ACTIVE LEARNING WITH UTILITY-TAILORED STRICTLY PROPER SCORES

Anonymous authors
Paper under double-blind review

ABSTRACT

Medical active learning must value labels according to the decisions that predictions support, while preserving accurate probability estimation. We introduce *Domain-Utility-Sensitive Active Learning* (DUSAL), which constructs a strictly proper score from a complete action–utility table. DUSAL-Hard adds normalized Bayes value to a strictly proper base potential. Its divergence decomposes exactly into predictive divergence and normalized decision regret; we characterize this augmentation up to affine terms. Under exact Bayesian updating, acquisition separates into Bayesian Estimate of Mean Proper Scores (BEMPS) and clinical expected value of sample information (EVSI). We also derive representation invariances, a smooth utility-gap variant, and predictive and decision convergence under explicit progress assumptions. Exact finite-model experiments demonstrate decision-regret reductions through boundary-sensitive acquisition and within-region smoothing, alongside prediction–decision trade-offs. Across five public medical tasks, DUSAL makes label selection sensitive to utility information, producing measurable differences from matched BEMPS. Although additional cost gains from utility augmentation remain unconfirmed under Holm correction, DUSAL-Div achieves Holm-controlled reductions in realized-cost area under the learning curve (AULC) on three of four tasks relative to top-b selection. Comparable gains with BEMPS-Div demonstrate the practical benefit of batch diversification.

1 INTRODUCTION

Medical annotation requires scarce specialist time (Budd et al., 2021). Pool-based active learning (AL) selects training labels within an annotation budget, but a label's predictive value need not match its value for a deployment decision. Entropy, Bayesian Active Learning by Disagreement (BALD), and BatchBALD target uncertainty or parameter information (Houlsby et al., 2011; Gal et al., 2017; Kirsch et al., 2019); core-set, Batch Active Learning by Diverse Gradient Embeddings (BADGE), and uncertainty herding also account for diversity (Sener & Savarese, 2018; Ash et al., 2020; Bae et al., 2025). These criteria do not explicitly represent the consequences of choosing one clinical action over another. A predictive update can change an action near a treatment threshold while leaving the action unchanged for an equally uncertain patient elsewhere.

Clinical utility compares complete action rows, not just class frequencies or outcome weights. Directly weighting a probability score by outcome costs can change the probability distribution it elicits. A decision-aware acquisition rule should instead preserve truthful probability estimation while assigning value to information that changes the optimal action (Hernández-Orallo et al., 2012). Strictly proper scores provide the first property through a convex potential (Savage, 1971; Gneiting & Raftery, 2007; Waghmare & Ziegel, 2026). Bayesian Estimate of Mean Proper Scores (BEMPS) uses their expected improvement for acquisition (Tan et al., 2021), while Bayesian decision analysis values experiments through the decisions they inform (Chaloner & Verdinelli, 1995; DeGroot, 1962). Our question is which utility-derived augmentation connects these two objectives and what that connection establishes about acquisition.

Domain-Utility-Sensitive Active Learning (DUSAL) augments a strictly proper base potential with normalized Bayes value. DUSAL-Hard uses the complete utility rows, including valid subgradients at action ties. Its utility contribution is positive when hypothetical updates reveal a better action and zero when all updates remain in one Bayes-action region. In contrast, fixed cost-weighted quadratic

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.

geometry leaves the binary exact-look-ahead ranking unchanged (Proposition 2). DUSAL-Soft smooths the maximum to retain utility-gap sensitivity within action regions, but only Hard has the exact decision-regret interpretation. Our contributions are:

1. **Canonical score construction and characterization.** Effective-action reduction and largest-regret normalization give representation invariance. Theorem 1 decomposes Hard's divergence into predictive divergence and normalized decision regret; Theorem 2 characterizes the finite convex augmentation with this property up to affine terms.
2. **Acquisition and decision consequences.** Theorem 3 separates exact acquisition into base-score learning and normalized clinical expected value of sample information (EVSI). Theorem 4 establishes convergence under candidate-process ignorability and policy-specific progress. Soft provides differentiable utility-gap geometry; Div selects batches using deployment-weighted score-change embeddings.
3. **Mechanism tests and public-data evaluation.** Exact finite models demonstrate decision-regret reductions and isolate boundary crossing, within-region learning, complete-row sensitivity, and prediction–decision trade-offs. Five public-data studies use matched BEMPS, random and entropy baselines, with utility and posterior diagnostics. The public-data studies show that utility information produces measurable changes in label selection, while batch diversification yields Holm-controlled realized-cost reductions for both DUSAL and BEMPS on three of four tasks.

2 Related Work

Predictive Bayesian acquisition and batch diversity. Uncertainty and model-change acquisition (Settles, 2009) includes BALD and BatchBALD (Houlsby et al., 2011; Gal et al., 2017; Kirsch et al., 2019). Expected predictive information gain (EPIG) averages predictive mutual information over a target distribution (Bickford Smith et al., 2023), analogous to our deployment distribution. BEMPS maximizes proper-score improvement and proves convergence (Tan et al., 2021); DUSAL derives the score from clinical utility. Core-set, BADGE, and uncertainty herding diversify representations, gradients, or coverage (Sener & Savarese, 2018; Ash et al., 2020; Bae et al., 2025); DUSAL-Div instead uses deployment-weighted BEMPS score-change clustering.

Decision-aware and loss-driven acquisition. DeGroot (1962)'s information criterion gives EVSI when loss is negative utility. Methods target decision regions (Javdani et al., 2014), treatment-effect decisions (Sundin et al., 2019), optimal-decision uncertainty (Filstroff et al., 2024), or operational cost through soft mean objective cost of uncertainty (Soft-MOCU) (Zhao et al., 2021a). Amortized design learns query and decision policies (Huang et al., 2024). Loss-Driven Bayesian Active Learning (LD-BAL) derives acquisition from a terminal loss (Huang et al., 2026). With matched Bayesian updating and target distribution, the negative DUSAL score yields our exact one-query objective. DUSAL supplies invariant utility-derived proper scores and Hard's exact-regret characterization. Subgradients handle action ties; LD-BAL's weighted-Bregman theorem assumes differentiable potentials.

Utility-tailored scores and calibration. Bayes acts induce proper scores and decision discrepancies (Dawid, 2007; Grünwald & Dawid, 2004), within the proper-score/Bregman framework (Gneiting & Raftery, 2007; Ovcharov, 2018; Waghmare & Ziegel, 2026). Decision-tailored scoring (Johnstone et al., 2011) includes proper losses and generalized variance for random probability estimates (Smith et al., 2026). DUSAL applies this Bregman-variance form to posterior predictives. Related work constructs task-specific proper losses through local curvature matching (Plaud et al., 2026), trains predictors through downstream optimization (Wilder et al., 2019; Elmachtoub & Grigas, 2022), or evaluates and improves decision-aware calibration (Zhao et al., 2021b; Hegazy et al., 2026). DUSAL instead uses a deployment utility table to select training labels. Its Hard construction combines a strictly proper predictive score with an exact decision-regret contribution, characterized up to affine terms. Under exact Bayesian updating, the resulting acquisition separates predictive score improvement from decision information, making the role of utility explicit.

Annotation cost and clinical utility. DUSAL values training labels by deployment utility, distinct from KnapsackBALD's unequal annotation costs (Dossou et al., 2026) and test acquisition at decision time (Piskorz et al., 2025; Ruhrberg Estévez et al., 2025). Imbalance-aware and biomedical acquisition address class coverage, redundancy, or uncertainty in pretrained models (Nuggehalli et al., 2025; Zhang et al., 2026; Lüth et al., 2026; Xie et al., 2026) and complement the deployment-utility objective studied here.

3 UTILITY-TAILORED PROPER-SCORE ACTIVE LEARNING

Let $\mathcal{X}$ be the input space, $\mathcal{Y} = \{1, \dots, K\}$ the outcome space, and $\Delta(\mathcal{Y}) = \{p \in \mathbb{R}^K : p_y \geq 0 \, \forall y \in \mathcal{Y}, \sum_{y \in \mathcal{Y}} p_y = 1\}$ the probability simplex. After n queries, the learner has a labeled set L_n, an unlabeled pool $\mathcal{U}_n$, a posterior $\pi_n(\theta) = p(\theta \mid L_n)$ over a model parameter $\theta \in \Theta$, and posterior predictive distribution $p_n(y \mid x) = \mathbb{E}_{\theta \sim \pi_n}[p(y \mid x, \theta)]$. Let $\mathcal{A} = \{1, \dots, M\}$ be a finite set of deployment actions and $U \in \mathbb{R}^{M \times K}$ an action–utility table, where $U(a, y)$ is the utility of action a under outcome y. The action count M may differ from K; deployment utility is distinct from labeling cost. Writing $u_a = (U(a, 1), \dots, U(a, K))^\top$, the Bayes value at predictive distribution p is $V_U(p) = \max_{a \in \mathcal{A}} u_a^\top p$. An action attaining this maximum is Bayes optimal, and its Bayes region contains all such p. Decision boundaries require jointly optimal actions, not merely pairwise utility equality (Appendix A). A fixed probability measure μ on $\mathcal{X}$ represents deployment; score improvement is averaged over $z \sim \mu$. An unlabeled reference cohort $E = \{z_m\}_{m=1}^N$ defines the empirical measure $\widehat{\mu} = \sum_{m=1}^N \omega_m \delta_{z_m}$, where δ_{z_m} is a point mass at z_m, $\omega_m \geq 0$, and $\sum_m \omega_m = 1$. Without an external cohort, the initial pool is a declared proxy; either way, reference points and weights are fixed before acquisition. Hard and Soft specify utility-derived potentials; Div is a batch-selection extension using their score-change profiles (Appendix Table 3). We use Hard and Soft as shorthand where unambiguous.

3.1 BAYES-VALUE AUGMENTATION OF A PROPER SCORE

Let $H : \mathcal{D}_H \to \mathbb{R}$ be a finite convex potential on a convex prediction domain $\mathcal{D}_H \subseteq \Delta(\mathcal{Y})$. At each $q \in \mathcal{D}_H$, choose a finite relative subgradient $h_H(q)$ satisfying $H(p) \geq H(q) + \langle h_H(q), p - q \rangle$ for every $p \in \mathcal{D}_H$. The resulting score and Bregman divergence are

$$S_H(q, y) = H(q) + \langle e_y - q, h_H(q) \rangle, \qquad D_H(p, q) = H(p) - H(q) - \langle h_H(q), p - q \rangle, \quad (1)$$

where $e_y \in \Delta(\mathcal{Y})$ places unit mass on outcome y. We assume $D_H(p, q) > 0$ whenever $p \neq q$, so S_H is strictly proper on $\mathcal{D}_H$. Algorithms 1–4 (Appendix C) default to the quadratic base $H_{\mathrm{quad}}(p) = \|p\|_2^2/2$ on the full simplex, for which $h_{H_{\mathrm{quad}}}(q) = q$ and $D_{H_{\mathrm{quad}}}(p, q) = \|p - q\|_2^2/2$. Other bases require finite relative subgradients throughout their declared prediction domains.

To prevent redundant actions from changing normalization and acquisition, first collapse rows a and b when $u_a^\top p = u_b^\top p$ for every $p \in \mathcal{D}_H$. Write $\mathrm{ri}(\mathcal{D}_H)$ for its relative interior. Among the remaining rows, retain action a exactly when some $p \in \mathrm{ri}(\mathcal{D}_H)$ satisfies $u_a^\top p > u_b^\top p$ for every $b \in \mathcal{A} \setminus \{a\}$; call this set $\mathcal{A}_{\mathrm{eff}}$. Discarded rows are envelope-redundant, including dominated actions and actions optimal only at ties. Lemma 1 ensures $\mathcal{A}_{\mathrm{eff}}$ is nonempty and preserves V_U on $\mathcal{D}_H$.

Because reduction can remove tied actions, we define the deployment tie rule afterward. All subsequent maxima use $\mathcal{A}_{\mathrm{eff}}$:

$$V_U(p) = \max_{a \in \mathcal{A}_{\mathrm{eff}}} u_a^\top p, \qquad a_U(p) \in \arg\max_{a \in \mathcal{A}_{\mathrm{eff}}} u_a^\top p,$$

The rule returns $a_U(p)$; regret $\mathrm{Reg}_U(q; p) = V_U(p) - u_{a_U(q)}^\top p$ evaluates the action selected by prediction q under distribution p. We normalize by the largest attainable regret:

$$L_U^{\mathrm{dec}} = \sup_{a \in \mathcal{A}_{\mathrm{eff}},\, p \in \mathcal{D}_H} \{V_U(p) - u_a^\top p\} = \sup_{p, q \in \mathcal{D}_H} \mathrm{Reg}_U(q; p). \quad (2)$$

The supremum allows $\mathcal{D}_H$ to be relatively open. On the full simplex it has the closed form

$$L_U^{\mathrm{dec}} = \max_{a, b \in \mathcal{A}_{\mathrm{eff}},\, y \in \mathcal{Y}} \{U(b, y) - U(a, y)\}.$$

Let $\bar{U}_y^{\mathrm{eff}} = |\mathcal{A}_{\mathrm{eff}}|^{-1} \sum_{a \in \mathcal{A}_{\mathrm{eff}}} U(a, y)$. If $L_U^{\mathrm{dec}} > 0$, define $\widetilde{U}(a, y) = \{U(a, y) - \bar{U}_y^{\mathrm{eff}}\}/L_U^{\mathrm{dec}}$ and let $\widetilde{u}_a$ denote its ath row as a column vector. If $L_U^{\mathrm{dec}} = 0$, set $\widetilde{U} = 0$. Centering removes outcome-specific offsets shared by all actions; scaling confines normalized regret to $[0, 1]$.

For a dimensionless utility strength $\lambda \geq 0$, DUSAL-Hard (Algorithm 1) augments the base potential by the normalized Bayes value:

$$G_{H,U,\lambda}(p) = H(p) + \lambda \max_{a \in \mathcal{A}_{\mathrm{eff}}} \widetilde{u}_a^\top p. \quad (3)$$

The maximum is convex and piecewise linear; it is nondifferentiable only on envelope-active tie sets. We call H m_H-strongly convex when some $m_H > 0$ satisfies $D_H(p, q) \geq m_H \|p - q\|_2^2/2$ for all $p, q \in \mathcal{D}_H$. Convex augmentation preserves this bound, so $G_{H,U,\lambda}$ remains m_H-strongly convex.

For prediction q, let $a_q = a_U(q)$. The maximizing row $\widetilde{u}_{a_q}$ is a valid subgradient even at ties, giving $g_{H,U,\lambda}(q) = h_H(q) + \lambda \widetilde{u}_{a_q}$. The generalized Savage construction (Savage, 1971; Gneiting & Raftery, 2007; Waghmare & Ziegel, 2026) then yields

$$S_{H,U,\lambda}(q, y) = G_{H,U,\lambda}(q) + \langle e_y - q, g_{H,U,\lambda}(q) \rangle = S_H(q, y) + \lambda \widetilde{U}(a_q, y). \tag{4}$$

Writing $S(q, p) = \sum_y p_y S(q, y)$ for the expected score under p, the selected Bregman divergence is

$$D_{H,U,\lambda}(p, q) = G_{H,U,\lambda}(p) - G_{H,U,\lambda}(q) - \langle g_{H,U,\lambda}(q), p - q \rangle.$$

The next theorem establishes strict propriety, including at action ties. It identifies this divergence with score regret: base-score regret plus the selected action's scaled decision regret.

Theorem 1 (Utility augmentation and exact regret decomposition)**.** *Let U have finite entries and $\lambda \geq 0$. If $L_U^{\mathrm{dec}} > 0$, then every $p, q \in \mathcal{D}_H$ satisfy*

$$S_{H,U,\lambda}(p, p) - S_{H,U,\lambda}(q, p) = D_{H,U,\lambda}(p, q) = D_H(p, q) + \frac{\lambda}{L_U^{\mathrm{dec}}} \mathrm{Reg}_U(q; p). \tag{5}$$

If $L_U^{\mathrm{dec}} = 0$, then $G_{H,U,\lambda} = H$ and $D_{H,U,\lambda} = D_H$. Thus $S_{H,U,\lambda}$ is strictly proper. Its potential, outcome-wise selected score, and divergence are unchanged under $U' = cU + \mathbf{1}_M \boldsymbol{o}^\top$ for every $c > 0$ and $\boldsymbol{o} \in \mathbb{R}^K$, where $\mathbf{1}_M$ is the M-vector of ones. On $\mathcal{D}_H$, its potential, expected scores $S_{H,U,\lambda}(q, p)$, selected divergence, and clinical regret are unchanged under adding or removing actions discarded by the effective-action reduction. If $\mathrm{aff}(\mathcal{D}_H) = \mathrm{aff}(\Delta(\mathcal{Y}))$, the outcome-wise selected score is unchanged under this reduction. If H is m_H-strongly convex, then $D_{H,U,\lambda}(p, q) \geq m_H \|p - q\|_2^2/2$.

Theorem 1 is proved in Appendix B. The regret term is classical decision-problem discrepancy (Grünwald & Dawid, 2004; Dawid, 2007). The converse shows that a convex augmentation whose divergence equals scaled decision regret must be scaled Bayes value plus an affine term.

Theorem 2 (Exact-regret characterization)**.** *Fix the effective decision problem and let $F : \mathcal{D}_H \to \mathbb{R}$ be finite and convex. For a selected finite relative subgradient $f(q)$, define $D_F(p, q) = F(p) - F(q) - \langle f(q), p - q \rangle$. Given $\kappa \geq 0$,*

$$D_F(p, q) = \kappa \, \mathrm{Reg}_U(q; p) \quad \textit{for every } p, q \in \mathcal{D}_H \tag{6}$$

if and only if $F = \kappa V_U + \ell$ on $\mathcal{D}_H$ for an affine function ℓ, with a compatible selection of relative subgradients. Thus Bayes value is the unique exact-regret augmentation up to affine terms, which do not affect score divergence or acquisition. The proof is in Appendix B.

The quadratic-base score and divergence are Brier quantities plus normalized utility and regret, respectively (Appendix A). For $\lambda > 0$, Eq. (5) gives $\mathrm{Reg}_U(q; p) \leq (L_U^{\mathrm{dec}}/\lambda) D_{H,U,\lambda}(p, q)$. Because $\mathrm{Reg}_U(q; p) \leq L_U^{\mathrm{dec}}$ and $\frac{1}{2}\|p - q\|_2^2 \leq 1$ on the simplex, λ directly controls the maximum normalized-regret contribution relative to the quadratic base. Unlike weighting $\log q_y$ by c_y, which elicits cost-distorted probabilities, Eq. (4) derives the outcome term and correction from one potential, preserving the optimum (Gneiting & Ranjan, 2011). For clinical costs C, set $U = -C$ and apply the same reduction. With Bayes risk $R_C(p) = \min_{a \in \mathcal{A}_{\mathrm{eff}}} c_a^\top p$, scale L_C^{dec}, and tie-broken Bayes action $a_C(q)$, the divergence for $L_C^{\mathrm{dec}} > 0$ is

$$D_{H,C,\lambda}(p, q) = D_H(p, q) + \tfrac{\lambda}{L_C^{\mathrm{dec}}} \{c_{a_C(q)}^\top p - R_C(p)\}. \tag{7}$$

The added term measures normalized excess clinical cost; it is set to zero for $L_C^{\mathrm{dec}} = 0$. With binary false-negative and false-positive costs, the Bayes-risk threshold is $t^\star = c_{\mathrm{FP}}/(c_{\mathrm{FN}} + c_{\mathrm{FP}})$. Updates within one action region give no utility gain; updates spanning $t^\star$ do. Appendix A gives the equivalent potential and score, the full binary derivation, and the fixed-quadratic comparison.

3.2 Soft utility-gap geometry

For differentiable H, DUSAL-Soft smooths the hard maximum with log-sum-exp over $\mathcal{A}_{\mathrm{eff}}$, avoiding curvature from redundant rows. For temperature $\tau > 0$, define

$$V_{U,\tau}(p) = \tau \log \sum_{a \in \mathcal{A}_{\mathrm{eff}}} \exp(\widetilde{u}_a^\top p/\tau), \qquad G_{H,U,\lambda,\tau}^{\mathrm{soft}}(p) = H(p) + \lambda V_{U,\tau}(p). \tag{8}$$

Let $\rho_a(p) \propto \exp(\widetilde{u}_a^\top p/\tau)$ be the resulting soft action probabilities. Additional curvature follows utility-gap directions $\widetilde{u}_a - \widetilde{u}_b$, normal to the corresponding indifference hyperplanes. It tends to zero as one action's margins dominate τ. Appendix A gives the gradient, Hessian, and local metric.

The resulting divergence has a closed form in terms of Kullback–Leibler (KL) divergence.

Proposition 1 (Soft divergence)**.** *The potential $G^{\mathrm{soft}}_{H,U,\lambda,\tau}$ generates a strictly proper score, has the representation invariances stated in Theorem 1, and has divergence*

$$D^{\mathrm{soft}}_{H,U,\lambda,\tau}(p,q) = D_H(p,q) + \lambda\tau\,\mathrm{KL}\{\rho(q)\,\|\,\rho(p)\}. \tag{9}$$

Here $\mathrm{KL}(r\|s) = \sum_{a\in\mathcal{A}_{\mathrm{eff}}} r_a \log(r_a/s_a)$. As $\tau \downarrow 0$, $V_{U,\tau}(p) \to \max_{a\in\mathcal{A}_{\mathrm{eff}}} \widetilde{u}_a^\top p$. For fixed $\tau > 0$, the KL term measures displacement between induced soft action distributions; unlike DUSAL-Hard, it is not literal clinical regret. The proof is in Appendix B.1.

3.3 Expected-score acquisition

Let $\mathcal{Z}_n$ collect all nonlabel information available before the round-n label, including candidate state and policy randomization; $\mathcal{Z}_0$ also contains the initial pool and, if random, the reference cohort. Define the pre-label sigma-field $\mathcal{F}_n = \sigma(L_n, \mathcal{Z}_0, \ldots, \mathcal{Z}_n)$. Thus the current candidate set and selected x_n are $\mathcal{F}_n$-measurable, while Y_{x_n} is not. For a candidate x and hypothetical label $\widetilde{y}$ with $p_n(\widetilde{y} \mid x) > 0$, the hypothetical posterior and predictive are $\pi_n^{x,\widetilde{y}}(\theta) \propto \pi_n(\theta)p(\widetilde{y} \mid x, \theta)$ and $p_n^{x,\widetilde{y}}(\cdot \mid z)$. Let $\widetilde{Y}$ denote the corresponding random hypothetical label. For either the fixed hard or soft potential G and its Savage score S_G, define the current deployment score and the gain from querying x by

$$Q_G(L_n) = \mathbb{E}_{z\sim\mu}\mathbb{E}_{\theta\sim\pi_n} \sum_y p(y \mid z, \theta) S_G(p_n(\cdot \mid z), y) = \mathbb{E}_{z\sim\mu} G(p_n(\cdot \mid z)), \tag{10}$$

$$\mathrm{Acq}_G(x \mid L_n) = \mathbb{E}_{\widetilde{Y}\sim p_n(\cdot|x)}[Q_G(L_n \cup \{(x, \widetilde{Y})\})] - Q_G(L_n). \tag{11}$$

The expressions for Q_G agree because p_n is the posterior mean predictive. Exact single-query DUSAL maximizes Eq. (11).

To express this gain as reduced posterior disagreement, let D_G be the selected Bregman divergence of G, $\mathbb{P}_\theta$ the joint law of candidate states, adaptive queries, labels, and policy randomization under model θ, and $\mathbb{P}$ its prior-predictive mixture under π_0. For $p_\theta^z = p(\cdot \mid z, \theta)$ and $p_n^z = p_n(\cdot \mid z)$, let $\mathcal{V}_{n+1}^{x,\widetilde{Y}}$ denote the dispersion after the hypothetical update at x, and define

$$\begin{aligned} \delta_n(z) &= \mathbb{E}_{\theta\sim\pi_n} D_G(p_\theta^z, p_n^z), & \mathcal{V}_n &= \mathbb{E}_{z\sim\mu}\delta_n(z), \\ \Delta_n(x) &= \mathcal{V}_n - \mathbb{E}_{\widetilde{Y}\sim p_n(\cdot|x)}[\mathcal{V}_{n+1}^{x,\widetilde{Y}}], \end{aligned} \tag{12}$$

Here $\delta_n(z)$ is local posterior score dispersion, $\mathcal{V}_n$ its deployment average, and $\Delta_n(x)$ its expected decrease. For fixed x, write $p = p_n(\cdot \mid z)$, $P' = p_n^{x,\widetilde{Y}}(\cdot \mid z)$, and $\mathbb{E}_{z,\widetilde{Y}}$ for averaging over $z \sim \mu$ and $\widetilde{Y} \sim p_n(\cdot \mid x)$. Exact Bayesian updating gives $\mathbb{E}_{\widetilde{Y}}[P'] = p$. The next theorem equates expected score improvement with expected reduction in posterior dispersion. For Hard, it isolates the value of decision information beyond base-score learning.

Theorem 3 (Acquisition and dispersion decomposition)**.** *Suppose all model and hypothetical posterior predictives lie in the declared prediction domain and the displayed expectations are finite. Under exact hypothetical Bayesian updating, every available candidate satisfies*

$$\mathrm{Acq}_G(x \mid L_n) = \mathbb{E}_{z,\widetilde{Y}} D_G(P', p) = \Delta_n(x). \tag{13}$$

For DUSAL-Hard with $L_U^{\mathrm{dec}} > 0$, define the utility EVSI by $\mathrm{EVSI}_U(x \mid L_n) := \mathbb{E}_z\{\mathbb{E}_{\widetilde{Y}} V_U(P') - V_U(p)\}$. Then this common quantity is

$$\begin{aligned} \mathrm{Acq}_{H,U,\lambda}(x \mid L_n) &= \mathbb{E}_{z,\widetilde{Y}}\left[D_H(P', p) + \frac{\lambda}{L_U^{\mathrm{dec}}}\,\mathrm{Reg}_U(p; P')\right] \\ &= \underbrace{\mathbb{E}_{z,\widetilde{Y}} D_H(P', p)}_{\mathrm{Acq}_H(x|L_n)} + \frac{\lambda}{L_U^{\mathrm{dec}}} \underbrace{\mathbb{E}_z\{\mathbb{E}_{\widetilde{Y}} V_U(P') - V_U(p)\}}_{\mathrm{EVSI}_U(x|L_n)}. \end{aligned} \tag{14}$$

For $L_U^{\mathrm{dec}} = 0$, it reduces to $\mathbb{E}_{z,\widetilde{Y}} D_H(P', p)$. For DUSAL-Soft,

$$\mathrm{Acq}^{\mathrm{soft}}_{H,U,\lambda,\tau}(x \mid L_n) = \mathbb{E}_{z,\widetilde{Y}}\left[D_H(P', p) + \lambda\tau\,\mathrm{KL}\{\rho(p) \,\|\, \rho(P')\}\right]. \tag{15}$$

The EVSI term is DeGroot's information for loss $-U$ (DeGroot, 1962): current p selects the action and updated P' evaluates it. Utility gain arises when the update reveals a better action; the base learns probabilities for the same reference patients. Exact acquisition and dispersion retain representation invariance and are tie-independent. Theorem 3 and Corollary 1 are proved in Appendix B.2.

3.4 Posterior-score convergence and decision consequences

Because finite pools are eventually exhausted, infinite-horizon analysis uses a continuing candidate process $\{\mathcal{C}_n\}$, such as a stream or replenished pool.

Assumption 1 (Finite informative exact-Bayesian process)**.** *$\mathcal{X}$, $\mathcal{Y}$, and Θ are finite, the true model $\theta^\star$ has $\pi_0(\theta^\star) > 0$, and Bayesian updates are exact. Every model predictive and posterior predictive lies in $\mathcal{D}_H$, $\mathcal{V}_0 < \infty$, and all inputs defining G remain fixed. The nonempty $\mathcal{C}_n$ and selected $x_n \in \mathcal{C}_n$ are $\mathcal{F}_n$-measurable. For every n, $\vartheta \in \Theta$, and $y \in \mathcal{Y}$, candidate generation and pre-label policy randomization are sequentially ignorable and labels follow the model:*

$$\mathbb{P}(\theta = \vartheta \mid \mathcal{F}_n) = \pi_n(\vartheta), \qquad \mathbb{P}_\vartheta(Y_n = y \mid \mathcal{F}_n) = p(y \mid x_n, \vartheta), \tag{16}$$

where $Y_n = Y_{x_n}$ and the first identity holds $\mathbb{P}$-almost surely. For every $\varepsilon \in \mathbb{Q}_{>0}$, a deterministic $c_\varepsilon > 0$ satisfies, $\mathbb{P}_{\theta^\star}$-almost surely for all n, $\mathcal{V}_n \geq \varepsilon \Rightarrow \max_{x\in\mathcal{C}_n} \Delta_n(x) \geq c_\varepsilon$. Moreover, either (i) some deterministic $\beta_{\mathrm{gr}} \in (0, 1]$ makes $\Delta_n(x_n) \geq \beta_{\mathrm{gr}} \max_x \Delta_n(x)$ eventually almost surely; or (ii), for every such ε, $\mathcal{V}_n \geq \varepsilon$ infinitely often implies $\Delta_n(x_n) \geq c_\varepsilon$ infinitely often, almost surely.

Condition (16) allows the process to adapt to L_n, but rules out additional information about θ. It is satisfied by fixed pools and by θ-invariant candidate and policy kernels. Theorem 3 then gives

$$\mathbb{P}(Y_n = y \mid \mathcal{F}_n) = p_n(y \mid x_n), \qquad \mathbb{E}_{\mathbb{P}}[\mathcal{V}_{n+1} \mid \mathcal{F}_n] = \mathcal{V}_n - \Delta_n(x_n). \tag{17}$$

Theorem 4 (Posterior-score convergence and decision consequences)**.** *Under Assumption 1, with $\mathbb{P}_{\theta^\star}$-probability one, the resulting exact single-query process satisfies $\mathcal{V}_n \to 0$ and $D_G(p^z_{\theta^\star}, p^z_n) \to 0$ simultaneously for every z with $\mu(z) > 0$. For DUSAL-Hard at each such z, regret is zero if $L_U^{\mathrm{dec}} = 0$; if $\lambda > 0$ and $L_U^{\mathrm{dec}} > 0$,*

$$\mathrm{Reg}_U(p^z_n; p^z_{\theta^\star}) \leq \frac{L_U^{\mathrm{dec}}}{\lambda} D_G(p^z_{\theta^\star}, p^z_n) \to 0.$$

If, additionally, H is m_H-strongly convex for some $m_H > 0$, then for either proposed potential $D_H(p^z_{\theta^\star}, p^z_n) \to 0$ and $p^z_n \to p^z_{\theta^\star}$, and clinical regret converges to zero for every $\lambda \geq 0$. Whenever regret converges to zero, a unique true Bayes action at z is eventually selected. Proof: Appendix B.3.

Under strong convexity, $\lambda = 0$ reaches the same endpoint when its own progress conditions hold. Utility changes acquisition and Hard directly controls regret, without a proved rate or label-complexity advantage. Random policies must satisfy Assumption 1(ii); Appendix Lemma 3 gives a sufficient condition. For batches, DUSAL-Div clusters candidates by N deployment-weighted score changes and selects the point nearest each of k centroids (Appendix C); convergence is unproved.

4 Experiments

We evaluate how DUSAL's theory translates into decision performance under finite annotation budgets. Theorem 3 identifies normalized EVSI as Hard's utility contribution, while Theorem 4 establishes convergence under its stated assumptions. Controlled experiments examine the acquisition mechanisms, and public-data studies assess their practical effects. We distinguish three outcomes: decision-sensitive scores, changes in label selection, and reductions in held-out decision cost.

4.1 Controlled decision problems

We use finite-hypothesis models with exact posteriors and hypothetical updates to isolate how utility shapes acquisition. Across 1,000 randomized instances, the theoretical identities and invariances hold

to within 7.65×10^{-16}. The following constructions connect these properties to query selection and decision regret; Appendix D.1 provides the detailed settings.

Hard prioritizes information that changes consequential decisions. In E1, a binary screening problem with a 19:1 false-negative-to-false-positive cost ratio, Hard selects queries that can change the optimal action over nuisance queries that produce larger probability changes without affecting decisions. Its sequential extension selects all 50 decision-relevant modules, compared with none for BEMPS, and eliminates terminal decision regret. E2 isolates sensitivity to decision stakes: when two queries induce equal predictive movement across different action boundaries, BEMPS assigns equal scores, whereas Hard favors the query with greater decision value.

Soft extends utility sensitivity to learning within an action region. In E4, neither candidate can change the optimal action in one update, so Hard's EVSI is zero. Soft nevertheless distinguishes the candidates through their utility gaps. Over 100 paired seeds, with $\lambda = 1$ and $\tau = 0.02$, Soft reduces mean cumulative regret from 1.870 for BEMPS to 1.446, a 22.7% reduction under nuisance-first tie-breaking. Reductions remain positive across all examined tie rules, ranging from 14.3% to 24.4%.

P3 demonstrates sensitivity to complete action rows. Across 100 paired perturbations of a three-action problem, Hard and Soft change their selected query when action rows change but column sums remain fixed. BEMPS and the column-weighted quadratic comparator keep their selections. E5 quantifies the predictive benefit of retaining the proper-score base. Across 100 paired finite worlds, Hard reduces Brier and negative-log-likelihood AULCs by 9.09% and 6.02% relative to utility-only acquisition. Decision-regret AULC increases at each of three evaluation cost ratios, by 25.06% under their equal mixture. Together, these controlled studies demonstrate decision-sensitive acquisition and regret reductions in the constructed settings, while quantifying the trade-off between probability accuracy and finite-budget decision performance.

4.2 Public-data protocol

Five studies cover five public tasks (Table 1). Binary studies use BreastMNIST, PneumoniaMNIST (Yang et al., 2023), and adverse drug event (ADE) classification, with five-member ResNet-18 ensembles for images and five MLP heads on frozen PubMedBERT embeddings for ADE. The respective (initial labels, additional budget, batch size) triples are $(40, 300, 20)$, $(100, 1000, 50)$, and $(100, 1500, 100)$, with retraining after each batch. Validation-only searches fix cost ratios before fresh-seed, held-out confirmation. The grouped-action study maps BloodMNIST's eight classes and PathMNIST's nine classes to two actions each, with batches of 50. The cost tables define research decision problems rather than clinician-elicited utilities.

Within studies, compared methods share initial labels, training rules, model and seed streams, and a fixed training-pool reference cohort. BEMPS ($\lambda = 0$) controls for utility augmentation. Replication adds random and entropy sampling (Appendix D.8); the grouped-action study adds Hard with a column-permuted acquisition table, evaluated under the true table (Appendix D.9). Laplace uses five last-layer posterior draws on frozen, dimension-reduced features (Appendix D.10). An outcome-blind second-batch Hard–BEMPS overlap gate admits BloodMNIST and ADE; PathMNIST is exploratory. Div replaces top-b selection for Hard and BEMPS by clustering deployment-weighted score-change profiles with k-means and querying candidates nearest the centroids. These arms join stored grouped-action and replication runs on four tasks with confirmed positive-control benefits, after exact-reproduction checks of environment compatibility (Appendix D.12).

The primary endpoint is budget-normalized held-out realized-cost area under the learning curve (AULC). At each budget, cost averages $C(a(q_i), y_i)/L_U^{\mathrm{dec}}$ over the same held-out outcomes; trapezoidal integration over label count gives AULC. Utility contrasts are DUSAL minus BEMPS, with negative values favoring DUSAL. We use one-sided exact paired signed-rank randomization, Holm control at $\alpha = 0.05$, and marginal paired-bootstrap 95% confidence intervals (CIs). Each binary family evaluates six contrasts. Grouped-action and Laplace families retain six planned contrasts, evaluating four confirmatory contrasts after gating; excluded contrasts remain unsupported, not observed nulls. The batch study tests two four-contrast families, one per score, reusing its parent studies' positive controls. Full protocols, exploratory analyses, and binary matched controls appear in Appendices D and D.5. Binary tables assign costs of zero to correct declarations, one to false positives, and the selected ratio to false negatives. Positive classes are malignancy, pneumonia, and an adverse-drug-event mention.

Table 1: Protocols fixed before each study's confirmatory runs. Superscripts: $+$ positive control, $-$ wrong-table negative control, $\circ$ cross-method agreement gate. Binary ratios are FN:FP; $\lambda = 1$ (binary), 8 (grouped actions); † exploratory; ‡ new arms paired with stored top-b runs from the grouped-action and replication studies after exact reproduction checks (Appendix D.12).

Study (posterior)	Tasks	Arms beyond BEMPS, Hard, Soft	Seeds / runs
Binary confirmation (ensemble)	Breast 1.5:1, Pneumonia 9:1, ADE 1.5:1	–	80 / 720
Replication (ensemble)$^{+}$	the same; other hardware	random, entropy	80 / 1,200
Grouped actions (ensemble)$^{+-}$	Blood 8→2 actions, Path 9→2	random, entropy, wrong table	80, 60 / 840
Laplace posterior$^{+-\circ}$	Blood, ADE; Path†	the same	80, 80, 60 / 1,320
Batch rule (ensemble)$^{+\ddagger}$	Blood, Path, Pneumonia, ADE	Div with BEMPS and Hard	80, 60, 80, 80 / 600

Table 2: Evaluated contrasts: paired relative difference from BEMPS in realized-cost AULC (negative favours DUSAL), 95% CI, Holm-adjusted p within each family. BEMPS $-$ random is the positive control (**bold blue**: resolved; otherwise the cell is descriptive; "–": not run). Orange: adverse interval. †: exploratory, raw p. The final block compares DUSAL-Div with top-b selection under each score, with Holm adjustment within two families of four. Its positive controls come from the studies supplying the stored top-b runs.

Task	BEMPS−random	**Hard** [95% CI]	p	**Soft** [95% CI]	p
Binary confirmation, ensemble					
Breast 1.5:1	–	+0.12% [−1.49%, +1.71%]	1.000	−0.19% [−1.84%, +1.41%]	1.000
Pneumonia 9:1	–	−0.04% [−0.77%, +0.69%]	1.000	+0.03% [−0.71%, +0.81%]	1.000
ADE 1.5:1	–	−0.44% [−1.23%, +0.30%]	1.000	−0.71% [−1.52%, +0.11%]	0.239
Replication, ensemble					
Breast	−0.01%	−0.53% [−2.09, +0.99]	1.000	+0.61% [−0.89, +2.11]	1.000
Pneumonia	**−7.82%**	−0.59% [−1.18, −0.00]	0.117	−0.21% [−0.93, +0.49]	1.000
ADE	**−8.88%**	+0.57% [−0.22, +1.36]	1.000	−0.03% [−0.77, +0.70]	1.000
Grouped actions, ensemble					
Blood	**−10.07%**	−0.67% [−1.82, +0.47]	0.909	+0.57% [−0.75, +1.92]	1.000
Path	**−3.64%**	+0.17% [−1.92, +2.30]	1.000	−0.55% [−2.61, +1.55]	1.000
Laplace posterior					
Blood	**−2.75%**	+0.64% [−0.29, +1.58]	1.000	+1.04% [+0.18, +1.92]	1.000
ADE	−0.50%	−0.03% [−0.28, +0.23]	1.000	−0.10% [−0.36, +0.16]	1.000
Path†	**−8.31%**	−0.33% [−1.45, +0.76]	0.412	−0.26% [−1.26, +0.75]	0.236
*Batch rule, ensemble: DUSAL-Div − top-b on **Hard** (left) and BEMPS (right)*					
Blood	**−10.07%**	−0.37% [−1.73, +0.97]	0.225	−1.12% [−2.50, +0.20]	0.138
Path	**−3.64%**	**−3.85%** [−6.11, −1.69]	0.002	**−3.10%** [−5.57, −0.65]	0.015
Pneumonia	**−7.82%**	**−1.09%** [−1.82, −0.34]	0.008	**−1.27%** [−2.06, −0.48]	0.008
ADE	**−8.88%**	**−1.16%** [−1.90, −0.41]	0.008	**−1.38%** [−2.23, −0.52]	0.006

4.3 HELD-OUT DECISION COST

DUSAL-Div improves held-out decision cost on three of four tasks (Table 2). Relative to Hard top-b selection, Hard-Div reduces realized-cost AULC by 3.85% on PathMNIST, 1.09% on PneumoniaMNIST, and 1.16% on ADE. All three reductions survive Holm correction within the prespecified four-task family. BloodMNIST shows a smaller estimated reduction of 0.37%, below the study's 1.73% minimum detectable effect.

The matched BEMPS comparisons show that these benefits extend across acquisition scores. BEMPS-Div achieves Holm-controlled reductions on the same three tasks, with estimated reductions across all four tasks ranging from 1.12% to 3.10%. This supports batch diversification as the source of the observed gains. Additional utility benefits remain unconfirmed in paired Hard-Div versus BEMPS-Div comparisons. The closest contrast to significance, ADE at +0.80% (95% CI [+0.05,+1.56]), favors BEMPS-Div in direction.

The broader comparisons also demonstrate the value of score-based acquisition: BEMPS reduces cost relative to random sampling in five of seven confirmatory settings, with reductions ranging from 2.75% to 10.07%. The exceptions are BreastMNIST in the replication and ADE under Laplace. Relative to matched BEMPS, Hard and Soft estimates vary in direction, and additional gains remain unconfirmed after Holm correction; these tests establish neither equivalence nor zero effect. Soft's adverse marginal interval on BloodMNIST under Laplace is not a multiplicity-adjusted finding of harm. Figure 4 and Appendix D.7 provide the contrasts and controls. Together, these results identify batch diversification as an effective way to improve finite-budget decision performance in the evaluated settings.

4.4 Acquisition diagnostics

Utility information measurably affects acquisition scores and label selection. First-round Hard utility shares are 0.560, 0.042, and 0.468 at the selected BreastMNIST, PneumoniaMNIST, and ADE costs; grouped-action shares are 0.621 on BloodMNIST and 0.554 on PathMNIST. Binary acquired-set differences from BEMPS average 19–29%. Varying $\lambda \in \{1, 4, 8\}$ changes the first batch, while reproducible cost improvements remain unconfirmed (Appendix D.14).

Hard values improvements in deployment actions. Under exact updating, its utility term vanishes if the current Bayes action remains optimal after every positive-probability hypothetical update at almost every reference input (Theorem 3). Base acquisition then remains intact, even before predictive uncertainty vanishes.

On BloodMNIST, permuted- and true-table acquisition have similar whole-run Jaccard overlaps with BEMPS (0.234 versus 0.233), so selection differences alone do not establish useful decision alignment. Median Hard–BEMPS batch overlap is 43/50 initially and 1/50 in the second ensemble batch; second-batch Laplace overlap is 41/50 (Figure 5**A**). This measures policy agreement along different acquisition histories, not hypothetical-update fidelity against matched refits.

Div changes first-batch composition substantially, sharing a median 0 of 50 items with Hard top-b on BloodMNIST and 0 on PathMNIST. Exploratory curves show early gains on PneumoniaMNIST and ADE diminishing by the final batch; PathMNIST retains gains throughout (Appendix D.12). This pattern is consistent with batch redundancy mattering most when labels are scarce.

Ensemble Hard/BEMPS acquisition takes 12.7/10.5 seconds per round on BloodMNIST and 18.7/13.8 on PathMNIST. Complete BloodMNIST runs with retraining take 27.6/27.1 minutes on one GPU, giving smaller proportional overhead (Appendix D.18). These timings describe the implementation, not exact Bayesian look-ahead.

Baseline ordering depends on the pipeline: Hard's relative cost against entropy on PathMNIST is +5.07% under the ensemble and -10.15% under Laplace. Representation and training also change, preventing attribution to posterior fidelity alone. Appendix D.18 connects these comparisons to refit audits of expected potential gain and posterior sensitivity (Appendices D.16 and D.17).

5 Limitations and Conclusion

DUSAL connects strictly proper scoring with deployment utility. Hard's divergence separates predictive error from normalized decision regret, while exact acquisition separates base-score improvement from decision information (Theorems 1–3). Controlled experiments demonstrate decision-sensitive selection and regret reductions, alongside prediction–decision trade-offs. On public data, utility augmentation measurably changes acquisition scores and label selection. Hard-Div achieves Holm-controlled reductions in realized-cost AULC on three of four tasks relative to Hard top-b selection. Additional utility gains over matched BEMPS remain unconfirmed, while comparable improvements with BEMPS-Div support batch diversification as an effective complement to score-based acquisition in the evaluated settings. The theory assumes fixed finite utilities and, for convergence, finite well-specified models, exact updating, ignorable candidates, and policy-specific progress; it gives no rate or batch guarantee. The experiments (Section 4) use research cost tables; posterior comparisons also change the learning pipeline, and batch agreement does not measure update fidelity. Further evaluations should test externally justified utilities with directly validated hypothetical updates and prespecified decision endpoints.

AI USE STATEMENT

We used generative AI tools to identify, summarize, and analyze relevant literature; provide feedback on experimental design; and assist with interpreting experimental results. Additional uses included polishing the manuscript; formatting references; creating or modifying figures; and creating, editing, and reviewing software code. The authors reviewed all AI-assisted material and take responsibility for the manuscript and associated artifacts, including text, mathematical claims, empirical findings, citations, figures, and code.

REFERENCES

Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. Deep batch active learning by diverse, uncertain gradient lower bounds. In *International Conference on Learning Representations*, 2020.

Wonho Bae, Danica J. Sutherland, and Gabriel L. Oliveira. Uncertainty herding: One active learning method for all label budgets. In *International Conference on Learning Representations*, 2025.

Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom Rainforth. Prediction-oriented bayesian active learning. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 7331–7348, 2023.

Samuel Budd, Emma C. Robinson, and Bernhard Kainz. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71:102062, 2021. doi: 10.1016/j.media.2021.102062.

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

A. Philip Dawid. The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics*, 59(1):77–93, 2007. doi: 10.1007/s10463-006-0099-8.

Morris H. DeGroot. Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics*, 33(2):404–419, 1962. doi: 10.1214/aoms/1177704567.

Bonaventure F. P. Dossou, Ines Arous, Audrey Durand, and Jackie Chi Kit Cheung. Active learning with non-uniform costs for African natural language processing. In *Findings of the Association for Computational Linguistics: EACL 2026*, pp. 6644–6656. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.findings-eacl.349.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, 2022. doi: 10.1287/mnsc.2020.3922.

Louis Filstroff, Iiris Sundin, Petrus Mikkola, Aleksei Tiulpin, Juuso Kylmäoja, and Samuel Kaski. Targeted active learning for bayesian decision-making. *Transactions on Machine Learning Research*, 2024.

Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1183–1192, 2017.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Tilmann Gneiting and Roopesh Ranjan. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422, 2011. doi: 10.1198/jbes.2010.08110.

Peter D. Grünwald and A. Philip Dawid. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *The Annals of Statistics*, 32(4):1367–1433, 2004. doi: 10.1214/009053604000000553.

Mahmoud Hegazy, Michael I. Jordan, and Aymeric Dieuleveut. Scalable utility-aware multiclass calibration. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 397–405, 2026.

José Hernández-Orallo, Peter Flach, and Cèsar Ferri. A unified view of performance metrics: Translating threshold choice into expected classification loss. *Journal of Machine Learning Research*, 13:2813–2869, 2012.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Daolang Huang, Yujia Guo, Luigi Acerbi, and Samuel Kaski. Amortized bayesian experimental design for decision-making. In *Advances in Neural Information Processing Systems*, volume 37, pp. 109460–109486, 2024. doi: 10.52202/079017-3475.

Zhuoyue Huang, Freddie Bickford Smith, and Tom Rainforth. Loss-driven bayesian active learning. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*, pp. 5140–5148, 2026.

Shervin Javdani, Yuxin Chen, Amin Karbasi, Andreas Krause, J. Andrew Bagnell, and Siddhartha Srinivasa. Near optimal bayesian active learning for decision making. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pp. 430–438, 2014.

David J. Johnstone, Victor Richmond R. Jose, and Robert L. Winkler. Tailored scoring rules for probabilities. *Decision Analysis*, 8(4):256–268, 2011. doi: 10.1287/deca.1110.0216.

Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. BatchBALD: Efficient and diverse batch acquisition for deep bayesian active learning. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388, 2010. doi: 10.1097/EDE.0b013e3181d61eeb.

Carsten T. Lüth, Jeremias Traub, Kim-Celine Kahl, Till Bungert, Lukas Klein, Lars Kraemer, Paul F. Jaeger, Klaus Maier-Hein, and Fabian Isensee. Finally outshining the random baseline: A simple and effective solution for active learning in 3d biomedical imaging. *Transactions on Machine Learning Research*, 2026.

Shyam Nuggehalli, Jifan Zhang, Lalit Jain, and Robert Nowak. Improved algorithm for deep active learning under imbalance via optimal separation. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 46815–46836, 2025.

Evgeni Y. Ovcharov. Proper scoring rules and bregman divergence. *Bernoulli*, 24(1):53–79, 2018. doi: 10.3150/16-BEJ857.

Julianna Piskorz, Nicolás Astorga, Jeroen Berrevoets, and Mihaela van der Schaar. Active feature acquisition for personalised treatment assignment. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 4330–4338, 2025.

Roman Plaud, Alexandre Perez-Lebel, Antoine Saillenfest, Thomas Bonald, Marine Le Morvan, Gaël Varoquaux, and Matthieu Labeau. Tailoring strictly proper scoring rules for downstream tasks: An application to causal inference. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026.

Silas Ruhrberg Estévez, Nicolás Astorga, and Mihaela van der Schaar. Timely clinical diagnosis through active test selection. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971. doi: 10.1080/01621459.1971.10482346.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*, 2018.

Burr Settles. Active learning literature survey. Technical Report 1648, University of Wisconsin–Madison, Department of Computer Sciences, 2009.

Zachary Smith, J. Eric Bickel, Richard Field, Jr., and Michael Darling. Decision-tailored uncertainty quantification for probability estimates. *Decision Analysis*, 23(3):248–259, 2026. doi: 10.1287/deca.2025.0421.

Iiris Sundin, Peter Schulam, Eero Siivola, Aki Vehtari, Suchi Saria, and Samuel Kaski. Active learning for decision-making from imbalanced observational data. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6046–6055, 2019.

Wei Tan, Lan Du, and Wray Buntine. Diversity enhanced active learning with strictly proper scoring rules. In *Advances in Neural Information Processing Systems*, volume 34, pp. 10906–10918, 2021.

Kartik Waghmare and Johanna Ziegel. Proper scoring rules for estimation and forecast evaluation. *Annual Review of Statistics and Its Application*, 13:271–296, 2026. doi: 10.1146/annurev-statistics-042424-050626.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658.

Zhuofan Xie, Zishan Lin, Jinliang Lin, Jie Qi, Shaohua Hong, and Shuo Li. Similarity-as-evidence: Calibrating overconfident vlms for interpretable and label-efficient medical active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20973–20984, 2026.

Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. MedMNIST v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10:41, 2023. doi: 10.1038/s41597-022-01721-8.

Jiancheng Zhang, Meiqing Li, Qi Zhang, and Yinglun Zhu. Active learning with foundation model priors: Efficient learning under class imbalance. In *Forty-third International Conference on Machine Learning*, 2026.

Guang Zhao, Edward R. Dougherty, Byung-Jun Yoon, Francis J. Alexander, and Xiaoning Qian. Bayesian active learning by soft mean objective cost of uncertainty. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 3970–3978, 2021a.

Shengjia Zhao, Michael Kim, Roshni Sahoo, Tengyu Ma, and Stefano Ermon. Calibrating predictions to decisions: A novel approach to multi-class calibration. In *Advances in Neural Information Processing Systems*, volume 34, 2021b.

Table 3: The score variants and batch extension of DUSAL. Hard and Soft augment the same strictly proper base potential H using the action–utility table U and dimensionless weight $\lambda \geq 0$; $\mathcal{A}_{\text{eff}}$ is the effective action set and $\widetilde{u}_a$ the ath normalized utility row. "Shorthand" is the name used in the experimental sections.

Name	Shorthand	Augmentation of H	What it is for
DUSAL-Hard	Hard	$+\lambda \max_{a \in \mathcal{A}_{\text{eff}}} \widetilde{u}_a^\top p$, the normalized Bayes value (Eq. (3))	Exact decision-regret decomposition (Thm. 1); acquisition splits into BEMPS plus clinical EVSI (Thm. 3)
DUSAL-Soft	Soft	$+\lambda\tau \log \sum_{a \in \mathcal{A}_{\text{eff}}} \exp(\widetilde{u}_a^\top p/\tau)$, a temperature-$\tau$ smoothing of that maximum (Eq. (8))	Differentiable utility-gap geometry; it separates updates that stay inside one Bayes-action region, which Hard values at zero
DUSAL-Div	Div	No change to H; a batch rule over deployment-weighted score-change profiles (Appendix C)	Batch diversity; lower cost on three of four public tasks, with similar gains using base scores (Section 4.3)

A Bayes-Value Score Details

Bayes regions and pairwise boundaries. For each $a \in \mathcal{A}_{\text{eff}}$, define its Bayes region on the declared prediction domain by

$$\mathcal{R}_a = \{p \in \mathcal{D}_H : u_a^\top p = V_U(p)\}.$$

For distinct $a, b \in \mathcal{A}_{\text{eff}}$, their pairwise indifference set and jointly optimal part are

$$\mathcal{H}_{ab} = \{p \in \mathcal{D}_H : (u_a - u_b)^\top p = 0\}, \qquad \mathcal{B}_{ab} = \mathcal{R}_a \cap \mathcal{R}_b = \{p \in \mathcal{D}_H : u_a^\top p = u_b^\top p = V_U(p)\}.$$

Thus $\mathcal{B}_{ab} \subseteq \mathcal{H}_{ab}$ contains exactly the points where a and b are jointly Bayes optimal. At a point in $\mathcal{H}_{ab} \setminus \mathcal{B}_{ab}$, the two actions have equal expected utility but a third action is strictly better. Their equality therefore does not mark a clinical decision boundary between a and b.

Lemma 1 (Effective actions and decision-regret scale)**.** *First collapse rows a and b for which $u_a^\top p = u_b^\top p$ at every $p \in \mathcal{D}_H$. Among the remaining rows, let $\mathcal{A}_{\text{eff}}$ contain action a exactly when there is a point $p \in \text{ri}(\mathcal{D}_H)$ such that $u_a^\top p > u_b^\top p$ for every $b \in \mathcal{A} \setminus \{a\}$. Then $\mathcal{A}_{\text{eff}}$ is nonempty and*

$$\max_{a \in \mathcal{A}} u_a^\top p = \max_{a \in \mathcal{A}_{\text{eff}}} u_a^\top p \quad \textit{for every } p \in \mathcal{D}_H.$$

The set, its envelope, and L_U^{dec} are unchanged by adding or deleting an action that fails this strict-optimality criterion. This statement preserves the Bayes value, not the full identity of actions tied on envelope-active sets. If $\text{aff}(\mathcal{D}_H) = \text{aff}(\Delta(\mathcal{Y}))$, rows collapsed in the first step agree outcome by outcome. On a lower-dimensional $\mathcal{D}_H$, they need only define the same affine functional on $\mathcal{D}_H$. Under $U' = cU + \mathbf{1}_M \boldsymbol{o}^\top$ with $c > 0$, $\mathcal{A}_{\text{eff}}$ is unchanged and $L_{U'}^{\text{dec}} = cL_U^{\text{dec}}$. On $\mathcal{D}_H = \Delta(\mathcal{Y})$, Eq. (2) has the finite closed form displayed in Section 3.1.

Proof. After duplicate collapse, the expected-utility functions $p \mapsto u_a^\top p$ are distinct on $\mathcal{D}_H$. For each pair, their equality set is therefore a proper hyperplane relative to the affine hull of $\mathcal{D}_H$ (or is empty). Finitely many such hyperplanes cannot cover $\text{ri}(\mathcal{D}_H)$, so away from their union all action utilities are distinct and the maximizing action is unique. This proves $\mathcal{A}_{\text{eff}} \neq \emptyset$. For any $p \in \mathcal{D}_H$, choose a sequence of such generic points in $\text{ri}(\mathcal{D}_H)$ converging to p. Each point has a maximizer in $\mathcal{A}_{\text{eff}}$; because the action set is finite, some action appears along a subsequence. Continuity shows that this action also maximizes at p, which proves envelope preservation. An envelope-redundant row cannot create a unique-maximizer region; if it agrees with another row on a relatively open set, duplicate collapse identifies them. It therefore changes neither $\mathcal{A}_{\text{eff}}$ nor the envelope.

Positive affine transformation preserves every strict action comparison and gives $V_{U'}(p) = cV_U(p) + \boldsymbol{o}^\top p$. Hence each regret and its supremum in Eq. (2) scale by c. On $\Delta(\mathcal{Y})$,

$$L_U^{\mathrm{dec}} = \max_{a\in\mathcal{A}_{\mathrm{eff}}} \sup_{p\in\Delta(\mathcal{Y})} \max_{b\in\mathcal{A}_{\mathrm{eff}}} (u_b - u_a)^\top p$$
$$= \max_{a,b\in\mathcal{A}_{\mathrm{eff}},\, y\in\mathcal{Y}} \{U(b,y) - U(a,y)\},$$

because a linear function on the simplex is maximized at a vertex. For a relatively open prediction domain the corresponding value is a supremum, which is why Eq. (2) does not assume attainment. □

For a polyhedral prediction domain and more than one distinct row, the reduction is computable by one linear program per action. Writing $\mathrm{cl}(\mathcal{D}_H)$ for the closure of the prediction domain, solve

$$\gamma_a^\star = \max_{p\in\mathrm{cl}(\mathcal{D}_H),\,\gamma} \gamma \quad \text{subject to} \quad (u_a - u_b)^\top p \geq \gamma \quad \text{for every } b \in \mathcal{A}\setminus\{a\}.$$

Action a is retained exactly when $\gamma_a^\star > 0$; continuity makes this equivalent to strict optimality somewhere in $\mathrm{ri}(\mathcal{D}_H)$. A value $\gamma_a^\star \leq 0$ therefore identifies a removable row. This criterion captures more than componentwise dominance. On the binary simplex, for example, let $u_1 = (2,0)^\top$, $u_2 = (0,2)^\top$, and $u_3 = (0.9, 0.9)^\top$. Neither u_1 nor u_2 dominates u_3 componentwise, yet $\max\{u_1^\top p, u_2^\top p\} \geq 1 > u_3^\top p$ for every p, so action 3 is envelope redundant. Replacing u_3 by $(1,1)^\top$ makes it optimal only at $p = (1/2, 1/2)$, where all three actions tie; it still defines no uniquely optimal region. This is why DUSAL fixes its tie rule after forming $\mathcal{A}_{\mathrm{eff}}$.

Canonicalization. The effective-action center satisfies $\bar{U}_y^{\mathrm{eff}} = |\mathcal{A}_{\mathrm{eff}}|^{-1} \sum_{a\in\mathcal{A}_{\mathrm{eff}}} U(a,y)$. Therefore $\bar{U}_y^{\mathrm{eff}\prime} = c\bar{U}_y^{\mathrm{eff}} + o_y$ and, when $L_U^{\mathrm{dec}} > 0$, $\widetilde{U}' = \widetilde{U}$. Lemma 1 also shows that envelope-redundant rows never enter the center or scale. If the representative of a collapsed equivalence class changes on a lower-dimensional domain, the ambient center can change, but the normalized Bayes-value function on $\mathcal{D}_H$ does not. If $L_U^{\mathrm{dec}} = 0$, the deployable effective actions have zero regret throughout $\mathcal{D}_H$, so the utility augmentation is set to zero. In particular, adding $(-\chi, -\chi)$ to binary rows $(1,0)$ and $(0,1)$ leaves $\mathcal{A}_{\mathrm{eff}}$, $L_U^{\mathrm{dec}} = 1$, and both DUSAL potentials unchanged for every $\chi > 0$.

Lower-dimensional prediction domains. Let $K = 3$, $\mathcal{D}_H = \{p \in \Delta(\mathcal{Y}) : p_1 = p_2\}$, $u_a = (1,0,0)^\top$, and $u_b = (0,1,0)^\top$. Then $u_a^\top p = u_b^\top p$ for every $p \in \mathcal{D}_H$, although $u_a \neq u_b$. Replacing one representative by the other can therefore change the realized utility term in $S_{H,U,\lambda}(q,y)$. It cannot change an expected score or selected divergence on $\mathcal{D}_H$, because $(u_a - u_b)^\top (p - q) = 0$ for every $p, q \in \mathcal{D}_H$. More generally, equivalent rows differ only by a direction annihilating the tangent space of $\mathcal{D}_H$. If $\mathrm{aff}(\mathcal{D}_H) = \mathrm{aff}(\Delta(\mathcal{Y}))$, equality of their functionals on $\mathcal{D}_H$ extends to every simplex vertex and forces equality of all outcome-wise entries.

Quadratic instance. For $H_{\mathrm{quad}}(p) = \|p\|_2^2/2$ and $L_U^{\mathrm{dec}} > 0$,

$$S_{H_{\mathrm{quad}},U,\lambda}(q,y) = q_y - \tfrac{1}{2}\|q\|_2^2 + \lambda\widetilde{U}(a_q,y), \quad D_{H_{\mathrm{quad}},U,\lambda}(p,q) = \tfrac{1}{2}\|p-q\|_2^2 + \tfrac{\lambda}{L_U^{\mathrm{dec}}}\mathrm{Reg}_U(q;p). \quad (18)$$

When $L_U^{\mathrm{dec}} = 0$, both reduce to their base-score counterparts.

Cost representation. For $C \in \mathbb{R}^{M\times K}$, let $c_a = (C(a,1), \ldots, C(a,K))^\top$, set $U = -C$, and apply the effective-action reduction. Define

$$R_C(p) = \min_{a\in\mathcal{A}_{\mathrm{eff}}} c_a^\top p, \qquad L_C^{\mathrm{dec}} = \sup_{a\in\mathcal{A}_{\mathrm{eff}},\, p\in\mathcal{D}_H} \{c_a^\top p - R_C(p)\}.$$

When $L_C^{\mathrm{dec}} > 0$, the potential and score are, up to an affine potential and its outcome-only score term,

$$G_{H,C,\lambda}(p) \equiv H(p) - \tfrac{\lambda}{L_C^{\mathrm{dec}}} R_C(p), \quad S_{H,C,\lambda}(q,y) \equiv S_H(q,y) - \tfrac{\lambda}{L_C^{\mathrm{dec}}} C_{a_C(q),y}, \quad (19)$$

where $a_C(q)$ is the tie-broken minimizer of $c_a^\top q$ over $\mathcal{A}_{\mathrm{eff}}$. Their divergence is Eq. (7). The symbol $\equiv$ omits terms independent of the prediction, which affect neither divergence nor acquisition. To verify Eq. (19), let $\bar{c}^{\mathrm{eff}} = |\mathcal{A}_{\mathrm{eff}}|^{-1} \sum_{a\in\mathcal{A}_{\mathrm{eff}}} c_a$. Then

$$\max_{a\in\mathcal{A}_{\mathrm{eff}}} \widetilde{u}_a^\top p = \frac{(\bar{c}^{\mathrm{eff}})^\top p - R_C(p)}{L_C^{\mathrm{dec}}}.$$

The first term is affine and does not affect Bregman divergence. It adds only the outcome-specific constant $\lambda\bar{c}_y^{\mathrm{eff}}/L_C^{\mathrm{dec}}$ to the score, which is independent of the prediction. This verifies Eq. (19) without changing propriety, acquisition rankings, or the divergence in Eq. (7).

Binary threshold behavior. Relabel binary outcomes and actions as $\{0,1\}$. For $C = \left(\begin{smallmatrix} 0 & c_{\mathrm{FN}} \\ c_{\mathrm{FP}} & 0 \end{smallmatrix}\right)$ and $t = \Pr(Y=1)$, the two action risks are $R_0(t) = c_{\mathrm{FN}} t$ and $R_1(t) = c_{\mathrm{FP}}(1-t)$. Their intersection is $t^\star = c_{\mathrm{FP}}/(c_{\mathrm{FN}} + c_{\mathrm{FP}})$. With current probability t and martingale update T', the utility part of acquisition is

$$\frac{\lambda}{L_C^{\mathrm{dec}}}\{R_C(t) - \mathbb{E}R_C(T')\}.$$

Because R_C is affine on either side of $t^\star$, this Jensen gap is zero if T' stays in one action region and positive if a nondegenerate update places mass on both sides of the threshold.

Proposition 2 (Binary fixed-quadratic degeneracy)**.** *Let $G_A(q) = q^\top Aq/2$ for a fixed symmetric matrix A that is positive on the binary simplex tangent space. Under exact hypothetical updating, the query ranking induced by G_A is identical to that induced by the unweighted quadratic potential.*

Proof. For $s, t \in [0,1]$, write $q(t) = (1-t, t) = e_0 + td$ with $d = (-1,1)^\top$. Along the simplex, $G_A(q(t))$ is a scalar quadratic whose second derivative is the positive constant $\alpha_A = d^\top Ad$. Hence

$$D_A(q(s), q(t)) = \tfrac{1}{2}\alpha_A(s-t)^2.$$

By Theorem 3, the acquisition for every candidate is the expectation of this divergence after its hypothetical update. Relative to the unweighted quadratic, it is therefore multiplied by the same positive constant $\alpha_A/(d^\top d)$ for every candidate, leaving all rankings unchanged. □

Soft derivatives. Let $r_a(p) = \exp(\widetilde{u}_a^\top p/\tau)$ and $\rho_a(p) = r_a(p)/\sum_{b\in\mathcal{A}_{\mathrm{eff}}} r_b(p)$. Standard log-partition differentiation gives

$$\nabla V_{U,\tau}(p) = \widetilde{U}_{\mathrm{eff}}^\top \rho(p), \qquad \nabla^2 V_{U,\tau}(p) = \frac{1}{\tau}\widetilde{U}_{\mathrm{eff}}^\top\{\operatorname{Diag}\rho(p) - \rho(p)\rho(p)^\top\}\widetilde{U}_{\mathrm{eff}}.$$

For the quadratic base, setting $\Sigma_\rho(p) = \operatorname{Diag}\rho(p) - \rho(p)\rho(p)^\top$ gives

$$\nabla G^{\mathrm{soft}}(p) = p + \lambda\widetilde{U}_{\mathrm{eff}}^\top\rho(p), \qquad \nabla^2 G^{\mathrm{soft}}(p) = I_K + \frac{\lambda}{\tau}\widetilde{U}_{\mathrm{eff}}^\top\Sigma_\rho(p)\widetilde{U}_{\mathrm{eff}}. \tag{20}$$

The covariance matrix has the pairwise form

$$\widetilde{U}_{\mathrm{eff}}^\top\{\operatorname{Diag}\rho(p) - \rho(p)\rho(p)^\top\}\widetilde{U}_{\mathrm{eff}} = \frac{1}{2}\sum_{a,b\in\mathcal{A}_{\mathrm{eff}}}\rho_a(p)\rho_b(p)(\widetilde{u}_a - \widetilde{u}_b)(\widetilde{u}_a - \widetilde{u}_b)^\top,$$

which gives the tangent-direction quadratic form

$$h^\top\nabla^2 G^{\mathrm{soft}}(p)h = \|h\|_2^2 + \frac{\lambda}{2\tau}\sum_{a,b\in\mathcal{A}_{\mathrm{eff}}}\rho_a(p)\rho_b(p)\{(\widetilde{u}_a - \widetilde{u}_b)^\top h\}^2. \tag{21}$$

If $|\mathcal{A}_{\mathrm{eff}}| > 1$ and $a^\star$ uniquely maximizes normalized utility at p, let

$$\gamma(p) = \min_{b\in\mathcal{A}_{\mathrm{eff}}\setminus\{a^\star\}}(\widetilde{u}_{a^\star} - \widetilde{u}_b)^\top p > 0.$$

Then

$$1 - \rho_{a^\star}(p) = \frac{\sum_{b\neq a^\star}\exp\{-[(\widetilde{u}_{a^\star} - \widetilde{u}_b)^\top p]/\tau\}}{1 + \sum_{b\neq a^\star}\exp\{-[(\widetilde{u}_{a^\star} - \widetilde{u}_b)^\top p]/\tau\}} \le (|\mathcal{A}_{\mathrm{eff}}| - 1)e^{-\gamma(p)/\tau}.$$

Thus the soft weights concentrate on $a^\star$, and the utility-curvature term vanishes, as $\gamma(p)/\tau \to \infty$. For $|\mathcal{A}_{\mathrm{eff}}| = 1$, that term is already zero.

B Formal Proofs

B.1 Hard and soft score properties

Proof of Theorem 1. If $L_U^{\mathrm{dec}} = 0$, canonicalization sets $\widetilde{U} = 0$, so every claim reduces to the corresponding base-score claim. Suppose $L_U^{\mathrm{dec}} > 0$. The maximum of finitely many affine functions is finite and convex. Because $a_q = a_U(q) \in \mathcal{A}_{\mathrm{eff}}$ attains that maximum, $\widetilde{u}_{a_q}$ is a valid subgradient

of $p \mapsto \max_{a \in \mathcal{A}_{\text{eff}}} \widetilde{u}_a^\top p$ at q. Hence $g_{H,U,\lambda}(q) = h_H(q) + \lambda \widetilde{u}_{a_q}$ is a valid selected subgradient of $G_{H,U,\lambda}$, including at action ties. Substitution in the generalized Savage construction gives

$$\begin{aligned} S_{H,U,\lambda}(q,y) &= H(q) + \lambda \widetilde{u}_{a_q}^\top q + \langle e_y - q, h_H(q) + \lambda \widetilde{u}_{a_q} \rangle \\ &= S_H(q,y) + \lambda \widetilde{U}(a_q, y), \end{aligned}$$

which proves Eq. (4).

Taking expectation under p and using $\max_{a \in \mathcal{A}_{\text{eff}}} \widetilde{u}_a^\top p = \widetilde{u}_{a_U(p)}^\top p$ yields

$$\begin{aligned} S_{H,U,\lambda}(p,p) - S_{H,U,\lambda}(q,p) &= D_H(p,q) + \lambda \{ \max_{a \in \mathcal{A}_{\text{eff}}} \widetilde{u}_a^\top p - \widetilde{u}_{a_q}^\top p \} \\ &= D_H(p,q) + \frac{\lambda}{L_U^{\text{dec}}} \{ V_U(p) - u_{a_q}^\top p \}. \end{aligned}$$

The effective-action center cancels in the second equality. The same expected-score difference equals the selected generalized Bregman divergence by its definition. It is strictly positive whenever $p \neq q$ because $D_H(p,q) > 0$; thus the augmented score is strictly proper. If H is m_H-strongly convex, its selected Bregman divergence is at least $m_H \|p - q\|_2^2/2$, and adding nonnegative regret proves the stated bound. Under a positive affine utility transformation, the canonicalization argument in Appendix A gives $\widetilde{U}' = \widetilde{U}$ exactly, proving outcome-wise score invariance. Adding or removing an envelope-redundant action does not change $\mathcal{A}_{\text{eff}}$. If instead the representative of a collapsed action class changes, Lemma 1 and the lower-dimensional-domain argument show that the normalized Bayes-value functions agree on $\mathcal{D}_H$ and that any difference between corresponding selected subgradients annihilates $p - q$ for $p, q \in \mathcal{D}_H$. Hence the potential, expected scores, selected divergence, and regret agree on $\mathcal{D}_H$. When $\operatorname{aff}(\mathcal{D}_H) = \operatorname{aff}(\Delta(\mathcal{Y}))$, equivalent rows agree outcome by outcome, which proves the final score-invariance claim. □

Proof of Theorem 2. For $a_q = a_U(q)$,

$$D_{V_U}(p,q) = V_U(p) - V_U(q) - u_{a_q}^\top (p - q) = V_U(p) - u_{a_q}^\top p = \operatorname{Reg}_U(q;p).$$

Let $f(q)$ be the selected relative subgradient of F and set $J = F - \kappa V_U$ and $j(q) = f(q) - \kappa u_{a_q}$. The assumed identity gives $J(p) - J(q) - \langle j(q), p - q \rangle = 0$ for every $p, q \in \mathcal{D}_H$. Fixing q_0 shows $J(p) = J(q_0) + \langle j(q_0), p - q_0 \rangle$ throughout $\mathcal{D}_H$, so J is affine there. Conversely, adding an affine function to κV_U and selecting its affine gradient plus κu_{a_q} leaves the divergence equal to $\kappa D_{V_U} = \kappa \operatorname{Reg}_U$. □

Proof of Proposition 1. For the log-partition term, the softmax definition gives

$$\log \frac{\rho_a(q)}{\rho_a(p)} = \frac{\widetilde{u}_a^\top (q - p)}{\tau} - \frac{V_{U,\tau}(q) - V_{U,\tau}(p)}{\tau}.$$

Multiplying by $\tau \rho_a(q)$, summing over $a \in \mathcal{A}_{\text{eff}}$, and using $\nabla V_{U,\tau}(q) = \sum_{a \in \mathcal{A}_{\text{eff}}} \rho_a(q) \widetilde{u}_a$ gives

$$\tau \operatorname{KL}\{\rho(q) \| \rho(p)\} = V_{U,\tau}(p) - V_{U,\tau}(q) - \langle \nabla V_{U,\tau}(q), p - q \rangle.$$

Adding the base divergence D_H proves Eq. (9). Since $D_H(p,q) > 0$ for $p \neq q$ and the KL term is nonnegative, the soft score is strictly proper. Canonicalization and $\mathcal{A}_{\text{eff}}$ are unchanged under positive affine utility transformations. Under effective-action reduction, corresponding normalized logits agree on $\mathcal{D}_H$; their gradients can differ only in directions annihilating $p - q$ for $p, q \in \mathcal{D}_H$. Thus the soft potential, expected scores, and divergence obey the invariance qualifications in Theorem 1; full-dimensionality also gives outcome-wise score invariance. Finally,

$$\max_{a \in \mathcal{A}_{\text{eff}}} \widetilde{u}_a^\top p \leq V_{U,\tau}(p) \leq \max_{a \in \mathcal{A}_{\text{eff}}} \widetilde{u}_a^\top p + \tau \log |\mathcal{A}_{\text{eff}}|,$$

which proves pointwise convergence as $\tau \downarrow 0$. □

B.2 ACQUISITION IDENTITY

Proof of Theorem 3. For fixed z, abbreviate $p_\theta^z = p(\cdot \mid z, \theta)$. Since $p_n^z = \mathbb{E}_{\theta\sim\pi_n} p_\theta^z$, the generalized Savage linear term cancels for any selected subgradient $g_G(p_n^z)$, and averaging this identity over $z \sim \mu$ gives

$$\begin{aligned}\mathbb{E}_{\theta\sim\pi_n} \sum_y p_\theta^z(y) S_G(p_n^z, y) &= G(p_n^z) + \langle \mathbb{E}_{\theta\sim\pi_n} p_\theta^z - p_n^z, g_G(p_n^z)\rangle \\ &= G(p_n^z), \\ \mathcal{V}_n &= \mathbb{E}_{z\sim\mu}\mathbb{E}_{\theta\sim\pi_n} G(p_\theta^z) - Q_G(L_n).\end{aligned}$$

After querying x and averaging its label under the posterior predictive, the Bayesian tower property leaves the conditional expectation of the first term unchanged, while the second term becomes $\mathbb{E}_{\widetilde{y}}[Q_G(L_n \cup \{(x, \widetilde{y})\})]$. Subtracting the expected $\mathcal{V}_{n+1}$ from $\mathcal{V}_n$ yields $\Delta_n(x) = \mathrm{Acq}_G(x \mid L_n)$.

For the pointwise acquisition form, let $P'_{x,z} = p_n^{x,\widetilde{Y}}(\cdot \mid z)$, abbreviate it as P', and set $p = p_n^z$. The Bayesian averaging identity gives $\mathbb{E}_{\widetilde{Y}}[P'] = p$, so

$$\begin{aligned}\mathbb{E}_{\widetilde{Y}} D_G(P', p) &= \mathbb{E}_{\widetilde{Y}} G(P') - G(p) - \langle g_G(p), \mathbb{E}_{\widetilde{Y}} P' - p\rangle \\ &= \mathbb{E}_{\widetilde{Y}} G(P') - G(p).\end{aligned}$$

Averaging over $z \sim \mu$ and substituting Eqs. (5) and (9) gives Eqs. (14) and (15), respectively. For the second line of Eq. (14), let $a_p = a_U(p)$. The averaging identity and optimality at p give

$$\mathbb{E}_{\widetilde{Y}} \mathrm{Reg}_U(p; P') = \mathbb{E}_{\widetilde{Y}} V_U(P') - u_{a_p}^\top \mathbb{E}_{\widetilde{Y}} P' = \mathbb{E}_{\widetilde{Y}} V_U(P') - V_U(p),$$

which is the expected value-of-sample-information representation. □

Corollary 1 (Tie independence of exact acquisition)**.** *For a fixed potential G, the exact quantities Q_G, Acq_G, $\mathcal{V}_n$, and Δ_n do not depend on which valid subgradient is selected at a nondifferentiable point of G. Individual scores and pairwise selected divergences can depend on that selection.*

Proof of Corollary 1. The cancellations in the preceding proof give $Q_G(L_n) = \mathbb{E}_z G(p_n^z)$ and $\mathcal{V}_n = \mathbb{E}_z\{\mathbb{E}_{\theta\sim\pi_n} G(p_\theta^z) - G(p_n^z)\}$. Neither expression contains the selected subgradient. Their conditional increments, Acq_G and Δ_n, are therefore also independent of that selection. Without posterior or hypothetical-label averaging, the linear term need not cancel, so an individual score or selected divergence can change at a nondifferentiable point of G. □

B.3 POSTERIOR MASS AND CONVERGENCE

Lemma 2 (True posterior mass is eventually bounded below)**.** *Under Assumption 1, there exist a random $\underline{\pi}_\star > 0$ and an almost surely finite $N_\star$ such that $\pi_n(\theta^\star) \geq \underline{\pi}_\star$ for all $n \geq N_\star$.*

Proof. Under the true-data law $\mathbb{P}_{\theta^\star}$, define for $\theta \neq \theta^\star$

$$\Lambda_n(\theta) = \prod_{i<n} \frac{p(Y_i \mid x_i, \theta)}{p(Y_i \mid x_i, \theta^\star)}.$$

Here x_i is the queried input and Y_i its observed label at round i. By Eq. (16), the nonlabel history contributes no additional parameter-dependent factor to posterior odds, so Λ_n is the remaining adaptive label-likelihood ratio. Each likelihood-ratio factor is evaluated only on the true-support event $p(Y_i \mid x_i, \theta^\star) > 0$. Values on histories containing a $\mathbb{P}_{\theta^\star}$-null outcome may be assigned arbitrarily. The adaptive x_n is $\mathcal{F}_n$-measurable, and the conditional label law in Assumption 1 gives

$$\mathbb{E}_{\mathbb{P}_{\theta^\star}}[\Lambda_{n+1}(\theta) \mid \mathcal{F}_n] = \Lambda_n(\theta) \sum_{y: p(y\mid x_n, \theta^\star) > 0} p(y \mid x_n, \theta) \leq \Lambda_n(\theta).$$

Thus $\Lambda_n(\theta)$ is a nonnegative supermartingale and converges almost surely to a finite limit. Bayes' rule on true-support histories and finiteness of Θ give

$$\pi_n(\theta^\star) = \frac{\pi_0(\theta^\star)}{\pi_0(\theta^\star) + \sum_{\theta\neq\theta^\star} \pi_0(\theta)\Lambda_n(\theta)} \to \frac{\pi_0(\theta^\star)}{\pi_0(\theta^\star) + \sum_{\theta\neq\theta^\star} \pi_0(\theta)\Lambda_\infty(\theta)} =: \pi_\infty(\theta^\star) > 0.$$

Take $\underline{\pi}_\star = \pi_\infty(\theta^\star)/2$ and use convergence to obtain $N_\star$. □

Lemma 3 (Randomized persistent-exploration criterion)**.** *Let $\{\mathcal{G}_n\}$ be the filtration immediately before each policy draw, so that $\mathcal{G}_n \subseteq \mathcal{F}_n \subseteq \mathcal{G}_{n+1}$, where $\mathcal{F}_n$ includes the selected x_n but precedes its label. Suppose $\mathcal{C}_n$, $\mathcal{V}_n$, and $\{\Delta_n(x) : x \in \mathcal{C}_n\}$ are $\mathcal{G}_n$-measurable. If every $\varepsilon \in \mathbb{Q}_{>0}$ has a deterministic $\eta_\varepsilon > 0$ such that, almost surely for all n,*

$$\mathbb{P}_{\theta^\star}(\Delta_n(x_n) \geq c_\varepsilon \mid \mathcal{G}_n) \geq \eta_\varepsilon \mathbf{1}\{\mathcal{V}_n \geq \varepsilon\}, \tag{22}$$

then policy branch (ii) in Assumption 1 holds. In particular, conditional uniform sampling satisfies this criterion when $|\mathcal{C}_n| \leq M < \infty$ uniformly and the candidate-progress condition in Assumption 1 holds.

Proof. Fix $\varepsilon \in \mathbb{Q}_{>0}$ and let $A_{n,\varepsilon} = \{\mathcal{V}_n \geq \varepsilon, \Delta_n(x_n) \geq c_\varepsilon\}$. On $\{\mathcal{V}_n \geq \varepsilon \text{ i.o.}\}$, the displayed lower bound implies $\sum_n \mathbb{P}_{\theta^\star}(A_{n,\varepsilon} \mid \mathcal{G}_n) = \infty$. The conditional Borel–Cantelli lemma therefore gives $A_{n,\varepsilon}$ infinitely often, proving branch (ii); countability of $\mathbb{Q}_{>0}$ gives a common probability-one event. Under uniform sampling, candidate progress makes at least one member of $\mathcal{C}_n$ c_ε-informative whenever $\mathcal{V}_n \geq \varepsilon$, so the conditional selection probability is at least $1/M$. □

Proof of Theorem 4. Let $\mathbb{P} = \sum_\theta \pi_0(\theta)\mathbb{P}_\theta$ be the Bayesian prior-mixture law. By Eq. (16), $\pi_n(\vartheta) = \mathbb{P}(\theta = \vartheta \mid \mathcal{F}_n)$ for every $\vartheta \in \Theta$, so the posterior weights are $\mathcal{F}_n$-martingales. Their fixed linear combinations $p_n^z = \sum_\vartheta \pi_n(\vartheta) p_\vartheta^z$ are martingales as well. For either fixed DUSAL potential, the linear term cancels and

$$\delta_n(z) = \sum_\theta \pi_n(\theta) G(p_\theta^z) - G(p_n^z).$$

The first term is a martingale and the second is a submartingale by convexity of G. Therefore

$$\mathbb{E}_{\mathbb{P}}[\delta_{n+1}(z) \mid \mathcal{F}_n] \leq \delta_n(z),$$

so $\mathcal{V}_n$ is a nonnegative supermartingale and converges almost surely. Write $\zeta_n = \Delta_n(x_n) \geq 0$ for the realized one-step progress. By Eq. (17),

$$\mathbb{E}_{\mathbb{P}}[\mathcal{V}_{n+1} \mid \mathcal{F}_n] = \mathcal{V}_n - \zeta_n.$$

Taking expectations and telescoping yields

$$\sum_{n=0}^{\infty} \mathbb{E}_{\mathbb{P}}[\zeta_n] \leq \mathbb{E}_{\mathbb{P}}[\mathcal{V}_0] < \infty.$$

Tonelli's theorem gives $\sum_n \zeta_n < \infty$ $\mathbb{P}$-almost surely. Since $\pi_0(\theta^\star) > 0$, $\mathbb{P}_{\theta^\star} \ll \mathbb{P}$, so convergence of $\mathcal{V}_n$ and summability also hold $\mathbb{P}_{\theta^\star}$-almost surely. Countability of $\mathbb{Q}_{>0}$ gives a common $\mathbb{P}_{\theta^\star}$-probability-one event for the progress and applicable policy conditions in Assumption 1. Work on its intersection with the convergence and summability events.

If $\mathbb{P}_{\theta^\star}(\mathcal{V}_\infty > 0) > 0$, then for some rational $\varepsilon > 0$ the event $\Omega_\varepsilon = \{\mathcal{V}_\infty \geq \varepsilon\}$ has positive probability. Its intersection with the common probability-one event above still has positive probability, and there convergence implies $\mathcal{V}_n \geq \varepsilon/2$ eventually. The progress condition then gives $\max_{x \in \mathcal{C}_n} \Delta_n(x) \geq c_{\varepsilon/2}$. In the β_{gr}-greedy branch, $\zeta_n \geq \beta_{\mathrm{gr}} c_{\varepsilon/2}$ eventually; in the persistent exploration branch, $\zeta_n \geq c_{\varepsilon/2}$ infinitely often. Either branch contradicts summability, so $\mathcal{V}_n \to 0$ $\mathbb{P}_{\theta^\star}$-almost surely.

For $z \in \mathrm{supp}(\mu)$, finiteness of $\mathcal{X}$ and nonnegativity give $0 \leq \delta_n(z) \leq \mathcal{V}_n/\mu(z) \to 0$. Lemma 2 then gives, eventually,

$$D_G(p_{\theta^\star}^z, p_n^z) \leq \frac{\delta_n(z)}{\pi_n(\theta^\star)} \leq \frac{\delta_n(z)}{\underline{\pi}_\star} \to 0.$$

For the hard potential, regret is identically zero when $L_U^{\mathrm{dec}} = 0$. If $\lambda > 0$ and $L_U^{\mathrm{dec}} > 0$, Eq. (5) and nonnegativity of D_H give

$$\mathrm{Reg}_U(p_n^z; p_{\theta^\star}^z) \leq (L_U^{\mathrm{dec}}/\lambda) D_G(p_{\theta^\star}^z, p_n^z) \to 0.$$

This route does not use strong convexity.

Now suppose H is m_H-strongly convex. For both proposed potentials, Theorem 1 and Proposition 1 give

$$D_G(p_{\theta^\star}^z, p_n^z) \geq D_H(p_{\theta^\star}^z, p_n^z) \geq \frac{m_H}{2} \|p_{\theta^\star}^z - p_n^z\|_2^2.$$

Table 4: Per-round acquisition complexity with streaming candidate scoring.

Method	Time	Additional memory
DUSAL-Hard	$O(P_n K[C_{\mathrm{upd}} + N(C_{\mathrm{pred}} + C_{\mathrm{pot}})])$	$O(NK + b_n + M_{\mathrm{hyp}})$
DUSAL-Div	$O(P_n K[C_{\mathrm{upd}} + N(C_{\mathrm{pred}} + C_{\mathrm{pot}})] + I_{\mathrm{km}} P_n b_n N)$	$O(P_n N + NK + b_n N + M_{\mathrm{hyp}})$
BatchBALD	$O(P_n S_{\mathrm{MC}} C_{\mathrm{pred}} + b_n P_n K S_{\mathrm{MC}} J_{b_n})$	$O(P_n S_{\mathrm{MC}} K + J_{b_n}(S_{\mathrm{MC}} + K))$
BADGE	$O(P_n C_{\mathrm{grad}} + P_n b_n d_{\mathrm{grad}})$	$O(P_n d_{\mathrm{grad}})$

Hence both base divergence and predictive error converge to zero. Let $a_n = a_U(p_n^z)$ and let $a_\star$ be optimal under $p_{\theta^\star}^z$. Optimality of a_n under p_n^z gives, for either potential and every $\lambda \geq 0$,

$$\begin{aligned}\mathrm{Reg}_U(p_n^z; p_{\theta^\star}^z) &\leq |u_{a_\star}^\top (p_{\theta^\star}^z - p_n^z)| + |u_{a_n}^\top (p_n^z - p_{\theta^\star}^z)| \\ &\leq 2 \max_{a \in \mathcal{A}_{\mathrm{eff}}} \|u_a\|_\infty \|p_n^z - p_{\theta^\star}^z\|_1 \to 0.\end{aligned}$$

Finally, whenever either route has established regret convergence, uniqueness of the true Bayes action gives a positive utility gap between it and every other effective action. A different selected action would incur at least this gap, so the true action is eventually selected. □

C Formal Algorithms

For reference point z_m, let $Q_{G,m}(L)$ denote Eq. (10) without the outer expectation over z. The weighted score-change embedding is

$$\begin{aligned} g_{G,m}(x \mid L_n) &= \sum_{\widetilde{y} \in \mathcal{Y}} p_n(\widetilde{y} \mid x)[Q_{G,m}(L_n \cup \{(x, \widetilde{y})\}) - Q_{G,m}(L_n)], \\ \boldsymbol{v}_G(x \mid L_n) &= \big(\sqrt{\omega_m}\, g_{G,m}(x \mid L_n)\big)_{m=1}^N. \end{aligned} \tag{23}$$

Uniform weights recover the unweighted embedding up to a common scale.

Computational cost. Table 4 excludes the fitted model, stored pool, and model retraining after acquisition, which are shared across methods. Let $P_n = |\mathcal{U}_n|$, $M_{\mathrm{eff}} = |\mathcal{A}_{\mathrm{eff}}|$, b_n be the realized batch size, and I_{km} the number of Lloyd iterations. Let C_{upd} be the cost of one candidate–label hypothetical update, C_{pred} one predictive evaluation, and $C_{\mathrm{pot}} = O(M_{\mathrm{eff}} K)$ one hard-potential evaluation; the temporary hypothetical state occupies M_{hyp} memory. For BatchBALD, the Monte Carlo (MC) sample count S_{MC} is the number of stochastic model samples and $J_{b_n} = \min\{K^{b_n}, J\}$ caps the enumerated or sampled joint-label configurations. For BADGE, $d_{\mathrm{grad}} = Kd$ is the last-layer gradient embedding dimension for representation width d, and C_{grad} includes forming one such embedding.

The first two rows of Table 4 follow because each of the $P_n K$ candidate–label hypotheses requires one hypothetical update and N reference predictions. Once $\mathcal{A}_{\mathrm{eff}}$ is fixed, direct hard-potential evaluation costs $C_{\mathrm{pot}} = O(M_{\mathrm{eff}} K)$ per reference prediction; the soft potential has the same order. DUSAL-Hard sums reference contributions as they are produced and need not store a pool-by-reference array. DUSAL-Div instead retains $P_n N$ expected score-change coordinates. Its full-pool k-means call with $k = b_n$ adds $O(I_{\mathrm{km}} P_n b_n N)$ time and $O(b_n N)$ centroid storage. These are streaming bounds; fully materializing every hypothetical class-probability tensor would require substantially more memory and is unnecessary. In deep implementations, the $P_n K$ hypothetical model updates may dominate reference scoring and clustering; Table 4 therefore keeps their cost explicit as C_{upd}.

The BatchBALD row uses its Monte Carlo-capped greedy bound (Kirsch et al., 2019). Its J_{b_n} factor grows exponentially with batch size before truncation, and cached stochastic pool predictions occupy $O(P_n S_{\mathrm{MC}} K)$ memory. The BADGE row assumes the original k-means++ seeding over d_{grad}-dimensional gradient embeddings (Ash et al., 2020); its b_n center updates cost $O(P_n b_n d_{\mathrm{grad}})$ and the embedding matrix costs $O(P_n d_{\mathrm{grad}})$ memory. Neither baseline depends on N. Thus fixing the reference cohort makes DUSAL linear in pool size, whereas taking $N = \Theta(P_n)$ makes its reference-scoring work quadratic; DUSAL-Div also incurs quadratic embedding memory in that regime. In practice, N should therefore be fixed independently of P_n or controlled by a declared reference-cohort subsampling rule.

Algorithm 1 Exact single-query DUSAL-Hard

Require: Pool $\mathcal{U}_0$, seed L_0, base potential H and selected relative subgradient h_H, utility table U, strength λ, tie order, reference measure μ, label budget B

Construct $\mathcal{A}_{\text{eff}}$, L_U^{dec}, and $\widetilde{U}$; construct $G_{H,U,\lambda}$ and $S_{H,U,\lambda}$ by Eqs. (3)–(4); freeze all inputs
for $n = 0, \dots, \min\{B, |\mathcal{U}_0|\} - 1$ **do**
  Update the exact posterior π_n and predictive p_n
  for all $x \in \mathcal{U}_n$ **do**
    Compute $\mathrm{Acq}_{H,U,\lambda}(x \mid L_n)$ by Eq. (11)
  end for
  $x_n \leftarrow \arg\max_{x \in \mathcal{U}_n} \mathrm{Acq}_{H,U,\lambda}(x \mid L_n)$
  Query y_n; set $L_{n+1} \leftarrow L_n \cup \{(x_n, y_n)\}$ and $\mathcal{U}_{n+1} \leftarrow \mathcal{U}_n \setminus \{x_n\}$
end for

Algorithm 2 Practical pool-based DUSAL

Require: Pool $\mathcal{U}_0$, seed L_0, base potential H (differentiable in soft mode), utility table U, λ, mode $\in \{\mathrm{hard}, \mathrm{soft}\}$ (default hard), τ if soft, tie order, reference pool E, weights $\{\omega_m\}_{m=1}^N$, label budget B, requested batch size b

Construct and freeze $\mathcal{A}_{\text{eff}}$, L_U^{dec}, and $\widetilde{U}$
if mode is hard **then**
  Construct $G_{H,U,\lambda}$ by Eq. (3)
else
  Construct $G^{\text{soft}}_{H,U,\lambda,\tau}$ over $\mathcal{A}_{\text{eff}}$ by Eq. (8)
end if
Freeze the score, utility parameters, tie order, and reference weights
$n \leftarrow 0$
while $|L_n| - |L_0| < B$ and $\mathcal{U}_n \neq \emptyset$ **do**
  $b_n \leftarrow \min\{b, B - (|L_n| - |L_0|), |\mathcal{U}_n|\}$
  Fit the declared probabilistic model or posterior approximation on L_n
  Estimate p_n on $\mathcal{U}_n$ and E
  Compute scalar acquisitions for plain DUSAL, or the embeddings in Algorithm 3 for DUSAL-Div
  Select the top-b_n candidates for plain DUSAL, or call Algorithm 4 with b_n for DUSAL-Div
  Query the oracle; form L_{n+1} and $\mathcal{U}_{n+1}$, then set $n \leftarrow n + 1$
end while
return the final labeled set and probabilistic predictor

Algorithm 3 Utility-tailored score-change embeddings

Require: L_n, pool $\mathcal{U}_n$, estimation pool $E = \{z_1, \dots, z_N\}$, weights $\omega_1, \dots, \omega_N$, score S_G

for all $m = 1, \dots, N$ **do**
  Estimate $\widehat{Q}_{G,m}(L_n)$
end for
for all $x \in \mathcal{U}_n$ **do**
  for all $\widetilde{y} \in \mathcal{Y}$ **do**
    Form $L_n^{x,\widetilde{y}} = L_n \cup \{(x, \widetilde{y})\}$ and estimate $\widehat{Q}_{G,m}(L_n^{x,\widetilde{y}})$ for every m
  end for
  Compute $\widehat{g}_{G,m}(x \mid L_n)$ for every m by Eq. (23) and set $\widehat{\boldsymbol{v}}_G(x \mid L_n) = (\sqrt{\omega_m}\widehat{g}_{G,m}(x \mid L_n))_{m=1}^N$
end for
return $\{\widehat{\boldsymbol{v}}_G(x \mid L_n) : x \in \mathcal{U}_n\}$

Algorithm 4 Diversity-enhanced batch selection (DUSAL-Div)

Require: Pool $\mathcal{U}_n$, embeddings $\widehat{\boldsymbol{v}}_G(x \mid L_n)$, realized batch size $b_n \le |\mathcal{U}_n|$

1: Run k-means with $k = b_n$ on $\{\widehat{\boldsymbol{v}}_G(x \mid L_n) : x \in \mathcal{U}_n\}$, reinitializing empty clusters until it produces nonempty clusters $\mathcal{K}_1, \dots, \mathcal{K}_{b_n}$ and centroids $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{b_n}$
2: **for all** $j = 1, \dots, b_n$ **do**
3: $\quad x_j \leftarrow \arg\min_{x \in \mathcal{K}_j} \|\widehat{\boldsymbol{v}}_G(x \mid L_n) - \boldsymbol{\xi}_j\|_2^2$
4: **end for**
5: **return** $\{x_1, \dots, x_{b_n}\}$

Algorithm 2 uses $b_n = b$ in full rounds and reduces it only for the remaining budget or pool. Algorithm 4 runs k-means on the full unlabeled pool with $k = b_n$.

D Experimental Details and Supporting Results

D.1 Controlled decision problems in detail

We first remove every approximation. In finite hypothesis models the posterior, the hypothetical updates, and both terms of Eq. (14) are available in closed form, so policies differ through their scores alone. Across 1,000 randomized instances the identities and invariances of Section 3 hold to within 7.65×10^{-16}, and every analytic expectation reported below agrees with an independent 10,000-draw Monte Carlo estimate. The constructions below isolate one property of the utility term each, and Table 5 collects their outcomes. We keep the labels of the frozen mechanism corpus: E0 verifies the identities and invariances themselves, E1 and E2 concern queries that can cross a decision boundary, E3 and its randomized extension P3 concern complete action rows, E4 concerns updates confined to one action region, E5 compares DUSAL-Hard with acquisition driven by utility alone, and E6 is a batch-selector fixture. E3 and E6 appear only in Appendix Table 6, which lists every one-step fixture.

The first property is sensitivity to boundary crossing, which constructions E1 and E2 isolate. A binary screening table with a 19:1 false-negative to false-positive cost has Bayes threshold 0.05. One candidate's label can carry a prediction across this threshold; a nuisance candidate produces a larger expected change in probability but can never change an action. BEMPS scores the pair $0.0016/0.0121$ and queries the nuisance candidate. The EVSI term raises only the first score, to $0.0216/0.0121$, and reverses the choice. In the analytic sequential extension, which repeats this one-step comparison over 50 decision-relevant and 50 nuisance modules, Hard spends its budget on $50/50$ decision-relevant modules and BEMPS on $0/50$, so terminal normalized regret is 0 versus 1. Construction E2 adds a three-action triage table in which two queries cross different boundaries with equal predictive movement: BEMPS ties them ($0.0049/0.0049$), whereas Hard separates them by the size of the decision consequence ($0.0399/0.0161$).

E1 also exposes the blind spot of Hard that motivates Soft: the Hard utility term is exactly zero when no single label can cross a boundary. Construction E4 places both candidates inside one action region, and Hard accordingly ties a decision-relevant and a nuisance query at $0.0001/0.0001$, while the utility-gap curvature of Soft (Eq. 21) separates them ($0.000629/0.000100$). In the associated shared-latent sequential simulation (100 paired seeds, $\lambda = 1$, $\tau = 0.02$, nuisance-first tie-breaking for BEMPS), Soft lowers mean cumulative regret from 1.870 to 1.446, a 22.7% reduction. Because BEMPS scores the two queries almost equally, the size of this reduction depends on how its near-ties are broken. Over nuisance-first, decision-relevant-first, and random tie-breaking, with and without a numerical tolerance, it ranges from 14.3% to 24.4%; its direction never changes.

The constructions so far are binary, where Proposition 2 shows that cost-weighting a quadratic score cannot alter exact rankings. Construction P3 probes the multi-action counterpart, in which a utility table is more than a set of outcome weights. In 100 paired worlds, generated as small martingale-preserving perturbations of one three-action construction, we change the action rows of the utility table while holding every column sum fixed, leaving the comparator's column-sum weights unchanged. Hard and Soft switch their selected query in 100 and 100 worlds; BEMPS and a column-weighted quadratic score switch in 0 and 0.

The last construction, E5, turns from what the utility term adds to what the proper-score base costs, since retaining that base is a design choice rather than a necessity. Against utility-only acquisition in

Table 5: Controlled decision problems under exact look-ahead. Each row isolates one property of the utility term; the comparator is BEMPS unless stated. E1 regret is normalized by the construction's total available decision opportunity and E4 regret is in native utility units, so rows are not comparable with each other or with the public-data endpoint. Arrows give the favorable direction. Appendix Table 6 lists all one-step scores and Appendix Table 7 all E5 endpoints.

	Property isolated	Variant	Endpoint	Comparator	DUSAL
E1	Boundary crossing	**Hard**	Decision-relevant modules queried ↑	0/50	50/50
			Terminal normalized regret ↓	1	0
E2	Utility at stake	**Hard**	One-step scores, low/high-threshold boundary	0.0049/0.0049	0.0399/0.0161
E4	Within-region learning	**Soft**	Mean cumulative regret ↓ (100 seeds)	1.870	1.446
P3	Complete action rows	**Hard/Soft**	Worlds in which the query follows the table ↑	0/100	100/100 each
E5	Price of propriety	**Hard**	Brier AULC ↓ (vs. utility-only)	0.305523	0.277747
	(comparator: utility-only)		Mixed-cost regret AULC ↓	0.100740	0.125989

Table 6: Exact mechanism gates. Scores separated by a slash follow the candidate order shown. All analytic expectations pass independent 10,000-draw Monte Carlo checks. These are constructed mechanism tests, not estimates of medical performance.

Gate	Candidate order	Comparator	Current construction
E1: boundary crossing	actionable/nuisance	BEMPS 0.0016/0.0121	Hard 0.0216/0.0121
E2: consequence magnitude	low/high boundary	BEMPS 0.0049/0.0049	Hard 0.0399/0.0161
E3: complete rows	x_A/x_B	BEMPS selects x_B under both U	Hard selects x_A/x_B under U_A/U_B
E4: within-region geometry	actionable/nuisance	Hard 0.0001/0.0001	Soft 0.000629/0.000100

100 paired finite worlds, Hard lowers the budget-normalized area under the learning curve (AULC) of the Brier score by 9.09% and of the negative log-likelihood (NLL) by 6.02%, but its decision-regret AULC is higher at each of three evaluation cost ratios, by 25.06% under their equal mixture (Appendix Table 7). DUSAL therefore trades some finite-budget decision regret for probability accuracy.

Together, the constructions answer the first question. Under exact look-ahead the utility term redirects queries where Theorem 3 says it should, and in E1 and E4 the redirection lowers decision regret. They are designed unit tests of the mechanism: they show what the utility term does when candidates carry boundary-crossing value (Hard) or utility-gap value (Soft) that the base score underrates, not how often such candidates occur. E5 adds that the proper-score base costs regret relative to utility-only acquisition. Whether the mechanism yields a benefit with learned, approximate posteriors on public medical tasks is the subject of the remaining sections.

D.2 Exact finite-model results

Full action rows preserve policy distinctions. E3 and its P3 extension hold column sums fixed while changing complete action rows. Hard and Soft change their selected query with the policy in all 100 and 100 paired P3 worlds, respectively; BEMPS and the legacy quadratic construction do not. Together with Proposition 2, this identifies a representational limitation of fixed quadratic weighting. It does not establish that DUSAL is the only remedy or that it improves public-data performance.

Boundary-crossing and within-region learning. The E1 sequential extension selects all 50 actionable modules with Hard and none with BEMPS. Its deterministic expected-opportunity curves end at regret 0 versus 1, normalized by the construction's total available opportunity, not by the public-data decision scale. Only E1's Random control uses repeated seeds. In the shared-latent E4 simulation, Soft at $\lambda = 1, \tau = 0.02$ reduces mean cumulative native regret from 1.870 to 1.446 over 100 paired seeds. That sequential comparison is Soft versus BEMPS; the Hard tie is a one-step result. These are constructed consequences, not general performance guarantees. Hard's utility term vanishes when all positive-probability hypothetical updates stay in one Bayes-action region; its base term remains. E1's nuisance query illustrates this condition with the identical full score 0.0121.

Prediction–decision trade-off. E5 compares Hard ($\lambda = 1$) with utility-only acquisition under the same binary 19:1 table, finite posterior, and normalized utility weight. Each of 100 paired worlds contains 32 hypotheses, 200 reference points, 25 candidates, and an eight-query budget; no neural

Table 7: E5 prediction–decision trade-off. Entries are mean AULCs over 100 paired finite worlds: each trajectory is integrated trapezoidally over steps 0–8 and divided by 8. All endpoints are lower-is-better. Regret uses each table's native cost units, and "mixture" averages the three regrets equally. Relative change is $(\overline{\text{Hard}} - \overline{\text{utility-only}})/\overline{\text{utility-only}}$. The last column counts worlds in which Hard is strictly better, not a significance test. Bold marks the lower unrounded mean within each row only; metrics have different scales.

Endpoint (AULC)	Hard	Utility-only	Relative change	Hard better / 100
Regret, 15:1	0.127594	**0.092814**	+37.47%	18
Regret, 19:1	0.126623	**0.099304**	+27.51%	19
Regret, 25:1	0.123750	**0.110103**	+12.39%	26
Regret, mixture	0.125989	**0.100740**	+25.06%	17
Brier	**0.277747**	0.305523	−9.09%	92
NLL	**0.430768**	0.458341	−6.02%	92

Table 8: Per-dataset protocol of the frozen cost search and confirmation, read back from stored run configurations. N_{val} is the discovery evaluation split; N_{test} is the sealed confirmation split; N_{ref} is the fixed, uniformly weighted reference cohort. The positive class carries the false-negative cost.

Dataset	Pool	N_{ref}	N_{val}	N_{test}	Init	Budget	Batch	Rounds	Backbone	Positive class
BreastMNIST	506	506	78	156	40	300	20	16	ResNet-18 ×5, 50 ep	label 0 (malignant)
PneumoniaMNIST	4,608	1,024	524	624	100	1,000	50	21	ResNet-18 ×5, 50 ep	label 1 (pneumonia)
ADE	4,101	1,024	899	899	100	1,500	100	16	MLP ×5, 30 ep	label 1 (event)

training loss differs between arms. Budget-averaged decision regret is evaluated under three cost ratios and their equal mixture. Table 7 reports all endpoints: Hard improves Brier and NLL, but incurs higher regret at every tested ratio, including the acquisition ratio. Thus preserving a strictly proper score is a principled probabilistic objective, not empirical robustness to cost misspecification or a guarantee of best finite-budget decisions. The example quantifies a trade-off that would be hidden by reporting only predictive metrics.

Batch-selector coverage. E6 is a separate 800-row fixture with manually specified scores and embeddings, not an end-to-end Bayesian experiment from a common generative model. Its Hard-based DIV selector covers 4/4 actionable modules versus 1/4 for scalar Hard. The stored "normalized regret" is defined as $1 - \text{coverage}/4$: the resulting 0 versus 0.75 is the same coverage result, not a second decision outcome. Comparisons with other selectors also change fabricated input geometry, so they cannot isolate a single component or establish public-data DUSAL-Div efficacy. The batch study in Appendix D.12 evaluates the latter directly.

D.3 Data, preprocessing, and declared decision problems

PneumoniaMNIST contains 4,708/524/624 official train/validation/test images, uses label 0 for normal and label 1 for pneumonia, and is distributed under CC BY 4.0 as part of MedMNIST v2 (Yang et al., 2023). BreastMNIST has 546/78/156 official images with label 0 for malignant and label 1 for normal/benign. We use the official 28×28 grayscale representation, convert to a tensor, and normalize with mean and standard deviation 0.5; no augmentation is applied. The ADE task is binary text classification of adverse-drug-event mentions on frozen PubMedBERT sentence embeddings (768-d) with an MLP head; its pool, validation, and test sizes are in Table 8. Test labels are used only for evaluation. Initial labels and all acquired labels come from the training split; the reference cohort is sampled once from the initial unlabeled training pool and fixed for the run.

The cost-discovery and confirmation problems are binary with action a declaring class a: $C(a, y) = c_{\text{FN}}$ when a is the negative declaration and y the positive class, $C(a, y) = c_{\text{FP}}$ for the reverse, and zero on the diagonal. With $c_{\text{FP}} = 1$ the sweep sets c_{FN} to the cost ratio, so the Bayes threshold on the positive-class probability is $t^\star = 1/(1 + c_{\text{FN}})$ and $L_U^{\text{dec}} = c_{\text{FN}}$. The ratios are research specifications; they were not elicited from clinicians and are not deployment recommendations. Positive scalar cost rescaling does not change the Bayes action or normalized endpoint; adding an outcome-only offset can translate the absolute endpoint while preserving paired contrasts on the same outcomes.

Table 9: Metric definitions after the endpoint correction. Reference-action regret is relative to a fitted model predictive, rather than the true conditional distribution or observed clinical outcomes.

Metric	Definition and role	Direction
Realized cost norm.	Held-out $N^{-1}\sum_i C(a(q_i), y_i)/L_U^{\mathrm{dec}}$; public-data endpoint	lower
Reference-action regret norm.	$N^{-1}\sum_i \mathrm{Reg}_U(q_i; p_{\mathrm{ref,i}})/L_U^{\mathrm{dec}}$ for a fully supervised CE predictive; optional diagnostic	lower
AUROC/AUPRC/macro-F1	Predictive discrimination summaries; secondary	higher
Brier/NLL/ECE	Probability accuracy/calibration summaries; secondary	lower
Hard EVSI	Expected Hard decision-regret term in acquisition; mechanism diagnostic	–
Soft action KL	Expected $\lambda\tau\,\mathrm{KL}\{\rho(q)\|\|\rho(p)\}$ contribution; not EVSI	–
Hard non-base share	Round-0 $\sum_x(\text{Hard} - \text{BEMPS})/\sum_x \text{Hard}$; normalized declared-cost EVSI share	–
Soft non-base share	Round-0 $\sum_x(\text{Soft} - \text{BEMPS})/\sum_x \text{Soft}$; soft-action $\lambda\tau$ KL share, not EVSI	–

R-CORE and the posterior-backend study use two additional stylized rectangular cost tables:

$$C_{\mathrm{Breast}} = \begin{pmatrix} 25 & 0 \\ 2 & 2 \\ 0 & 8 \end{pmatrix}, \qquad C_{\mathrm{Retina}} = \begin{pmatrix} 0 & 0 & 4 & 12 & 20 \\ 1 & 1 & 0 & 3 & 8 \\ 6 & 6 & 2 & 0 & 0 \end{pmatrix}.$$

RetinaMNIST has 1,080/120/400 official images and five ordered grades (Yang et al., 2023). All three actions in each table are effective on the declared simplex. These tasks test rectangular action semantics, not clinical validity of the numerical costs.

D.4 Metric semantics

Secondary summaries use area under the receiver operating characteristic curve (AUROC), area under the precision–recall curve (AUPRC), macro-F1, the Brier score, negative log-likelihood (NLL), and expected calibration error (ECE).

AULC is the trapezoidal integral over the observed increasing label budgets, divided by the budget range. Thus it is an average over the measured learning curve, not an unnormalized area. The estimand is conditional on the fixed benchmark tasks; bootstrap intervals therefore resample seeds, not datasets. The one-sided limits below bound the paired difference. They are not equivalence tests, and nonsignificance alone does not establish equivalence.

D.5 Matched-control protocol of the binary confirmation

Moving to public data replaces the exact posterior by a retrained deep ensemble and the exact update by an approximation. The protocol is built so that, despite these changes, a difference between arms can still be attributed to the utility term.

We use BreastMNIST and PneumoniaMNIST (Yang et al., 2023) and binary adverse drug event (ADE) classification. The image tasks use five-member ResNet-18 ensembles; ADE uses five MLP heads on frozen PubMedBERT embeddings. Label budgets of 300, 1,000, and 1,500 are added to small initial sets, and the ensemble is retrained after every acquired batch (Appendix Table 8).

The comparison is a matched control. BEMPS is DUSAL with $\lambda = 0$, and we compare it with DUSAL-Hard ($\lambda = 1$) and DUSAL-Soft ($\lambda = 1$, $\tau = 0.05$). Within a seed all arms share cross-entropy (CE) training, ensemble and seed streams, initial labels, and a fixed reference cohort. They consequently start from the same fitted model and score the same candidates, and at the first round their score difference is exactly the utility term. The term is cheap: on BloodMNIST acquisition takes 12.7 s per round for DUSAL-Hard against 10.5 s for BEMPS, and on PathMNIST 18.7 s against 13.8 s, while retraining the ensemble takes about 67 s per round for every arm, so a full run costs 27.6 against 27.1 minutes on one GPU (Appendix Table 29 lists every arm).

A matched control still requires a decision problem. The utility table is an input to DUSAL, and the benchmarks do not supply one. Instead of fixing costs by hand, a validation-only search compares

Table 10: Complete validation cost sweep: relative change in realized-cost AULC, DUSAL versus BEMPS, over 20 paired seeds per cell (negative favors DUSAL). $t^\star$ is the Bayes threshold on the positive-class probability. Bold marks the 3 cells whose one-sided 95% UCB lies below zero; the selected cell per dataset is the eligible cell with the lowest UCB.

FN:FP	$t^\star$	BreastMNIST		PneumoniaMNIST		ADE	
		Hard	Soft	Hard	Soft	Hard	Soft
1	.500	−2.47%	−4.45%	−0.62%	−1.17%	+0.14%	−0.07%
1.5	.400	−2.96%	**−3.67%**	−1.55%	−1.09%	−0.31%	−1.04%
2	.333	+0.01%	−1.24%	−1.56%	−1.63%	+0.06%	+0.21%
3	.250	−0.14%	−0.13%	−1.30%	−1.33%	+0.50%	−0.18%
4	.200	+1.68%	+0.25%	−1.14%	−1.48%	+1.47%	+0.87%
6	.143	−1.40%	+1.02%	−2.14%	−0.75%	+0.82%	+1.12%
9	.100	−0.87%	+0.91%	**−3.85%**	**−4.05%**	+0.81%	+0.96%
13	.071	+1.64%	+2.53%	−0.41%	−2.13%	+1.14%	+0.75%
19	.050	+2.11%	+3.30%	−2.73%	−2.69%	+0.77%	+0.30%
29	.033	+3.71%	+2.26%	+2.13%	+1.81%	+2.18%	+1.96%
49	.020	+3.54%	+9.17%	+0.96%	+4.59%	−0.80%	+2.23%
99	.010	+7.72%	+11.29%	+0.32%	+5.20%	+5.47%	+1.49%
Spearman(log ratio, rel. change)		+0.853	+0.972	+0.406	+0.315	+0.476	+0.825

the three arms at 12 false negative (FN) to false positive (FP) cost ratios (FN:FP) from 1:1 to 99:1, with 20 paired seeds per cell (2,160 runs) and test labels unread. A rule fixed beforehand selects, per dataset, the cell with the lowest one-sided 95% upper confidence bound on the paired DUSAL−BEMPS difference: 1.5:1 for BreastMNIST, 9:1 for PneumoniaMNIST, and 1.5:1 for ADE (Appendix Table 10). The search thus places the confirmation where validation evidence was most favorable to DUSAL, although that evidence is weak: only 3 of the 72 validation bounds lie below zero, about the number expected by chance (3.6). The ratios define research decision problems and are not clinical recommendations. All three are binary two-action problems; Appendix D.9 separately studies multiclass outcomes grouped into deployment actions.

Confirmation then uses data that played no part in this selection. All three arms are rerun at the selected costs on 80 fresh seeds, disjoint from the validation seeds, and evaluated on held-out test splits (720 runs). The endpoint is the budget-normalized area under the learning curve (AULC) of held-out realized cost divided by the decision scale L_U^{dec} (Appendix Table 9); the estimand is DUSAL minus BEMPS, so negative values favor DUSAL. The 6 contrasts (three tasks, Hard and Soft) form one family, tested by one-sided exact paired signed-rank randomization with Holm control at $\alpha = 0.05$ and accompanied by 100,000-replicate paired bootstrap 95% confidence intervals (CIs). The analysis plan, including a commitment not to search further costs after an unsupported result, was fixed before confirmation (Appendix D.6).

D.6 Frozen cost search and fresh-seed confirmation

Selection rule. Discovery evaluated every cell on the validation split with test labels sealed (the selection record stores `test_set_accessed=false`). A cell is eligible if at least 80% of seeds are nondegenerate in action, at least 50% of seeds change their acquired set relative to BEMPS, the mean acquired-set Jaccard distance is at least 0.01, and the positive action carries at least 1% mass. Among eligible cells the rule ranks by one-sided 95% UCB of the paired AULC difference (20,000 paired-bootstrap replicates), then paired mean, then cost-grid order, Hard before Soft, and selects the top cell per dataset; the analyzer refuses to select from an unsealed split. Every one of the 72 cells was eligible. Only 3 validation UCBs fall below zero, each by less than 10^{-4}, versus the nominal count 3.6; dependence among cells means this count is not a calibrated null test.

Confirmation protocol. The confirmation manifest (SHA-256 prefix `7404a84271572668`) binds the selection record (`44c7bebf645dd1da`), launcher (`dfc1ff6fd57d073c`), source commit `72d0a96`, and code fingerprint `9ecb25c0e508a12b`; all 3600 per-run scientific files are indexed by content hash (`b2613801a28f4ac5`). Seeds 100–179 are disjoint from the discovery seeds 60–79. Inference is one-sided exact paired signed-rank randomization with Holm control over the family of 6 at $\alpha = 0.05$, with 100,000 paired bootstrap replicates for intervals; the success rule is a negative paired mean with Holm $p < 0.05$. The frozen claim limit restricts any supported contrast to

Table 11: Fresh-seed effects at the validation-selected costs (80 paired fresh test seeds per row). "AULC B/D" lists the BEMPS/DUSAL normalized realized-cost AULC means; lower is better. Δ = DUSAL − BEMPS, and the bracketed values are its two-sided marginal paired-bootstrap 95% CI. Relative change is Δ/BEMPS, so negative favors DUSAL. "Fav." counts seeds with $\Delta < 0$; the final column gives one-sided exact raw/Holm p values over the six prespecified contrasts. No contrast meets the success rule.

Dataset (FN:FP)	Arm	AULC B / D	Δ [95% CI] (absolute)	Rel. change	Fav.	p raw / Holm
BreastMNIST	**Hard**	0.186518/0.186733	$+0.000215$ $[-0.002784, +0.003190]$	$+0.12\%$	39/80	0.623/1.000
1.5:1	**Soft**	0.186518/0.186164	-0.000354 $[-0.003434, +0.002638]$	-0.19%	39/80	0.490/1.000
PneumoniaMNIST	**Hard**	0.024758/0.024748	-0.000010 $[-0.000192, +0.000172]$	-0.04%	39/80	0.519/1.000
9:1	**Soft**	0.024758/0.024766	$+0.000009$ $[-0.000177, +0.000200]$	$+0.03\%$	45/80	0.345/1.000
ADE	**Hard**	0.136882/0.136276	-0.000606 $[-0.001684, +0.000415]$	-0.44%	43/80	0.228/1.000
1.5:1	**Soft**	0.136882/0.135917	-0.000965 $[-0.002076, +0.000156]$	-0.71%	45/80	0.040/0.239

Table 12: What the confirmation rules out, and what it could have detected. "Benefit bound" is the one-sided 95% paired-bootstrap lower limit on Δ expressed as a relative reduction, i.e., an upper limit on the benefit: any larger DUSAL relative reduction is excluded at the stated marginal one-sided 95% level by the 80-seed test-split data. SD is the paired-difference standard deviation as a fraction of the BEMPS baseline; MDE is the relative effect detectable at 80% power, one-sided $\alpha = 0.05$, at the stated seed count. The discovery sweep ran at 20 seeds.

Dataset	Arm	Benefit bound	SD	MDE_{20}	MDE_{40}	MDE_{80}	MDE_{160}
BreastMNIST	Hard	1.21%	7.35%	4.08%	2.89%	2.04%	1.44%
	Soft	1.57%	7.45%	4.14%	2.93%	2.07%	1.46%
PneumoniaMNIST	Hard	0.66%	3.36%	1.87%	1.32%	0.93%	0.66%
	Soft	0.59%	3.49%	1.94%	1.37%	0.97%	0.69%
ADE	Hard	1.09%	3.52%	1.96%	1.38%	0.98%	0.69%
	Soft	1.39%	3.74%	2.08%	1.47%	1.04%	0.74%

its dataset–cost pair under the executed acquisition-only CE protocol, and the no-rescue rule forbids further cost search for this claim once the family is unsupported.

These aggregate bounds summarize whole learning curves. Figure 1 shows the BreastMNIST contrast across budgets; the validation advantage is not recovered on the test seeds.

D.7 Binary confirmation and replication: full results

We begin with selection, the second question. In all 1440/1440 first-round comparisons the arms score identical candidate indices, which confirms the matched design. From that common start the arms diverge: at the selected costs, on average 19–29% of the labels acquired by DUSAL over a run are absent from the matched BEMPS run (Figure 2A). The pattern is not specific to the selected costs. Across all 72 validation cells the paired-seed mean Jaccard similarity of the acquired sets lies between 0.55 and 0.70. For scale, two independent uniformly random acquisitions of the same size would differ by $1 - b/N$, that is 40.7%, 78.3%, and 63.4% on the three tasks for budget b and pool size N, so matched DUSAL and BEMPS runs remain far more alike than unrelated runs.

The fractions are measured on the validation cohort. This study has no utility-free, magnitude-matched perturbation control; the permuted-table diagnostic in Appendix D.9 is a different comparison. Full-run overlap changes little on the utility weight: on BreastMNIST it is 0.579, 0.573, and 0.570 for $\lambda = 1, 4, 8$ (Appendix Table 22), whereas the first batch keeps changing with λ (Appendix Table 20). The answer to the second question is therefore a qualified yes. The utility term changes the first batch in proportion to its weight, and the runs then diverge; what fraction of the full-run difference reflects decision-relevant selection, rather than trajectory divergence after any early change, is not identified.

We now turn to the third question, whether the changed selection lowers held-out decision cost. Table 13 reports every prespecified contrast, and none is supported. The six point estimates lie within -0.71% to $+0.12\%$ of the BEMPS mean, four favoring DUSAL and two favoring BEMPS, and all

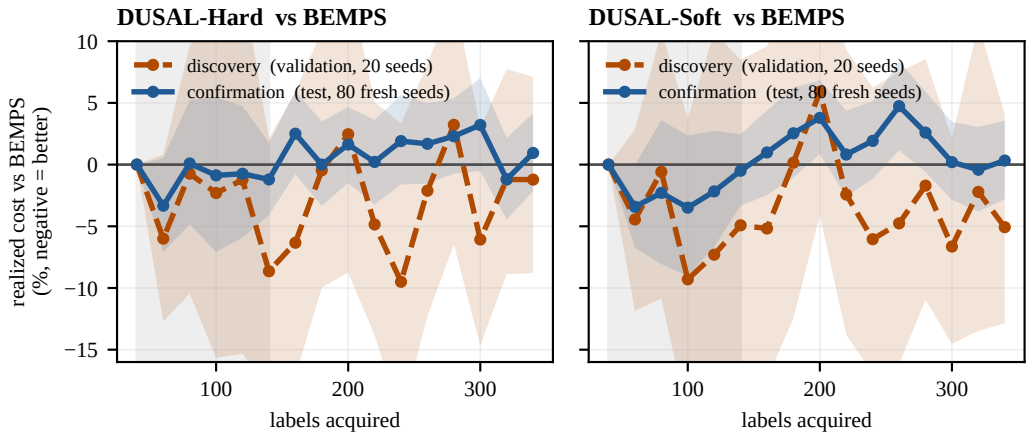

Figure 1: Per-budget paired difference in realized cost on BreastMNIST at 1.5:1, discovery (validation, 20 seeds) against confirmation (test, 80 fresh seeds), with paired-bootstrap 95% bands; relative effects use the ratio of means, matching the AULC convention. The discovery curve lies below zero across most of the budget; the confirmation curve is slightly negative early and drifts positive after about 150 labels, so no sustained or phase-specific advantage is visible. The shaded first third is the window later pre-registered as the λ studies' secondary endpoint.

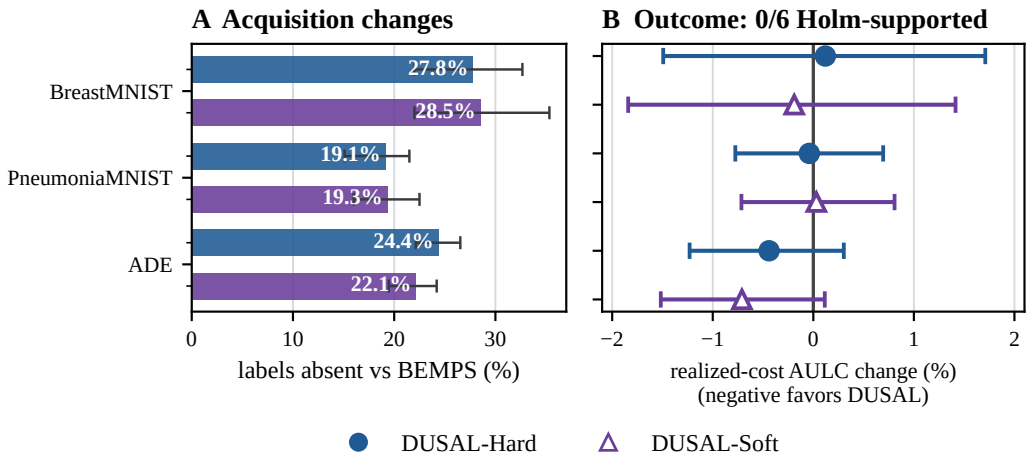

Figure 2: Acquisition and cost results from separate cohorts. Within each dataset the upper, blue row is DUSAL-Hard and the lower, purple row is DUSAL-Soft. **(A)** Per-seed absent fraction $(1-J)/(1+J)$ for equal-sized DUSAL/BEMPS sets at selected costs; bars are means and whiskers min–max over 20 validation seeds. **(B)** Realized-cost AULC $\overline{\Delta}/\overline{\text{BEMPS}}$ over 80 fresh test seeds; intervals are marginal paired-bootstrap 95% absolute-Δ CIs divided by the observed BEMPS mean. Negative favors DUSAL; all intervals cross zero and no contrast passes Holm.

six intervals contain zero. The smallest raw one-sided p value, 0.040 for ADE Soft, becomes 0.239 after Holm correction (Figure 2B).

The reported precision limits the unresolved effects. With 80 paired seeds a single one-sided test detects relative effects of 0.93%–2.07% with 80% power (the minimum detectable effect (MDE)) (before Holm adjustment; at the most stringent Holm step, $\alpha/6$, they are about 1.3 times larger), and the data exclude, at the marginal one-sided 95% level, relative benefits larger than 0.59%–1.57% (Appendix Table 12). These limits are contrast-specific; they are neither simultaneous bounds nor equivalence tests. The answer to the third question on these tasks and budgets is that any cost

Table 13: Fresh-seed confirmation at the validation-selected costs: normalized realized-cost AULC on held-out test splits, mean over 80 paired seeds (lower is better). Δ is DUSAL minus BEMPS with a two-sided marginal paired-bootstrap 95% CI; "Rel." is Δ over the BEMPS mean, so negative favors DUSAL. "Bound" is the largest relative benefit not excluded at the marginal one-sided 95% level, and the minimum detectable effect (MDE) is the relative effect detectable with 80% power at this seed count. Holm-adjusted p values cover all 6 contrasts. Raw p values and favorable-seed counts are in Appendix Table 11. No interval excludes zero, so no entry is highlighted.

Dataset (FN:FP)	Variant	BEMPS ↓	DUSAL ↓	Δ [95% CI]	Rel. ↓	Bound	MDE	Holm p
BreastMNIST (1.5:1)	**Hard**	0.186518	0.186733	$+0.000215\ [-0.002784, +0.003190]$	$+0.12\%$	1.21%	2.04%	1.000
	Soft	0.186518	0.186164	$-0.000354\ [-0.003434, +0.002638]$	-0.19%	1.57%	2.07%	1.000
PneumoniaMNIST (9:1)	**Hard**	0.024758	0.024748	$-0.000010\ [-0.000192, +0.000172]$	-0.04%	0.66%	0.93%	1.000
	Soft	0.024758	0.024766	$+0.000009\ [-0.000177, +0.000200]$	$+0.03\%$	0.59%	0.97%	1.000
ADE (1.5:1)	**Hard**	0.136882	0.136276	$-0.000606\ [-0.001684, +0.000415]$	-0.44%	1.09%	0.98%	1.000
	Soft	0.136882	0.135917	$-0.000965\ [-0.002076, +0.000156]$	-0.71%	1.39%	1.04%	0.239

advantage of DUSAL over matched BEMPS is either absent or smaller than the contrast-specific bounds of Table 13. Precision is not the same as assay sensitivity, however: the confirmation family contains no arm known to beat BEMPS on these tasks. A replication with two further arms tests whether acquisition benefits against random sampling are detectable on the same endpoints.

The replication reruns all three arms together with random and entropy sampling on the same three decision problems, costs, 80 seeds, endpoint and Holm family (1,200 runs), with the analysis fixed before launch (Appendix D.8). It runs on different hardware and library versions from the confirmation, and every arm is rerun rather than only the two new ones, because runs from different environments do not pair: the same seed, initial labels and random streams give a different first-round model on the two machines, and pairing across them would have injected an environment offset larger than the effect under test and of either sign (Appendix D.8). The replication therefore also asks whether the null depends on the environment it was measured in.

The BEMPS–random control passes on two tasks, but not the third. On PneumoniaMNIST BEMPS lowers realized-cost AULC relative to random sampling by -7.82% [-9.21,-6.43] (71/80 seeds favourable, $p < 10^{-4}$) and on ADE by -8.88% [-9.89,-7.85] ($p < 10^{-4}$), so the prespecified positive control passes on both. On BreastMNIST it fails: BEMPS and random sampling differ by -0.01% [-2.37,+2.37], with 42/80 seeds favorable. The estimate and its interval do not establish an exactly zero effect. Entropy sampling separates from random (-2.48%, $p = 0.019$), so the task admits a detectable acquisition effect, but the BEMPS control remains unresolved. The BreastMNIST utility contrasts still estimate differences against BEMPS; they do not establish equivalence or detectable headroom above that baseline.

The replication also finds no Holm-supported utility benefit among its 6 contrasts (Appendix Table 14); the point estimates lie between -0.59% and +0.61%, and the smallest raw one-sided p value, 0.020 for PneumoniaMNIST Hard (-0.59% [-1.18,-0.00]), becomes 0.117 after Holm correction. What the passing controls add is a scale for this null. On PneumoniaMNIST and ADE the family's MDE is 0.75% and 1.01%, so the design resolves effects 10.4 and 8.8 times smaller than the acquisition advantage it measures on the same seeds, and the utility term shows none at that resolution. The BEMPS means differ from the original confirmation by +1.31%, +0.21% and -1.05% across the two environments (Appendix D.8), so the null is not a property of one machine.

One further pattern in the replication bears on what a utility term refines. Plain entropy sampling beats BEMPS on all three tasks (BEMPS − entropy +2.53%, +1.39% and +1.15%, all three surviving Holm correction over the three tasks with largest adjusted $p = 0.026$), and DUSAL-Hard inherits the deficit (+1.73% against entropy on ADE, $p < 10^{-4}$). Appendix D.9 finds the same ordering on PathMNIST and the reverse on BloodMNIST, and Appendix D.10 shows that the ordering changes with the posterior, representation, and training pipeline.

D.8 Binary replication with random and entropy arms

The replication of Section 4.3 reruns the confirmation's three decision problems, costs and seeds 100–179 with five arms: random and entropy sampling, BEMPS, DUSAL-Hard and DUSAL-Soft at $\lambda = 1$ (1,200 runs, commit `64ee8bc`). Its analysis was fixed before launch: a positive control

Table 14: Binary replication with random and entropy arms: paired relative difference in normalized realized-cost AULC over the comparator mean (negative favors the first arm) with marginal paired-bootstrap 95% CIs, $n = 80$ seeds per task. Family rows show Holm-adjusted p values; control and secondary rows show raw two-sided p values. The BEMPS − entropy row is exploratory. **Bold blue** marks an interval that excludes zero in the favorable direction.

	Contrast	BreastMNIST		PneumoniaMNIST		ADE	
		Rel. ↓ [95% CI]	p	Rel. ↓ [95% CI]	p	Rel. ↓ [95% CI]	p
Controls	BEMPS − random	-0.01% $[-2.37, +2.37]$	0.959	**-7.82% $[-9.21, -6.43]$**	$< 10^{-4}$	**-8.88% $[-9.89, -7.85]$**	$< 10^{-4}$
	entropy − random	-2.48% $[-5.12, +0.24]$	0.019	**-9.09% $[-10.28, -7.90]$**	$< 10^{-4}$	**-9.91% $[-10.83, -8.99]$**	$< 10^{-4}$
Holm family	**Hard** − BEMPS	-0.53% $[-2.09, +0.99]$	1.000	-0.59% $[-1.18, -0.00]$	0.117	$+0.57\%$ $[-0.22, +1.36]$	1.000
	Soft − BEMPS	$+0.61\%$ $[-0.89, +2.11]$	1.000	-0.21% $[-0.93, +0.49]$	1.000	-0.03% $[-0.77, +0.70]$	1.000
Secondary	**Hard** − entropy	$+1.99\%$ $[-0.37, +4.40]$	0.100	$+0.79\%$ $[+0.10, +1.50]$	0.061	$+1.73\%$ $[+1.01, +2.44]$	$< 10^{-4}$
	BEMPS − entropy	$+2.53\%$ $[+0.37, +4.72]$	0.026	$+1.39\%$ $[+0.61, +2.19]$	0.002	$+1.15\%$ $[+0.36, +1.93]$	0.010

per task (BEMPS minus random, gate passed when the bootstrap interval's upper bound is below zero), the six Hard and Soft contrasts against BEMPS as a Holm family, the BEMPS means against the stored confirmation as a descriptive environment comparison, and the entropy comparisons as secondaries. The unit of work was a seed block: all five arms of a seed ran in one worker on one device, so every paired contrast is formed within a single environment, and seeds rather than arms were distributed across machines.

The family's MDE at this seed count is 1.96%, 0.75% and 1.01% for Hard (1.91%, 0.91% and 0.94% for Soft), so on PneumoniaMNIST and ADE the positive control is 10.4 and 8.8 times the smallest effect the family could have detected.

Every arm was rerun, rather than only the two new ones, because runs from different environments do not pair. The replication ran on different hardware and library versions from the stored confirmation, and under the same seed, initial labels and random streams the first-round model differs between the two. The paired correlation of DUSAL-Hard with BEMPS falls from 0.64 within one environment to 0.25 across the two on BreastMNIST (0.62 to 0.18 on PneumoniaMNIST, 0.34 to 0.20 on ADE), inflating the MDE by 1.45, 1.44 and 1.10 times; and the environment shifts the BEMPS mean by +1.31%, +0.21% and -1.05%, which on two of the three tasks exceeds the effect under test. Pairing the new DUSAL-Hard runs against the stored BEMPS runs would have reported +0.78%, -0.39% and -0.49% in place of the within-environment -0.53%, -0.59% and +0.57%, reversing the sign on BreastMNIST and ADE; the offset is systematic, so more seeds would not remove it. The aggregate BEMPS means nevertheless agree across environments (0.1890 against 0.1865 on BreastMNIST, 0.0248 against 0.0248 on PneumoniaMNIST and 0.1354 against 0.1369 on ADE), which is the sense in which the null does not depend on the machine.

D.9 Grouped-action confirmation with acquisition controls

This study evaluates multiclass outcomes mapped to fewer deployment actions. BloodMNIST has eight classes and PathMNIST nine, each mapped to two actions with $L_U^{\mathrm{dec}} = 10$. A proposed three-action DermaMNIST condition was excluded by the pilot gate. Six arms share the seed, initial labels, training rules, ensemble construction, and reference cohort: random, entropy, BEMPS, Hard, Soft, and Hard acquiring under a permuted cost table. Evaluation always uses the true table.

The wrong-table arm permutes outcome columns. On BloodMNIST it moves the penalty for auto-accepting an immature granulocyte onto neutrophils and the erythroblast penalty onto monocytes; on PathMNIST it moves the tumour penalty onto mucus and the stroma penalty onto smooth muscle. Column permutation preserves the multiset of columns and, on the full simplex, the effective-action structure and decision scale. It need not preserve utility-score magnitudes on the observed predictive distributions. The comparison is a specification diagnostic motivated by negative-control reasoning (Lipsitch et al., 2010) and randomization-based sensitivity checks (Adebayo et al., 2018), not an identification result establishing a unique causal mechanism.

An outcome-blind pilot selected the smallest $\lambda \in \{1, 2, 4, 8\}$ whose mean first-batch Jaccard overlap with BEMPS was at most 0.8. BloodMNIST and PathMNIST entered at $\lambda = 8$; DermaMNIST was dropped by the delivery rule at overlap 0.827. The two excluded contrasts remain in the planned

Table 15: Multi-action confirmation: normalized realized-cost AULC on held-out test splits at $\lambda = 8$, mean over paired fresh seeds (lower is better). The top block is the prespecified positive control and the wrong-table control, which acquires under a cost table that penalizes auto-accepting a neutrophil instead of an immature granulocyte (BloodMNIST) and leaving mucus unflagged instead of tumour (PathMNIST); the middle block is the Holm family of Appendix D.9. "Rel." is the paired difference over the comparator mean, so negative favors the first arm, and brackets are marginal paired-bootstrap 95% CIs. **Bold blue** marks an interval that excludes zero in the favorable direction and orange one that excludes zero adversely; plain entries are unresolved. DermaMNIST was dropped by the outcome-blind delivery rule and its two contrasts count as unsupported.

	Contrast	BloodMNIST ($n = 80$)		PathMNIST ($n = 60$)	
		Rel. ↓ [95% CI]	p	Rel. ↓ [95% CI]	p
Controls	BEMPS − random	**−10.07%** [−11.55, −8.61]	$< 10^{-4}$	**−3.64%** [−6.89, −0.47]	0.035
	wrong table − BEMPS	−0.27% [−1.57, +1.02]	0.810	−0.06% [−2.16, +2.02]	0.978
Holm family	**Hard** − BEMPS	−0.67% [−1.82, +0.47]	0.909	+0.17% [−1.92, +2.30]	1.000
	Soft − BEMPS	+0.57% [−0.75, +1.92]	1.000	−0.55% [−2.61, +1.55]	1.000
Secondary	entropy − random	**−3.41%** [−5.05, −1.73]	0.000	**−8.14%** [−10.65, −5.69]	$< 10^{-4}$
	Hard − entropy	**−7.53%** [−9.06, −6.01]	$< 10^{-4}$	+5.07% [+2.80, +7.37]	0.000

family of 6 but are not empirical null observations. Confirmation used 80 and 60 fresh seeds, respectively, with the primary endpoint and inference procedure of Section 4.2.

Decision-cost comparisons. BEMPS improves on random by -10.07% [-11.55,-8.61] on BloodMNIST (77/80 favorable seeds, $p < 10^{-4}$) and -3.64% [-6.89,-0.47] on PathMNIST ($p = 0.035$). None of the four evaluated Hard/Soft contrasts passes the planned Holm correction; the smallest adjusted value is 0.909. Their estimates span -0.67% to +0.57%. The controls detect acquisition benefits; they establish neither a zero utility effect nor scope for improvement over BEMPS.

The permuted-table estimates against BEMPS are -0.27% ($p = 0.810$) and -0.06% ($p = 0.978$). Hard's whole-run Jaccard overlap with BEMPS is 0.233 and 0.043; the corresponding wrong-table values are 0.234 and 0.043. Similar overlap summaries do not establish identical rankings, acquired examples, or mechanisms. They show that acquisition divergence alone cannot demonstrate useful alignment with the true decision problem.

Batch agreement and score activity. BloodMNIST scores all 11,749 candidates. Hard and BEMPS share a median 43 of 50 first-batch items (range 35–47); Hard and the wrong-table arm share 41. After retraining on labeled sets differing by at most 15 items in 210, second-batch Hard–BEMPS overlap is 1 (range 0–5; chance 0.2), and Hard–wrong-table overlap is 1. These comparisons concern different policies with different updated training sets. They do not directly measure agreement between a hypothetical update and the subsequent refit on the same data, nor show that acquisition stops influencing later rounds.

The Hard utility term is active for 4.3% of candidates initially and 0.5% in the last round. Its reported score shares are 0.621 and 0.554 on BloodMNIST and PathMNIST. Nevertheless, the fractions of acquired items whose two most probable classes belong to the same action group are close: Hard/BEMPS/wrong-table shares are 0.385/0.384/0.386 on BloodMNIST and 0.442/0.447/0.442 on PathMNIST. This statistic captures only one aspect of decision-relevant acquisition.

Secondary endpoints and baselines. Across 64 Hard/Soft contrasts on 16 secondary endpoints, 0 survive Holm correction (Table 29). The most expensive error is -1.40% rarer for BloodMNIST Hard ($p = 0.178$) and -2.49% rarer for PathMNIST Soft ($p = 0.147$), without resolved improvements. Nominal $p < 0.05$ counts are 4, 1, and 6 for Hard, Soft, and the wrong table. These exploratory counts illustrate multiplicity and do not identify beneficial mechanisms.

Entropy improves on random by -8.14% [-10.65,-5.69] on PathMNIST. Hard is +5.07% [+2.80,+7.37] worse than entropy there ($p = 0.000$), but -7.53% better on BloodMNIST ($p < 10^{-4}$). Thus baseline ordering depends on the evaluated task and pipeline. The Laplace comparison below changes the learning pipeline as well as the posterior approximation.

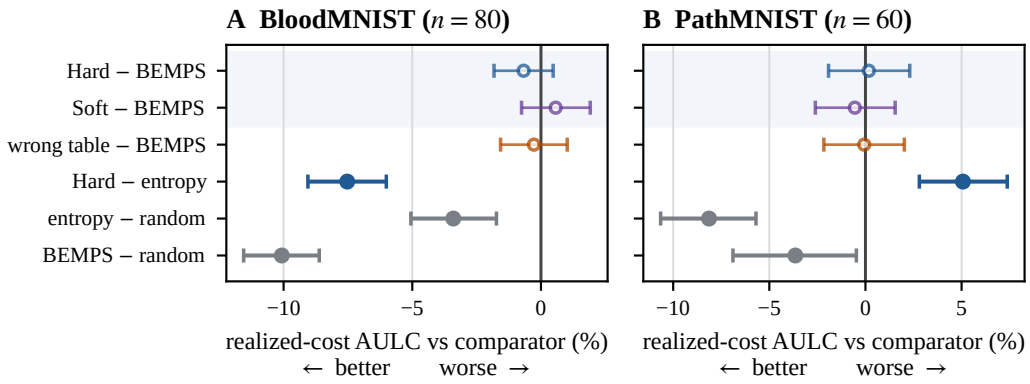

Figure 3: Evaluated contrasts in grouped-action confirmation, as a paired relative difference in realized-cost AULC with a marginal paired-bootstrap 95% CI. Filled markers denote an interval that excludes zero. The shaded band is the Holm family of Table 15; the rows below it are the controls and the entropy comparisons, "wrong table" being DUSAL-Hard run with the tumour and immature-granulocyte penalties moved onto mucus and neutrophils. On the same seeds the design detects a -10.07% advantage of BEMPS over random sampling and a -7.53% advantage of Hard over entropy, but no Hard/Soft advantage over matched BEMPS passes Holm correction.

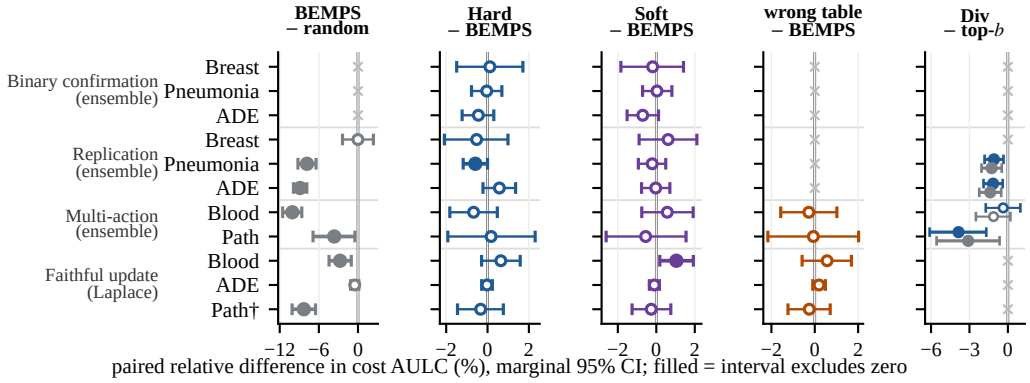

Figure 4: Contrasts and controls from Table 2, including the batch-rule comparisons paired with stored top-b runs. Filled: interval excludes zero; cross: arm not run; †: exploratory. In the final panel, the upper blue marker denotes Div applied to Hard and the lower grey marker Div applied to BEMPS. Their similar gains support the contribution of batch diversity without utility augmentation. No confirmatory Hard/Soft benefit over BEMPS passes Holm correction. Intervals are marginal, not equivalence bounds; wrong-table arms diagnose utility specification.

D.10 Laplace-posterior confirmation

This study changes the retrained ensemble to a Laplace last-layer posterior on frozen features, with five draws used for finite posterior reweighting. It retains the grouped-action decision problems, costs and $\lambda = 8$, the binary ADE problem, and the six acquisition arms. Feature extraction, dimensionality reduction, and the fitted model also change; the comparison does not isolate posterior approximation alone. Appendix D.11 gives the pipeline, seeds, and predeclared gate (1,320 runs).

The outcome-blind pilot uses second-round overlap between the Hard and BEMPS batches as an operational agreement gate. Thresholds are 30 of 50 items for the image tasks and 60 of 100 for ADE. BloodMNIST and ADE pass at 34 and 74; PathMNIST reaches 26 and is excluded from confirmation but evaluated separately as exploratory. Its two planned contrasts count as unsupported,

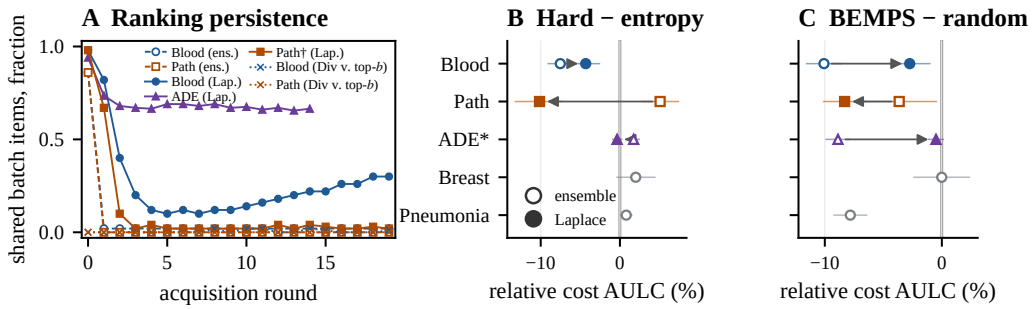

Figure 5: **(A)** Items shared by two batches per round, median over seeds, as a fraction of the batch. DUSAL-Hard against BEMPS: under the ensemble (hollow, dashed) agreement drops after the first batch; under the Laplace pipeline (filled, solid) agreement is higher; colour and shape identify the task. On both ensemble tasks, DUSAL-Div and Hard top-b (dotted, crosses) have zero median overlap at every round, unlike the high initial overlap between Hard and BEMPS (Section 4.4). **(B, C)** Base-rule ordering under two learning pipelines on the same tasks, seeds and costs (95% CIs; arrows from ensemble to Laplace); asterisk: replication on different seeds; †: exploratory condition.

not as completed confirmatory tests. Passing this gate establishes cross-method batch agreement, not fidelity of hypothetical updates to actual refits or exact Bayesian updating.

On the confirmatory seeds, second-batch Hard–BEMPS overlap is 41 of 50 on BloodMNIST (range 18–50) and 74 of 100 on ADE. Sixth-round values are 5 and 69. Agreement can therefore change substantially over a run even after the initial gate passes.

Decision cost. The planned family has 0/6 supported contrasts; the smallest adjusted p among the evaluated confirmatory contrasts is 1.000. The grouped-action and binary-replication families have 0/6 and 0/6 supported contrasts, respectively. Planned but excluded contrasts are included in these denominators, not treated as completed tests.

BloodMNIST resolves a BEMPS advantage over random sampling of -2.75% [-4.44,-1.03] ($p = 0.002$). Hard is +0.64% [-0.29,+1.58] relative to BEMPS, and Soft is +1.04% [+0.18,+1.92]. Soft's marginal interval excludes zero adversely, but the reported primary tests are one-sided tests of benefit, not multiplicity-adjusted tests of harm. The ADE control is unresolved (-0.50% [-1.13,+0.13], $p = 0.143$); Hard and Soft estimates are -0.03% and -0.10%. None of the evaluated confirmatory contrasts passes Holm correction. Exploratory PathMNIST resolves the random control at -8.31%, with unresolved Hard and Soft estimates of -0.33% and -0.26%.

The wrong-table estimates against BEMPS are +0.57%, +0.19%, and -0.25%; Hard against the wrong table is +0.07%, -0.22%, and -0.08%. BloodMNIST second-batch Hard–wrong-table overlap is 36, close to Hard–BEMPS. Together with the acquired-composition summaries in Appendix D.11, these results do not establish useful alignment with the true table, but neither do they prove that decision content is unused.

Pipeline dependence. On PathMNIST, Hard's relative cost against entropy changes from +5.07% under the ensemble to -10.15% under Laplace, a 15.2-point change. BloodMNIST retains its favorable sign (-7.53% versus -4.32%). The BEMPS–random contrasts change from -10.07% to -2.75% on BloodMNIST and from -3.64% to -8.31% on PathMNIST; the ADE comparison changes from -8.88% to -0.50% on different seeds. These observations establish sensitivity to the evaluated learning pipeline, not a unique causal role for posterior fidelity. BloodMNIST simultaneously passes the agreement gate and shows an acquisition benefit over random, while neither utility variant establishes an additional benefit.

D.11 Laplace posterior and cross-method agreement gate

The posterior of Section D.10 is a Laplace approximation over the weights of a multinomial logistic last layer on frozen features: ImageNet ResNet-18 penultimate features for the image tasks and

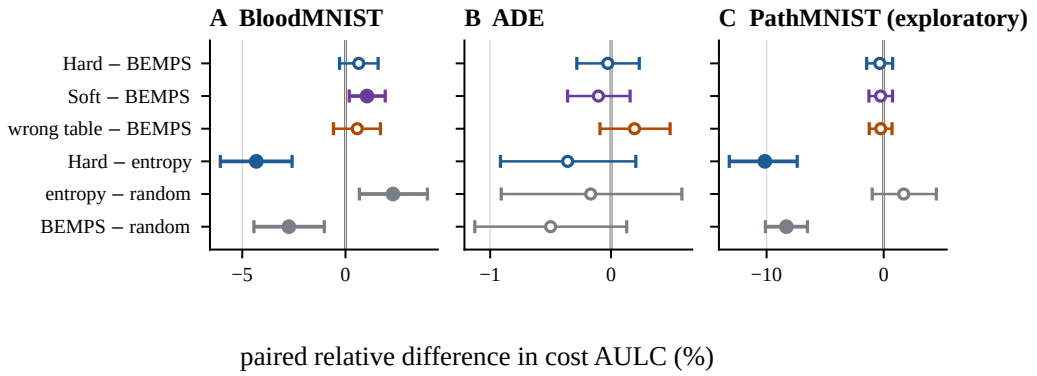

Figure 6: Evaluated contrasts of the Laplace-posterior study, in the format of Figure 3: paired relative difference in realized-cost AULC with a marginal paired-bootstrap 95% CI, filled markers marking an interval that excludes zero. Panels A and B are the Holm family; panel C is the exploratory PathMNIST condition that failed the agreement gate and is never pooled. The BloodMNIST design detects the positive control and a -4.32% advantage of Hard over entropy, and Soft is above BEMPS with a marginal interval excluding zero, not a Holm-adjusted test of harm. On PathMNIST the sign of the Hard–entropy comparison is the reverse of Figure 3B on the same seeds.

Table 16: Laplace-posterior confirmation: normalized realized-cost AULC on held-out test splits under a Laplace last-layer posterior, mean over paired fresh seeds (lower is better), in the format of Table 15. BloodMNIST and ADE form the Holm family; PathMNIST failed the outcome-blind agreement gate, is exploratory, and shows raw p values. "Rel." is the paired difference over the comparator mean, so negative favors the first arm; brackets are marginal paired-bootstrap 95% CIs. **Bold blue** marks an interval excluding zero favorably; orange marks adverse exclusion.

	Contrast	BloodMNIST ($n = 80$)		ADE ($n = 80$)		PathMNIST, exploratory ($n = 60$)	
		Rel. ↓ [95% CI]	p	Rel. ↓ [95% CI]	p	Rel. ↓ [95% CI]	raw p
Controls	BEMPS − random	**−2.75%** [−4.44, −1.03]	0.002	−0.50% [−1.13, +0.13]	0.143	**−8.31%** [−10.10, −6.51]	$< 10^{-4}$
	wrong table − BEMPS	+0.57% [−0.58, +1.69]	0.272	+0.19% [−0.09, +0.49]	0.347	−0.25% [−1.24, +0.71]	0.921
Holm family	**Hard** − BEMPS	+0.64% [−0.29, +1.58]	1.000	−0.03% [−0.28, +0.23]	1.000	−0.33% [−1.45, +0.76]	0.412
	Soft − BEMPS	+1.04% [+0.18, +1.92]	1.000	−0.10% [−0.36, +0.16]	1.000	−0.26% [−1.26, +0.75]	0.236
Secondary	entropy − random	+2.30% [+0.67, +3.97]	0.019	−0.17% [−0.91, +0.59]	0.502	+1.71% [−0.98, +4.50]	0.419
	Hard − entropy	**−4.32%** [−6.06, −2.59]	$< 10^{-4}$	−0.36% [−0.91, +0.21]	0.181	**−10.15%** [−13.19, −7.39]	$< 10^{-4}$

PubMedBERT embeddings for ADE, standardized and projected to 32 principal components plus a bias. The layer is fit to its maximum-a-posteriori weights by Newton's method under a Gaussian prior of unit precision, and $K = 5$ posterior members are drawn from the Gaussian centred at that fit with the inverse Hessian as covariance, using common random numbers per member seed so that a hypothetical update reweights a fixed draw rather than resampling it. No network is retrained. Decision problems, costs, λ, endpoint, family, test and Holm control are those of Sections D.9 and 4.3 (ADE at the replication's 1.5:1 cost and $\lambda = 1$); seeds are 300–379 for BloodMNIST and ADE and 300–359 for PathMNIST (1,320 runs, commit `e415d20`). The protocol, including the gate below and the status of PathMNIST, was frozen before any confirmatory run.

The gate measures agreement between methods, not update fidelity. On pilot seeds 290–294 with two acquisition rounds, a task entered the family only if the median second-round overlap between the DUSAL-Hard and BEMPS batches reached 30 of 50 (60 of 100 on ADE). BloodMNIST gave 34 and ADE 74 and entered; PathMNIST gave 26 and was dropped, its two contrasts counting as unsupported in the family of 6. Before any confirmatory outcome was read, PathMNIST was nevertheless scheduled in full as an exploratory intermediate-persistence condition, to be reported in its own block and never pooled; that is its status in Table 16 and Figure 6. Table 17 reports the resulting batch overlaps. BloodMNIST and ADE ran on one machine; the PathMNIST seed blocks were distributed over four, each seed's six arms on one of them.

Table 17: Cross-method batch agreement under the Laplace last-layer posterior: items shared between the batches two arms select in the same round, as median [min, max] over seeds, out of a batch of 50 (100 on ADE). Round 0 is the common first batch; from round 1 onward each arm scores candidates with its own updated posterior. Under the retrained ensemble the corresponding BloodMNIST round-1 median was 1.

	BloodMNIST			ADE			PathMNIST (exploratory)		
Arms	round 0	round 1	round 5	round 0	round 1	round 5	round 0	round 1	round 5
Hard, BEMPS	49 [46, 50]	41 [18, 50]	5 [0, 10]	94 [8, 100]	74 [19, 97]	69 [36, 83]	49 [46, 50]	34 [8, 50]	1 [0, 5]
Hard, wrong table	48 [45, 50]	36 [16, 49]	5 [1, 17]	94 [69, 100]	74 [37, 96]	65 [42, 83]	49 [46, 50]	24 [7, 50]	1 [0, 4]
Soft, BEMPS	49 [46, 50]	43 [18, 50]	5 [0, 18]	94 [8, 100]	75 [20, 98]	71 [42, 87]	49 [47, 50]	37 [9, 50]	1 [0, 7]

Table 18: Complete confirmatory batch-selection results. Δ is the paired relative difference in realized-cost AULC from top-b selection under the same score; negative values favour DUSAL-Div. "fav" counts seeds with lower cost. MDE is the minimum detectable effect under the study's convention $2.486\,\mathrm{sd}/\sqrt{n}$. Holm correction is applied separately within each four-task family.

Task	n	Δ [95% CI]	fav	raw p	Holm p	MDE
Primary family: Hard-Div − Hard top-b						
Blood	80	-0.37% $[-1.73, +0.97]$	44/80	0.225	0.225	1.73%
Path	60	-3.85% $[-6.11, -1.69]$	42/60	0.0004	0.002	2.83%
Pneumonia	80	-1.09% $[-1.82, -0.34]$	48/80	0.003	0.008	0.94%
ADE	80	-1.16% $[-1.90, -0.41]$	48/80	0.003	0.008	0.96%
Second family: BEMPS-Div − BEMPS top-b						
Blood	80	-1.12% $[-2.50, +0.20]$	39/80	0.138	0.138	1.72%
Path	60	-3.10% $[-5.57, -0.65]$	40/60	0.008	0.015	3.14%
Pneumonia	80	-1.27% $[-2.06, -0.48]$	45/80	0.003	0.008	1.01%
ADE	80	-1.38% $[-2.23, -0.52]$	49/80	0.002	0.006	1.09%

The within-action share of acquired items, the composition statistic of Section D.9, is 0.469, 0.415, 0.503, 0.502, 0.503 and 0.509 for random, entropy, BEMPS, Hard, Soft and the wrong table on BloodMNIST, and 0.555, 0.482, 0.680, 0.677, 0.677 and 0.687 on PathMNIST; it is undefined for the two-class ADE problem. Every arm that carries a utility term, correct or scrambled, acquires the same composition as BEMPS.

D.12 Batch selection: results and reproducibility checks

DUSAL-Div applies Algorithm 3 followed by Algorithm 4. Each candidate has one score-change coordinate per frozen reference point, with coordinate m scaled by $\sqrt{\omega_m}$. Lloyd k-means runs in double precision over the candidate set, with k equal to the realized batch size. Empty clusters are repaired by the geometric farthest-member rule, and the candidate nearest each recomputed centroid is selected. The scalar acquisition value does not enter the within-cluster choice. To isolate the batch rule, each task retains its parent experiment's candidate set: the full unlabelled pool on BloodMNIST (11,799), PneumoniaMNIST (4,608), and ADE (4,101), and the original 20,000-candidate subsample on PathMNIST, whose pool is larger. Algorithm 4 thus operates on the full pool for three tasks and a prefiltered set for PathMNIST. The realized k is 50 on the three batch-50 tasks and 100 on ADE.

The study adds two DUSAL-Div arms, comprising 600 runs on seeds 100–179 (binary), 300–379 (multi-action). Each run is paired with the stored top-b run for the same seed from Sections D.9 and D.8. The paired corpora share the code fingerprint `cbae67f775ea1732` (commit `64ee8bc`); the new arms use configuration overrides on the same checkout. Before confirmation, we reran one stored arm per task on a different machine. The reproducibility gate required identical code fingerprints, realized costs at every round, acquired indices, and endpoints. The gate passes on 4 of 4 tasks. BreastMNIST was excluded before any DUSAL-Div run because its positive control was inconclusive (Table 14). This exclusion follows the study's eligibility rule, not evidence that batch selection cannot help.

Table 18 reports both families. The paired comparisons do not establish an additional benefit from utility augmentation: Hard-Div minus BEMPS-Div is +0.08%, -0.61%, -0.41%, and +0.80% on the

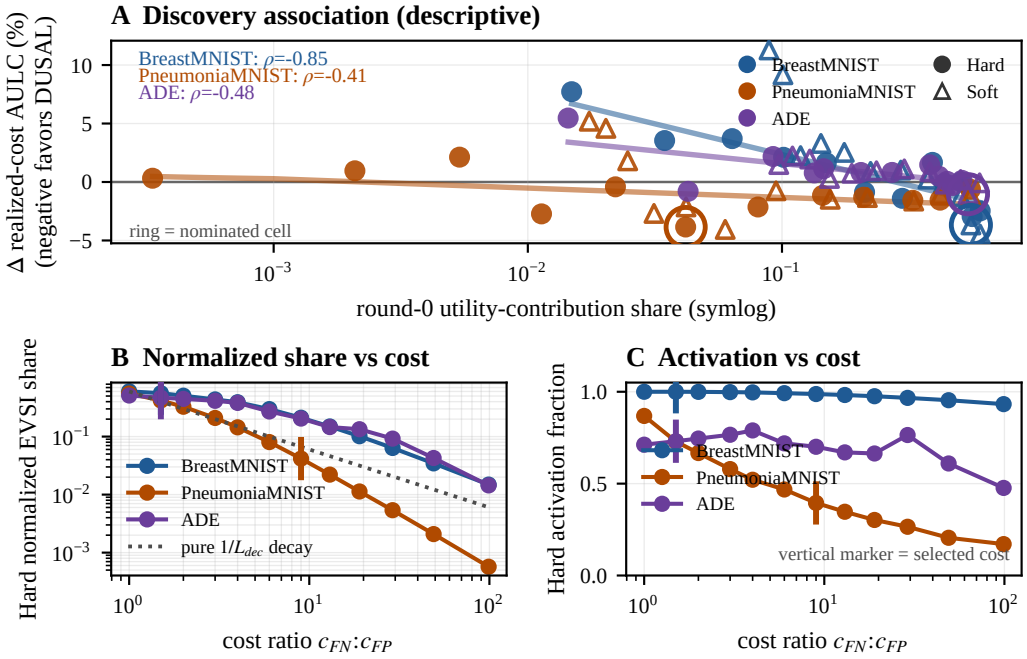

Figure 7: Utility contribution across the exploratory discovery sweep. **(A)** Validation relative realized-cost AULC versus the round-0 non-base acquisition share for 12 cost ratios × three tasks × two variants (72 cells; 20 paired seeds each). For Hard, the non-base term is normalized declared-cost EVSI; for Soft, it is the soft-action $\lambda\tau$ KL term and is not decision-regret EVSI. Filled circles denote Hard and open triangles Soft; lines and ρ values are descriptive Hard-only fits, not confirmatory predictors or causal estimates. Rings mark the nominated Hard or Soft cell in panel A; panel B's dotted curve shows pure $1/L_U^{\mathrm{dec}}$ decay; panel C reports the fraction of candidates whose normalized utility contribution exceeds 10^{-15}, with vertical markers at each selected cost, and panel B's horizontal axis is the FN:FP ratio, which equals L_U^{dec}. Hard activation at the selected PneumoniaMNIST ratio is 0.395.

four tasks, with p values of 0.930, 0.734, 0.389, and 0.053, respectively. The ADE contrast is closest to significance and favors BEMPS-Div in direction. Against the standard BEMPS top-b baseline, Hard-Div gives -3.69% [-6.29,-1.07] on PathMNIST and -1.67% [-2.46,-0.89] on PneumoniaMNIST; the other two contrasts are inconclusive.

Three exploratory analyses describe the result's scope. First, Div changes batch composition substantially. Median first-batch overlap with Hard top-b is 0 of 50 on BloodMNIST and 0 on PathMNIST. On ADE, overlap falls from 35 of 100 to 6 by the sixth batch. Second, on PneumoniaMNIST and ADE, the gain peaks before the tenth batch and returns to zero by the final batch; only PathMNIST retains the gain throughout the schedule. This pattern is consistent with batch redundancy being most costly when the labelled set is small. Third, gains occur in both within-machine and cross-machine comparisons. The 80 within-machine ADE pairs give -1.16%, and the 80 cross-machine PneumoniaMNIST pairs give -1.09%. PathMNIST gives -4.20% across machines and -3.14% within machines. These exploratory analyses have no multiplicity correction.

D.13 Mechanism characterization

How large the utility term is, and whether a larger term goes with a better outcome, can be read off the discovery sweep. Figure 7 plots both across all 72 cells, and Table 19 gives the same quantities by cost ratio.

Round-0 candidate sets agreed exactly in 1440/1440 arm comparisons, allowing us to extract the utility term directly. A probability-space boundary-mass proxy, $|p - t^\star| < \epsilon$, is misleading here:

Table 19: Round-0 EVSI share of the Hard acquisition (top) and fraction of candidates with a utility contribution above 10^{-15} (bottom), by cost ratio. Both quantities are computed directly from stored round-0 traces; every arm shares the fitted state and candidate indices, so the arm difference is the utility term itself. The validation-selected costs are marked. With $\lambda = 1$, the share contains an explicit $1/L_U^{\mathrm{dec}}$ factor; if the raw EVSI-to-BEMPS ratio were fixed, its odds $s/(1-s)$ would decay as $1/L_U^{\mathrm{dec}}$. For each seed, the raw $\sum EVSI / \sum$ BEMPS ratio equals $L_U^{\mathrm{dec}}\, s/(1-s)$. We transform each seed before averaging, not the mean share. The ratio of these raw-ratio means is $0.993\times$ on BreastMNIST and $1.057\times$ on ADE from 1:1 to 99:1, so a small normalized share need not mean an inactive mechanism. On PneumoniaMNIST it falls to $0.045\times$, alongside lower activation. These are descriptive, posterior-conditional diagnostics, not predictors of downstream efficacy.

FN:FP	1	1.5	2	3	4	6	9	13	19	29	49	99
EVSI share												
BreastMNIST	0.602	**0.560**	0.506	0.437	0.391	0.297	0.211	0.149	0.101	0.064	0.035	0.015
PneumoniaMNIST	0.552	0.419	0.327	0.210	0.144	0.080	**0.042**	0.022	0.011	0.005	0.002	0.001
ADE	0.511	**0.468**	0.446	0.414	0.381	0.271	0.204	0.146	0.134	0.092	0.043	0.014
Activation rate (fraction with normalized utility contribution $> 10^{-15}$*)*												
BreastMNIST	1.000	**1.000**	1.000	0.998	0.997	0.992	0.988	0.982	0.976	0.967	0.954	0.933
PneumoniaMNIST	0.868	0.729	0.667	0.579	0.521	0.468	**0.395**	0.347	0.302	0.265	0.205	0.171
ADE	0.712	**0.731**	0.745	0.766	0.790	0.719	0.701	0.670	0.664	0.765	0.610	0.477

Table 20: Round-0 top-b selection under a hypothetical utility weight, computed without new runs. Because the utility term enters linearly, $\mathrm{value}(\lambda) = \mathrm{BEMPS} + \lambda\,[(1/L_U^{\mathrm{dec}})\,EVSI]$ and both terms are recoverable from the stored round-0 traces, so the batch selected under any λ is arithmetic. Entries are the Jaccard similarity of the top-b batch to the $\lambda = 0$ batch. At the 1.5:1 cells used by the λ studies, selection saturates by $\lambda \approx 8$, which fixed their levels; at PneumoniaMNIST 9:1, where few candidates are active, the batch keeps changing up to $\lambda = 32$.

Dataset	FN:FP	b	λ=0.25	0.5	1	2	4	8	16	32
BreastMNIST	1.5	20	0.91	0.89	0.84	0.81	0.80	0.79	0.77	0.77
PneumoniaMNIST	1.5	50	0.96	0.92	0.89	0.85	0.84	0.82	0.81	0.81
PneumoniaMNIST	9	50	0.98	0.96	0.93	0.89	0.82	0.74	0.67	0.61
ADE	1.5	100	0.73	0.60	0.50	0.41	0.35	0.33	0.31	0.31

with saturated predictions, a band around a tail threshold can include many points that remain far from the boundary in log-odds and do not cross it under a one-label update. We therefore report the directly extracted term. Over whole trajectories the acquired-index Jaccard similarity between DUSAL and BEMPS is between 0.55 and 0.70 across all 72 cells; at the selected cells it is 0.566/0.557 (BreastMNIST Hard/Soft), 0.679/0.676 (PneumoniaMNIST), and 0.608/0.638 (ADE). The base acquisition value $\mathbb{E}[D_H(P', p)]$ measures how far one label moves the reference predictive. Its per-round median candidate, summarized by the median over rounds and seeds at the selected ratio, is $1.6e-8$ on PneumoniaMNIST (per-round maximum candidate, same summary, $7.0e-4$), $4.8e-8$ on ADE (maximum $7.9e-6$), and $3.6e-4$ on BreastMNIST (maximum $8.7e-3$); at round 0 alone the medians are larger but the ordering is the same. When score separations are small, finite-reweighting or retraining variation may change acquisition rankings. Section D.15 measures ranking agreement directly. This diagnostic characterizes update sensitivity without identifying the cause of the confirmation null.

D.14 Utility-weight mechanism studies

Registration and scope. Both studies were registered before the confirmation family completed. Their primary analyses, secondary endpoints, and relation to the confirmation were fixed in advance: they vary λ at a fixed cost ratio to characterize the mechanism, report every direction of effect, and cannot revise the confirmation conclusion. They ran on separate hardware and a different framework build from the confirmation campaign. Comparisons therefore remain within each study and are not pooled with Table 11. Levels are $\lambda \in \{0, 1, 4, 8\}$, with $\lambda = 0$ the paired BEMPS baseline and the levels bounded by Table 20; the arm is DUSAL-Hard with the Hard potential; the cost ratio is 1.5:1, chosen by the registration's mechanism criterion (round-0 activation near one and EVSI share above 0.3, Table 19) rather than by observed performance; protocols otherwise follow Table 8; evaluation

Table 21: Utility-weight studies. Each cohort pairs every λ with $\lambda = 0$ within seed; entries are relative change in realized-cost AULC (negative favors the larger weight), favorable seed counts (seeds with a negative paired difference; for the slope, seeds with a negative slope), and two-sided paired-bootstrap 95% CIs on the paired difference (on the slope for the slope rows). The primary is the within-seed slope on $\log(1 + \lambda)$ with its two-sided bootstrap p; "first third" is the pre-registered secondary, partial AULC over the first third of the budget grid. The BreastMNIST $n = 60$ cohort is the pre-declared independent replication of the $n = 20$ cohort; the pooled estimate is descriptive and carries no p value by design.

Cohort	Contrast	Rel. change	Favorable	95% CI
ADE, $n = 20$	λ=1 vs 0	$+0.27\%$	9/20	$[-0.002026, +0.002736]$
	λ=4 vs 0	$+0.81\%$	5/20	$[-0.000763, +0.002810]$
	λ=8 vs 0	-0.12%	9/20	$[-0.001807, +0.001440]$
	slope on $\log(1 + \lambda)$	$+0.05\%$	10/20	$[-0.000657, +0.000748]$; p: 0.815
	first third, λ=8 vs 0	$+1.61\%$	5/20	$[+0.000679, +0.005161]$
BreastMNIST, $n = 20$	λ=1 vs 0	-4.51%	15/20	$[-0.013105, -0.002422]$
	λ=4 vs 0	-1.36%	10/20	$[-0.008447, +0.003860]$
	λ=8 vs 0	-3.89%	12/20	$[-0.012237, -0.000570]$
	slope on $\log(1 + \lambda)$	-1.01%	12/20	$[-0.004337, +0.001089]$; p: 0.233
	first third, λ=8 vs 0	-1.12%	11/20	$[-0.013803, +0.007564]$
BreastMNIST replication, $n = 60$	λ=1 vs 0	$+0.30\%$	32/60	$[-0.005171, +0.006360]$
	λ=4 vs 0	$+0.35\%$	32/60	$[-0.005375, +0.006600]$
	λ=8 vs 0	-0.07%	28/60	$[-0.005658, +0.005406]$
	slope on $\log(1 + \lambda)$	-0.01%	28/60	$[-0.002413, +0.002321]$; p: 0.988
	first third, λ=8 vs 0	$+0.34\%$	25/60	$[-0.010328, +0.011809]$
BreastMNIST pooled, $n = 80$	λ=1 vs 0	-0.89%	47/80	$[-0.005962, +0.003241]$
	λ=4 vs 0	-0.07%	42/80	$[-0.004884, +0.004635]$
	λ=8 vs 0	-1.02%	40/80	$[-0.006138, +0.002744]$
	slope on $\log(1 + \lambda)$	-0.26%	40/80	$[-0.002376, +0.001488]$; p: not reported (escalated sample)
	first third, λ=8 vs 0	-0.00%	36/80	$[-0.008804, +0.008590]$

Table 22: What λ changes and what it does not. Jaccard is the similarity of the full acquired index set to the $\lambda = 0$ set; "differ" is the mean number of acquired labels not in the $\lambda = 0$ set (of 1,500 on ADE, 300 on BreastMNIST). AUROC and ECE are AULC means over seeds.

	ADE, $n = 20$				BreastMNIST, $n = 20$				BreastMNIST replication, $n = 60$			
λ	Jaccard	differ	AUROC	ECE	Jaccard	differ	AUROC	ECE	Jaccard	differ	AUROC	ECE
0	1.000	0	0.8995	0.0990	1.000	0	0.8509	0.1187	1.000	0	0.8462	0.1252
1	0.613	360	0.9004	0.0981	0.579	80	0.8518	0.1132	0.574	82	0.8506	0.1260
4	0.594	382	0.8987	0.0998	0.573	82	0.8432	0.1240	0.564	84	0.8508	0.1236
8	0.586	392	0.8997	0.0977	0.570	82	0.8507	0.1188	0.570	83	0.8504	0.1234

is on the validation split. Seeds are 200–219 (ADE) and 200–219 (BreastMNIST), disjoint from discovery and confirmation. The primary is the within-seed OLS slope of AULC on $\log(1 + \lambda)$, null slope zero, two-sided; the pre-registered secondary is partial AULC over the first third of the budget grid. After the BreastMNIST result at 20 seeds proved unresolved, the sample was extended by a cohort of 60 seeds (220–279); the extension was declared before it ran as an independent replication analyzed first and on its own, with the pooled 80-seed estimate descriptive only, no p value, and no further extension in either direction.

Effects of utility weight. On ADE the slope is $+0.000073$ per unit $\log(1 + \lambda)$ ($+0.05\%$ of baseline; $p = 0.815$; 10/20 seeds negative), and the pre-registered early-budget secondary is adverse and monotone in λ ($\lambda = 8$: $+1.61\%$, two-sided bootstrap $p = 0.009$, which survives a Holm correction over the three levels compared, although a paired t test does not); because the registration fixed the window but not the level, it is reported as a descriptive direction rather than a prespecified contrast. On BreastMNIST the first cohort's $\lambda = 1$ contrast, -4.51% with 15/20 favorable seeds, sits at that cohort's own MDE of 4.22% (paired SD 7.60% of baseline), the three levels are pairwise inseparable ($|t| = 1.38$, 0.30, 1.30 for 1 vs 4, 1 vs 8, 4 vs 8), and the ordering is non-monotone. The independent replication cohort is flat (slope -0.000024, $p = 0.988$; $\lambda = 1$: $+0.30\%$), which is also what the 80-seed confirmation of the same dataset–cost pair found for $\lambda = 1$ ($+0.12\%$; the λ study evaluates on the validation split and the confirmation on the test split). In every cohort λ changes about a quarter of the acquired labels; on ADE and in the replication cohort realized cost, AUROC, and ECE show little descriptive change and no consistent dose response (Table 22), while the first BreastMNIST cohort's marginal intervals at $\lambda = 1$ and 8 exclude zero without a resolved

Table 23: Median ranking agreement relative to converged-refit Bregman plug-in displacement. Each cell summarizes ten seeds. ρ is Spearman correlation, "top-10" is set overlap, and "selected pct." is the converged-refit plug-in percentile of the approximation's selected candidate. The converged-refit plug-in self-comparison is 1.0 by construction. These are update-sensitivity diagnostics, not refit expected-score-gain acquisition estimates.

Task	Method	Finite reweight			One-step refit		
		ρ	top-10	selected pct.	ρ	top-10	selected pct.
Breast U2	BEMPS	0.407	0.50	0.750	0.038	0.35	0.344
	Hard	0.385	0.50	0.734	0.036	0.40	0.359
	Soft	0.387	0.50	0.719	0.050	0.35	0.375
Retina U3	BEMPS	0.064	0.35	0.422	0.189	0.40	0.656
	Hard	0.069	0.30	0.484	0.208	0.30	0.797
	Soft	0.074	0.40	0.391	0.180	0.35	0.688
ADE	BEMPS	0.241	0.45	0.813	0.157	0.30	0.594
	Hard	0.245	0.45	0.813	0.157	0.30	0.594
	Soft	0.243	0.45	0.813	0.153	0.30	0.594

slope. At equal run budget a dose-response is 1.85× less precise for the range contrast than a two-arm comparison (2.11× at six levels). The design trades precision for information about the response shape and a possible optimum; it does not offer a power advantage.

D.15 Bregman plug-in update sensitivity

The practical acquisition exactly reweights the current finite ensemble by a hypothetical label. Between active-learning rounds, however, the code trains a new ensemble from scratch. This is not the fixed-Θ sequential posterior recursion assumed by the convergence theorem. The diagnostic compares the finite-support acquisition with expected Bregman displacement from the current predictive after one-step and converged hypothetical refits on the same frozen candidate sets. A refit need not satisfy the predictive-martingale identity, so the latter displacement is not generally equal to expected potential gain and must not be called refit acquisition. Its 30 blocks (three tasks, ten seeds) and 270 method/update rows pass all artifact checks. The three audited tasks are Breast U2, Retina U3, and the binary ADE task; the PneumoniaMNIST confirmation task is not among them, so the diagnostic characterizes the update operator, not that specific confirmation cell.

Table 23 gives the per-task agreements. The frozen overall median-ρ gate was 0.40, with a dataset–method floor of 0.15. Overall finite-reweight medians are 0.194/0.210/0.217 for BEMPS/Hard/Soft; all fail the overall gate, and all Retina finite-reweight rows fail the floor, giving the classification POSTERIOR-SENSITIVE. This establishes disagreement with the refit Bregman plug-in, rather than with a refit expected-score-gain ranking; the exact finite-support identity is unchanged. Both BEMPS and DUSAL fail the agreement gate. Table 11 compares the executed finite-ensemble/retraining pipelines, but the diagnostic limits how directly their results can be attributed to the exact acquisitions in Theorem 3. This earlier diagnostic has no refit-versus-refit replication control; the subsequent direct-Q study below supplies one for a frozen last-layer model. Neither diagnostic alone identifies why the full-network public-data pipelines have similar decision cost.

D.16 Direct expected-potential-gain refit audit

The 30-block REFIT-Q study evaluates direct expected potential gain on a fixed cohort,

$$\widehat{A}_{Q,G}(x) = \mathbb{E}_{Y\sim\bar{p}_n(\cdot|x)}\mathbb{E}_{z\sim\mu}\left[G\Big(\bar{p}^{x,Y}_{\mathrm{refit}}(\cdot \mid z)\Big) - G(\bar{p}_n(\cdot \mid z))\right],$$

where each $\bar{p}$ averages members at the individual reference point z before applying G; it does not average reference predictions before applying the potential or substitute a Bregman displacement. The three tasks are Breast U2, Retina U3, and binary ADE at 19:1, with ten fresh seeds 30–39. Frozen features feed a CE-trained last-layer model with a Gaussian prior and Laplace approximation: each block uses 20 posterior draws, 32 candidates, and 128 reference items. These are hypothetical-update diagnostics, not full-network active-learning trajectories.

All 360 method/update rows pass the immutable-artifact and direct-Q recomputation audit. Converged refits A and B use different initial states and independent antithetic draw banks, providing a replication

Table 24: Direct-Q ranking audit, distinct from the Bregman plug-in audit. Each entry is a median Spearman correlation over all 30 task–seed blocks for the indicated method, not a median of dataset medians. The left column compares finite reweighting with converged refit A; the right compares converged refit replicas B and A. Higher agreement is better, but neither column is an efficacy endpoint or comparison against an exact posterior.

Method	Finite reweight vs. A	Refit B vs. A
BEMPS	-0.028	0.600
Hard	-0.002	0.491
Soft	-0.007	0.466

control that the earlier plug-in diagnostic lacks. Table 24 shows both comparisons. The pooled A/B median correlation 0.498 and top-10 overlap 0.55 miss the frozen $0.80/0.70$ gates, yielding REFIT-MC-SENSITIVE, not a fidelity pass. Consequently the refit reference itself is insufficiently stable for attributing disagreement solely to finite reweighting. This finding motivates separating initialization and posterior-sampling variability before testing a full-network correction; it does not invalidate the finite-model identity or establish a cause of the public-data null.

D.17 POSTERIOR-BACKEND SENSITIVITY: SELECTION CHANGES, NO ESTABLISHED BENEFIT

A separately frozen family contains 270 runs: Breast U2, Retina U3, and binary ADE at 19:1, crossed with three posterior configurations and three acquisition methods on seeds 260–269. Within every backend, BEMPS, Hard ($\lambda = 1$), and Soft ($\lambda = 1, \tau = 0.05$) share CE training, initial labels, reference cohort, and seed streams. E5 is a five-network ensemble without dropout; M20/M50 train one network with dropout probability 0.3 and use respectively 20/50 replayable head masks, locked across candidate, reference, and test predictions within a round. E5 versus MC therefore changes both training regularization and posterior approximation, not only the sample count. Total labeled-set sizes range from 10–110, 25–216, and 100–1,600; fixed reference sizes are 536, 1,024, and 1,024.

This family was executed independently, out of the parent campaign's frozen stage order; that deviation is recorded in its standalone provenance. Its scientific manifest, all 270 runs, and all associated prediction, target-measure, support, and acquisition files pass the complete audit. It is not a replication or extension of the original six dataset-specific confirmation contrasts: tasks, costs, budgets, and the aggregate estimand differ.

No aggregate cost contrast passes Holm correction. For each seed, the primary effect is the equal-weight mean of the three task-specific normalized realized-cost AULC differences. The six Hard/Soft $\times$ backend contrasts use ten paired seed-level units and one Holm family, not 30 independent task–seed units. None passes. Hard–M20 has a nominal one-sided $p = 0.0322$ but Holm $p = 0.1934$; its marginal interval does not override the family rule (Table 25). The four prespecified backend interactions also remain unresolved; nonsignificance is not backend equivalence. Tables 26 and 27 retain all task-specific results rather than promoting favorable cells.

Initial-batch sensitivity. We additionally compare initial batches within this same family, using the same ordered candidate IDs, seed blocks, and overlap definition on both axes (Table 28). Within a backend, changing acquisition replaces a minority of the first batch; changing posterior configuration replaces substantially more in these tasks. Even M20 versus M50 replaces 39.76% on average across the fixed tasks and methods. This is a post-hoc descriptive result, not a test that posterior approximation causes the efficacy null: E5–MC includes a training change, MC supports are not a same-parent nested-draw experiment, and no full-network retraining noise floor is measured here. Its practical implication is narrower: reproducible acquisition requires the posterior/support policy as well as the score formula to be specified.

D.18 SCOPE OF THE ACQUISITION DIAGNOSTICS

Utility augmentation modestly increases runtime in the reported ensemble runs. Acquisition takes 12.7 seconds per round for Hard and 10.5 for BEMPS on BloodMNIST, compared with 18.7 and 13.8

Table 25: Complete frozen posterior-sensitivity inference. Top: DUSAL minus BEMPS normalized realized-cost AULC, averaged equally over the three fixed tasks within each of ten seeds; negative favors DUSAL. Bottom: MC-minus-E5 differences of those effects, not new efficacy contrasts. Intervals are marginal paired-bootstrap 95% CIs. Primary p values are one-sided with Holm-6; interaction p values are two-sided with Holm-4. No adjusted test rejects. The two families are separate from the original 720-run confirmation.

Backend	Method	Mean Δ	Marginal 95% CI	Raw p	Holm p
E5	Hard	-0.003778	$[-0.008668, +0.001605]$	0.1162	0.4648
E5	Soft	-0.001645	$[-0.006673, +0.003241]$	0.3125	0.9375
M20	Hard	-0.003868	$[-0.007053, -0.000673]$	0.0322	0.1934
M20	Soft	-0.001788	$[-0.003837, +0.000561]$	0.0654	0.3271
M50	Hard	$+0.000632$	$[-0.004154, +0.005678]$	0.6523	0.9375
M50	Soft	-0.000324	$[-0.002791, +0.002334]$	0.3848	0.9375
M20–E5	Hard	-0.000090	$[-0.005441, +0.005709]$	0.9219	1.0000
M20–E5	Soft	-0.000143	$[-0.005328, +0.004545]$	0.7695	1.0000
M50–E5	Hard	$+0.004409$	$[-0.003942, +0.012575]$	0.4316	1.0000
M50–E5	Soft	$+0.001321$	$[-0.004796, +0.007331]$	0.6953	1.0000

Table 26: All task–backend absolute results in the independent posterior family. Entries are mean $\pm$ sample standard deviation of normalized realized-cost AULC over ten seeds; lower is better. All columns use CE training. Comparisons of acquisition methods hold the backend fixed; comparisons across backends also change the model configuration. Per-task summaries are descriptive, not the frozen equal-task primary estimand.

Task	Backend	BEMPS	Hard	Soft
Breast U2	E5	0.144562 ± 0.022723	0.135464 ± 0.026745	0.143477 ± 0.027324
	M20	0.199485 ± 0.021565	0.190745 ± 0.012975	0.195699 ± 0.019637
	M50	0.201186 ± 0.017890	0.202124 ± 0.016231	0.198896 ± 0.015577
Retina U3	E5	0.088495 ± 0.006853	0.088042 ± 0.009180	0.086446 ± 0.006635
	M20	0.103725 ± 0.008410	0.101207 ± 0.008553	0.102981 ± 0.007248
	M50	0.100076 ± 0.007147	0.100403 ± 0.010482	0.100896 ± 0.010056
ADE 19:1	E5	0.042050 ± 0.002994	0.040267 ± 0.003801	0.040248 ± 0.002853
	M20	0.035391 ± 0.003422	0.035044 ± 0.003600	0.034558 ± 0.002316
	M50	0.034228 ± 0.002725	0.034857 ± 0.002645	0.034726 ± 0.001569

on PathMNIST. Complete BloodMNIST runs take 27.6 and 27.1 minutes, respectively, on one GPU. These timings describe the evaluated implementation and retraining, not general complexity.

The following diagnostics address possible explanations for the unsupported public-data improvements. They are not an exhaustive identification of causes, and descriptive agreement is distinct from a test of decision benefit.

Utility activity and weight. At the selected binary costs, first-round utility shares are 0.560, 0.042, and 0.468, with nonzero contributions for 1.000, 0.395, and 0.731 of candidates. Thus inactive EVSI is not a general explanation, although it can matter at particular costs or rounds. PneumoniaMNIST at 9:1 has BEMPS sensitivity 99.65% and normalized cost 0.024758. Weight studies up to $\lambda = 8$ do not find a reproducible improvement. This bounds the evidence for the tested weights, not every possible weighting scheme.

Cost asymmetry and detectable effects. In validation, relative cost increases with the ratio on BreastMNIST (Spearman +0.853 for Hard and +0.972 for Soft) and ADE Soft (+0.825). BreastMNIST estimates at 99:1 are +7.72% and +11.29%. These exploratory trends do not support a general claim that stronger asymmetry improves utility-aware acquisition. The positive controls detect larger acquisition effects in several settings, but do not establish equivalence between DUSAL and BEMPS or quantify remaining achievable improvement.

Table 27: All task-specific DUSAL-minus-BEMPS contrasts from Table 26. Negative means lower normalized realized-cost AULC. These post-hoc CIs use 40,000 paired-seed bootstrap resamples at ten seeds and are marginal 95% intervals conditional on the fixed test cohort; they do not replace the six aggregate Holm tests or support selection of a best dataset/backend after inspection.

Task	Backend	Method	Mean Δ	Marginal 95% CI
Breast U2	E5	Hard	-0.009097	$[-0.020854, +0.003600]$
		Soft	-0.001085	$[-0.014046, +0.011726]$
	M20	Hard	-0.008740	$[-0.018400, +0.001000]$
		Soft	-0.003786	$[-0.009295, +0.002587]$
	M50	Hard	$+0.000938$	$[-0.011964, +0.014064]$
		Soft	-0.002290	$[-0.006742, +0.001888]$
Retina U3	E5	Hard	-0.000453	$[-0.004228, +0.003249]$
		Soft	-0.002048	$[-0.004360, +0.000230]$
	M20	Hard	-0.002518	$[-0.005016, +0.000836]$
		Soft	-0.000745	$[-0.003645, +0.001875]$
	M50	Hard	$+0.000327$	$[-0.004867, +0.005152]$
		Soft	$+0.000820$	$[-0.004688, +0.007140]$
ADE 19:1	E5	Hard	-0.001783	$[-0.003874, +0.000249]$
		Soft	-0.001803	$[-0.003873, +0.000275]$
	M20	Hard	-0.000347	$[-0.001919, +0.001160]$
		Soft	-0.000833	$[-0.002249, +0.000627]$
	M50	Hard	$+0.000630$	$[-0.000468, +0.001770]$
		Soft	$+0.000498$	$[-0.000614, +0.001667]$

Table 28: Same-family round-0 batch replacement. For equal-sized batches A, B, replacement is $1 - |A \cap B|/|A|$, not $1-$Jaccard. All saved selections were replayed from stable descending raw scores and matched candidate IDs. Panel A averages over three fixed tasks and ten seeds; panel B uses the same units, stratifies by fixed acquisition method, and then averages methods in the last column. These percentages describe selection sensitivity, not efficacy or causal variance.

A: change acquisition, hold backend fixed

Backend	Hard vs. BEMPS	Soft vs. BEMPS
E5	7.47%	4.40%
M20	10.30%	5.63%
M50	9.53%	6.93%

B: change backend, hold acquisition fixed

Backend pair	BEMPS	Hard	Soft	Mean
M20 / E5	97.37%	96.80%	96.70%	96.96%
M50 / E5	96.37%	95.33%	95.17%	95.62%
M50 / M20	41.43%	38.73%	39.10%	39.76%

Update approximations and batch agreement. In the separate frozen-feature Laplace audit, finite-reweighting scores correlate with converged-refit gains at -0.028, -0.002, and -0.007. Independent refits correlate at 0.600, 0.491, and 0.466. These are direct update-comparison diagnostics for that setting, not measurements of the deep-ensemble pipeline as a whole. Conversely, the confirmation studies' cross-arm batch overlaps compare different policies after different labels. They do not validate hypothetical updates against matched refits. The Laplace gate therefore leaves update fidelity unresolved despite improved cross-method agreement.

Interpretation. The wrong-table controls show that acquired-set divergence alone is insufficient evidence of decision benefit. Similar aggregate overlaps and composition statistics do not prove identical decision content. Backend comparisons change representation and learning choices as well as posterior approximation. Their baseline reversals are pipeline-dependent findings, not evidence that posterior fidelity uniquely determines acquisition performance. Finally, lower declared cost need not imply fewer missed cases: the unresolved BreastMNIST Soft estimate combines more false

Table 29: Secondary endpoints of the multi-action confirmation: paired relative change against BEMPS (%) with the raw one-sided p value in parentheses; "high-stakes share" is two-sided. Negative favors the arm for cost-type endpoints and positive for the others. No Hard or Soft entry survives Holm correction over the 64-test scan.

	BloodMNIST ($n = 80$)			PathMNIST ($n = 60$)		
Endpoint (AULC)	Hard	Soft	Wrong table	Hard	Soft	Wrong table
Recall on penalized classes	+0.06% (0.271)	−0.13% (0.861)	+0.04% (0.415)	−0.13% (0.605)	+0.32% (0.117)	+0.02% (0.562)
Net benefit	+0.26% (0.030)	−0.02% (0.514)	+0.16% (0.124)	−0.39% (0.712)	+0.10% (0.353)	+0.00% (0.501)
Accuracy	+0.10% (0.019)	+0.03% (0.257)	+0.10% (0.005)	−0.07% (0.440)	−0.31% (0.793)	+0.04% (0.375)
Macro-F1	+0.09% (0.070)	+0.05% (0.178)	+0.14% (0.001)	+0.19% (0.356)	−0.23% (0.729)	+0.01% (0.461)
AUROC	+0.02% (0.145)	−0.00% (0.789)	+0.02% (0.040)	+0.21% (0.166)	+0.10% (0.409)	+0.15% (0.220)
AUPRC	+0.12% (0.052)	+0.06% (0.317)	+0.19% (0.002)	+0.36% (0.295)	−0.29% (0.836)	+0.14% (0.398)
Brier score	−0.50% (0.028)	−0.16% (0.368)	−0.57% (0.009)	−0.04% (0.361)	+0.21% (0.616)	−0.10% (0.367)
NLL	−0.34% (0.270)	+0.04% (0.788)	−0.69% (0.037)	−0.44% (0.249)	−0.19% (0.513)	−0.48% (0.311)
ECE	−1.32% (0.121)	+0.05% (0.741)	−1.15% (0.304)	−0.45% (0.256)	−0.26% (0.449)	+0.58% (0.519)
Realized cost (primary)	−0.67% (0.152)	+0.57% (0.755)	−0.27% (0.405)	+0.17% (0.522)	−0.55% (0.245)	−0.06% (0.512)
Realized cost, first third	+0.03% (0.530)	+0.82% (0.783)	+0.35% (0.658)	+0.12% (0.585)	−0.88% (0.164)	+0.30% (0.746)
Realized cost, last third	−1.16% (0.042)	+0.10% (0.630)	−0.75% (0.230)	−0.54% (0.423)	−0.29% (0.398)	−0.29% (0.411)
Realized cost, final round	−0.65% (0.309)	−0.93% (0.089)	−1.08% (0.155)	−0.72% (0.260)	+0.13% (0.536)	−0.74% (0.238)
High-stakes share of acquired labels	−0.25% (0.841)	−0.09% (0.892)	−0.12% (0.702)	+0.86% (0.114)	+1.41% (0.020)	+0.75% (0.278)
Most expensive error, rate	−1.40% (0.178)	+2.15% (0.853)	−0.28% (0.561)	−0.57% (0.353)	−2.49% (0.147)	−0.68% (0.501)
Any costly error, rate	−0.32% (0.273)	−0.52% (0.177)	−0.17% (0.466)	+0.21% (0.717)	+0.91% (0.863)	+0.22% (0.741)

negatives with fewer false positives, while ADE Soft shows the reverse (Appendix D.21). None of these diagnostics establishes clinical safety or efficacy.

D.19 Decision opportunities and limits of extrapolation

Theorem 3 identifies when Hard can add decision value: a hypothetical observation must reveal a better action on a set of reference inputs with positive deployment mass. In the binary case, updates with mass strictly on both sides of the threshold give a positive utility gap; updates confined to one action region give zero. This is a statement about the acquisition component, not a sufficient condition for better realized learning curves. Relative candidate rankings also depend on the base term, λ, the posterior, and the reference population.

Changing a cost ratio changes both the normalization and the action boundary. It need not monotonically change normalized EVSI or performance, because hypothetical updates and the available decision opportunities can change too. In the validation sweep, mean relative differences over 1:1–2:1 favor DUSAL on all six dataset–variant pairs: -1.81%, -3.12%, -1.24%, -1.30%, -0.04%, and -0.30%. Over 19:1–99:1, the corresponding values are +4.27%, +6.50%, +0.17%, +2.23%, +1.91%, and +1.49%. These exploratory summaries use 20 seeds; they do not establish a favorable clinical cost regime. Fresh confirmation at the selected costs does not resolve the proposed improvement.

Complete-row utility can distinguish decision problems that a column-weighted quadratic summary cannot, as P3 demonstrates. Fewer actions than outcomes are neither required by DUSAL nor proved necessary for an advantage. The grouped-action public tasks exercise this setting but do not establish a benefit. Likewise, E5's 9.09% Brier improvement over utility-only acquisition is a controlled prediction–decision trade-off, not evidence of a clinical advantage.

Future tests can target reference predictions near consequential boundaries and compare candidate rankings under directly validated hypothetical updates. Utilities should be justified externally, and endpoints fixed before evaluation. These are testable design considerations, not conditions under which the present experiments prove that DUSAL will outperform its base rule.

D.20 Secondary-endpoint scan of the multi-action confirmation

Every scalar the runs recorded per round was evaluated as a paired budget-normalized AULC contrast against BEMPS. The scan is exploratory: the confirmatory family is closed, and the no-rescue rule forbids using these endpoints to revive it. All 64 Hard and Soft contrasts were corrected together by Holm and 0 survive. The wrong-table arm is shown beside them as the negative control.

Table 30: Post-hoc error decomposition of the original 720 confirmation runs. Entries are DUSAL-minus-BEMPS differences in budget-averaged errors per 1,000 fixed test examples, with marginal 95% paired-seed bootstrap intervals (80 seeds, 40,000 resamples). Negative means fewer errors; all intervals include zero. These are integrated rates, not cumulative event counts, distinct patients, or projected hospital outcomes; ADE units are text items. Both arms use CE training.

Dataset / FN:FP	Method	Δ FN / 1,000 [95% CI]	Δ FP / 1,000 [95% CI]
BreastMNIST 1.5:1	Hard	-0.34 $[-3.06, +2.46]$	$+0.84$ $[-5.62, +7.18]$
	Soft	$+1.35$ $[-1.83, +4.55]$	-2.55 $[-9.94, +4.55]$
PneumoniaMNIST 9:1	Hard	$+0.02$ $[-0.14, +0.18]$	-0.27 $[-1.59, +1.01]$
	Soft	-0.01 $[-0.17, +0.15]$	$+0.20$ $[-1.10, +1.46]$
ADE 1.5:1	Hard	-0.74 $[-2.04, +0.55]$	$+0.20$ $[-1.12, +1.55]$
	Soft	-1.03 $[-2.35, +0.28]$	$+0.10$ $[-1.07, +1.28]$

Table 31: Outcome-identified PneumoniaMNIST R-FULL results. Each row contains 20 paired seeds and reports normalized realized-cost AULC (lower is better). Δ is DUSAL minus BEMPS; relative change is Δ/BEMPS, matching Table 11; negative favors DUSAL. Intervals are two-sided marginal (multiplicity-unadjusted) paired-bootstrap 95% CIs; raw p values are one-sided exact paired sign flips. Hard's Holm values belong to its locked two-ratio primary family; Soft is a separate secondary family. Blue bold marks favorable point-estimate direction only. Soft at 4:1 is the sole marginal interval excluding zero, but neither Holm-corrected family rejects a contrast.

Arm	FN:FP	BEMPS	DUSAL	Δ	Rel. change	95% CI	Raw p	Holm p
Hard	4:1	0.050384	0.049693	**−0.000691**	−1.37%	[−0.001415, 0.000106]	0.0488	0.0975
Hard	9:1	0.025249	0.024896	**−0.000353**	−1.40%	[−0.000789, 0.000092]	0.0722	0.0975
Soft	4:1	0.050384	0.049515	**−0.000868**	−1.72%	[−0.001747, −0.000040]	0.0344	0.0688
Soft	9:1	0.025249	0.024935	**−0.000314**	−1.24%	[−0.000636, 0.000016]	0.0430	0.0688

D.21 Error decomposition: lower declared cost is not a sensitivity guarantee

The finite cost objective trades errors; it does not impose a hard sensitivity constraint. We reconstruct FN and FP components from all saved confirmation rounds and independently replay the final operating point from saved probabilities. For fixed positive-class prevalence π and ratio r:1, normalized realized cost is $[r\pi(1-\text{sensitivity}) + (1-\pi)(1-\text{specificity})]/r$. Integrating each term over the original label budgets exactly reconstructs the primary AULC contrasts. Table 30 gives all six comparisons, without changing the frozen efficacy tests.

At PneumoniaMNIST's selected 9:1 cost, BEMPS already has 99.65% budget-average sensitivity, and FP accounts for 91.11% of its normalized cost. High FN penalty therefore need not mean that FN dominates the residual observed cost. BreastMNIST Soft's small favorable cost point estimate combines more FN with fewer FP; ADE Soft shows the reverse trade-off. Neither component change is resolved by its interval. Reporting both errors makes the assumed cost objective interpretable without translating a cost point estimate into fewer missed patients or clinical safety.

D.22 Earlier PneumoniaMNIST replication (R-FULL)

R-FULL is the frozen 20-seed PneumoniaMNIST replication that preceded the cost search. It compares the same three arms at two research-specified ratios, 4:1 and 9:1, starting from 100 labels and acquiring 1,000 labels in batches of 50 with a fixed 1,024-point uniformly weighted reference cohort; the 120 artifacts record acquired indices, observed labels, class and action counts, acquisition ranks, source commit, source-content fingerprint, and the declared cost table. It was frozen as an independent replication after an exploratory cost sweep and is not represented as the original preregistration. Hard has a locked two-ratio family tested by one-sided exact paired sign flips with Holm control at 0.05; Soft is a separately labeled secondary family; 20,000 paired-seed bootstrap resamples give the displayed two-sided marginal percentile intervals.

All four R-FULL point estimates favor DUSAL over matched BEMPS, with descriptive relative changes of −1.37% and −1.40% for Hard and −1.72% and −1.24% for Soft at 4:1 and 9:1, respectively, and favorable paired-seed fractions from 13/20 to 14/20. The Hard raw exact sign-flip values

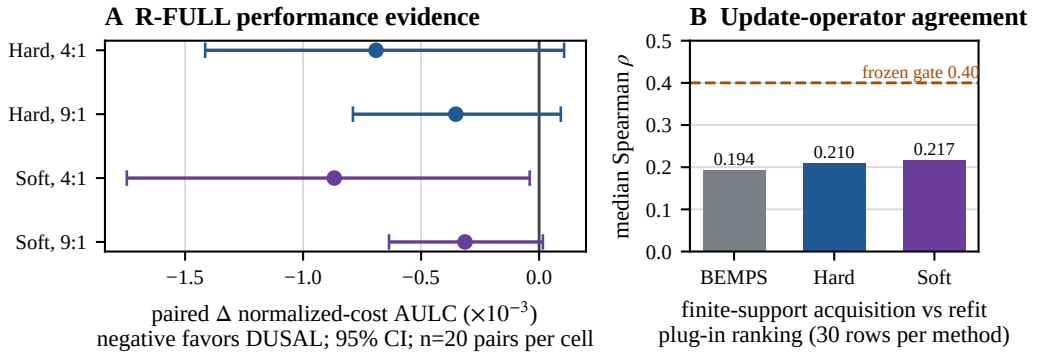

Figure 8: R-FULL and update-sensitivity evidence. **Left:** the four paired realized-cost AULC effects from Table 31; negative values favor DUSAL and bars are marginal, multiplicity-unadjusted paired-bootstrap 95% intervals. Family-wise decisions use the Holm values in Table 31. **Right:** median Spearman agreement between the exact finite-support acquisition ranking and the converged-refit Bregman plug-in displacement ranking for BEMPS, Hard, and Soft, aggregated over three audited tasks and ten seeds per task. The dashed line is the frozen 0.40 overall gate. The audit is provenance-valid but classified POSTERIOR-SENSITIVE; a passed integrity audit is not a passed agreement gate, and this panel does not estimate refit expected score gain.

are 0.0488 and 0.0722, becoming 0.0975 at both ratios after its locked Holm-2 correction; the Soft values are 0.0344 and 0.0430, becoming 0.0688 in its separate Holm-2 family. Neither family rejects a contrast (0/2 each). The consistent direction at 20 seeds is the same pattern the cost search showed on validation, and the same $9{:}1$ cost table entered the confirmation family through the discovery rule with 80 fresh seeds, where the relative changes are -0.04% and $+0.03\%$ (Table 11). The favorable R-FULL direction was not sustained at the larger sample. The 20-seed contrasts remain unresolved. Figure 8 shows these four effects beside the update-operator agreement of Appendix D.15, the two diagnostics that together motivated the larger confirmation.

D.23 RETROSPECTIVE MATCHED-LOSS FACTORIAL UNDER TWO ESTIMANDS

Table 32 summarizes ten current Bayes-value benchmarks, including ADE, using held-out normalized realized-cost AULC. The retrospective study crossed acquisition $\{\text{BEMPS}, \text{Hard}\}$ with either CE training or the Bayes-value Hard training score (BH). Its four cells are A: BEMPS+CE, B: Hard+CE, C: BEMPS+BH, and D: Hard+BH. Thus B–A holds CE fixed, while D–C holds BH fixed; A versus C and B versus D instead change the training loss while holding acquisition fixed. As noted in Section D.26, BH has the same gradient as quadratic/Brier training almost everywhere and is not evidence for a smooth utility-dependent Hard training gradient. The displayed S: Soft+CE arm comes from the separately frozen 100-run Soft acquisition extension at $\tau = 0.1$; it is not a fifth factorial cell and reuses A as its matched control. A–D contain 400 runs and S contains 100, with the same ten seeds per displayed cell. Because the outcome-identified endpoint was reconstructed after these artifacts were observed, both tables are descriptive, not confirmatory.

The lowest descriptive mean varies across datasets and settings; no one arm dominates this retrospective overview. Tables 32 and 33 summarize the same trajectories, not independent replications. The leading two descriptive ranks change between estimands on DermaMNIST, ISIC2020, OCTMNIST, PneumoniaMNIST, and RetinaMNIST. The outcome-based realized-cost endpoint and the model-relative reference-action diagnostic therefore answer different questions.

D.24 RETROSPECTIVE R-CORE ANALYSIS

R-CORE's frozen endpoint used the fully supervised reference-model proxy. The 240 raw artifacts are complete; the recorded "blocked" state came from an analysis import failure, not a failed scientific audit. We reconstructed the outcome-identified endpoint from every stored round. Because that

Table 32: Cross-dataset current Bayes-value results: retrospective held-out normalized realized-cost AULC by dataset and setting (lower is better). Entries are mean ± sample standard deviation over ten matched seeds. Bold and underlining mark the lowest and second-lowest unrounded means within each dataset; these are descriptive ranks, not significance claims. The endpoint was adopted after the artifacts were observed, so the table does not establish confirmatory superiority.

Dataset	A: BEMPS+CE	B: Hard+CE	C: BEMPS+BH	D: Hard+BH	S: Soft+CE
ADE	0.081421 ± 0.004177	0.080639 ± 0.003694	**0.075141 ± 0.003956**	0.076277 ± 0.005371	0.080476 ± 0.003309
BloodMNIST	0.039071 ± 0.001170	0.038870 ± 0.000786	0.038391 ± 0.001114	**0.038283 ± 0.001428**	0.038966 ± 0.000943
BreastMNIST	0.110188 ± 0.005824	0.109226 ± 0.004056	0.110949 ± 0.005018	0.111474 ± 0.004941	**0.107440 ± 0.005025**
DermaMNIST	0.074503 ± 0.002473	0.075388 ± 0.002738	**0.073755 ± 0.003502**	0.074092 ± 0.003069	0.075246 ± 0.001999
ISIC2020	**0.021928 ± 0.000621**	0.021930 ± 0.000542	0.022951 ± 0.000673	0.022882 ± 0.000534	0.022032 ± 0.000514
MedAbstracts	0.083637 ± 0.003222	0.083903 ± 0.002594	0.082269 ± 0.002521	**0.081470 ± 0.001819**	0.085106 ± 0.004116
OCTMNIST	0.127636 ± 0.006221	**0.126392 ± 0.006921**	0.127360 ± 0.006969	0.127117 ± 0.008246	0.126593 ± 0.007658
PathMNIST	0.042662 ± 0.001457	0.042672 ± 0.001883	**0.041304 ± 0.002216**	0.041674 ± 0.002619	0.042427 ± 0.001495
PneumoniaMNIST	0.041175 ± 0.000776	0.041720 ± 0.001517	**0.040554 ± 0.001473**	0.041070 ± 0.001673	0.040808 ± 0.001563
RetinaMNIST	0.154980 ± 0.003600	0.156919 ± 0.004694	0.156743 ± 0.003490	0.156460 ± 0.004639	**0.153736 ± 0.004315**

Table 33: Historical fully-supervised-reference action-regret AULC for the same trajectories and settings (lower is better). Each value measures the regret of the model-selected action relative to the action selected by a separately trained fully supervised CE predictive, normalized by the decision scale. That predictive is neither the true $p^*(y \mid x)$ nor an observed clinical outcome, so this is a model-relative diagnostic rather than population or clinical regret. Entries follow Table 32.

Dataset	A: BEMPS+CE	B: Hard+CE	C: BEMPS+BH	D: Hard+BH	S: Soft+CE
ADE	0.025590 ± 0.003049	0.025150 ± 0.002867	**0.020337 ± 0.001784**	0.022295 ± 0.003943	0.024681 ± 0.002638
BloodMNIST	0.030006 ± 0.001599	0.030008 ± 0.001408	0.029610 ± 0.001377	**0.029383 ± 0.001576**	0.030361 ± 0.001484
BreastMNIST	0.152769 ± 0.091319	0.149493 ± 0.093213	0.160734 ± 0.115269	0.160177 ± 0.108589	**0.148982 ± 0.095641**
DermaMNIST	**0.046186 ± 0.001291**	0.048015 ± 0.001834	0.047126 ± 0.002720	0.047049 ± 0.002232	0.047647 ± 0.002065
ISIC2020	0.012680 ± 0.001177	**0.012491 ± 0.001023**	0.014269 ± 0.001087	0.013863 ± 0.001009	0.012666 ± 0.001090
MedAbstracts	0.016714 ± 0.002696	0.017231 ± 0.002248	0.014996 ± 0.001690	**0.014978 ± 0.001287**	0.016961 ± 0.002107
OCTMNIST	0.109581 ± 0.005486	0.108214 ± 0.004005	0.107145 ± 0.003836	**0.103641 ± 0.004714**	0.110059 ± 0.007053
PathMNIST	0.037258 ± 0.001462	0.037719 ± 0.001991	**0.036033 ± 0.002626**	0.036526 ± 0.003290	0.037216 ± 0.001780
PneumoniaMNIST	0.017522 ± 0.001482	0.017962 ± 0.001426	0.017465 ± 0.002512	**0.016769 ± 0.002719**	0.018486 ± 0.002150
RetinaMNIST	0.057042 ± 0.005304	0.056113 ± 0.006554	**0.054610 ± 0.003722**	0.055243 ± 0.005875	0.056384 ± 0.005589

Table 34: R-CORE retrospective normalized realized-cost AULC, mean ± standard deviation over 20 seeds. All arms use CE. Lower is better. These absolute summaries include every frozen arm and do not confer confirmatory status on the corrected endpoint.

Task	Random	Entropy	BALD	BEMPS	Hard	Soft
Breast U2	.15345 ± .03850	.12970 ± .02109	.13182 ± .02810	.13647 ± .01910	.14022 ± .01950	.13623 ± .01680
Retina U3	.08616 ± .00724	.08975 ± .01120	.08963 ± .00732	.09067 ± .00986	.09038 ± .00802	.09083 ± .00759

correction was made after observing the artifacts, Table 34 is descriptive, and the paired tests are two-sided rather than a new superiority family.

Against BEMPS, Hard's paired mean differences are $+0.003752$ on Breast (95% CI $[-0.007326, 0.014258]$) and -0.000293 on Retina ($[-0.002017, 0.001464]$). Soft's are -0.000241 ($[-0.010172, 0.010218]$) and $+0.000155$ ($[-0.002026, 0.002431]$), respectively. All retrospective Holm-adjusted two-sided sign-flip values equal 1.0. The Breast Hard contrast changes direction relative to the old reference-predictive proxy, illustrating why the proxy cannot be treated as the public-data estimand. On Retina U3, random sampling has the lowest descriptive mean among all six arms. On Breast U2, entropy sampling has the lowest mean and Hard the second-highest. These differences are unresolved at this seed count: the Breast paired-difference standard deviation implies a relative MDE near 10% at 20 seeds. The table therefore establishes neither a reliable ranking of the six methods nor an advantage of BEMPS over these simpler controls.

D.25 Historical raw-calibration gains: a distinct training-pipeline finding

Historical pipeline. The earliest diagonal-quadratic pipeline produced lower raw calibration error on several datasets. We report this observation separately from the current acquisition tests and call the archived pipeline *Legacy-Q*. It uses the superseded diagonal-quadratic acquisition and a weighted quadratic proper-score training loss ($\lambda_{\text{acq}} = 1$, $\lambda_{\text{loss}} = 2$). Its comparators are random acquisition with CE, quadratic BEMPS with CE, and BEMPS with class-weighted CE. The comparison therefore changes training as well as acquisition; the observed calibration gap cannot be assigned to acquisition

Table 35: Historical raw ECE AULC, not current Bayes-value acquisition. Entries are mean ± sample standard deviation over the common seeds shown in n; lower is better. Random/BEMPS use CE, WCE is BEMPS with class-weighted CE, and Legacy-Q uses the archived weighted quadratic training loss. Bold denotes the lowest unrounded mean among these four pipelines within a row, not significance or superiority to all training objectives. Relative reduction is $100(\overline{\text{BEMPS}} - \overline{\text{Legacy-Q}})/\overline{\text{BEMPS}}$; negative means worse. Executable-source identity is unverified for this historical corpus.

Dataset	n	Random+CE	BEMPS+CE	WCE	Legacy-Q	Reduction (%)
PathMNIST	10	0.0519 ± 0.0042	0.0613 ± 0.0103	0.0493 ± 0.0051	$\mathbf{0.0426 \pm 0.0054}$	30.5
OCTMNIST	5	0.1935 ± 0.0127	0.1964 ± 0.0082	0.2036 ± 0.0107	$\mathbf{0.1764 \pm 0.0079}$	10.2
DermaMNIST	10	0.1401 ± 0.0059	0.1298 ± 0.0073	0.1386 ± 0.0075	$\mathbf{0.1103 \pm 0.0058}$	15.0
BloodMNIST	10	0.0372 ± 0.0037	0.0253 ± 0.0023	0.0261 ± 0.0024	$\mathbf{0.0214 \pm 0.0029}$	15.5
RetinaMNIST	5	0.3140 ± 0.0080	0.3244 ± 0.0241	0.3104 ± 0.0267	$\mathbf{0.3008 \pm 0.0185}$	7.3
BreastMNIST	10	0.1149 ± 0.0190	0.1229 ± 0.0129	0.1253 ± 0.0195	$\mathbf{0.1142 \pm 0.0121}$	7.1
ISIC2020	5	$\mathbf{0.0318 \pm 0.0018}$	0.0368 ± 0.0082	0.0499 ± 0.0060	0.0341 ± 0.0055	7.6
PneumoniaMNIST	10	0.1200 ± 0.0132	0.1124 ± 0.0054	0.1232 ± 0.0051	$\mathbf{0.1085 \pm 0.0039}$	3.5
ADE	10	0.0926 ± 0.0082	0.0832 ± 0.0072	0.0993 ± 0.0071	$\mathbf{0.0554 \pm 0.0052}$	33.4
MedAbstracts	10	0.2027 ± 0.0112	0.2173 ± 0.0125	0.1994 ± 0.0081	$\mathbf{0.1784 \pm 0.0085}$	17.9

Table 36: Historical final-round raw ECE on exactly the same runs and common seeds as Table 35. Mean ± sample standard deviation; lower is better. Bold is a within-row descriptive rank among four pipelines, not an inferential claim. This table reports the terminal prediction, whereas the preceding table summarizes calibration throughout the label budget.

Dataset	Random+CE	BEMPS+CE	WCE	Legacy-Q	Reduction (%)
PathMNIST	0.0454 ± 0.0044	0.0458 ± 0.0066	0.0437 ± 0.0063	$\mathbf{0.0336 \pm 0.0069}$	26.7
OCTMNIST	0.2070 ± 0.0153	0.1922 ± 0.0130	0.2059 ± 0.0112	$\mathbf{0.1676 \pm 0.0082}$	12.8
DermaMNIST	0.1396 ± 0.0085	0.1218 ± 0.0040	0.1214 ± 0.0066	$\mathbf{0.1003 \pm 0.0043}$	17.7
BloodMNIST	0.0304 ± 0.0045	$\mathbf{0.0150 \pm 0.0016}$	0.0182 ± 0.0026	0.0163 ± 0.0019	−8.8
RetinaMNIST	0.3107 ± 0.0196	0.3381 ± 0.0190	$\mathbf{0.3102 \pm 0.0283}$	0.3102 ± 0.0100	8.2
BreastMNIST	$\mathbf{0.1022 \pm 0.0174}$	0.1024 ± 0.0134	0.1031 ± 0.0145	0.1115 ± 0.0173	−8.8
ISIC2020	0.0182 ± 0.0021	0.0172 ± 0.0005	0.0156 ± 0.0015	$\mathbf{0.0093 \pm 0.0022}$	45.7
PneumoniaMNIST	0.1199 ± 0.0154	0.1106 ± 0.0062	0.1144 ± 0.0064	$\mathbf{0.1046 \pm 0.0055}$	5.4
ADE	0.0777 ± 0.0089	0.0666 ± 0.0039	0.0677 ± 0.0056	$\mathbf{0.0412 \pm 0.0061}$	38.2
MedAbstracts	0.2157 ± 0.0132	0.2419 ± 0.0120	0.2279 ± 0.0074	$\mathbf{0.2047 \pm 0.0119}$	15.4

alone. Both unweighted CE and Brier scoring are strictly proper, so this is not a proper-versus-improper explanation of the CE comparison.

Endpoint and retrospective reconstruction. We use all ten datasets from the original core suite and the intersection of available seeds across the four specified arms within each dataset (340 runs). Extra runs in only one arm do not enlarge its effective sample. Raw top-label ECE uses the ensemble-mean prediction: with 15 equal-width confidence bins, it is the sum of bin mass times the absolute difference between bin accuracy and mean confidence. No post-hoc calibration is applied. Each trajectory's ECE AULC is its trapezoidal integral over labelled-set size divided by the budget span; the final-round ECE is a separate endpoint. We reconstruct the former from all saved round metrics and replay the latter from saved predictions.

Calibration improves during learning, but not on every task or endpoint. Legacy-Q has lower mean ECE AULC than BEMPS+CE on 10 of 10 datasets (Table 35). The reductions on PathMNIST, DermaMNIST, and ADE are respectively 30.5%, 15.0%, and 33.4% relative to the paired baseline mean. These quantify observed differences in raw confidence–accuracy agreement, not percentage-point changes in predictive accuracy. The final endpoint favors Legacy-Q on 8 datasets; Table 36 retains every dataset and all four arms, including reversals. Figure 9 directly compares the two endpoints without choosing the more favorable one for each dataset.

Recorded training and acquisition controls. The archived component controls hold either the acquisition or the training objective fixed in their recorded configurations (Table 37). On PathMNIST, BEMPS+CE and Legacy-Q acquisition+CE yield ECE AULCs 0.0613 and 0.0611, whereas switching to the legacy loss at fixed BEMPS acquisition yields 0.0419. This supports reading the historical result as a training-pipeline calibration finding rather than a large acquisition effect. The weighted-quadratic

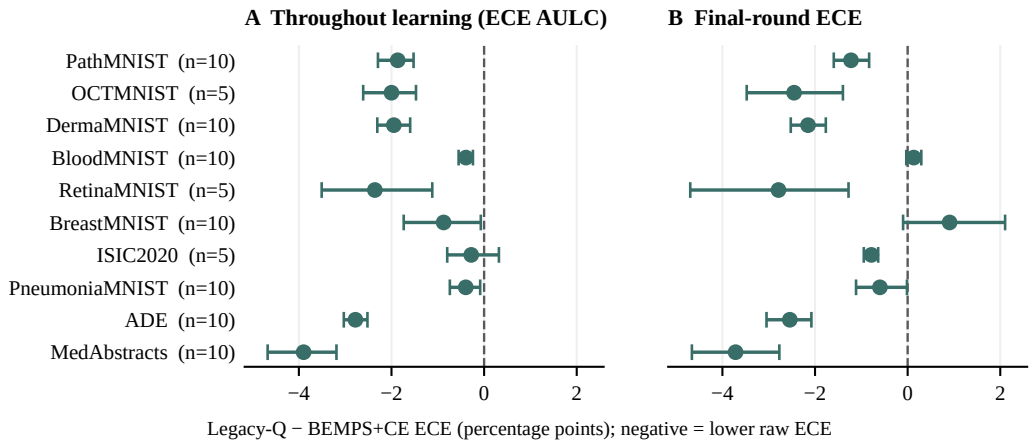

Figure 9: Historical raw-calibration differences across all ten core datasets. Points are Legacy-Q minus BEMPS+CE ECE in percentage points (100Δ, not relative reduction); negative favors Legacy-Q. The panels separate budget-normalized AULC from final-round ECE. Error bars are post-hoc marginal 95% paired-seed bootstrap intervals (40,000 resamples), conditional on the recorded test cohorts. They do not adjust for multiplicity or quantify historical environment/source uncertainty, and are not current Hard/Soft efficacy tests. Dataset order is fixed across panels, not ranked by effect size.

Table 37: Historical ECE AULC controls on common seeds within each dataset. Entries are paired mean differences; negative means lower ECE. "Acq@CE" switches BEMPS to legacy acquisition with CE fixed; "Acq@QL" makes the same switch with the legacy quadratic loss (QL) fixed. "QL–CE" compares QL with CE at fixed BEMPS acquisition. "QL–Brier" changes training curvature from $\lambda_{\text{loss}} = 0$ to 2 under legacy acquisition. The last column is not a Hard/Soft comparison. Source identity remains unverified; no cell is promoted to a new significant or causal effect.

Dataset	Acq@CE	Acq@QL	QL–CE	QL–Brier
PathMNIST	−0.0002	+0.0007	−0.0193	+0.0008
OCTMNIST	+0.0005	+0.0013	−0.0213	+0.0048
DermaMNIST	−0.0010	−0.0034	−0.0161	−0.0054
BloodMNIST	+0.0002	+0.0002	−0.0041	+0.0007
RetinaMNIST	−0.0079	+0.0108	−0.0344	−0.0090
BreastMNIST	+0.0000	+0.0000	−0.0087	−0.0010
ISIC2020	+0.0000	+0.0000	−0.0028	+0.0027
PneumoniaMNIST	+0.0000	+0.0000	−0.0040	+0.0013
ADE	+0.0000	+0.0000	−0.0278	+0.0002
MedAbstracts	−0.0032	−0.0061	−0.0329	−0.0037

versus unweighted-Brier contrast has mixed directions across tasks: the result is not evidence that utility curvature uniquely improves calibration. These are retrospective contrasts of recorded runs, not source-verified causal components.

Interpretation. When a workflow consumes probabilities without a separate recalibration step, raw confidence–accuracy agreement is a distinct empirical objective from acquisition efficiency or realized decision cost. The archived results show that training-objective choice can accompany appreciable improvements in that quantity; they motivate testing training and acquisition as separate components. Top-label ECE alone does not establish class-specific calibration, lower false-negative rates, or clinical benefit. Nor do these tables establish superiority to all training losses or to a properly held-out recalibrated baseline. Earlier evaluation-set-fitted foils do not provide that comparison.

Provenance boundary. All historical numbers here are retrospective. Stored configuration fields identify the intended pipeline, but the original records lack a verifiable source-commit/content-fingerprint pair; matching round-0 metrics cannot repair that omission. Recorded

worker/configuration differences also preclude treating every historical contrast as a fully controlled intervention. The new script `scripts/gen_numbers_legacy_calibration.py` records present-day result/prediction hashes and recomputes the metrics; this verifies numerical reconstruction, not execution provenance. We retain the raw-ECE results as historical observations without carrying over their earlier significance labels or treating them as evidence for the current method.

D.26 Training-loss and mechanism-diagnostic corrections

The literal Hard Bayes-value Savage loss contains an action-dependent affine term, but its gradient with respect to predicted probabilities equals the quadratic/Brier gradient almost everywhere. Consequently the historical Hard-factorial C/D arms test Brier-like versus CE training; they do not identify a utility-tailored Hard training gradient or a "complete utility-tailored pipeline." Soft's utility-dependent curvature permits a differentiable, matched training-loss experiment in future work.

Hard and Soft diagnostics must also remain semantically disjoint. For Hard, the non-base acquisition contribution is literal EVSI and licenses $\mathrm{Reg}_U(q;p) \le (L_U^{\mathrm{dec}}/\lambda) D_{H,U,\lambda}(p,q)$. For finite-temperature Soft, the contribution is $\lambda\tau$ KL between soft action distributions; it is neither clinical regret nor EVSI, and the Hard sharp bound does not apply. Hypothetical outcomes are enumerated exactly. Reference uncertainty is therefore computed across the R reference units after summing over outcomes. Uniform cohorts use the ordinary reference-unit standard error (SE); future boundary-mixture cohorts use the mixture ratio estimator's influence function, while Kish effective sample size is descriptive only. The old $K \times R$ pseudo-sample SE was anticonservative. Historical Soft `evsi_*` labels and Hard/Soft `frac_pos`/SE values are not pooled with corrected diagnostics; the round-0 EVSI share and activation rate of Table 19 are computed from stored acquisition traces and do not require the unstored per-round member predictives.

For the same fixed-potential reason, an automatically calibrated Soft temperature is resolved from the initial posterior/reference cohort once and then frozen for the run. Re-solving τ each round would define a changing potential G_n and fall outside the acquisition and convergence statements. All historical τ sweeps remain exploratory sensitivity analyses; no temperature is selected after inspection. The cost-search and confirmation manifests fixed $\tau = 0.05$ before execution.

D.27 Artifact verification and study scope

We include an artifact as evidence for the current Bayes-value method only when its clean source revision contains the executed full-row implementation and its stored method, score, cost orientation, label schedule, posterior policy, fixed reference cohort, and source-content fingerprint agree with its manifest. Table 38 separates this verification from each study's inferential role. The historical calibration analysis reconstructs recorded metrics but cannot verify the original executable source. We report it as evidence about the earlier pipeline, separate from current acquisition results.

The five retrospective studies of the current Bayes-value method contain 2,000 trajectories; with R-CORE they total 2,240. Every round stores a finite expected utility (EU), a positive decision scale, and a label budget, so normalized realized-cost AULC is reconstructed as $-\mathrm{EU}/L_U^{\mathrm{dec}}$ without refitting. This endpoint correction was made after those artifacts were observed. Their exhaustive cell summaries and input-file hashes are released as the machine-readable `dusal_current_bv_realized_cost_reanalysis_v1` supplement; they remain exploratory and separate from the frozen confirmation family. The frozen files contain Python's nonstandard not-a-number (`NaN`) token for unavailable ancillary metrics; the audit records all such tokens and separately requires every quantity consumed by the corrected analysis to be finite.

D.28 Provenance and rerun boundary

Table 39 records the code revision and source-content fingerprint used for each study's runs.

The R-FULL audit matched all 600 per-run scientific files (120 each of `result.json`, resolved configuration, prediction sidecar, acquisition trace, and target-measure sidecar) to the frozen backup checksum index. The first analysis attempt reported 240 semantic-hash errors because its validator renormalized already normalized 624-element evaluation weights, changing each stored value by one floating-point unit before recomputing the byte-level hash. The recovery validator preserves a

Table 38: Study inventory and verification status. "PASS" means the declared files and numerical identities passed all required checks; inferential status is reported separately. The 2,000 historical current Bayes-value trajectories can be re-expressed on held-out realized cost without graphics processing unit (GPU) reruns. The 3,225 legacy artifacts use superseded score maps, mixed protocols, or inadequate executable provenance and are excluded from current-method claims.

Corpus	Units	Artifacts	Audit/role	Licensed conclusion
Exact mechanisms E0–E4	$1{,}000 + 4 \times 10{,}000 + 2 \times 100$	1	PASS, exact	Algebra, invariances, and constructed rankings
Finite-world/selector E5–E6/P3	1,800 rows	1	PASS, constructed	Prediction–decision trade-off, full-row policies, artificial batch coverage
Cost discovery (validation only)	20 paired seeds, 72 cells	2,160	PASS; exploratory selection, test sealed	Per-dataset cost selection; no test-set claim
Fresh-seed confirmation	80 paired seeds, 6 contrasts	720	PASS; primary Holm family	0/6 supported; bounded null; no-rescue rule in force
λ mechanism studies	20+80 paired seeds	80+320	registered exploratory; separate environment	No reproducible improvement in two registered Hard-weight studies; never pooled with the family
R-FULL Pneumonia replication	20 paired seeds	120	PASS integrity; Hard 0/2, Soft 0/2	Favorable point estimates at 20 seeds; not sustained at 80
Binary replication with controls	80 paired seeds, 3 tasks	1,200	PASS; separate Holm family	0/6 supported; confirmed positive-control benefit on two of three tasks
Multi-action confirmation	80, 60 paired seeds	840	PASS; separate Holm family	0/6 supported; similar outcomes under wrong-table and true-table acquisition
Laplace-posterior confirmation	80, 80, 60 paired seeds	1,320	PASS; separate Holm family	0/6 supported; one task exploratory after its gate
Batch rule	80, 60, 80, 80 paired seeds	600	PASS; two Holm families of 4	3/4 supported in each; paired with stored top-b runs after reproduction checks
Bregman refit sensitivity	270 rows	30 blocks	PASS; POSTERIOR-SENSITIVE	Plug-in update agreement, not refit acquisition or efficacy
Direct-Q refit sensitivity	360 rows	30 blocks	PASS; REFIT-MC-SENSITIVE	Direct potential gain and A/B refit replication; limited to frozen last-layer models
Posterior-backend sensitivity	10 seeds, 3 tasks	270	PASS; separate Holm family	0/6 supported; independent execution with disclosed stage-order deviation
R-CORE S1	20 paired seeds	240	PASS; retrospective	Descriptive realized-cost summaries only
Hard-factorial	10 paired seeds, 10 tasks	400	PASS; retrospective	Acquisition sensitivity; Hard-score training is Brier-gradient-equivalent
Initial Soft and τ extension	10 paired seeds, 10 tasks	100+300	PASS; exploratory	Temperature sensitivity; no confirmatory replacement
Breast and three-task cost sweeps	20 paired seeds	300+900	PASS; exploratory	Cost sensitivity; no post hoc primary
Superseded diagonal/mixed corpus	–	3,225	excluded	No current Hard/Soft claim
Historical calibration subset	10 tasks, paired seeds	340 core runs	retrospective; source-unverified	Raw ECE of the earlier pipeline only; additional loss controls in Appendix D.25

Table 39: Key immutable provenance identifiers. Hashes identify executable content or manifests; mutable coordinator logs and queue state are not scientific artifacts.

Corpus	Source commit	Content hash	Manifest/raw/audit SHA-256 prefix
Exact E0–E4	`0edc8ea`	`2982f8fdf0db698e`	`5e010bcb82065a73`
R-CORE exact/S1	`ddd074b`	`9f79af20404bfbd1`	stored with each audit
R-FULL scientific runs	`bbbc9eb`	`4966d7d7399501c2`	`62719c0c7fd61972/f217ebf101f88015`
R-FULL analysis recovery	`81507b3`	`9e5d8f5e44adabe3`	`526c572c265a9386/0899f6f374238d60`
Update sensitivity	`bbbc9eb`	`4966d7d7399501c2`	`ebcb9b9c4a116652`
Cost discovery	`72d0a96`	`9ecb25c0e508a12b`	selection `44c7bebf645dd1da`
Fresh-seed confirmation	`72d0a96`	`9ecb25c0e508a12b`	manifest `7404a84271572668`/index `b2613801a28f4ac5`/audit `661fcff701602e27`
λ studies	`72d0a96` (remote checkout)	–	registered protocol; separate build, never pooled
Binary replication, five arms	`64ee8bc`	`cbae67f775ea1732`	report `analysis/binary_pc_v1/final_report.json`
Faithful-update confirmation	`e415d20`	`01a867b1e3feda06`	report `analysis/faithful_v1/final_report.json`
Multi-action confirmation	`64ee8bc`	`cbae67f775ea1732`	reported with the study
Batch rule	`64ee8bc`	`cbae67f775ea1732`	same fingerprint as the paired top-b runs

persisted, hash-addressed vector after checking its mass and semantic digest. A length-624 regression and the full test suite pass; a fresh independent replay returns 120 artifacts and zero errors, and an independent implementation reproduces every paired statistic. The raw artifacts and frozen ledger were not modified. Mutable coordinator state, including its log and `experiment_todo.json`, is excluded from the scientific checksum claim. The cost-search and confirmation stages index 10800 and 3600 per-run scientific files respectively; every number in Tables 11, 10, 12, 19, 20, 21, and 22 is regenerated from those artifacts by a deterministic script with fixed bootstrap seeds and byte-matches the released binding file.

The additional E5/P3, posterior-family, initial-batch, error-decomposition, and direct-Q bindings are regenerated by `scripts/gen_numbers_findings.py`. It checks the exact trace digest, all 1350 posterior-family files, 1440 confirmation result/prediction files, and 66 direct-Q indexed files. It recomputes E5 AULCs, the six frozen MC primary tests and four interactions, all initial selections, the error-component contrasts, and pooled direct-Q medians. The generated block records source digests and separates post-hoc from frozen analyses; the original artifacts are unchanged.

No GPU rerun is needed to correct the public-data endpoint for the 2,240 complete current Bayes-value trajectories: their held-out expected utility and decision scale identify realized cost. New runs are necessary only for new scientific claims: stable full-network refit expected-score-gain acquisition and its downstream efficacy, a utility-gradient training factorial, a posterior class whose one-step update is verifiably faithful, or external tasks with clinically elicited utility provenance. The completed studies include the Laplace posterior evaluation (Appendix D.10) and the DUSAL-Div comparison

(Appendix D.12). Each new study needs a clean commit, an immutable manifest, fixed worker and member-seed policies, acquired-index sidecars, and prospective realized-cost tests.

E Reproducibility and Reporting Checklist

Reproducing an empirical study requires the base potential H, prediction domain, and selected subgradient; hard or soft mode; raw and canonical utility tables; action semantics and utility provenance; L_U^{dec}, tie order, $\mathcal{A}_{\mathrm{eff}}$, λ, and τ; discarded actions; and tests of positive rescaling, outcome offsets, and irrelevant-action invariance. The record should also include train/validation/test splits, seed sets, selected query indices, posterior approximation and hypothetical-update code, reference-cohort construction and weights, candidate-pool or stream law, the basis for sequential ignorability, martingale residuals, acquisition time, training time, and stopping criteria. Batch reports must state that k-means used the full pool with k equal to the realized batch size and must identify any prefilter as an ablation. Validation labels used for calibration or hyperparameter selection count toward the expert-label budget or are reported separately. The convergence theorem must not be invoked for an exhausted finite pool, approximate updates, nonignorable candidates, or the batch algorithm.