MARGINAL OPTIMIZATION VALUE OF EVIDENCE FOR CONSTRAINED ALLOCATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Allocating limited resources under incomplete information requires choosing which evidence to acquire before choosing an allocation. Yet reducing uncertainty need not resolve which feasible allocation is preferable. We introduce Marginal Optimization Value of Evidence (MOVE), an acquisition rule that targets uncertainty in utility differences between feasible allocations. MOVE pairs predicted covariance reduction with the sensitivity of an entropy-smoothed allocation value, so that constraints shape which observations are useful. We establish an exact allocation-contrast interpretation and connect the implemented score to reoptimization value through explicit virtual-update and prediction-error terms, without assuming that the learned model is a coherent posterior. A masked-evidence model supplies query estimates, and an exact discrete solver returns the final allocation. Across 96 held-out allocation instances on public county data, spanning 32 county subsets and three objectives, MOVE achieves higher overall utility than all evaluated same-budget baselines, with component and objective-wise analyses supporting allocation-sensitive acquisition.

1 INTRODUCTION

An allocation system with limited resources often has two decisions to make: which evidence to obtain and which candidates to select. Consider an outreach program that must choose a fixed-size group of counties while meeting budget and coverage requirements. Evidence that resolves which of two feasible groups has higher utility can improve the allocation; an update that merely raises every group's utility equally cannot. Thus, a query can be informative about utilities without being informative about the choice. The relevant comparison is between complete feasible allocations.

Valuing observations by their downstream consequences is a classical idea in knowledge-gradient methods and goal-oriented experimental design (Frazier et al., 2009; Kandasamy et al., 2019; Neiswanger et al., 2022). Active feature acquisition provides conditional models and policies for purchasing missing measurements (Ma et al., 2019; Covert et al., 2023; Valancius et al., 2024), while sufficient-information analyses identify objective directions that determine an optimization decision (Bennouna et al., 2025; Ye et al., 2026). Our focus is a concrete design question at their intersection: *how should a query be valued when it reveals a block of correlated evidence about a candidate, but its benefit depends on competition among constrained allocations?* A candidate's marginal uncertainty alone does not describe how that observation changes the comparison.

We approach this question through *feasible-allocation contrasts*: utility differences between allocations that satisfy all constraints. This perspective imposes two requirements on acquisition. A rule should reflect which feasible alternatives are competitive at the current information state, and how a candidate observation could reduce uncertainty about their relative utilities. Constraints determine the alternatives; the joint utility model determines what a query can reveal about them. The value of the same evidence block can therefore change when a different coverage requirement changes the feasible alternatives. Our hypothesis is that targeting these contrasts is more useful than reducing utility uncertainty without regard to the feasible choices.

We propose MOVE, short for Marginal Optimization Value of Evidence, to implement this perspective. At each step, the method pairs the predicted utility-covariance reduction from a query with the sensitivity of an entropy-smoothed allocation value. The resulting score averages uncertainty reductions along feasible-allocation contrasts, weighting each reduction by its relevance to the current

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.

decision. Beyond the ideal-posterior case, we relate the implemented score to reoptimization value through errors restricted to feasible-allocation contrasts. This separates the role of decision geometry from the accuracy of the learned updates. A learned masked-evidence model estimates the moments needed for acquisition; smoothing is used only to choose evidence, and the final allocation is solved exactly for the predicted utilities.

We evaluate MOVE across a suite of county-subset allocation problems assembled from public health and demographic data. The held-out evaluation combines 32 county subsets with three objectives, producing 96 allocation instances across 324 counties and 17 states, with eight subsets per Census region. Counties are disjoint across training, development, and test partitions. At the same query budget, MOVE achieves higher overall utility than all evaluated external acquisition baselines. Component comparisons, including a variance-only MOVE variant, test the score's incremental contribution separately from the full-system comparison. Objective-wise results, constraint slices, and source-use analyses examine where gains occur and how queries change with the objective.

Our contributions are threefold:

- **An allocation-contrast perspective.** We formulate policy-conditioned evidence acquisition around uncertainty in the utility differences between feasible allocations, making the acquisition target depend on the choices available to the allocator.
- **A sensitivity-based acquisition rule.** We introduce MOVE, which combines predicted covariance reduction with allocation sensitivity to score evidence queries. A value bound relates the learned covariance updates to reoptimization gains and one-step decision value.
- **Empirical validation and analysis.** Across the public-data allocation suite, MOVE attains the highest overall utility among the evaluated same-budget methods. Component, objective-wise, and query-use analyses assess the score's contribution and the method's behavior across the tested objectives.

2 Problem formulation

We study *evidence acquisition for constrained allocation*: choosing what to observe before allocating a limited resource. An input instance consists of n candidates, a policy descriptor p specifying the utility objective, and a nonempty feasible family $\mathcal{F} \subseteq \{0,1\}^n$ encoding resource and coverage constraints. A vector $s \in \mathcal{F}$ indicates the selected candidates and is called an allocation or slate. Each candidate has partially observed evidence blocks; the state $\mathcal{I}_k$ after k queries contains the observed values, their mask, policy/source descriptions, and public constraint metadata. A legal query $g = (i, b)$ reveals candidate i's hidden block of type b, returning Y_g at cost $c_g > 0$. Given an acquisition budget, the task is to select queries sequentially and output a feasible allocation $\hat{s}$ to maximize the expected value of $(u^\star)^\top \hat{s}$, where $u^\star \in \mathbb{R}^n$ is the unknown policy-conditioned utility vector. The primary protocol permits K unit-cost queries. The acquisition budget is separate from the resource budget defining $\mathcal{F}$, which stays fixed during querying. For evaluation, $s^\star$ denotes an oracle maximizer of $(u^\star)^\top s$ over $\mathcal{F}$; the outcome labels used to compute utility are not queryable evidence.

3 MOVE: acquiring evidence about competing allocations

Overview. A useful query resolves uncertainty that can change a comparison between feasible allocations. Figure 1 shows MOVE's four modules: *utility inference* estimates joint utility moments; *allocation sensitivity* identifies relevant utility directions; *virtual reveals* estimate query-specific covariance reductions; and *query selection* pairs the latter two quantities to choose an observation. Each real reveal updates the evidence state. Once the query budget is spent, an exact solver selects an allocation from the updated predicted utilities.

3.1 Inferring utilities from masked evidence

Independent candidate uncertainties can miss how a query changes beliefs about other candidates. We therefore predict *joint* utility moments. A frozen encoder embeds policy/source text; trainable projections fuse it with observed-block encodings, explicit mask bits, missing-block embeddings,

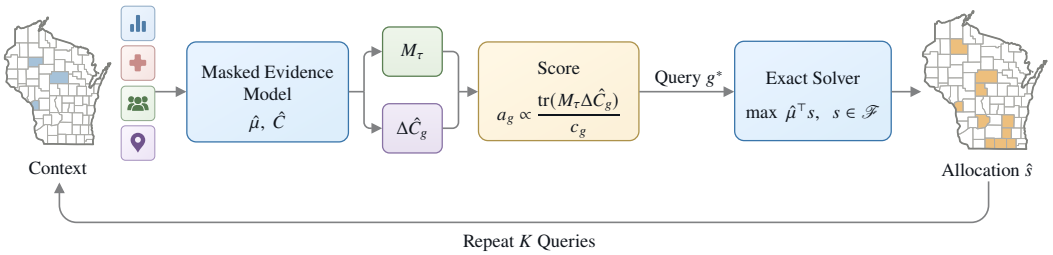


Figure 1: The MOVE acquisition loop. The masked model supplies utility estimates and predicts hidden evidence; virtual reveals reuse that model without accessing real hidden blocks. The feasible family determines allocation sensitivity and constrains the final solver. Only the selected query reveals real evidence. Smoothing guides acquisition, whereas the final allocation uses the updated predicted mean and an exact discrete solver.

and public metadata. Mask bits distinguish missing entries from observed values, including genuine zero-valued measurements. With trainable parameters θ, the model outputs

$$\hat{\mu}_\theta(\mathcal{I}) \in \mathbb{R}^n, \qquad C_\theta(\mathcal{I}) = D_\theta(\mathcal{I}) + L_\theta(\mathcal{I}) L_\theta(\mathcal{I})^\top \in \mathbb{R}^{n\times n}. \tag{1}$$

Here $\hat{\mu}_\theta$ is the predicted utility mean, C_θ the joint utility-covariance estimate, D_θ a positive diagonal matrix, and $L_\theta \in \mathbb{R}^{n\times r}$ a factor of rank at most $r = 4$. The off-diagonal entries of C_θ represent cross-candidate dependence. For candidate i, $D_{\theta,ii} = \mathrm{softplus}(\eta_i) + 10^{-6}$, where η_i is a scalar model output and $\mathrm{softplus}(z) = \log(1 + e^z)$. The model also predicts hidden evidence. Training masks separate observed inputs from hidden evidence and utility targets; evaluation labels never become inputs. These outputs provide a shared utility representation for allocation sensitivity and hypothetical evidence updates. Appendix C gives implementation details.

3.2 Identifying allocation-sensitive utility directions

The exact optimizer is locally constant away from decision boundaries. We obtain sensitivity by smoothing the value over *feasible allocations*, so budget and coverage constraints enter the score directly. For utility vector $u \in \mathbb{R}^n$ and temperature $\tau > 0$, define

$$V_\tau(u) = \tau \log \sum_{s\in\mathcal{F}} e^{u^\top s/\tau}, \qquad \pi_\tau(s;u) = \frac{e^{u^\top s/\tau}}{\sum_{s'\in\mathcal{F}} e^{u^\top s'/\tau}}. \tag{2}$$

The value V_τ smooths the hard maximum $V_0(u) = \max_{s\in\mathcal{F}} u^\top s$; π_τ assigns Gibbs weights to feasible allocations. Writing $\tilde{s}_\tau(u) = \mathbb{E}_{s\sim\pi_\tau}[s]$ for their mean, the Hessian with respect to u is

$$M_\tau(u) = \nabla_u^2 V_\tau(u) = \frac{\mathbb{E}_{s\sim\pi_\tau}[ss^\top] - \tilde{s}_\tau(u)\tilde{s}_\tau(u)^\top}{\tau} \succeq 0. \tag{3}$$

For direction $v \in \mathbb{R}^n$, $v^\top M_\tau v = \mathrm{Var}_{s\sim\pi_\tau}(v^\top s)/\tau$: weight increases when feasible allocations disagree along v. When several feasible allocations have similar predicted utility, the Hessian combines their disagreements across utility directions. We evaluate $\widehat{M} = M_\tau(\hat{\mu}_\theta(\mathcal{I}))$ once per state, fixing τ for differentiation and query comparison (Appendix C). The Gibbs weights define the acquisition sensitivity. The final decision is obtained by discrete optimization of the updated predicted utilities.

3.3 Estimating what each query would reveal

Uncertainty is useful to acquisition only when a query can reduce it. We insert hypothetical evidence into the queried slot and rerun the *same* masked model. Each candidate reveal is evaluated in a separate copy of the current observed state. Write $C_\theta(\mathcal{I}, Y_g)$ for the resulting covariance and $\widehat{\mathbb{E}}_g$ for numerical integration over predicted responses Y_g. The estimated reduction is

$$\widehat{\Delta C}_g = \Pi_{\succeq 0}\Big(\mathrm{sym}\Big[C_\theta(\mathcal{I}) - \widehat{\mathbb{E}}_g C_\theta(\mathcal{I}, Y_g)\Big]\Big). \tag{4}$$

For a square matrix A, $\mathrm{sym}(A) = (A + A^\top)/2$; $\Pi_{\succeq 0}$ replaces negative eigenvalues by zero. This produces a positive-semidefinite update whose pairing with allocation sensitivity yields a nonnegative query score. Virtual states are discarded after scoring, so only the selected query accesses real hidden evidence. Appendix C describes virtual covariance evaluation.

3.4 SELECTING EVIDENCE AND RETURNING AN ALLOCATION

To prioritize uncertainty reduction that affects the decision, we pair the query-specific matrix with the shared allocation sensitivity:

$$\widehat{a}_g = \frac{1}{2c_g} \operatorname{tr}\left(\widehat{M\Delta C}_g\right). \tag{5}$$

Here $\widehat{a}_g$ is a score per unit cost and tr sums diagonal entries. Let $\mathcal{G}(\mathcal{I}_k)$ contain the still-hidden candidate–block pairs. We query $g_k \in \arg\max_{g\in\mathcal{G}(\mathcal{I}_k)} \widehat{a}_g$ and update $\mathcal{I}_{k+1}$ with its response and mask. After K unit-cost queries, the loop returns

$$\hat{s} \in \arg\max_{s\in\mathcal{F}} \hat{\mu}_\theta(\mathcal{I}_K)^\top s. \tag{6}$$

The solver is exact for the *predicted* objective. Algorithm 1 in Appendix C gives the complete loop. Cost normalization defines a heuristic for unequal-cost affordable queries; the experiments and exact-ranking result use equal costs.

Computation. With $m = |\mathcal{F}|$ cached feasible indicators, dense sensitivity costs $O(mn^2)$. Each candidate query additionally requires virtual model evaluations and an $O(n^3)$ eigendecomposition. The evaluated instances have $n = 15$ and at most $\binom{15}{4} = 1365$ slates before filtering. The feasible indicators are cached once and reused for sensitivity calculations and terminal optimization within an instance.

3.5 WHY THE SCORE TARGETS ALLOCATION-RELEVANT EVIDENCE

We first characterize the geometry of the implemented score, then analyze its ideal-posterior counterpart. Fix a mean vector $\mu \in \mathbb{R}^n$ and any symmetric covariance-change matrix ΔC_g, and write $a_g = \operatorname{tr}(M_\tau(\mu)\Delta C_g)/(2c_g)$.

Proposition 1 (Allocation contrasts)**.** *For independent $s, s' \sim \pi_\tau(\cdot;\mu)$,*

$$a_g = \frac{1}{4\tau c_g}\mathbb{E}_{s,s'}\left[(s-s')^\top \Delta C_g (s-s')\right]. \tag{7}$$

For the contrast space $\mathcal{D}_\mathcal{F} = \operatorname{span}\{s - s' : s, s' \in \mathcal{F}\}$, $\ker M_\tau(\mu) = \mathcal{D}_\mathcal{F}^\perp$.

Here ker is the null space and $\mathcal{D}_\mathcal{F}^\perp$ contains directions constant across feasible allocations. Substituting $\mu = \hat{\mu}_\theta(\mathcal{I})$ and $\Delta C_g = \widehat{\Delta C}_g$ gives the same identity for the implemented score: its contrast-based form holds after virtual approximation and PSD projection. Thus the weighting excludes decision-invariant utility directions. For fixed-size slates, a covariance reduction purely along $\mathbf{1}$ gets zero weight; a common-factor observation can still matter if it informs relative utilities.

Why more variance reduction can have zero decision value. Consider a one-query instance that selects one of two candidates, with $\mathcal{F} = \{(1,0)^\top, (0,1)^\top\}$ and utilities $u^\star = (u_0 + A + B, u_0 + A - B)^\top$. The offset $u_0 > a + b$ is known; A and B are independent, uniform on $\{-a, a\}$ and $\{-b, b\}$, with $0 < b < a$. Two unit-cost queries reveal either A or B exactly. Total-variance acquisition and ideal MOVE use the same exact prior and covariance reductions. A query's decision value is the expected increase in the best posterior-mean allocation value.

Table 1: Exact query values in the two-candidate construction at the initial tie; $\tau > 0$ is the smoothing temperature.

Query	Total variance reduction	Decision value	Ideal MOVE score
Reveal A (common shift)	$2a^2$	0	0
Reveal B (contrast)	$2b^2$	b	$b^2/(2\tau)$

Revealing A removes more total variance but shifts both posterior means equally, leaving the choice unresolved. Revealing B identifies the better candidate and increases expected allocation utility by b. For every $0 < b < a$ and $\tau > 0$ in this construction, total-variance acquisition selects a

zero-value query, whereas ideal MOVE selects the informative contrast (Table 1). *Even with exact beliefs, total-variance reduction need not improve the decision*: allocation sensitivity distinguishes shared uncertainty from uncertainty about the choice. Appendix A.1 gives the calculation; Section 4.3 evaluates learned estimates.

Local information value. For the VOI analysis, assume that all conditional means and covariances come from the same posterior over $u^\star$, with finite second moments. Define $\mu = \mathbb{E}[u^\star \mid \mathcal{I}]$, $\mu_g^+ = \mathbb{E}[u^\star \mid \mathcal{I}, Y_g]$, $C = \mathrm{Cov}(u^\star \mid \mathcal{I})$, and $C_g^+ = \mathrm{Cov}(u^\star \mid \mathcal{I}, Y_g)$. Total covariance gives $\Delta C_g = C - \mathbb{E}[C_g^+ \mid \mathcal{I}] = \mathrm{Cov}(\mu_g^+ \mid \mathcal{I}) \succeq 0$. For either the hard value V_0 or the smooth value V_τ, the one-step value of information is

$$\mathrm{VOI}_t(g \mid \mathcal{I}) = \mathbb{E}[V_t(\mu_g^+) \mid \mathcal{I}] - V_t(\mu), \qquad t \in \{0, \tau\}. \tag{8}$$

The index t selects hard or smoothed value. This is expected gross gain, before deducting cost; the hard version also equals expected Bayes-regret reduction (Appendix A.2).

Proposition 2 (Local information value)**.** *Under this posterior model, fix $\mathcal{I}$, $\mathcal{F}$, and $\tau > 0$. If the Hessian of V_τ is L_3-Lipschitz in operator norm along the segments from μ to μ_g^+, and $\mathbb{E}[\|\mu_g^+ - \mu\|_2^3 \mid \mathcal{I}] < \infty$, then*

$$\left|\mathrm{VOI}_\tau(g \mid \mathcal{I}) - \tfrac{1}{2}\,\mathrm{tr}(M_\tau(\mu)\Delta C_g)\right| \le \frac{L_3}{6}\mathbb{E}[\|\mu_g^+ - \mu\|_2^3 \mid \mathcal{I}]. \tag{9}$$

Here L_3 bounds Hessian variation and $\|\cdot\|_2$ is the Euclidean norm. The posterior-mean martingale cancels the linear term in a Taylor expansion, leaving the trace as a local approximation to one-step expected value.

The same posterior model yields $\mathrm{VOI}_\tau(g) = \tau\mathbb{E}[\mathrm{KL}(\pi_\mu \| \pi_{\mu_g^+}) \mid \mathcal{I}]$ and $|\mathrm{VOI}_\tau(g) - \mathrm{VOI}_0(g)| \le \tau \log|\mathcal{F}|$. Appendix A.3 gives the derivations. Section 3.6 extends this analysis to learned updates by making their prediction and integration errors explicit.

Exact ranking in a restricted case. The local rule has an exact ranking interpretation when only two allocations are feasible.

Proposition 3 (Two-allocation Gaussian ranking)**.** *Under the posterior model above, let $\mathcal{F} = \{s_a, s_b\}$ with $s_a \neq s_b$ and common query cost $c > 0$. Set $d = s_a - s_b$, $\delta = d^\top \mu$, and $q_g = d^\top \Delta C_g d$. If $d^\top(\mu_g^+ - \mu) \mid \mathcal{I} \sim \mathcal{N}(0, q_g)$ for every legal query, ideal MOVE and hard one-step VOI induce identical query rankings, including ties.*

Here d is the allocation contrast and q_g is its posterior-mean update variance. Both criteria increase with q_g; Appendix A.4 provides the closed form and proof.

3.6 From learned updates to decision value

The practical rule uses learned means and a projected covariance update. To analyze these quantities directly, fix $\mathcal{I}, \mathcal{F}, \tau$ and let Q_g be the normalized, nonnegative virtual-response measure used by numerical integration, including a single-point rule. Write $\hat{\mu} = \hat{\mu}_\theta(\mathcal{I})$, $z_g(y) = \hat{\mu}_\theta(\mathcal{I}, y) - \hat{\mu}$, and $S_g = \mathbb{E}_{Q_g}[z_g z_g^\top]$. Let P project onto $\mathcal{D}_\mathcal{F}$. The smooth value of reoptimizing after a virtual reveal, relative to retaining the current Gibbs allocation, is

$$J_g = \frac{\mathbb{E}_{Q_g}[V_\tau(\hat{\mu} + z_g) - V_\tau(\hat{\mu}) - \tilde{s}_\tau(\hat{\mu})^\top z_g]}{c_g} = \frac{\tau}{c_g}\mathbb{E}_{Q_g}\mathrm{KL}(\pi_{\hat{\mu}} \| \pi_{\hat{\mu}+z_g}). \tag{10}$$

Subtracting the value change of the current allocation isolates the benefit of reoptimization and permits nonzero mean drift in learned updates. MOVE continues to select queries using Eq. (5).

Theorem 1 (Learned-score value bridge)**.** *For finite nonempty $\mathcal{F}$, $c_g > 0$, and finite third projected update moments, let L_3 be a Hessian Lipschitz constant of V_τ along projected update segments. With $\widehat{M} = M_\tau(\hat{\mu})$ and the actual $\widehat{\Delta C}_g$ from Eq. (4),*

$$|J_g - \widehat{a}_g| \le \epsilon_g := \frac{\|\widehat{M}\|_F \, \|P(\widehat{\Delta C}_g - S_g)P\|_F}{2c_g} + \frac{L_3}{6c_g}\mathbb{E}_{Q_g}\|Pz_g\|_2^3. \tag{11}$$

For true posterior means μ, μ_g^+, set $e_0 = \|P(\hat{\mu} - \mu)\|_2$ and $e_g = \mathbb{E}[\|P(\hat{\mu}_\theta(\mathcal{I}, Y_g) - \mu_g^+)\|_2 \mid \mathcal{I}]$. Let ξ_g be the absolute virtual-integration error in the unnormalized gain in Eq. (10), relative to the actual response law. Then, for $D_{\mathcal{F}} = \max_{s,s' \in \mathcal{F}} \|s - s'\|_2$,

$$\left| \frac{\mathrm{VOI}_0(g \mid \mathcal{I})}{c_g} - \widehat{a}_g \right| \leq B_g := \epsilon_g + \frac{\tau \log |\mathcal{F}| + D_{\mathcal{F}}(e_0 + e_g) + \xi_g}{c_g}. \tag{12}$$

The first bound measures agreement between the implemented covariance reduction and virtual mean changes *in allocation-relevant directions*. The second adds prediction, integration, and smoothing errors. If $g^\star$ maximizes true one-step VOI per cost and $\widehat{g}$ maximizes MOVE, the selection loss is at most $B_{g^\star} + B_{\widehat{g}}$ under the stated error bounds. Appendix A.6 proves these results for arbitrary learned updates with finite moments.

From query value to the learned allocation. Let $\hat{s}_0$ and $\hat{s}_g^+(Y_g)$ be the discrete allocations obtained from the learned means before and after one reveal, using the solver in Eq. (6). Their expected utility gain under the actual response law is $A_g = \mathbb{E}[(\mu_g^+)^\top \hat{s}_g^+ \mid \mathcal{I}] - \mu^\top \hat{s}_0$. Under the conditions of Theorem 1,

$$\left| \frac{A_g}{c_g} - \widehat{a}_g \right| \leq B_g + \frac{D_{\mathcal{F}}(e_0 + e_g)}{c_g}. \tag{13}$$

The additional term accounts for prediction error in the terminal decisions along feasible-allocation contrasts. This extends the score–value connection to the discrete allocation returned by the learned model. Appendix A.6 derives the result.

4 Experiments

We test allocation-sensitive acquisition across county subsets, policy objectives, and constraint conditions. We first compare final allocation quality, then examine score components and objective-conditioned query behavior.

4.1 Evaluation setting

Allocation problems and public data. We aggregate CDC surveillance and vaccination records, CDC/ATSDR and CDC PLACES attributes, and NCHS and Census population, rurality, and adjacency records. A *county subset* (the base context) specifies 15 candidate counties and a reference week. Combining it with each of three objectives—incidence, severe outcomes, and equity—defines three allocation instances with public costs and coverage requirements. Each instance selects four counties and queries surveillance, severe-risk, equity/access, or spillover blocks. Appendix B gives source releases, eligibility and aggregation rules, and objective definitions.

Coverage and protocol. The collection contains 96 county subsets and 288 objective-conditioned instances, covering 960 unique counties across 50 states or equivalents. County-disjoint training/development/test partitions contain 48/16/32 subsets. The 32 test subsets cover 324 counties in 17 states, span all four Census regions equally, and yield 96 held-out instances. Counties can recur across subsets within one partition. Resampling keeps each subset and its three objective variants together. These construction and split requirements determine the evaluated county pool. Initially one block per county is observed. All acquisition methods share $K = 4$ additional unit-cost queries, initial masks, public information, and the exact final solver. Future two-week outcomes enter utility supervision and evaluation, never queryable inputs. Appendix B reports the split composition and geographic overlap.

Baselines and comparison design. We compare Uncertainty, EDDI, Dynamic MI, ACO, ReAct, and Toolformer (Ma et al., 2019; Covert et al., 2023; Valancius et al., 2024; Yao et al., 2023; Schick et al., 2023) through the same candidate–block query interface. The language-model baselines, ReAct and Toolformer, share the `gemma-3-4b-it` backbone. Ridge and decision-focused learning (DFL) plug-ins use the initial evidence without querying (Hoerl and Kennard, 1970; Donti et al., 2017; Wilder et al., 2019). Table 2 compares complete systems; the MOVE-family variants in Table 3 examine its acquisition components. Appendix C.2 specifies the predictor and routing implementations.

Metrics and uncertainty. The primary metric is mean episode-wise utility ratio, $(u^\star)^\top \hat{s}/(u^\star)^\top s^\star$. Oracle-set Jaccard measures agreement with the oracle selection. Policy response measures the mean

pairwise slate dissimilarity across a context's objectives. Paired 95% intervals use 500 resamples of the 32 test subsets, keeping all three objectives together. They quantify subset-resampling uncertainty; component results are point estimates. Appendix C.3 defines the metrics and statistical units.

Table 2: Main results across the public-data allocation suite. Utility ratio is primary, Jaccard secondary, and policy response diagnostic. The reference plug-ins use no additional evidence; acquisition methods share budget K (Section 4.1).

Method	Utility ratio ↑	Jaccard ↑	Policy response	Budget
Ridge plug-in	0.791	0.352	0.772	0
DFL plug-in	0.772	0.346	0.772	0
Uncertainty	0.814	0.384	0.760	K
EDDI	0.797	0.374	0.741	K
Dynamic MI	0.809	0.379	0.737	K
ACO	0.797	0.344	0.814	K
ReAct	0.811	0.363	0.834	K
Toolformer	0.795	0.361	0.813	K
MOVE	**0.878**	**0.486**	0.801	K

4.2 Allocation quality

MOVE achieves the highest utility among the evaluated systems: 0.878 versus 0.814 for Uncertainty (Table 2). The paired difference is 0.064 with 95% interval [0.033, 0.096]. Oracle-set Jaccard also increases, from 0.384 to 0.486, with difference interval [0.045, 0.164]. These paired intervals support a gain across held-out subsets; Table 15 reports the bootstrap intervals for individual methods and paired differences. The two metrics capture complementary aspects of the output: objective value and agreement on which counties receive the allocation. The system comparison establishes the end-to-end result; the following within-family comparisons examine the contribution of the acquisition design.

4.3 Ablations of allocation-sensitive acquisition

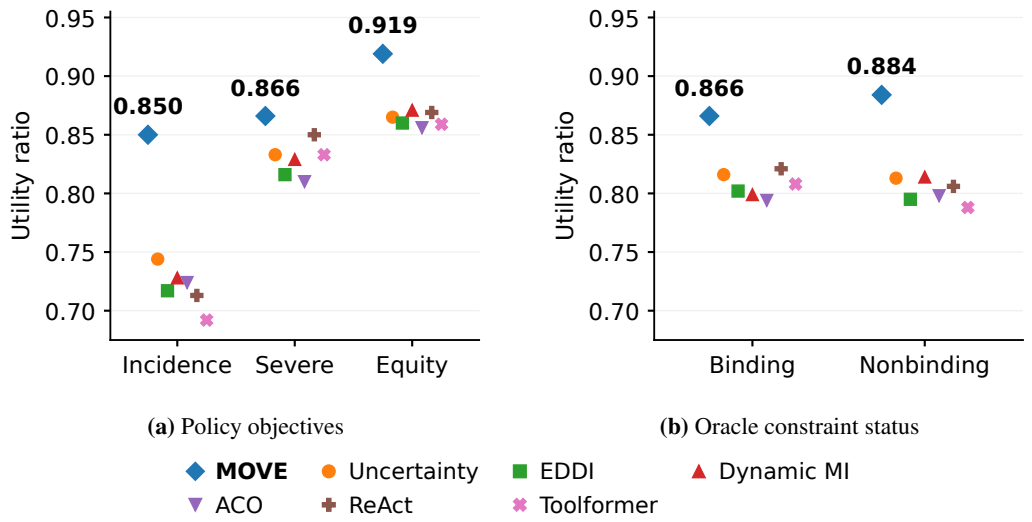
Table 3: MOVE-family ablations under the shared budget K. Change is the utility-ratio difference from full MOVE, computed from the reported means. Policy response measures changes in the selected sets across objectives. Component entries are point estimates.

Variant	Utility ratio ↑	Change	Policy response
MOVE	**0.878**	—	0.801
Without allocation sensitivity	0.851	−0.027	0.800
Without covariance contraction	0.860	−0.018	0.796
Without semantic descriptors	0.852	−0.026	0.808
Variance-only acquisition	0.859	−0.019	0.793

Allocation information beyond variance. MOVE is designed to prioritize uncertainty relevant to feasible alternatives, not uncertainty alone. The variance-only variant tests this distinction within the MOVE family: it reaches 0.859, compared with 0.878 for the full score (Table 3). This 0.019 increase supports using allocation relevance in addition to within-model uncertainty, complementing the complete-system comparison.

Where uncertainty matters and what a query changes. Equation (5) pairs two quantities: $\widehat{M}$ weights utility directions by allocation sensitivity, while $\widehat{\Delta C}_g$ predicts the reduction associated with query g. The corresponding removal variants give 0.851 without sensitivity and 0.860 without contraction, versus 0.878 for MOVE. The lower utility of both removal variants supports using decision relevance together with query-specific information gain.

Conditioning the evidence model on the objective. Policy and source descriptors give the masked model context for interpreting a reveal. Removing them yields utility 0.852, a decrease of 0.026 from MOVE, supporting their role in the complete acquisition pipeline. Yet policy response is slightly higher without descriptors (0.808 versus 0.801; Table 3). The descriptors improve allocation utility even though they produce slightly less change in the selected sets across objectives.



(a) Policy objectives **(b)** Oracle constraint status

Figure 2: Allocation quality across evaluation conditions. Both panels show MOVE and all six external acquisition baselines, with a common utility-ratio scale and budget K. Each objective contains 32 episodes; binding/nonbinding contains 31/65. Markers show group means for two partitions of the same test set.

These ablations support combining allocation-sensitive weighting, query-specific covariance reduction, and semantic conditioning of utility prediction. Section 3.6 separately bounds how update-consistency and prediction errors affect the score's agreement with decision value.

4.4 Performance across objectives and constraints

Different objectives on the same county subsets. We test whether the aggregate advantage extends across all three objectives. MOVE achieves the highest mean utility ratio for each of the three objectives: 0.850 for incidence, 0.866 for severe outcomes, and 0.919 for equity (Figure 2(a)). The strongest external comparators are respectively Uncertainty (0.744), ReAct (0.850), and Dynamic MI (0.871). The positive but uneven gaps (0.106, 0.016, 0.048) show that the overall advantage extends across all three evaluated objectives. Because each objective is evaluated on the same county subsets, this comparison changes the allocation priorities while retaining the candidate setting.

When constraints change the oracle choice. The evaluator labels an episode *oracle-binding* when constraints change the oracle set relative to unconstrained top-four selection. This identifies cases where feasibility changes the target decision. On its 31 episodes in 21 contexts, MOVE reaches 0.866 versus 0.821 for ReAct, with paired difference interval [0.011, 0.074]. On the 65 nonbinding episodes, MOVE reaches 0.884 versus the strongest comparator's 0.814 (Dynamic MI; Figure 2(b)). Appendix Table 13 also reports higher oracle-set agreement, with Jaccard 0.472/0.493 for MOVE on the binding/nonbinding groups. MOVE improves both utility and oracle-set agreement in the two groups, with larger gains in the nonbinding group. Group membership is determined by the final oracle selection.

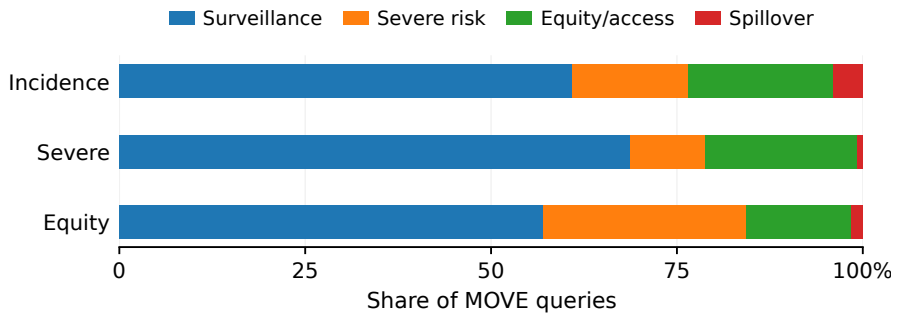


Figure 3: Distribution of MOVE queries across evidence sources for each objective. Surveillance receives the largest share under all three objectives; the remaining shares vary by objective.

4.5 Objective-conditioned query behavior

Query allocation across objectives. Figure 3 shows how MOVE distributes queries across evidence sources. Surveillance receives the largest share under each objective, while the other shares vary. Utility and responsiveness capture different aspects of this behavior: ReAct changes its selected sets more across objectives (policy response 0.834 versus 0.801), but attains lower utility (0.811 versus 0.878). The descriptor ablation exhibits the same distinction between changes in selection and allocation quality.

5 Related work

Evidence acquisition for a downstream decision. EDDI and dynamic feature selection use information about prediction targets (Ma et al., 2019; Covert et al., 2023); target information alone need not distinguish feasible allocations. A closer rule is Perets et al. (2026), which weights pairwise feature gains by influence on a top-k frontier. Its ranking weights do not encode resource budgets or coverage quotas: a competitive pairwise exchange may be infeasible under those constraints. MOVE instead derives weights from *complete feasible allocations* and combines them with joint utility-covariance changes. The resulting query weights reflect which alternative allocations satisfy the resource and coverage constraints.

Knowledge-gradient, goal-oriented design, and decision-theoretic entropy already value observations through terminal choices (Frazier et al., 2009; Kandasamy et al., 2019; Neiswanger et al., 2022); ACO also supports general decision-making and nongreedy acquisition (Valancius et al., 2024). MOVE instantiates these decision-aware objectives for constrained allocation through a finite-slate sensitivity matrix and virtual block updates. This yields a computable local score and an analysis of its relation to one-step decision value.

Optimization-sensitive learning and information. Decision-focused learning aligns *training* with downstream optimization (Donti et al., 2017; Wilder et al., 2019), using sensitivities supplied by differentiable layers (Amos and Kolter, 2017; Agrawal et al., 2019). Optimization-margin active learning selects training labels (Liu et al., 2025), while decision-sufficient representations characterize information that preserves optimal choices (Bennouna et al., 2025; Ye et al., 2026). These approaches address training labels, prediction objectives, or sufficient representations. MOVE uses decision geometry at *acquisition time*, weighting the predicted covariance change for each candidate–block reveal by the Hessian of feasible-allocation value.

The closest algebraic analogue is RMWorld (Wang et al., 2026), which also couples task sensitivity with covariance reduction. Its weighting matrix measures the reduction in locally linearized squared rate-prediction risk under a linear-Gaussian model. MOVE instead measures sensitivity of the optimized allocation value through the covariance of feasible allocation indicators. This excludes decision-invariant directions and emphasizes utility differences among competing allocations. Recomputing the sensitivity after each observation links the next query to the updated allocation problem.

6 Conclusion

We introduced MOVE, an evidence-acquisition rule that values observations through utility differences between feasible allocations. By pairing predicted covariance reduction with allocation sensitivity, MOVE directs queries toward uncertainty that matters to the final decision. Our analysis preserves the score's allocation-contrast interpretation for learned updates and bounds its relation to reoptimization value through explicit contrast-restricted errors. Experiments across county subsets and three allocation objectives show higher utility than the evaluated same-budget baselines, with component and objective-wise analyses supporting the proposed design. While the current formulation assumes fixed objectives and constraints within each acquisition episode, its explicit link between query value and feasible alternatives motivates future work on acquisition under evolving priorities and resource availability.

AI USE STATEMENT

Generative AI tools were used for assistance with manuscript editing and code-related tasks. All scientific contributions, experimental design, analysis, and conclusions were developed and verified by the authors.

ETHICS STATEMENT

Generative AI tools were used for assistance with manuscript editing and code-related tasks. All scientific contributions, experimental design, analysis, and conclusions were developed and verified by the authors.

REPRODUCIBILITY STATEMENT

Appendix A contains assumptions and proofs. Appendix B specifies public sources, preprocessing, splits, evidence masks, and utility definitions. Appendix C describes the algorithm, baseline implementations, and statistical protocol; Appendix D provides detailed results. Code will accompany the submission.

REFERENCES

Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and J. Zico Kolter. Differentiable Convex Optimization Layers. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/9ce3c52fc54362e22053399d3181c638-Abstract.html.

Brandon Amos and J. Zico Kolter. OptNet: Differentiable Optimization as a Layer in Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 136–145. PMLR, 2017. URL https://proceedings.mlr.press/v70/amos17a.html.

Omar Bennouna, Amine Bennouna, Saurabh Amin, and Asuman Ozdaglar. What data enables optimal decisions? an exact characterization for linear optimization. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 166020–166049. Curran Associates, Inc., 2025. doi: 10.52202/085713-5534. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/f2865c76810f9f702f52fd3e4150a296-Paper-Conference.pdf.

Centers for Disease Control and Prevention. Weekly United States COVID-19 Cases and Deaths by County - ARCHIVED, 2023a. URL https://data.cdc.gov/d/yviw-z6j5. Dataset yviw-z6j5. Issued 2023-06-06; accessed 2026-04-19.

Centers for Disease Control and Prevention. COVID-19 Vaccinations in the United States,County, 2023b. URL https://data.cdc.gov/d/8xkx-amqh. Dataset 8xkx-amqh. Updated 2023-05-12; accessed 2026-04-19.

Centers for Disease Control and Prevention. PLACES: Local Data for Better Health, County Data, 2025 release, 2025. URL https://data.cdc.gov/d/swc5-untb. 2025 release; long-format county data. Accessed 2026-04-19.

Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry/Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index 2022 Database U.S., 2024. URL https://www.atsdr.cdc.gov/place-health/php/svi/svi-data-documentation-download.html. 2022 county-level database. Released 2024-05-21; accessed 2026-04-19.

Ian Connick Covert, Wei Qiu, Mingyu Lu, Na Yoon Kim, Nathan J White, and Su-In Lee. Learning to Maximize Mutual Information for Dynamic Feature Selection. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 6424–6447. PMLR, 2023. URL https://proceedings.mlr.press/v202/covert23a.html.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based End-to-end Model Learning in Stochastic Optimization. In *Advances in Neural Information Processing Systems*, volume 30, pages 5484–5494, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3fc2c60b5782f641f76bcefc39fb2392-Abstract.html.

Peter Frazier, Warren Powell, and Savas Dayanik. The Knowledge-Gradient Policy for Correlated Normal Beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009. doi: 10.1287/ijoc.1080.0314.

Arthur E. Hoerl and Robert W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67, 1970. doi: 10.1080/00401706.1970.10488634.

Kirthevasan Kandasamy, Willie Neiswanger, Reed Zhang, Akshay Krishnamurthy, Jeff Schneider, and Barnabás Póczos. Myopic Posterior Sampling for Adaptive Goal Oriented Design of Experiments. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3222–3232. PMLR, 2019. URL https://proceedings.mlr.press/v97/kandasamy19a.html.

Mo Liu, Paul Grigas, Heyuan Liu, and Zuo-Jun Max Shen. Active learning for contextual linear optimization: A margin-based approach, 2025. URL https://arxiv.org/abs/2305.06584.

Chao Ma, Sebastian Tschiatschek, Konstantina Palla, José Miguel Hernández-Lobato, Sebastian Nowozin, and Cheng Zhang. EDDI: Efficient Dynamic Discovery of High-Value Information with Partial VAE. In *Proceedings of the 36th International Conference on Machine Learning*,

volume 97 of *Proceedings of Machine Learning Research*, pages 4234–4243. PMLR, 2019. URL https://proceedings.mlr.press/v97/ma19c.html.

National Center for Health Statistics. NCHS Urban-Rural Classification Scheme for Counties, 2024. URL https://www.cdc.gov/nchs/data-analysis-tools/urban-rural.html. Documentation page dated 2024-09-17; 2023 classification. Accessed 2026-04-19.

Willie Neiswanger, Lantao Yu, Shengjia Zhao, Chenlin Meng, and Stefano Ermon. Generalizing Bayesian Optimization with Decision-theoretic Entropies. In *Advances in Neural Information Processing Systems*, volume 35, pages 21016–21029, 2022. doi: 10.52202/068431-1528. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/8451a20c5a7e0ee5671dda28f7daf7f3-Abstract-Conference.html.

Binyamin Perets, Natalie Mendelson, Shiran Vainberg, Yehuda Chowers, Shai Shen-Orr, and Shie Mannor. Zero-Shot Active Feature Acquisition via LLM-Elicitation. arXiv:2606.18933, 2026. URL https://arxiv.org/abs/2606.18933.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Advances in Neural Information Processing Systems*, volume 36, pages 68539–68551, 2023. doi: 10.52202/075280-2997. URL https://proceedings.neurips.cc/paper/2023/hash/d842425e4bf79ba039352da0f658a906-Abstract-Conference.html.

U.S. Census Bureau. County Adjacency File, 2024. URL https://www.census.gov/geographies/reference-files/2024/geo/county-adjacency.html. 2024 edition. Accessed 2026-04-19.

U.S. Census Bureau. Annual County and Puerto Rico Municipio Resident Population Estimates by Selected Age Groups and Sex: April 1, 2020 to July 1, 2024, 2025. URL https://www2.census.gov/programs-surveys/popest/datasets/2020-2024/counties/asrh/. Vintage 2024, CC-EST2024-AGESEX. Released 2025-06-26; accessed 2026-04-19.

U.S. Census Bureau. County Population Totals and Components of Change: 2020-2025, 2026. URL https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html. Vintage 2025, CO-EST2025-POP. Released March 2026; accessed 2026-04-19.

Michael Valancius, Maxwell Lennon, and Junier Oliva. Acquisition Conditioned Oracle for Nongreedy Active Feature Acquisition. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 48957–48975. PMLR, 2024. URL https://proceedings.mlr.press/v235/valancius24a.html.

Xiucheng Wang, Nan Cheng, and Junxi Huan. RMWorld: Task-Aware Radio World Models with Value-of-Information Guided Multi-Trial Learning for Multi-UAV Communication Control. arXiv:2608.20126, 2026. URL https://arxiv.org/abs/2608.20126.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the Data-Decisions Pipeline: Decision-Focused Learning for Combinatorial Optimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658. URL https://ojs.aaai.org/index.php/AAAI/article/view/3982.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing Reasoning and Acting in Language Models. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=WE_vluYUL-X.

Yuhan Ye, Saurabh Amin, and Asuman Özdağlar. Learning decision-sufficient representations for linear optimization. In Steve Hanneke and Tor Lattimore, editors, *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 6924–6975. PMLR, 29 Jun–03 Jul 2026. URL https://proceedings.mlr.press/v336/ye26a.html.

A Analysis of the acquisition score

Throughout this appendix, the information state, finite nonempty feasible family $\mathcal{F}$, and temperature $\tau > 0$ are fixed. Algebraic results apply to any symmetric covariance-change estimate. The ideal-posterior results use conditional moments from one posterior; Appendix A.6 analyzes arbitrary learned means and virtual updates with explicit prediction and integration errors.

A.1 Feasible allocations, smoothing, and contrasts

Write $\mathcal{F} = \{s_1, \dots, s_m\}$ and let Δ_m be the probability simplex. For $H(\pi) = -\sum_j \pi_j \log \pi_j$,

$$V_\tau(u) = \max_{\pi \in \Delta_m} \left\{ \sum_j \pi_j u^\top s_j + \tau H(\pi) \right\}. \tag{14}$$

Indeed, substituting $\log \pi_j^* = u^\top s_j/\tau - \log Z(u)$ gives the objective as $\tau \log Z(u) - \tau\,\mathrm{KL}(\pi\|\pi^*)$. Its unique maximizer is the Gibbs distribution in Eq. (2). The entropy regularizes a distribution over complete feasible allocations. Differentiating the finite sum gives

$$\nabla \pi_j = \frac{\pi_j}{\tau}(s_j - \tilde{s}_\tau), \qquad \nabla V_\tau = \tilde{s}_\tau, \qquad M_\tau = \frac{1}{\tau}\sum_j \pi_j (s_j - \tilde{s}_\tau)(s_j - \tilde{s}_\tau)^\top.$$

Proof of Proposition 1. For independent s, s' with the Gibbs law,

$$\mathbb{E}[(s - s')(s - s')^\top] = 2\{\mathbb{E}[ss^\top] - \mathbb{E}[s]\mathbb{E}[s]^\top\} = 2\tau M_\tau.$$

Multiplying by ΔC_g, taking the trace, and dividing by $4\tau c_g$ proves Eq. (7). For any v, $v^\top M_\tau v = \mathrm{Var}(v^\top s)/\tau$. Every $s \in \mathcal{F}$ has strictly positive Gibbs probability. The variance is therefore zero precisely when $v^\top s$ is constant on $\mathcal{F}$, or equivalently when $v^\top(s - s') = 0$ for every pair. Since M_τ is symmetric PSD, this proves $\ker M_\tau = \mathcal{D}_\mathcal{F}^\perp$. □

If P is the orthogonal projector onto $\mathcal{D}_\mathcal{F}$, then

$$M_\tau = P M_\tau P, \qquad \mathrm{tr}(M_\tau \Delta C_g) = \mathrm{tr}(M_\tau P \Delta C_g P). \tag{15}$$

Thus only the covariance change projected onto allocation contrasts contributes. The relevance of differences between feasible optima is consistent with sufficient-data geometry for linear optimization (Bennouna et al., 2025; Ye et al., 2026); Eq. (15) states how this geometry enters the present smooth acquisition score.

For fixed-size slates, $\mathbf{1} \in \mathcal{D}_\mathcal{F}^\perp$. More generally, if $As = b$ for *every* $s \in \mathcal{F}$, then $\mathrm{range}(A^\top) \subseteq \ker M_\tau$. These equalities must hold throughout $\mathcal{F}$. An observation of a common factor can still be useful through its correlation with allocation contrasts; its weight is determined by the projected covariance change in Eq. (15).

Calculation for the two-candidate separation. For the one-query construction in Table 1, let $e_1 = (1,0)^\top$, $e_2 = (0,1)^\top$, $\mathbf{1} = (1,1)^\top$, and $d = e_1 - e_2 = (1,-1)^\top$. The known offset is $u_0 > a + b$ and the independent signs have amplitudes $0 < b < a$. Hence $\mu = u_0\mathbf{1}$ and

$$C = a^2 \mathbf{1}\mathbf{1}^\top + b^2 dd^\top, \qquad \Delta C_A = a^2 \mathbf{1}\mathbf{1}^\top, \qquad \Delta C_B = b^2 dd^\top.$$

Revealing A leaves covariance $b^2 dd^\top$; revealing B leaves covariance $a^2\mathbf{1}\mathbf{1}^\top$. Neither remaining covariance depends on the observed sign, so these are exact expected covariance reductions. Total-variance acquisition prefers A because $\mathrm{tr}(\Delta C_A) = 2a^2 > 2b^2 = \mathrm{tr}(\Delta C_B)$.

Initially both candidates have mean u_0. After observing A, their means are both $u_0 + A$, whose maximum has expectation u_0; thus the exact one-step decision value is zero. After observing B, their means are $u_0 + B$ and $u_0 - B$, whose maximum is $u_0 + |B| = u_0 + b$; thus the decision value is b. The unknown common shift does not prevent identification of the better candidate after B is revealed.

At the initial tie, the two feasible allocations have Gibbs weights $1/2$, giving $M_\tau = dd^\top/(4\tau)$. Since $d^\top \mathbf{1} = 0$ and $d^\top d = 2$, the unit-cost ideal MOVE scores are

$$\frac{1}{2}\mathrm{tr}(M_\tau \Delta C_A) = 0, \qquad \frac{1}{2}\mathrm{tr}(M_\tau \Delta C_B) = \frac{b^2}{2\tau} > 0.$$

In this construction, ideal MOVE agrees with exact one-step decision value, while total variance reverses the query ordering. The analytical comparison isolates the weighting criterion under identical exact beliefs.

A.2 Local information value

Assume $u^\star$ has finite conditional second moments and define

$$\mu = \mathbb{E}[u^\star \mid \mathcal{I}], \quad \mu_g^+ = \mathbb{E}[u^\star \mid \mathcal{I}, Y_g], \quad \delta_g = \mu_g^+ - \mu.$$

The tower property and law of total covariance give

$$\mathbb{E}[\delta_g \mid \mathcal{I}] = 0, \qquad C - \mathbb{E}[C_g^+ \mid \mathcal{I}] = \mathbb{E}[\delta_g \delta_g^\top \mid \mathcal{I}] = \Delta C_g \succeq 0. \tag{16}$$

Positive semidefiniteness in Eq. (16) applies to the response-averaged posterior reduction. The implemented estimator imposes this structure by projection; Appendix A.6 analyzes its agreement with learned mean updates.

Hard decision value and Bayes regret. For $V_0(u) = \max_{s \in \mathcal{F}} u^\top s$, define the conditional Bayes decision regret

$$R(\mathcal{I}) = \mathbb{E}[V_0(u^\star) \mid \mathcal{I}] - V_0(\mu). \tag{17}$$

The first term is the expected value of an oracle that knows $u^\star$; the second is the best value available from the current posterior mean. Iterated expectation cancels the oracle term before and after a query, giving $\mathrm{VOI}_0(g \mid \mathcal{I}) = R(\mathcal{I}) - \mathbb{E}[R(\mathcal{I}, Y_g) \mid \mathcal{I}]$. Thus one-step information value equals the expected reduction in conditional decision regret.

Proof of Proposition 2. Taylor's theorem at μ gives

$$V_\tau(\mu + \delta_g) = V_\tau(\mu) + \nabla V_\tau(\mu)^\top \delta_g + \tfrac{1}{2} \delta_g^\top M_\tau(\mu) \delta_g + R_g,$$

where $|R_g| \le L_3 \|\delta_g\|_2^3 / 6$. The conditional expectation of the linear term is zero. The expected quadratic term is

$$\mathbb{E}[\delta_g^\top M_\tau(\mu) \delta_g \mid \mathcal{I}] = \mathrm{tr}\big(M_\tau(\mu) \mathbb{E}[\delta_g \delta_g^\top \mid \mathcal{I}]\big) = \mathrm{tr}(M_\tau(\mu) \Delta C_g).$$

Subtracting $V_\tau(\mu)$ and bounding $|\mathbb{E}[R_g \mid \mathcal{I}]|$ proves the proposition. □

Explicit regularity for a finite family. If $D = \max_{s, s' \in \mathcal{F}} \|s - s'\|_2$, then

$$\|M_\tau(u)\|_{\mathrm{op}} \le \frac{D^2}{4\tau}, \qquad \|M_\tau(u) - M_\tau(v)\|_{\mathrm{op}} \le \frac{D^3}{\tau^2} \|u - v\|_2. \tag{18}$$

The first inequality follows because the range of $w^\top s$ has width at most D for unit w, and its variance is at most one quarter of the squared width. For the second, the third derivative of V_τ applied to three directions is their third centered Gibbs moment divided by τ^2. Each centered indicator has norm at most D. This bounds the trilinear operator norm by D^3/τ^2; integrating along the segment from u to v gives Eq. (18). Thus $L_3 = D^3/\tau^2$ is a valid global bound. It becomes loose as the temperature decreases.

A.3 An exact divergence identity and smoothing error

For finite vectors μ, ν, define the Bregman divergence

$$D_{V_\tau}(\nu, \mu) = V_\tau(\nu) - V_\tau(\mu) - \nabla V_\tau(\mu)^\top (\nu - \mu).$$

Writing $\pi_\mu = \pi_\tau(\cdot; \mu)$, direct substitution of the Gibbs probabilities gives

$$D_{V_\tau}(\nu, \mu) = \tau \, \mathrm{KL}(\pi_\mu \| \pi_\nu). \tag{19}$$

Taking $\nu = \mu_g^+$ and using the mean martingale property yields

$$\mathrm{VOI}_\tau(g) = \mathbb{E}[D_{V_\tau}(\mu_g^+, \mu) \mid \mathcal{I}] = \tau \mathbb{E}[\mathrm{KL}(\pi_\mu \| \pi_{\mu_g^+}) \mid \mathcal{I}]. \tag{20}$$

The KL divergence is directed from current to updated Gibbs weights. These distributions arise from the entropy-smoothed optimization problem, and their divergence gives the value whose local expansion yields the trace score.

To relate this analysis to the hard final decision, let $m = |\mathcal{F}|$. Factoring the largest exponential out of Eq. (2) gives

$$0 \le V_\tau(u) - V_0(u) \le \tau \log m.$$

Set $h(u) = V_\tau(u) - V_0(u)$. Since $h \in [0, \tau \log m]$,

$$|\mathrm{VOI}_\tau(g) - \mathrm{VOI}_0(g)| = |\mathbb{E}[h(\mu_g^+) \mid \mathcal{I}] - h(\mu)| \le \tau \log m.$$

Consequently,

$$\left|\mathrm{VOI}_0(g) - \tfrac{1}{2}\,\mathrm{tr}(M_\tau(\mu)\Delta C_g)\right| \le \tau \log m + \frac{L_3}{6}\mathbb{E}[\|\delta_g\|_2^3 \mid \mathcal{I}]. \tag{21}$$

The bound separates smoothing error from Taylor remainder. Decreasing τ reduces the first term but can increase the Hessian regularity constant in the second.

A.4 Exact query ranking for two feasible allocations

For $\delta = d^\top \mu$, $\sigma_g = \sqrt{q_g} > 0$, and the standard normal density ϕ and cdf Φ, the hard information value is

$$\mathrm{VOI}_0(g) = \sigma_g \phi(|\delta|/\sigma_g) - |\delta|\Phi(-|\delta|/\sigma_g), \tag{22}$$

with continuous extension zero at $\sigma_g = 0$.

Proof of Proposition 3. Let $\sigma_g = \sqrt{q_g}$. The hard value is $V_0(\mu) = s_b^\top \mu + (\delta)_+$, where $(x)_+ = \max(x, 0)$. The expected change in its linear term is zero. Integrating the positive part of a normal variable gives Eq. (22) for $\sigma_g > 0$. To obtain that expression, use $\mathbb{E}[(\delta + \sigma Z)_+] = \sigma\phi(\delta/\sigma) + \delta\Phi(\delta/\sigma)$ and subtract $(\delta)_+$. The continuous extension at $\sigma = 0$ is zero, and differentiation with respect to $\sigma > 0$ gives $\phi(|\delta|/\sigma) > 0$. Thus exact VOI increases strictly with q_g.

For the two-slate Gibbs law, $\rho = \pi_\tau(s_a; \mu) = (1 + e^{-\delta/\tau})^{-1}$ and

$$M_\tau = \frac{\rho(1-\rho)}{\tau} dd^\top, \qquad a_g = \frac{\rho(1-\rho)}{2\tau c} q_g.$$

The prefactor is positive and does not depend on the query. The two criteria therefore induce the same ordering, including ties. □

The Gaussian positive-part calculation also appears in knowledge-gradient analysis (Frazier et al., 2009). Jointly Gaussian utilities and linear-Gaussian observations satisfy the update assumption. Proposition 3 concerns one-step ranking with exactly two feasible allocations, equal query costs, and Gaussian posterior-mean contrast updates.

A.5 Error in the estimated score

Let $M = M_\tau(\mu)$, $a_g = \mathrm{tr}(M\Delta C_g)/(2c_g)$, and $\widehat{a}_g = \mathrm{tr}(\widehat{M}\widehat{\Delta C}_g)/(2c_g)$. If the Frobenius-norm errors obey $\|\widehat{M} - M\|_F \le \epsilon_M$ and $\|\widehat{\Delta C}_g - \Delta C_g\|_F \le \epsilon_{C,g}$, then

$$|\widehat{a}_g - a_g| \le \frac{\epsilon_M \|\Delta C_g\|_F + \epsilon_{C,g}\|M\|_F + \epsilon_M \epsilon_{C,g}}{2c_g}. \tag{23}$$

Expand $(M + E_M)(\Delta C_g + E_{C,g}) - M\Delta C_g$ and apply the Frobenius trace inequality to the three terms. Symmetry justifies using $\mathrm{tr}(AB)$ in place of $\mathrm{tr}(A^\top B)$. If this bound is at most ϵ for all legal queries and $\widehat{g} \in \arg\max_g \widehat{a}_g$, then

$$a_{g^*} \le \widehat{a}_{g^*} + \epsilon \le \widehat{a}_{\widehat{g}} + \epsilon \le a_{\widehat{g}} + 2\epsilon,$$

where $g^* \in \arg\max_g a_g$. This controls selection loss relative to the ideal trace criterion. The score error includes errors in the moments, virtual integration, and PSD projection.

A.6 LEARNED-SCORE APPROXIMATION OF REOPTIMIZATION VALUE

Fix the current information state, a finite nonempty feasible family $\mathcal{F}$, a temperature $\tau > 0$, and positive query costs. All expectations below are conditional on this state. Q_g is the probability measure represented by the numerical reveal rule: its weights are nonnegative and sum to one. It may be a point mass or a finite weighted collection of virtual responses. The learned update $\hat{\mu}_g^+(y) = \hat{\mu}_\theta(\mathcal{I}, y)$ and projected matrix $\widehat{\Delta C}_g$ are arbitrary model outputs with the stated finite moments. When Q_g is random, the first bound is conditional on its realization.

Exact reoptimization identity for learned means. Let $\bar{s} = \tilde{s}_\tau(\hat{\mu})$ and $z_g(y) = \hat{\mu}_g^+(y) - \hat{\mu}$. The current Gibbs distribution is feasible for the variational problem in Eq. (14) after any update. Its objective at $\hat{\mu} + z_g$ is $V_\tau(\hat{\mu}) + \bar{s}^\top z_g$. Thus

$$f_g(y) = V_\tau(\hat{\mu}_g^+(y)) - V_\tau(\hat{\mu}) - \bar{s}^\top z_g(y)$$

is the gain from reoptimizing the smooth allocation distribution relative to retaining its current weights. By Eq. (19), $f_g(y) = \tau \mathrm{KL}(\pi_{\hat{\mu}} \| \pi_{\hat{\mu}_g^+(y)}) \geq 0$ pointwise, proving Eq. (10). The identity holds pointwise and permits nonzero mean drift. The *uncentered* predicted smooth-value change equals

$$\mathbb{E}_{Q_g}[V_\tau(\hat{\mu}_g^+) - V_\tau(\hat{\mu})] = c_g J_g + \bar{s}^\top \mathbb{E}_{Q_g} z_g.$$

This distinguishes reoptimization value from drift in the predicted value of an unchanged allocation. Under coherent posterior updates and the actual response law the drift vanishes, recovering the ideal smooth VOI identity. For general learned updates, J_g measures reoptimization value under the virtual-response measure Q_g.

Proof of the learned-score bound. Let P be the orthogonal projector onto $\mathcal{D}_\mathcal{F}$ and $N = I - P$. For any $v \in \ker P$, $s^\top v$ is constant across $s \in \mathcal{F}$. Consequently $V_\tau(x + v) = V_\tau(x) + s_0^\top v$ for any $s_0 \in \mathcal{F}$, while Gibbs weights and the Hessian do not change. Linear terms cancel from the Bregman difference, so $f_g(y) = D_{V_\tau}(\hat{\mu} + Pz_g(y), \hat{\mu})$. Taylor expansion only along the projected update gives

$$f_g(y) = \tfrac{1}{2} z_g(y)^\top \widehat{M} z_g(y) + r_g(y), \qquad |r_g(y)| \leq \tfrac{L_3}{6} \|P z_g(y)\|_2^3.$$

The finite-family bound in Eq. (18) permits $L_3 = D_\mathcal{F}^3/\tau^2$. Taking the Q_g expectation, subtracting $c_g \widehat{a}_g$, and using $\widehat{M} = P\widehat{M}P$ yields

$$c_g(J_g - \widehat{a}_g) = \tfrac{1}{2} \operatorname{tr}\Big(\widehat{M} P (S_g - \widehat{\Delta C}_g) P\Big) + \mathbb{E}_{Q_g} r_g.$$

The Frobenius trace inequality proves Eq. (11). Here $S_g = \mathbb{E}_{Q_g}[z_g z_g^\top]$ is the *uncentered second moment*, including the contribution of mean drift. The bound depends on its disagreement with the implemented contraction after projection onto allocation contrasts.

Proof of the true-decision-value bound. Let $\mathbb{P}_g$ be the actual conditional response law. Define true means $\mu = \mathbb{E}[u^\star \mid \mathcal{I}]$, $\mu_g^+(y) = \mathbb{E}[u^\star \mid \mathcal{I}, Y_g = y]$ from one underlying joint distribution. Define

$$\xi_g = \left| \mathbb{E}_{Q_g} f_g(Y_g) - \mathbb{E}_{\mathbb{P}_g} f_g(Y_g) \right|, \quad e_0 = \|P(\hat{\mu} - \mu)\|_2, \quad e_g = \mathbb{E}_{\mathbb{P}_g} \|P(\hat{\mu}_g^+ - \mu_g^+)\|_2.$$

Assume these quantities are finite. Set $h(v) = V_\tau(v) - \bar{s}^\top v$. Its gradient is $\tilde{s}_\tau(v) - \bar{s}$, a difference of two points in $\mathrm{conv}(\mathcal{F})$, so it lies in $\mathcal{D}_\mathcal{F}$ with norm at most $D_\mathcal{F}$. Hence

$$|h(v) - h(w)| \leq D_\mathcal{F} \|P(v - w)\|_2.$$

Since the *true* posterior mean has zero expected update,

$$\mathrm{VOI}_\tau(g) = \mathbb{E}_{\mathbb{P}_g}[h(\mu_g^+) - h(\mu)].$$

Comparing this expression to $\mathbb{E}_{\mathbb{P}_g}[h(\hat{\mu}_g^+) - h(\hat{\mu})]$ incurs at most $D_\mathcal{F}(e_0 + e_g)$. Changing the response measure to Q_g incurs ξ_g. Finally, the smoothing bound in Appendix A.3 and Eq. (11) give Eq. (12). The true conditional means define the decision-value target, and the errors e_0, e_g quantify its difference from the learned predictions.

For $b_g = \mathrm{VOI}_0(g)/c_g$, $g^\star \in \arg\max_g b_g$, and $\widehat{g} \in \arg\max_g \widehat{a}_g$,

$$b_{g^\star} - b_{\widehat{g}} \leq B_{g^\star} + \widehat{a}_{g^\star} - \widehat{a}_{\widehat{g}} + B_{\widehat{g}} \leq B_{g^\star} + B_{\widehat{g}}.$$

In particular the loss is at most $2B$ when all $B_g \leq B$. A score gap larger than the sum of the relevant error bounds certifies that the preferred query retains its ordering for true one-step VOI per cost. The ordering certificate is conditional on the stated one-step error bounds.

PSD projection and update consistency. The theorem uses the projected matrix from Eq. (4). Projection also satisfies the following full-matrix Frobenius bound. Let A be the symmetric unprojected update, $A_+ = \Pi_{\succeq 0}(A)$, and let $S \succeq 0$ (in particular, $S = S_g$). Writing $N = A - A_+ \preceq 0$, with $\operatorname{tr}(A_+ N) = 0$, gives

$$\|A - S\|_F^2 - \|A_+ - S\|_F^2 = \|N\|_F^2 - 2\operatorname{tr}(SN) \geq 0.$$

Clipping therefore reduces or preserves the full-matrix Frobenius distance to S. The score bound separately uses the contrast-restricted mismatch of the projected matrix; query ordering is governed by that mismatch and the other error terms.

From a query to the learned terminal decision. Let $\hat{s}_0 \in \arg\max_{s \in \mathcal{F}} \hat{\mu}^\top s$ and $\hat{s}_g^+(y) \in \arg\max_{s \in \mathcal{F}} \hat{\mu}_g^+(y)^\top s$. Under the actual law, define the one-reveal gain of the learned terminal rule by

$$A_g = \mathbb{E}_{\mathbb{P}_g}[(\mu_g^+)^\top \hat{s}_g^+] - \mu^\top \hat{s}_0.$$

If $r_0 = V_0(\mu) - \mu^\top \hat{s}_0$ and $r_g^+(y) = V_0(\mu_g^+(y)) - (\mu_g^+(y))^\top \hat{s}_g^+(y)$, then $A_g = \mathrm{VOI}_0(g) + r_0 - \mathbb{E}_{\mathbb{P}_g} r_g^+$. Optimality for the predicted utilities gives $0 \leq r_0 \leq D_{\mathcal{F}} e_0$ and $0 \leq \mathbb{E} r_g^+ \leq D_{\mathcal{F}} e_g$. Therefore

$$\left| \frac{A_g}{c_g} - \widehat{a}_g \right| \leq B_g + \frac{D_{\mathcal{F}}(e_0 + e_g)}{c_g}.$$

The extension bounds the learned terminal decision using the same prediction and integration errors as the query-value analysis.

B PUBLIC-DATA AGGREGATION AND ALLOCATION SUITE

B.1 ALLOCATION INSTANCES AND DATA SPLITS

The allocation suite combines the public releases in Table 10 (Centers for Disease Control and Prevention, 2023a;b; Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry/Geospatial Research, Analysis, and Services Program, 2024; Centers for Disease Control and Prevention, 2025; National Center for Health Statistics, 2024; U.S. Census Bureau, 2024; 2025; 2026). Each allocation instance specifies evidence groups, an initial mask, a utility objective, and resource and coverage constraints. County subsets lie within a Census region and may span several states. Construction uses seed 20260424.

Table 4: Data splits. Rows count county–policy pairs; county and state counts are unique within each split. The total state count is taken over the union of splits.

Split	Contexts	Episodes	Rows	Counties	States/equivalents
Train	48	144	2160	478	27
Development	16	48	720	158	18
Test	32	96	1440	324	17
Total	96	288	4320	960	50

A base context comprises 15 counties and a reference week. Its three policy objectives yield three episodes, giving 288 episodes, 1440 county–context slots, and 4320 county–policy rows overall. Counties may recur within their assigned split and are disjoint across splits. Reference weeks span 2020-W50 through 2022-W06. All 96 held-out instances in Table 4 are evaluated.

B.2 GEOGRAPHIC COVERAGE

Table 5: Cross-split geographic overlap. Counties are disjoint across splits; shared states are listed.

Split pair	Shared counties	Number of states	State identifiers
Train–development	0	8	LA, MO, ND, NJ, NM, OK, OR, RI
Train–test	0	4	ND, NJ, PA, WI
Development–test	0	2	ND, NJ

Each Census region contributes 12 training, four development, and eight test contexts. Table 6 groups contexts by population and rurality: the small-population threshold is 180,000, and the rurality threshold is 0.60 on the score derived from NCHS codes 1–6.

Table 6: Context counts by rurality and population size.

Split	Rural, small	Rural, non-small	Nonrural, small	Nonrural, non-small
Train	12	0	12	24
Development	3	0	4	9
Test	8	0	8	16
Total	23	0	24	49

B.3 Queryable evidence and initial masks

Table 7: The four evidence blocks available to acquisition, containing 14 fields. Outcome labels are reserved for supervision and evaluation.

Block	Fields	Names
Surveillance	5	`cases_recent_per100k`, `cases_lag_1`, `cases_lag_2`, `cases_trend_2w`, `cases_trend_delay_1w`
Severe vulnerability	4	`deaths_recent_per100k`, `deaths_lag_1`, `pct_age_65_plus`, `chronic_rate_pct`
Equity/access	3	`svi_overall`, `access_gap`, `vaccination_gap_pct`
Spatial spillover	2	`neighbor_case_pressure`, `neighbor_count`

We compute the two surveillance trend features as follows.

$$\text{cases_trend_2w} = \text{cases_recent_per100k} - \tfrac{1}{2}(\text{cases_lag_1} + \text{cases_lag_2}), \tag{24}$$

$$\text{cases_trend_delay_1w} = \text{cases_lag_1} - \text{cases_lag_2}. \tag{25}$$

The initial mask exposes one block per county. Mask seed 13, the context identifier, and county FIPS determine an ordering that cycles through the four blocks, giving counts 4/4/4/3. Each context therefore starts with 15 visible and 45 hidden candidate–block slots. Its three objectives share the mask. Queries reveal previously hidden slots without replacement, leaving $15 + K$ observed slots after acquisition.

Inference uses observed evidence, its mask, and public constraint metadata. Selection costs and quota-membership flags remain visible even when their underlying evidence blocks are hidden, so a flag may disclose partial information about an attribute. Target-normalization columns, future outcomes, oracle utilities, and binding flags are available only to the evaluator.

B.4 Policy utility and feasible allocations

For public costs a, resource budget β_p, quota-membership vector q_p, and quota h_p, the feasible family is

$$\mathcal{F}_p = \{s \in \{0,1\}^{15} : \mathbf{1}^\top s = 4,\ a^\top s \le \beta_p,\ q_p^\top s \ge h_p\}. \tag{26}$$

Incidence and equity require at least one and two high-vulnerability counties, respectively; severe-outcome allocation requires at least two counties with high older-adult shares. Quota membership is public even when the corresponding evidence block is masked. These constraints remain fixed throughout acquisition.

For a context, let z_{ij} denote the value of utility attribute j for county i, and let $\bar{z}_{ij}$ denote its within-context min–max normalization. For nonconstant attributes,

$$\bar{z}_{ij} = \frac{z_{ij} - \min_{i'} z_{i'j}}{\max_{i'} z_{i'j} - \min_{i'} z_{i'j}}, \qquad u^\star_{i,p} = \sum_j w_{pj} \bar{z}_{ij}.$$

Table 8 gives the weights. Constant public utility components, including vaccination-gap and rurality columns, are assigned zero. This within-context normalization defines the evaluation target; model inputs use the scaling described below.

The final evaluator enumerates the size-four combinations, applies the supplied budget and quota filters, and maximizes summed utility over the remaining slates. It uses predicted utility for a method's decision and private target utility for the oracle. Oracle-binding status is an evaluator-side slice label based on whether constrained and unconstrained oracle slates differ.

Table 8: Utility weights by objective. Columns sum to one, and zero weights omit an attribute. Future cases and deaths are evaluator-only outcome labels at test time.

Attribute	Incidence	Severe outcomes	Equity
`future_cases_2w`	0.55	0.20	0.35
`future_deaths_2w`	0.10	0.45	0.10
`vaccination_gap_pct`	0.20	0.00	0.10
`neighbor_case_pressure`	0.10	0.00	0.00
`svi_overall`	0.05	0.05	0.25
`pct_age_65_plus`	0.00	0.15	0.00
`chronic_rate_pct`	0.00	0.10	0.00
`access_gap`	0.00	0.05	0.15
`rurality`	0.00	0.00	0.05

Table 9: Coverage constraints; all policies select four counties. Cost vectors and resource budgets are supplied per episode.

Policy	Public membership flag	Minimum count
Incidence	`high_svi`	1
Severe outcomes	`high_age65`	2
Equity	`high_svi`	2

B.5 Public sources and preprocessing

The harmonized panel includes the District of Columbia, excludes Puerto Rico and territories, and omits Connecticut because its county-equivalent identifiers could not be matched consistently across the joined releases.

Preprocessing standardizes FIPS, joins static attributes, and retains records with valid regions and positive populations. Public features are median-imputed in the master table; lagged case and death values retain missingness indicators. Model inputs use `log1p` for nonnegative skewed features, `asinh` for signed trends, and identity transforms for bounded attributes. The subsequent median/IQR scaling is fitted on training data.

Evaluation period. Dynamic reference weeks span 2020-W50 to 2022-W06 and are paired with the later static releases in Table 10. The resulting evaluation compares retrospective allocations on a fixed data assembly. Future two-week outcomes provide supervision and evaluation targets; acquisition accesses only the evidence in Table 7.

Table 10: Public sources and release versions used in the allocation suite.

Source	Data product and role
CDC county COVID-19 records	Historical case/death records covering 2020–May 2023; weekly surveillance and future-outcome targets.
CDC county vaccination	Historical county vaccination records transformed to weekly vaccination-gap features.
CDC/ATSDR SVI	2022 Social Vulnerability Index; county vulnerability and access-related attributes.
CDC PLACES	County release 2025; chronic-condition features.
NCHS urban–rural codes	2023 county classification; rurality features and stratification.
Census population totals	2020–2025 product, registry file `co-est2025-pop.xlsx`; population denominators.
Census population characteristics	2020–2024 product, registry file `cc-est2024-agesex-all.csv`; older-adult shares.
Census county adjacency	2024 adjacency product; neighboring-county evidence.

C Implementation and evaluation details

C.1 Masked utility model and acquisition

The frozen `e5-mistral-7b-instruct` encoder represents policy and source descriptions. Trainable projections combine these representations with the observed candidate–block values, explicit masks, missing-block embeddings, and public metadata. The utility head predicts $\hat{\mu}_\theta$ and $C_\theta = D_\theta + L_\theta L_\theta^\top$, with $L_\theta \in \mathbb{R}^{15\times 4}$ and $D_{\theta,ii} = \mathrm{softplus}(\eta_i) + 10^{-6}$. Training samples partial masks on the training contexts and supervises hidden evidence and policy-conditioned utility; development contexts support model selection.

Virtual covariance evaluation. A hypothetical response is inserted into a temporary copy of the observed state and evaluated by the same masked model. The operator $\widehat{\mathbb{E}}_g$ in Eq. (4) numerically approximates the expected updated covariance. Evaluating the model at the predicted response mean is a one-point quadrature; sample-based integration instead averages hypothetical-response evaluations. Both operate on predicted responses, so only the selected real query consumes evidence budget. We compute the virtual covariance reduction, symmetrize it, and project it onto the positive-semidefinite cone.

Temperature is computed from the current predicted utility vector and held fixed across virtual comparisons and differentiation, then recomputed after a real query. The same feasible family is used for sensitivity and the final discrete allocation.

Algorithm 1 MOVE for K unit-cost queries

Require: Initial state $\mathcal{I}_0$, feasible family $\mathcal{F}$, trained model θ, $K \le |\mathcal{G}(\mathcal{I}_0)|$

1: **for** $k = 0, \dots, K-1$ **do**
2: Predict $\hat{\mu}_\theta(\mathcal{I}_k), C_\theta(\mathcal{I}_k)$; set the local temperature τ.
3: Form $\widehat{M} = M_\tau(\hat{\mu}_\theta(\mathcal{I}_k))$ using Eq. (3).
4: **for** $g \in \mathcal{G}(\mathcal{I}_k)$ **do**
5: Estimate $\widehat{\Delta C}_g$ by Eq. (4); score $\widehat{a}_g$ by Eq. (5).
6: **end for**
7: Execute $g_k \in \arg\max_{g\in\mathcal{G}(\mathcal{I}_k)} \widehat{a}_g$; update $\mathcal{I}_{k+1}$ with its real response.
8: **end for**
9: **return** the allocation in Eq. (6) by exact enumeration.

C.2 Baselines and shared evaluation protocol

Acquisition methods share the candidate–block interface, initial information, query budget, and exact final solver. Ridge and DFL use only initial evidence. Table 2 compares complete systems with their respective predictors; Table 3 compares MOVE-family variants.

ReAct and Toolformer use Gemma-3-4B-IT with deterministic generation, at most 96 new tokens per response, and the shared query budget K. Query scores or rankings are parsed from JSON, score

Table 11: Predictor and acquisition roles in the evaluation.

Method	Implementation role
Ridge plug-in	Supervised utility prediction from the initial evidence.
DFL plug-in	Decision-focused utility prediction from the initial evidence.
Uncertainty	Uncertainty-based acquisition pipeline, distinct from the MOVE-family variance-only variant.
EDDI	Generative information acquisition over grouped evidence.
Dynamic MI	Conditional-information selection of candidate–block queries.
ACO	Acquisition-conditioned nongreedy policy over the block interface.
ReAct	Language-model observe/query routing over legal evidence blocks.
Toolformer	Language-model tool scoring over legal evidence blocks.

lines, or ranked mentions. Invalid outputs trigger a deterministic fallback based on public descriptors. Both use the initial evidence in Appendix B.3. The names identify task-specific prompting and routing implementations: the Toolformer row uses tool scoring without the original self-supervised tool-training stage. MOVE uses a separate frozen evidence encoder.

The component table reports descriptive point estimates for the evaluated MOVE variants; equivalence of all other training factors is unspecified. Paired confidence intervals are reported for the system comparisons in Table 15.

C.3 Metrics and statistical units

For episode t, selected support S_t, and oracle support $S_t^\star$,

$$\mathrm{UR}_t = \frac{(u_t^\star)^\top \hat{s}_t}{(u_t^\star)^\top s_t^\star}, \qquad \mathrm{J}_t = \frac{|S_t \cap S_t^\star|}{|S_t \cup S_t^\star|}.$$

Utility ratio requires a positive oracle denominator. For base context c and its three policy objectives $\mathcal{P}$, policy response is

$$\mathrm{PR}_c = \frac{2}{|\mathcal{P}|(|\mathcal{P}|-1)} \sum_{p<p'} \left(1 - \frac{|S_{c,p} \cap S_{c,p'}|}{|S_{c,p} \cup S_{c,p'}|}\right).$$

Utility and Jaccard are averaged over episodes; policy response is averaged over contexts. Equal-size groups of three objectives give the same overall mean under equal-weight context averaging. Slice averages use only eligible episodes.

Paired intervals use 500 base-context bootstrap resamples, preserving method pairing and all three objectives of each context. The binding comparison contains 31 episodes in 21 contexts. Table 15 gives marginal 95% intervals for individual contrasts under this subset-resampling scheme. The uncertainty analysis holds the trained systems fixed and treats base contexts as the resampling units.

D Additional results

Tables 2 and 3 report aggregate and component results. The tables below provide the full objective and constraint breakdowns and the corresponding uncertainty intervals. Figures 2 and 3 show the same utility means and query-source proportions.

D.1 Component values

Table 3 reports the complete MOVE-family component results, including policy response. Its variance-only variant shares the MOVE family; Uncertainty is the separate system in Table 2.

D.2 Policy objectives

Table 12: Complete objective-wise test results. Overall utility and Jaccard repeat Table 2 to make the breakdown self-contained.

Method	Overall	Incidence	Severe	Equity	Jaccard
Ridge plug-in	0.791	0.709	0.807	0.858	0.352
DFL plug-in	0.772	0.720	0.778	0.818	0.346
Uncertainty	0.814	0.744	0.833	0.865	0.384
EDDI	0.797	0.717	0.816	0.860	0.374
Dynamic MI	0.809	0.728	0.829	0.871	0.379
ACO	0.797	0.724	0.810	0.856	0.344
ReAct	0.811	0.713	0.850	0.869	0.363
Toolformer	0.795	0.692	0.833	0.859	0.361
MOVE	**0.878**	**0.850**	**0.866**	**0.919**	**0.486**

The strongest external baseline differs by objective: uncertainty acquisition for incidence, ReAct for severe outcomes, and dynamic MI for equity. The three fixed objectives each contribute 32 test episodes.

D.3 Oracle-binding status

Table 13: Utility and overlap on oracle-binding and nonbinding episodes. Both groups use the same constrained final solver.

	Utility ratio		Jaccard	
Method	Binding	Nonbinding	Binding	Nonbinding
Ridge plug-in	0.798	0.788	0.382	0.337
DFL plug-in	0.771	0.772	0.359	0.339
Uncertainty	0.816	0.813	0.400	0.377
EDDI	0.802	0.795	0.391	0.366
Dynamic MI	0.799	0.814	0.400	0.369
ACO	0.794	0.798	0.355	0.339
ReAct	0.821	0.806	0.389	0.351
Toolformer	0.808	0.788	0.399	0.342
MOVE	**0.866**	**0.884**	**0.472**	**0.493**

The evaluator assigns binding status when the constrained oracle selection differs from the unconstrained top-four selection. The two groups contain 31 and 65 episodes. MOVE attains higher utility and overlap in both groups, with larger gains in the nonbinding group. Acquisition uses the same information protocol in both groups.

D.4 Compact objective and slice summary

Table 14: Utility across policy objectives and oracle-binding episodes. The best comparator is selected separately within each column; no single baseline achieves all four comparator values. Full method-wise results appear in Tables 12 and 13.

	Incidence	Severe outcomes	Equity	Binding
Best same-budget baseline	0.744	0.850	0.871	0.821
MOVE	**0.850**	**0.866**	**0.919**	**0.866**
Absolute difference	+0.106	+0.016	+0.048	+0.045

Each comparator value is the maximum over the corresponding column in Tables 12 and 13. Figure 2 shows all six same-budget external methods.

D.5 Reported uncertainty intervals

Table 15: Marginal 95% subset-bootstrap intervals for system means and paired contrasts, using 500 resamples of 32 base contexts. The binding contrast uses 31 episodes in 21 contexts.

Method or contrast	Mean/difference	95% interval
Uncertainty utility	0.814	[0.788, 0.841]
ReAct utility	0.811	[0.784, 0.835]
Toolformer utility	0.795	[0.777, 0.818]
MOVE utility	0.878	[0.852, 0.902]
MOVE minus uncertainty: utility	+0.064	[0.033, 0.096]
MOVE minus uncertainty: Jaccard	+0.102	[0.045, 0.164]
MOVE minus ReAct: binding utility	+0.045	[0.011, 0.074]

Each interval applies to its listed mean or paired contrast. Component ablations in Table 3 are reported as point estimates.

D.6 Source usage

Figure 3 in the main text visualizes the source mix across the three policy cards.

The source proportions describe how queries are distributed under each objective. Removing semantic descriptors yields slightly higher policy response (0.808 versus 0.801) and lower utility (0.852 versus 0.878). These two metrics separate changes in the selected sets from their allocation utility; the proportions alone do not identify the utility contribution of each source.