatment–control pseudobulk differences define evaluator-only sign labels. Targets, including every guide and cell assigned to a target, are partitioned before prior fitting, policy training, and evaluation, so no target appears in more than one split. Only training-target distributions initialize response particles and hyperparameters. A held-out target's response records remain in the replay environment and enter the posterior only after the policy selects that target; no cross-target posterior update imports another held-out target's responses. Guide efficiency supplies the observed failure likelihood, and sampling burden normalized within cell line supplies relative cost. Rank stability is the bootstrap frequency with which a method retains its full-cohort ordering across the three readouts. Appendix J specifies replay aggregation and the evidence/truth roles.

All matched comparisons use four train seeds and 500 independent evaluation episodes per seed, the budget grid $\{5, 10, 15, 20, 25, 35, 50\}$, and common episode instances and simulator randomness. Table intervals use 10,000 hierarchical paired resamples: train seeds are sampled first and paired episodes are then sampled within each selected seed. Perturb-seq resampling clusters targets within cell line before replay episodes; Figure 2 gives pointwise binomial intervals for campaign incidence, and raw harmful-release counts use the 2,000 pooled episodes. Acquisition hyperparameters minimize validation cost-to-ϵ, with Loss@25 as tie-breaker; native score thresholds are selected for each requested α on validation and fixed before testing. CertiPlan and BOED share the four-layer encoder, AdamW (Loshchilov & Hutter, 2019), 200,000 training episodes, teacher-state histories, the number of objective-specific teacher targets, tuning trials, and a 325 GPU-hour cap per family on four NVIDIA A100 80GB GPUs. Their teacher objectives differ necessarily: terminal decision risk for CertiPlan and model mutual information for BOED.

Random samples affordable actions uniformly. Greedy IG maximizes one-step full-model information gain (Tong & Koller, 2001). BOED is a DAD-style sequential policy trained for Monte Carlo model mutual information and stops when the posterior entropy $H[p(Z \mid \mathcal{H}_t)]$ crosses a validation-tuned threshold (Foster et al., 2021). Myopic VoI selects Eq. 2 directly and stops at a separately tuned threshold on one-step posterior decision risk. Thompson samples one posterior hypothesis and chooses its preferred intervention (Thompson, 1933). The "MI acquisition" swap trains the CertiPlan backbone on BOED's objective while retaining Eq. 3; "entropy stop" retains the CertiPlan action policy and substitutes BOED's entropy functional; "no decision posterior" supplies the same learner with full-model summaries instead of the closed effect state. Parity therefore means identical episode instances, posterior inputs, action masks, budgets, capacity, optimization counts, tuning trials, and compute, with only the objective-specific labels changed.

Cost-to-ϵ uses a fixed 500-particle evaluator shared by every policy. If $T_\epsilon = \inf\{C_t : \tilde{p}_{\mathrm{err},t} \le \epsilon\}$ on the shadow history, the reported estimand is the capped first-passage cost $C_\epsilon = \min\{T_\epsilon, B\}$, with $C_\epsilon = B$ when no crossing occurs. Shadow rollouts continue affordable actions to the budget boundary after a native stop; every failed action contributes its paid cost, and cost ties receive half a paired win. Thus the mean, interval, and paired win refer to cap-imputed cost rather than a survival-model estimate. Loss@B=25 is terminal 0–1 error on the same shadow history. Native harmful voluntary release is $\{\tau_\pi < \tau_{\mathrm{budget}},\ \hat{Z}(\mathcal{H}_{\tau_\pi}) \neq Z\}$, with all 2,000 campaigns in the denominator. This common denominator directly evaluates the joint event of voluntary release and error. The indifference margin δ defines eligible sign episodes, ϵ defines shadow acquisition, and α indexes the native release-score threshold; all equal 0.05 in headline experiments but govern different events. For best-intervention selection, Cost-to-95% is defined by the same capped first-passage construction with $1 - \hat{p}_{\mathrm{err},t} \ge 0.95$, and top-1 error and simple regret are evaluated on the shadow history at $B = 25$.

Table 2: Parity-controlled sign release over four seeds and 2,000 held-out episodes per family. Metric cells stack CertiPlan over BOED as mean ± paired-bootstrap 95% half-width; harmful voluntary-release incidence uses $\alpha = 0.05$, cost wins count episodes favoring CertiPlan, and GPU-hours list CertiPlan/BOED.

Family	Cost-to-ϵ	Loss@25	Harmful release	Cost wins	GPU-h
Linear-Gaussian	13.1 ± 0.7 17.6 ± 0.8	0.055 ± 0.007 0.086 ± 0.008	0.052 ± 0.006 0.092 ± 0.008	1381	318/317
Sparse additive	16.0 ± 0.8 21.0 ± 0.9	0.079 ± 0.008 0.115 ± 0.009	0.058 ± 0.007 0.098 ± 0.009	1334	324/323
Sachs	11.7 ± 0.7 16.2 ± 0.7	0.046 ± 0.007 0.081 ± 0.008	0.050 ± 0.006 0.085 ± 0.008	1362	311/310

Table 3: Frozen-stack component swaps from an independent matched rerun. Means ± 95% half-width cover four seeds and 2,000 episodes; harmful release uses $\alpha = 0.05$, and all rows share the same encoder and training schedule.

Family	Policy	Cost-to-ϵ	Loss@25	Harmful release
LG	CertiPlan	12.7 ± 0.6	0.051 ± 0.006	0.050 ± 0.006
LG	MI acquisition	17.5 ± 0.8	0.082 ± 0.007	0.057 ± 0.007
LG	Entropy stop	12.9 ± 0.7	0.055 ± 0.007	0.109 ± 0.009
SA	CertiPlan	15.6 ± 0.7	0.074 ± 0.007	0.055 ± 0.007
SA	MI acquisition	20.4 ± 0.9	0.109 ± 0.008	0.063 ± 0.008
SA	Entropy stop	16.0 ± 0.8	0.079 ± 0.008	0.124 ± 0.010

5.2 Matched efficiency and mechanism isolation

Under the complete matched interface, CertiPlan lowers capped cost-to-ϵ by 25.6%, 23.8%, and 27.8%; Loss@25 by 36.0%, 31.3%, and 43.2%; and harmful voluntary-release incidence by 43.5%, 40.8%, and 41.2% for linear-Gaussian, sparse-additive, and Sachs. It wins the paired cost comparison in 1,334–1,381 of 2,000 episodes per family. The ordering appears under matched backbone capacity, teacher-state exposure, target count, tuning budget, and offline compute. The broader six-method comparison in Appendix C places this result against direct heuristics, including Myopic VoI.

The swaps separate action value from release. Replacing decision-risk acquisition with mutual information raises acquisition cost by 37.8% in linear-Gaussian and 30.8% in sparse additive while largely retaining CertiPlan's release behavior. Replacing the shared posterior statistic by entropy leaves the independently rerun acquisition metrics within uncertainty but produces an approximately twofold harmful-release incidence. Removing the decision posterior degrades both acquisition outcomes in the complete matrix (Table 7). On tractable libraries, CertiPlan also stays within 0.006–0.009 objective units of finite-tree DP and agrees with 90.2–93.4% of its affordable continuation actions (Table 13), tying the amortized policy to its teacher rather than to backbone capacity alone.

5.3 Harmful-release operating curves

Figure 2 compares native rules at $\alpha \in \{0.01, 0.02, 0.05, 0.10, 0.20\}$. CertiPlan has the lowest unconditional harmful-release incidence in every one of the 15 family–threshold cells. The curves directly show how each score threshold changes the campaign-level joint event of voluntary release and error. Sparse-additive rates move from above the two smallest thresholds to below the upper values; the widest intervals occur at $\alpha = 0.01$. Table 4 gives the operational $\alpha = 0.05$ counts.

5.4 Regime behavior and query transfer

For the regime grid, d_ψ is the number of variables in the generated ancestor subgraph whose structural coefficients can change the terminal library-effect label. We control d_ψ/d before sampling the

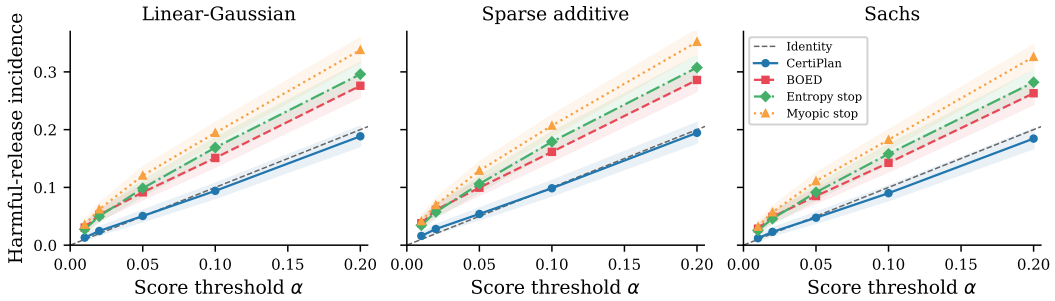

Figure 2: Unconditional harmful voluntary-release incidence by score threshold. Lines report four train seeds and 2,000 held-out episodes per method, family, and threshold; bands are pointwise normal-approximation binomial 95% intervals.

Table 4: Unconditional harmful-release incidence at the operational score threshold $\alpha = 0.05$, with raw count/2,000.

Family	Method	Harmful release@0.05
LG	CertiPlan	0.0505 (101/2000)
LG	BOED	0.0915 (183/2000)
LG	Entropy stop	0.0985 (197/2000)
LG	Myopic stop	0.1205 (241/2000)
SA	CertiPlan	0.0540 (108/2000)
SA	BOED	0.0995 (199/2000)
SA	Entropy stop	0.1060 (212/2000)
SA	Myopic stop	0.1290 (258/2000)
Sachs	CertiPlan	0.0475 (95/2000)
Sachs	BOED	0.0850 (170/2000)
Sachs	Entropy stop	0.0915 (183/2000)
Sachs	Myopic stop	0.1110 (222/2000)

remaining nuisance edges, while holding mean intervention cost and failure probability fixed within each cost-CV column.

The interaction map exposes the mechanism's operating regime. At $d_\psi/d = 0.10$ and cost CV 1.25, cost falls by 38.9% in linear-Gaussian and 33.1% in sparse additive. At $d_\psi/d = 1$ and CV 0.10, the reductions narrow to 2.6% and 1.9%. The endpoint contrast is the predicted consequence of removing both sources of acquisition leverage—query-local nuisance uncertainty and price dispersion—and the two-factor trend remains stable across $d \in \{10, 20, 50\}$ (Appendix F).

The categorical query preserves CertiPlan's architecture and yields smaller, informative margins. It reduces top-1 error by 30.3% and 22.4% relative to BOED while retaining substantially lower harmful-release incidence (Table 11). The result shows that decision-closed posterior alignment transfers from a binary sign to a multiway terminal decision.

5.5 External perturbation evidence

On target-disjoint K562 and RPE1 perturbations, CertiPlan reduces acquisition cost by 21.7% and fixed-budget sign error by 26.0% relative to BOED, with the highest bootstrap rank stability across all three readouts (Table 14). The cell-cycle readout has the widest Loss@25 intervals: 0.012, 0.015, and 0.014 for CertiPlan, BOED, and Myopic VoI. The same ordering under Sachs cost remappings (Appendix K) links the synthetic mechanism tests to two observed perturbation settings.

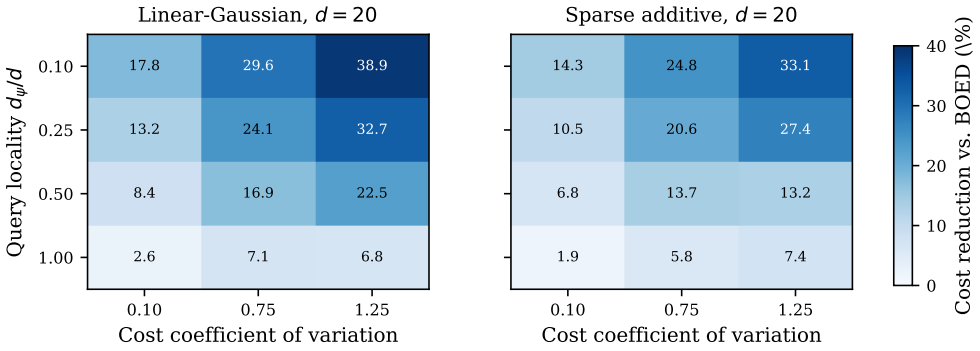

Figure 3: CertiPlan capped cost-to-ϵ reduction relative to parity-matched BOED as query locality and cost heterogeneity vary at $d = 20$. Each cell averages four seeds and 2,000 paired episodes. Targeted acquisition is most valuable when a small effect subspace and dispersed costs make graph-wide information expensive.

6 Discussion

Two properties explain the gains. First, separating graph hypotheses that induce the same terminal label has no direct Bayes-risk value; the locality-by-cost grid shows that decision targeting matters most for local queries and dispersed prices. Second, predictive closure carries successful outcomes into the effect posterior and observed failures into the reliability-and-budget state used by future actions. Component swaps assign the large cost change to the acquisition objective and the large harmful-release change to posterior alignment. Agreement with finite-tree planning and small nonzero objective gaps connect the amortized policy to the finite-budget recursion rather than to backbone capacity.

The operating evidence and its scope are complementary. The ESS adjustment is an empirically selected finite-particle release score, not a distribution-free confidence bound; harmful release is reported per campaign rather than conditional on release, while capped shadow cost and fixed-budget loss evaluate acquisition on a common continuation. Model-level closure is exact for the stated observation map, and Lemma 1 isolates the additional retained-posterior approximation assumption. The cost margin contracts when queries are global and prices nearly uniform, Sachs prices are protocol-relative, and Perturb-seq is a target-disjoint observed-response replay. Within these regimes, the operating curves, 12 matched misspecification cells (Appendix H), categorical transfer, and observed perturbation tests consistently connect the state construction to lower decision error and spend.

The practical consequence is a clean division of labor: causal modeling supplies a posterior closed under the known library, terminal loss defines what information is useful, and the release rule reads the resulting decision labels directly. This interface preserves the scientific endpoint through planning, amortization, and stopping, which is the source of the separable acquisition and release improvements.

7 Conclusion

CertiPlan turns a known library's terminal causal decision into a posterior interface shared by prediction, non-myopic acquisition, and release. With architecture, teacher-state exposure, tuning, and compute matched, it lowers capped shadow cost-to-ϵ by 23.8–27.8%, fixed-budget sign error by 31.3–43.2%, and harmful voluntary-release incidence by 40.8–43.5% relative to BOED; action and stopping swaps locate the two effects. The 33.1–38.9% advantage for local, price-heterogeneous queries and the 21.7% Perturb-seq cost reduction show where decision-closed experimental design most strongly changes the operating point.

REFERENCES

Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal Bayesian optimization. In *AISTATS*, pp. 3155–3164, 2020. URL https://proceedings.mlr.press/v108/aglietti20a.html.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. In *NeurIPS*, 2020.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995. doi: 10.1214/ss/1177009939.

Chen Cheng and Xun Huan. Optimal stopping for sequential Bayesian experimental design. *arXiv preprint arXiv:2509.21734*, 2025. URL https://arxiv.org/abs/2509.21734.

Mathieu Chevalley, Yusuf H. Roohani, Arash Mehrjou, Jure Leskovec, and Patrick Schwab. A large-scale benchmark for network inference from single-cell perturbation data. *Communications Biology*, 8(1):412, 2025. doi: 10.1038/s42003-025-07764-y.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002. URL https://jmlr.org/papers/v3/chickering02b.html.

Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential Bayesian experimental design. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3384–3395. PMLR, 2021. URL https://proceedings.mlr.press/v139/foster21a.html.

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory*, volume 49, pp. 998–1027. PMLR, 2016. URL https://proceedings.mlr.press/v49/garivier16a.html.

AmirEmad Ghassami, Saber Salehkaleybar, Negar Kiyavash, and Elias Bareinboim. Budgeted experiment design for causal structure learning. In *ICML*, 2018. URL https://proceedings.mlr.press/v80/ghassami18a.html.

Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13(79):2409–2464, 2012. URL https://jmlr.org/papers/v13/hauser12a.html.

Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal bandits: Learning good interventions via causal inference. In *NeurIPS*, 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/hash/b4288d9c0ec0a1841b3b3728321e7088-Abstract.html.

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In *NeurIPS*, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/c0a271bc0ecb776a094786474322cb82-Abstract.html.

Dennis V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956. doi: 10.1214/aoms/1177728069.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal bandits with unknown graph structure. In *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/d010396ca8abf6ead8cacc2c2f2f26c7-Abstract.html.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.

Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern Bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024. doi: 10.1214/23-STS915.

Joseph M. Replogle, Reuben A. Saunders, Angela N. Pogson, Jeffrey A. Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J. Wagner, Karen Adelman, Gila Lithwick-Yanai, Nika Iremadze, Florian Oberstrass, Doron Lipson, Jessica L. Bonnar, Marco Jost, Thomas M. Norman, and Jonathan S. Weissman. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 185(14):2559–2575.e28, 2022. doi: 10.1016/j.cell.2022.05.013.

Tom Rossa, Angus Phillips, and Tom Rainforth. Action-BED: Task-driven Bayesian experimental design with singly intractable objectives. *arXiv preprint arXiv:2606.23662*, 2026. URL https://arxiv.org/abs/2606.23662.

Elizabeth G. Ryan, Christopher C. Drovandi, James M. McGree, and Anthony N. Pettitt. A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1):128–154, 2016. doi: 10.1111/insr.12107.

Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005. doi: 10.1126/science.1105809.

Nino Scherrer, Olexa Bilaniuk, Yashas Annadani, Anirudh Goyal, Patrick Schwab, Bernhard Schölkopf, Michael C Mozer, Yoshua Bengio, Stefan Bauer, and Nan Rosemary Ke. Learning neural causal models with active interventions. *arXiv preprint arXiv:2109.02429*, 2021. URL https://arxiv.org/abs/2109.02429.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, 2 edition, 2000.

Scott Sussex, Anastasiia Makarova, and Andreas Krause. Model-based causal Bayesian optimization. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Vk-34OQ7rFo.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. doi: 10.1093/biomet/25.3-4.285.

Simon Tong and Daphne Koller. Active learning for structure in Bayesian networks. In *IJCAI*, 2001.

Christian Toth, Lars Lorch, Christian Knoll, Andreas Krause, Franz Pernkopf, Robert Peharz, and Julius von Kügelgen. Active Bayesian causal inference. In *Advances in Neural Information Processing Systems*, volume 35, pp. 16261–16275, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/675e371eeeea99551ce47797ed6ed33e-Paper-Conference.pdf.

Zheyu Zhang, Jiayuan Dong, Jie Liu, and Xun Huan. Goal-oriented sequential Bayesian experimental design for causal learning. *arXiv preprint arXiv:2507.07359*, 2025. URL https://arxiv.org/abs/2507.07359.

AI USE STATEMENT

The authors used language-model assistance for copyediting and presentation. The scientific ideas, methods, data, experiments, proofs, code, analyses, and interpretations are the authors' work; the authors reviewed the edited text and remain responsible for the manuscript.

REPRODUCIBILITY STATEMENT

Section 5 specifies datasets, splits, models, baselines, budgets, hardware, seeds, and uncertainty construction. Appendices A–K provide update details, protocol-specific tables, raw counts, stress settings, and additional diagnostics.

A RETAINED-STATE AND UPDATE DETAILS

The family prior factorizes as $p(G, \Theta, \rho) = p(G)p(\Theta \mid G)p(\rho)$. For synthetic SCMs, $p(G)$ is the acyclic graph generator with edge probability 0.3, $p(\Theta \mid G)$ is the corresponding linear-Gaussian or sparse-additive mechanism-and-noise generator, and $p(\rho)$ is $\mathrm{Beta}(2, 10)$. These same episode priors initialize every compared posterior. For empirical families, particles bootstrap intervention-indexed measured-response distributions, and the protocol annotations determine ρ. The policy receives library metadata and observations, not the sampled graph.

For every action I, the closure $\mathcal{C}^{(k)}$ contains the ancestors of the observed coordinates O_I and outcome Y in the intervened graph, together with intervention targets. Linear-Gaussian mechanisms have $X_j = \sum_{h \in \mathrm{pa}(j)} \beta_{hj} X_h + \varepsilon_j$, so the induced batch likelihood is multivariate Gaussian and interventional means follow a topological linear solve. Sparse-additive mechanisms have $X_j = \sum_{h \in \mathrm{pa}(j)} g_{hj}(X_h) + \varepsilon_j$; their likelihood is the product of retained-node conditionals and their effect refresh is a topological Monte Carlo expectation. Sachs observes the 11 measured proteins and uses intervention-indexed response-density particles. Perturb-seq observes a target pseudobulk vector and guide-success flag, with non-targeting pseudobulk as control. These observation definitions fix both the likelihood and effect operator rather than deriving them from the terminal label alone.

For successful batch D_t and failure flag $F_t = 0$,

$$w_{t+1}^{(k)} \propto w_t^{(k)} (1 - \rho_{I_t}^{(k)}) p(D_t \mid I_t, \Theta_{\mathrm{ret}}^{(k)}), \tag{4}$$

$$\mu_{\mathcal{I}^+}^{(k)} \leftarrow \mathrm{Refresh}_{\mathrm{family}}(\Theta_{\mathrm{ret}}^{(k)}). \tag{5}$$

For $F_t = 1$, $w_{t+1}^{(k)} \propto w_t^{(k)} \rho_{I_t}^{(k)}$; the paid cost and empty batch are recorded, and each particle's effect vector remains fixed. Under the evaluated factorized prior, the exact marginal effect posterior is unchanged while the reliability posterior updates; finite ensembles can exhibit vanishing sampling correlation between the two particle coordinates.

Proposition 1 (Predictive decision closure for graph SCMs)**.** *Under the modular graph-SCM factorization above, the exact posterior projected onto the retained coordinates is sufficient for the joint law of $(Z, U_{t:\tau})$ under any future library-action sequence. Consequently, exact Bayes updating, library-effect refresh, one-step value, and the Bellman recursion depend on the full posterior only through this projected state.*

Proof. For an intervention I, every ancestor of $O_I \cup \{Y\}$ in the mutilated graph lies in $\mathcal{C}^{(k)}$. Integrating variables outside this ancestral set removes normalized conditional factors and leaves the same marginal law for D, while $\rho_I^{(k)}$ gives the failure branch. The intervention-effect vector and hence Z are measurable from the retained mechanisms. Multiplying the retained likelihood by the exact projected posterior density, normalizing, and refreshing the effects therefore gives the same transition kernel as marginalizing the full posterior. Induction over future actions proves the joint-law statement. Particle resampling and clipping are numerical operators applied after this projection and are not used in the sufficiency argument. □

For Sachs and Perturb-seq, protocol closure is direct: $\mathcal{S}_t$ stores every target-indexed response and failure density queried by the library, so Bayes updates and decision labels require no latent graph coordinate.

B POSTERIOR-PERTURBATION BOUND

Lemma 1. *Let P and Q be distributions over a finite decision variable Z with $\|P - Q\|_{\mathrm{TV}} \le \eta$. For intervention I, let P^I and Q^I be joint predictive distributions of (Z, U) and suppose $\|P^I - Q^I\|_{\mathrm{TV}} \le$*

η. Under 0–1 loss, $|R_P - R_Q| \le \eta$ and $|\Delta_P(I) - \Delta_Q(I)| \le 2\eta$. With shared price c_I, cost-normalized scores differ by at most $2\eta/c_I$.

Proof. For a fixed action a, its error event $E_a = \{Z \neq a\}$ satisfies $|P(E_a) - Q(E_a)| \le \eta$. Let a_P and a_Q minimize these probabilities. Then $R_Q \le Q(E_{a_P}) \le P(E_{a_P}) + \eta = R_P + \eta$, and symmetry gives $|R_P - R_Q| \le \eta$. The expected Bayes risk after observing U can be written

$$\mathcal{R}(P^I) = \inf_{\gamma:\mathcal{U}\to\mathcal{Z}} \mathbb{E}_{(Z,U)\sim P^I}[\mathbf{1}\{\gamma(U) \neq Z\}]\,.$$

For every decision rule γ, the two expectations under P^I and Q^I differ by at most η because the loss lies in $[0, 1]$. Taking the infimum preserves this uniform bound, so $|\mathcal{R}(P^I) - \mathcal{R}(Q^I)| \le \eta$. Since $\Delta_P(I) = R_P - \mathcal{R}(P^I)$ and likewise for Q, the triangle inequality gives $|\Delta_P(I) - \Delta_Q(I)| \le 2\eta$. Division by shared c_I yields the score bound. □

Here Δ_Z in Eq. 2 is Δ_P under the full success–failure predictive law. If failure is particle-independent with common probability $\bar{\rho}_I$, and both the Z marginals and the success-conditioned joint laws over (Z, D) satisfy the same TV bound η, then factoring out the unchanged failure branch tightens the score bound to $2(1 - \bar{\rho}_I)\eta/c_I$. For any two actions, a full-posterior score gap larger than the sum of their individual error bounds preserves their ranking. Wrong-target and likelihood shifts change the decision posterior and therefore η; inaccurate prices perturb cost normalization separately; non-additive setup fees alter trajectory cost while preserving the one-step posterior calculation.

C Breadth and uncertainty comparison

The broad comparison places the parity-controlled BOED result among direct acquisition heuristics. The following two tables report point estimates first and paired uncertainty on the same histories second.

Table 5: Six-method causal sign-release comparison under the original matched online interface. Entries are means over 500 held-out episodes; lower is better.

	Linear-Gaussian		Sparse additive		Sachs	
Method	Cost	Loss	Cost	Loss	Cost	Loss
Random	38.2	0.294	42.7	0.341	29.1	0.218
Greedy IG	25.4	0.132	29.8	0.178	21.3	0.124
Myopic VoI	22.1	0.118	26.4	0.157	19.7	0.108
Thompson	21.8	0.121	25.9	0.162	19.2	0.112
BOED	18.4	0.083	21.7	0.114	16.8	0.079
CertiPlan	12.4	0.048	15.1	0.071	11.2	0.042

Table 6: Paired uncertainty for Table 5. Each cell stacks CertiPlan over BOED; the two win counts are episode-level cost / Loss@25 comparisons on the same 500 histories.

Family	Cost-to-ϵ	Loss@25	CertiPlan wins
Linear-Gaussian	12.4 ± 0.6 18.4 ± 0.7	0.048 ± 0.007 0.083 ± 0.009	402 / 348
Sparse additive	15.1 ± 0.7 21.7 ± 0.9	0.071 ± 0.008 0.114 ± 0.011	391 / 332
Sachs	11.2 ± 0.6 16.8 ± 0.7	0.042 ± 0.006 0.079 ± 0.010	387 / 339

This original-interface comparison differs from the parity protocol in Table 2: it matches the online action interface and encoder width, while Table 2 additionally matches teacher availability, tuning trials, train steps, and GPU-hour caps. The margin contracts under the stricter protocol and remains consistent across every reported outcome.

D FROZEN-STACK MECHANISM SWAPS

Each row changes one state, acquisition, or stopping component while preserving the encoder and training schedule. Cost and Loss@25 use shadow trajectories; the release column uses the row's native stopping rule.

Table 7: Complete component-swap study. Metrics report mean ± 95% half-width over four seeds and 2,000 held-out episodes.

Family	Variant	Cost	Loss@25	Harmful release	Wins
LG	CertiPlan	12.7 ± 0.6	0.051 ± 0.006	0.050 ± 0.006	–
LG	Entropy stop	12.9 ± 0.7	0.055 ± 0.007	0.109 ± 0.009	1429
LG	MI acquisition	17.5 ± 0.8	0.082 ± 0.007	0.057 ± 0.007	1316
LG	No decision posterior	16.4 ± 0.7	0.069 ± 0.007	0.066 ± 0.008	1279
SA	CertiPlan	15.6 ± 0.7	0.074 ± 0.007	0.055 ± 0.007	–
SA	Entropy stop	16.0 ± 0.8	0.079 ± 0.008	0.124 ± 0.010	1458
SA	MI acquisition	20.4 ± 0.9	0.109 ± 0.008	0.063 ± 0.008	1354
SA	No decision posterior	18.3 ± 0.8	0.093 ± 0.008	0.071 ± 0.008	1308

E HARMFUL-RELEASE COUNTS AND INTERVALS

At $\alpha = 0.05$, Table 4 gives every method's raw count. Across the full threshold grid, Figure 2 uses pointwise normal-approximation binomial intervals. The interval half-widths at $\alpha = 0.01$ are 0.0050, 0.0075, 0.0072, and 0.0081 in linear-Gaussian; 0.0055, 0.0084, 0.0080, and 0.0087 in sparse additive; and 0.0048, 0.0073, 0.0070, and 0.0077 in Sachs. These widest-tail intervals preserve the separation between CertiPlan and detached rules.

Table 8: Cross-method harmful-release incidence at the four score thresholds not repeated in Table 4, with raw counts in parentheses. Each denominator is 2,000 held-out episodes; Figure 2 supplies pointwise binomial 95% intervals.

Family	α	CertiPlan	BOED	Entropy stop	Myopic stop
LG	.01	0.0130 (26)	0.0305 (61)	0.0275 (55)	0.0355 (71)
LG	.02	0.0245 (49)	0.0525 (105)	0.0500 (100)	0.0620 (124)
LG	.10	0.0940 (188)	0.1510 (302)	0.1690 (338)	0.1945 (389)
LG	.20	0.1885 (377)	0.2760 (552)	0.2960 (592)	0.3380 (676)
SA	.01	0.0160 (32)	0.0380 (76)	0.0345 (69)	0.0415 (83)
SA	.02	0.0280 (56)	0.0610 (122)	0.0575 (115)	0.0695 (139)
SA	.10	0.0985 (197)	0.1615 (323)	0.1790 (358)	0.2070 (414)
SA	.20	0.1945 (389)	0.2860 (572)	0.3075 (615)	0.3520 (704)
Sachs	.01	0.0120 (24)	0.0285 (57)	0.0255 (51)	0.0320 (64)
Sachs	.02	0.0230 (46)	0.0490 (98)	0.0460 (92)	0.0570 (114)
Sachs	.10	0.0900 (180)	0.1425 (285)	0.1580 (316)	0.1825 (365)
Sachs	.20	0.1845 (369)	0.2630 (526)	0.2820 (564)	0.3260 (652)

F REGIME ENDPOINTS

At the most local, heterogeneous endpoint, CertiPlan / BOED / Myopic VoI cost-to-ϵ is 12.9 / 21.1 / 23.4 in linear-Gaussian and 15.6 / 23.3 / 26.1 in sparse additive. At the global, near-uniform endpoint the corresponding values are 18.7 / 19.2 / 20.1 and 21.2 / 21.6 / 22.4. These endpoint cells anchor the interaction plotted in Figure 3.

Table 9: Capped cost-to-ϵ reduction relative to parity-matched BOED at the two mechanism endpoints. Four seeds and 2,000 paired episodes per cell.

Family	d	d_ψ/d=0.10, CV 1.25	d_ψ/d=1.00, CV 0.10
Linear-Gaussian	10	36.4%	3.1%
Linear-Gaussian	20	38.9%	2.6%
Linear-Gaussian	50	41.2%	2.2%
Sparse additive	20	33.1%	1.9%

Table 10: Endpoint outcomes at $d = 20$. Harmful release reports campaign incidence (count/2,000); paired 95% half-widths span 0.6–1.0 for cost and 0.006–0.011 for rates.

Family	Endpoint	Method	Cost-to-ϵ	Loss@25	Harmful release
LG	local / heterogeneous	CertiPlan	12.9	0.049	0.050 (100)
LG	local / heterogeneous	BOED	21.1	0.089	0.093 (186)
LG	local / heterogeneous	Myopic VoI	23.4	0.116	0.127 (254)
LG	global / near-uniform	CertiPlan	18.7	0.093	0.082 (164)
LG	global / near-uniform	BOED	19.2	0.096	0.087 (174)
LG	global / near-uniform	Myopic VoI	20.1	0.104	0.111 (222)
SA	local / heterogeneous	CertiPlan	15.6	0.073	0.055 (110)
SA	local / heterogeneous	BOED	23.3	0.118	0.102 (204)
SA	local / heterogeneous	Myopic VoI	26.1	0.153	0.133 (266)
SA	global / near-uniform	CertiPlan	21.2	0.122	0.092 (184)
SA	global / near-uniform	BOED	21.6	0.125	0.098 (196)
SA	global / near-uniform	Myopic VoI	22.4	0.134	0.124 (248)

G Categorical best-action results

Table 11: Best-intervention selection under the same known libraries. Simple regret and top-1 error are measured at $B = 25$; harmful voluntary release uses $\alpha = 0.05$. Results average four seeds and 2,000 episodes; rate entries include raw counts/2,000.

Family	Method	Simple regret	Top-1 error	Cost-to-95%	Harmful release
LG	CertiPlan	0.034	0.083 (166)	14.6	0.052 (104)
LG	BOED	0.049	0.119 (238)	18.8	0.089 (178)
LG	Myopic VoI	0.041	0.102 (204)	16.5	0.118 (236)
SA	CertiPlan	0.052	0.111 (222)	17.8	0.058 (116)
SA	BOED	0.069	0.143 (286)	21.9	0.096 (192)
SA	Myopic VoI	0.061	0.129 (258)	19.7	0.127 (254)

H Cross-method misspecification

The matched shifts replace 18%, 22%, and 20% of target entries in linear-Gaussian, sparse-additive, and Sachs libraries. Cost bias inflates the logged price of the expensive library half by 8%, 18%, and 12%, respectively. The batching shift charges one shared setup fee per three-action assay group, with family-specific fees 1.4, 1.8, and 1.1 in protocol cost units.

Wrong-target mismatch is the hardest shift for every method, and its cost penalty is largest in sparse additive. CertiPlan preserves the lowest harmful-release incidence in all 12 family–shift cells. Across the table, paired 95% half-widths span 0.6–1.3 for cost, 0.006–0.014 for Loss@25, and 0.006–0.012 for harmful release. Shared setup fees raise spend more sharply than fixed-budget sign error, matching the cost-side perturbation in Eq. 2.

Table 12: Matched misspecification by family. Each cell reports capped cost-to-ϵ (extra cost relative to the method's nominal row) / Loss@25 / harmful-release incidence (count/2,000). Four seeds and 500 episodes per seed, family, method, and shift.

Setting	CertiPlan	BOED	Myopic VoI
LG nominal	13.0 (0.0%) / 0.053 / 0.051 (102)	17.6 (0.0%) / 0.087 / 0.092 (184)	20.8 (0.0%) / 0.116 / 0.121 (242)
LG wrong target	15.7 (20.8%) / 0.071 / 0.066 (132)	22.0 (25.0%) / 0.112 / 0.119 (238)	26.4 (26.9%) / 0.143 / 0.151 (302)
LG cost bias	14.5 (11.5%) / 0.059 / 0.058 (116)	20.2 (14.8%) / 0.096 / 0.101 (202)	23.7 (13.9%) / 0.125 / 0.131 (262)
LG setup fee	15.2 (16.9%) / 0.063 / 0.064 (128)	20.9 (18.8%) / 0.101 / 0.109 (218)	25.1 (20.7%) / 0.133 / 0.142 (284)
SA nominal	15.7 (0.0%) / 0.076 / 0.055 (110)	20.9 (0.0%) / 0.113 / 0.099 (198)	25.2 (0.0%) / 0.154 / 0.129 (258)
SA wrong target	19.5 (24.2%) / 0.099 / 0.078 (156)	26.7 (27.8%) / 0.151 / 0.137 (274)	32.6 (29.4%) / 0.185 / 0.166 (332)
SA cost bias	18.3 (16.6%) / 0.085 / 0.063 (126)	25.0 (19.6%) / 0.129 / 0.112 (224)	29.9 (18.7%) / 0.166 / 0.139 (278)
SA setup fee	19.0 (21.0%) / 0.092 / 0.074 (148)	25.8 (23.4%) / 0.138 / 0.124 (248)	31.5 (25.0%) / 0.174 / 0.154 (308)
Sachs nominal	11.5 (0.0%) / 0.045 / 0.049 (98)	16.4 (0.0%) / 0.080 / 0.086 (172)	19.3 (0.0%) / 0.108 / 0.121 (242)
Sachs wrong target	13.3 (15.7%) / 0.058 / 0.059 (118)	19.4 (18.3%) / 0.103 / 0.106 (212)	23.2 (20.2%) / 0.137 / 0.142 (284)
Sachs cost bias	12.6 (9.6%) / 0.049 / 0.054 (108)	18.3 (11.6%) / 0.088 / 0.095 (190)	21.4 (10.9%) / 0.119 / 0.132 (264)
Sachs setup fee	13.1 (13.9%) / 0.053 / 0.061 (122)	19.1 (16.5%) / 0.094 / 0.101 (202)	22.7 (17.6%) / 0.128 / 0.139 (278)

I GAP TO FINITE-TREE PLANNING

Finite-tree dynamic programming uses the recursion in Section 4 on small libraries. At each action node, the fixed set of $L = 32$ common-random-number pseudo-outcomes defines the outgoing support and posterior transitions; backward induction enumerates this sampled tree to the budget boundary and is exact relative to that finite support. This construction provides a numerical teacher for the continuous predictive model through a fixed, shared outcome support. Agreement compares only affordable continuation actions. Excess objective is the difference from tree DP in Eq. 1's expected paid cost plus λ times budget-exhaustion loss, using the same validation-fixed λ.

Table 13: Tractable linear-Gaussian SCMs with small libraries. Excess objective is relative to finite-tree DP; agreement is the affordable-action match with the tree planner; harmful release gives incidence (count/1,000). Four seeds and 1,000 held-out episodes per setting.

d	Method	Excess objective	Cost	Harmful release	Agreement
5	Tree DP	0.000	9.8	0.049 (49)	1.000
5	CertiPlan	0.006	10.2	0.052 (52)	0.934
5	Teacher only	0.014	10.9	0.058 (58)	0.887
5	PPO, no teacher	0.027	12.6	0.071 (71)	0.801
5	BOED	0.036	13.8	0.091 (91)	0.742
5	Myopic VoI	0.031	12.9	0.119 (119)	0.779
10	Tree DP	0.000	11.7	0.050 (50)	1.000
10	CertiPlan	0.009	12.4	0.054 (54)	0.902
10	Teacher only	0.019	13.3	0.061 (61)	0.846
10	PPO, no teacher	0.034	15.1	0.077 (77)	0.752
10	BOED	0.043	16.5	0.096 (96)	0.701
10	Myopic VoI	0.037	15.4	0.126 (126)	0.728

Paired 95% half-widths span 0.003–0.009 for excess objective, 0.4–0.9 for cost, 0.007–0.012 for harmful release, and 0.018–0.035 for action agreement. CertiPlan stays within 0.006–0.009 objective units of finite-tree planning and matches 90.2–93.4% of its actions. Teacher-only policies lose efficiency, and removing teacher initialization raises both excess objective and seed variability.

J PERTURB-SEQ READOUT RESULTS

The replay unit is a target-level pseudobulk for one cell line and readout. All cells and guides for a target remain in one target-disjoint split; non-targeting guides provide the control distribution. Training-split targets alone fit the empirical prior over response-density parameters and guide efficiency. At evaluation, the response pool for each held-out target belongs to the hidden replay environment. An adaptive query draws a batch only from the selected target, records guide success or failure, and updates that target's response particles; observations are not propagated between held-out targets. The full held-out pseudobulk fixes the evaluator's treatment–control sign before replay, but neither that label nor unqueried target responses enter the policy state. Episode pools reset independently. Hierarchical resampling preserves target and cell-line identity before sampling episodes, and the aggregate is the equal-weight mean over the three modules.

Table 14: Held-out Perturb-seq targets averaged over three biological readouts. Each readout uses 500 replay episodes per seed over four seeds (2,000 total); harmful release averages the readout-level rates.

Method	Cost-to-ϵ	Loss@25	Harmful release	Rank stability
CertiPlan	14.8	0.094	0.057	0.82
BOED	18.9	0.127	0.096	0.61
Myopic VoI	17.2	0.113	0.128	0.56

Table 15: Target-disjoint Perturb-seq replay by readout. Each cell is capped cost-to-ϵ / Loss@25 (error count/2,000) / harmful-release count out of 2,000.

Readout	CertiPlan	BOED	Myopic VoI
Ribosomal stress	13.9 / 0.081 (162) / 108	18.0 / 0.115 (230) / 182	16.3 / 0.103 (206) / 248
Cell cycle	15.7 / 0.102 (204) / 122	19.8 / 0.136 (272) / 202	18.1 / 0.121 (242) / 266
Mitochondrial	14.8 / 0.099 (198) / 112	18.9 / 0.130 (260) / 192	17.2 / 0.115 (230) / 254

The cell-cycle readout yields the widest fixed-budget uncertainty, while the ordering remains stable on all three readouts without relying on one unusually large effect.

K SACHS ROBUSTNESS

The four mappings preserve the measured-protein response particles and alter only protocol cost, failure annotation, or perturbation split. For the held-out split, non-held-out perturbations fit the response-density prior; a held-out perturbation's measurements enter its posterior only when selected, while its frozen response-derived label remains evaluator-only. Rank stability summarizes ordering preservation over paired bootstrap resamples.

Table 16: Sachs robustness under proxy remapping and a held-out perturbation split. Each stacked cell orders CertiPlan, BOED, and Myopic VoI; entries are mean $\pm$ 95% half-width.

Setting	Cost	Loss@25	Harmful release	Rank stability
Protocol costs	11.5 ± 0.6 16.4 ± 0.7 19.3 ± 0.8	0.044 ± 0.006 0.080 ± 0.008 0.109 ± 0.009	0.049 ± 0.006 0.086 ± 0.008 0.121 ± 0.010	0.91 ± 0.03
Conservative failure	12.1 ± 0.7 16.9 ± 0.8 19.8 ± 0.9	0.046 ± 0.006 0.082 ± 0.008 0.111 ± 0.010	0.052 ± 0.006 0.089 ± 0.008 0.126 ± 0.010	0.89 ± 0.04
Time-dominated	12.8 ± 0.8 17.8 ± 0.8 20.7 ± 0.9	0.048 ± 0.007 0.083 ± 0.008 0.113 ± 0.010	0.051 ± 0.006 0.087 ± 0.008 0.123 ± 0.010	0.87 ± 0.04
Held-out perturbation	11.9 ± 0.7 16.7 ± 0.8 19.6 ± 0.9	0.047 ± 0.007 0.081 ± 0.009 0.110 ± 0.010	0.050 ± 0.006 0.085 ± 0.009 0.119 ± 0.011	0.90 ± 0.04