TARGETED ACTIVE LEARNING FOR PREFERENCE-BASED TREATMENT EFFECTS ON MULTIVARIATE OUTCOMES

Anonymous authors
Paper under double-blind review

ABSTRACT

Treatment efficacy is traditionally demonstrated on the basis of a single primary outcome. However, clinical decision-making usually requires consideration of multiple outcomes, balancing expected benefits against potential risks. The relative value assigned to these outcomes varies substantially from one patient to another. Given a *preference rule* over outcome profiles, treatment effects and optimal policies can be defined and estimated. Such a rule is rarely available in practice: it must itself be estimated from pairwise comparisons of outcome profiles, which are costly to collect from clinical experts. We propose an active learning framework that selects which comparisons to query. Standard criteria maximize the information gained on the preference rule itself. We instead target the quantities of interest, and select the query that most reduces uncertainty on the treatment effect and on the optimal policy induced by the learned rule. Under a Gaussian process model of the preference rule, we derive a closed-form approximation of this criterion. On semi-synthetic data built from a Parkinson's disease cohort with 13 clinical outcomes, our criterion achieves lower treatment effect estimation error and lower policy regret than existing criteria at equal query budget.

1 INTRODUCTION

A large literature (Rubin, 1974; Angrist & Imbens, 1995; Nie & Wager, 2021) answers whether a treatment increases or decreases one quantity of interest, such as the outcome of a disease. However many clinical decisions involve several outcomes at once: a treatment may improve one symptom while worsening another, or bring a clinical benefit at the cost of a side effect. While one can estimate an effect for each outcome on its own, the problem becomes more difficult when several separate and competing effects have to be turned into a single treatment decision. Standard practice avoids this by reducing the outcome to a scalar before the causal question is even posed, either by picking a single primary endpoint or by combining outcomes into a composite score (Freemantle et al., 2003). Such a score settles the trade-offs in advance for the whole population, while the relative importance of each outcome varies across patients and clinical contexts. Parkinson's disease illustrates this well, by combining motor symptoms, such as tremor, rigidity and akinesia, with non-motor symptoms such as sleep disorders, depression, cognitive decline and dysautonomia (Kalia & Lang, 2015). It is usually treated with dopamine replacement therapy, which improves motor symptoms but comes with its own costs: motor complications such as fluctuations and dyskinesia, non-motor complications such as impulse control disorders and hallucinations, and a worsening of existing non-motor symptoms such as dysautonomia. Treatments act differently on each of these dimensions, and no standard scale says how to weigh them against each other. Clinicians naturally balance them when selecting a therapy but classical treatment effect estimation methods don't. Parkinson's disease is used as a running example throughout the paper.

Instead of measuring how much a treatment moves some fixed score, Even & Josse (2025) ask which complete outcome profile a clinician would judge to be better overall. They assume access to a known preference rule, which quantifies for any two outcome profiles how likely it is that the first would be judged preferable to the second. The preference rule captures both that a single clinician is not perfectly consistent and that different clinicians may disagree. Parnas et al. (2026) then define *conditional preference treatment effect* (CPTE) as the probability that, among patients sharing the

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.

same characteristics, the outcome experienced under treatment would be preferred to the outcome experienced under control. This yields a treatment effect for outcomes with several dimensions, or with categories that can be ranked but not subtracted.

1.1 Contributions and outline of the paper

The preference rule is known in some cases, for instance for hierarchical outcomes (Pocock et al., 2012), but in general the trade-offs between symptoms, side effects and quality of life have to be elicited. Clinicians or other experts are shown pairs of outcome profiles and asked which one they prefer, and the rule is learned from these labeled pairs. Each comparison takes clinician time, so only a small and somewhat noisy set of judgments can be collected. Two difficulties follow. First, the uncertainty in the learned preference rule propagates to the CPTE, which becomes a random quantity rather than a fixed number. Second, the choice of which comparisons to ask becomes crucial. Existing selection methods target either the preference model itself or the answers to future comparisons, but not the causal estimand, which integrates preferences over the distributions of potential outcomes. In this paper, we address both difficulties with, to our knowledge, the first active learning method for preference elicitation that targets a causal estimand. We treat the CPTE as a random quantity under the posterior of the preference model, and we select the comparisons that are expected to be most informative about it. Clinician time is then spent on the trade-offs that decide which treatment is better for the target patients, rather than on learning the preference rule everywhere. After presenting the background (Section 2) and formalizing the active learning problem (Section 3), we make the following contributions:

- We formulate the CPTE induced by a probabilistic preference model as a posterior random variable Ψ_x, which captures the uncertainty that limited preference data leaves in the estimand (Section 4). We then linearise this estimand around the posterior mean utility to obtain a tractable, Gaussian characterisation of its law, with an error bound computable at every round from the current posterior (Section 4.1).
- We introduce the *CPTE – Expected Information Gain* (CPTE-EIG), an acquisition criterion that scores candidate comparisons by their expected information gain about the causal estimand in a target population. Under the linearised CPTE, we derive a computationally efficient expression for it, in closed form up to a one-dimensional quadrature (Section 4.2).
- We evaluate the framework in controlled experiments and on a semi-synthetic Parkinson's disease cohort with thirteen outcomes, and show that at equal query budget it reduces both CPTE estimation error and the regret of the resulting treatment policy (Section 5).

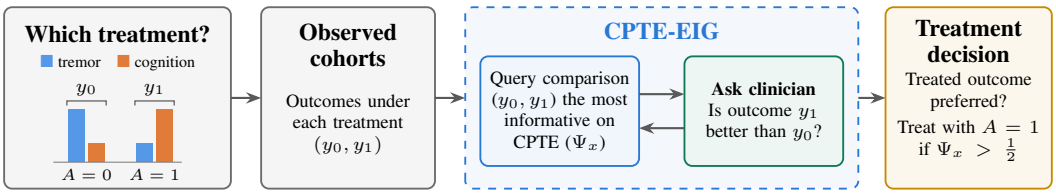


Figure 1: Overall workflow presented in the paper: CPTE-EIG asks the clinician only the comparisons that matter for the treatment decision.

1.2 Related works

Comparing treatments on several outcomes. Weighing benefits and risks across several outcomes at once is a recurrent problem in clinical research and drug regulation. Multi-criteria decision analysis offers a structured way to combine several outcomes into a single recommendation (Gongora-Salazar et al., 2023). Composite endpoints do this by combining outcomes into one weighted sum (the composite outcome), so everything sits on the same scale (Freemantle et al., 2003). Composite outcomes can also be learned on data in a covariate-dependent way (Butler et al., 2018). Generalized pairwise comparisons (GPC) (Buyse, 2010) and the win ratio (Pocock et al., 2012) avoid this by ranking outcomes by priority and comparing them hierarchically, one

pair of patients at a time, moving to a lower-priority outcome only when higher-priority ones are tied. Either way, a clinician or any specialist, has to fix the weights or the ranking in advance, and the result is usually a single population-level number rather than a causal effect that varies with a patient's characteristics. Even & Josse (2025) show this matters for the win ratio: the historical estimand can reverse treatment recommendations under heterogeneous populations, and they propose an individual-level fix, though the ranking itself stays fixed. Petit et al. (2025) and Parnas et al. (2026) push this individual-level view from estimation to decision-making, defining an optimal individualized treatment rule within the same prioritized-outcome GPC framework.

Learning preferences from comparisons. A standard way to model pairwise preferences is to assume each option has a hidden value, and that people prefer whichever option is worth more (Bradley & Terry, 1952). Chu & Ghahramani (2005a) turned this into a Bayesian model: a Gaussian process describes the hidden value function, and each comparison updates a posterior over it, giving not just which item is preferred but how confident that ordering is. Later work extends this in several directions: handling clinicians who may disagree with each other (Houlsby et al., 2012), noisy or self-contradictory judgments (Chau et al., 2022), and answers that depend on the context in which a comparison is asked (Bengs et al., 2022). In every case, the object being learned is the preference function itself.

Choosing which comparisons to ask. Since each comparison takes clinician time, the comparisons should be chosen carefully rather than at random. Much of the work on active preference learning aims at the best option. Brochu et al. (2007) ask as few questions as possible to find the item a person values most, and preferential Bayesian optimisation (González et al., 2017) searches for the maximiser of a latent function from pairwise comparisons. In these works, the goal is the best option or the preference function itself. Our method belongs to Bayesian experimental design (Lindley, 1956; Chaloner & Verdinelli, 1995; Rainforth et al., 2024), which selects the query with the largest expected information gain about a quantity of interest. Existing criteria differ in the quantity they target. Bayesian active learning by disagreement (BALD) (Houlsby et al., 2011), a widely used criterion for preference learning, targets the model parameters. Expected predictive information gain (EPIG) (Smith et al., 2023) instead targets the predictions on inputs drawn from a chosen target distribution. For treatment effects, Causal-BALD (Jesson et al., 2021) and Causal-EPIG (Gao et al., 2026) choose the individuals whose outcomes are measured so as to reduce uncertainty respectively in the conditional average treatment effect estimator. Our setting differs from both in two ways. First, we collect noisy pairwise judgments rather than outcome measurements, and use them to learn a preference function rather than an outcome model. Second, we target neither the preference function nor a treatment effect on a scalar outcome, but a treatment effect that is defined only once the preference function is combined with the distribution of the multivariate potential outcomes.

2 BACKGROUND

Throughout the paper we use two different sources of data answering different questions: *(i)* a cohort of patients tells us how treatments change outcomes, while *(ii)* a set of pairwise comparisons, in which clinicians state which of two outcome profiles they prefer, tells us how outcomes should be compared. The first is fixed and available in advance, and is subject to the assumptions we describe next in Section 2.1. The second is collected sequentially by our method, is indexed independently of the cohort, and is introduced in Section 3.

2.1 NOTATIONS AND ASSUMPTIONS

Each participant in the cohort is described by covariates $X \in \mathcal{X}$, a binary treatment $A \in \{0, 1\}$ and an outcome $Y \in \mathcal{Y}$, with potential outcomes $Y(0), Y(1)$. The outcome is multivariate, $\mathcal{Y} \subseteq \mathbb{R}^p$, each coordinate being a bounded ordinal score for one clinical dimension. These coordinates are not required to be on a common scale. Table 1 specifies the quantities we introduce for Parkinson's disease as a concrete example. We work under the following standard causal assumptions (Rosenbaum & Rubin, 1983; Rubin, 1980).

Table 1: The quantities of the paper instantiated for Parkinson's disease (PD), chosen in accordance with specialists. MDS-UPDRS (Goetz et al., 2008) is the standard clinical scale for rating PD.

Quantity	Parkinson's disease instantiation
X	Full covariate. Includes comorbidities, PD subtype/staging, cognitive and psychiatric scales, anxiety, fatigue, biomarkers/genetics, ethnicity, race, education, family history, BMI.
$\tilde{X} = \gamma(X)$	sex, age group (<50 to $80+$, 5-year bins), disease duration
$Y \in \mathcal{Y} \subseteq \mathbb{R}^{13}$	13 MDS-UPDRS items. *Non-motor* (7): cognition, hallucinations, mood, apathy, dopamine dysregulation, sleep, orthostatic symptoms. *Motor* (4): self-care, tremor, walking/balance, freezing. *Motor complications* (2): dyskinesia and fluctuation impact
A	0 (levodopa) or 1 (dopamine agonists)

Assumption 1 (Causal Assumptions) *(Consistency & No interference)* $Y = Y(A)$*; (Unconfoundedness)* $\big(Y(0), Y(1)\big) \perp A \mid X$*; (Positivity)* $0 < \mathbb{P}(A = 1 \mid X = x) < 1$ *for all* x*.*

Under Assumption 1, the conditional laws of $Y(0)$ and $Y(1)$ given $X = x$ are identifiable from observed data (Hernan & Robins, 2025). We write $\hat{\mu}_a(\cdot \mid x)$, $a \in \{0, 1\}$ for their estimates. Let

$$\nu_x(\mathrm{d}z) = \hat{\mu}_1(\mathrm{d}y_1 \mid x) \otimes \hat{\mu}_0(\mathrm{d}y_0 \mid x)\,, \qquad z = (y_1, y_0)\,, \tag{1}$$

be the joint law of an independent treated–control pair sharing same covariates x. Each $\hat{\mu}_a$ is a joint distribution over the full outcome vector: replacing it with a product of coordinate marginals would discard within-arm dependence between clinical dimensions and change the law of z.

2.2 Comparing outcome profiles

Informing about how two outcome profiles y_0 and y_1 compare is a clinical judgment, elicited directly from clinicians, who are shown the two profiles without the treatment that produced them. A clinical pairwise comparison may require context knowledge about the patient for whom the comparison is made, such as a set of features. Showing a clinician the full covariate x of a patient together with the compared outcomes is neither feasible nor right in principle, since some components of x affect how well a treatment works without affecting how outcomes should be traded off against each other. We therefore elicit preferences using only a fixed, clinically justified summary of the patient, keeping x for identification alone. Let $\gamma : \mathcal{X} \to \tilde{\mathcal{X}}$ be a fixed, known map, and write $\tilde{x} = \gamma(x) \in \tilde{\mathcal{X}}$ for the resulting *query covariate*. Table 1 gives γ for Parkinson's disease as a concrete example: tolerance for a side effect legitimately changes with age, whereas ethnicity may help predict how well a treatment works, and so remains in x, but has no place in deciding whether tremor is preferable to cognitive impairment.

Following Parnas et al. (2026), we summarise how a clinician ranks two outcome profiles by a *preference rule* w, and extend it to query covariates. In Parnas et al. (2026), w accounts both for a single clinician not being deterministic and for disagreement between clinicians. In this work, we model a single rule describing an average clinician, and do not model individual raters separately.

Definition 1 (Preference rule) *The* preference rule w *is defined as the probability taken over clinicians' judgments that* y_1 *is preferred to* y_0 *when both are shown at query covariate* $\tilde{x}$*: for* $y_1, y_0 \in \mathcal{Y}$ *and* $\tilde{x} \in \tilde{\mathcal{X}}$*,* $w(y_1 \mid y_0, \tilde{x}) = \mathbb{P}(y_1 \succ y_0 \mid \tilde{x})$*. We assume the no ties assumption:* $w(y_1 \mid y_0, \tilde{x}) + w(y_0 \mid y_1, \tilde{x}) = 1$*, for every* $y_0, y_1, \tilde{x}$*.*

2.3 The conditional preference treatment effect

With w fixed, we can now state the causal target of the paper.

Definition 2 (CPTE and Ψ_x) *For a fixed rule* w*, the conditional preference treatment effect (CPTE) (Parnas et al., 2026) at covariate* $x \in \mathcal{X}$ *is:*

$$\mathrm{CPTE}_w(x) = \mathbb{E}\big[\, w\big(Y_i(1) \mid Y_j(0), \gamma(x)\big) \;\big|\; X_i = X_j = x \,\big], \tag{2}$$

with $i \neq j$. The expectation is over two independent copies of the potential outcomes at x. Under Assumption 1, the CPTE is identified from observed data as $\mathbb{E}[w(Y_i \mid Y_j, \gamma(x)) \mid X_i = X_j = x,\ A_i = 1,\ A_j = 0]$ (Parnas et al., 2026), or equivalently as the integral of $w(\cdot \mid \cdot, \gamma(x))$ against ν_x, previously defined in Equation (1)*:*

$$\Psi_x \;=\; \mathrm{CPTE}_w(x) \;=\; \int_{\mathcal{Y}^2} w(y_1 \mid y_0, \gamma(x))\, \nu_x(dz). \tag{3}$$

Ψ_x is the probability that a treated outcome is preferred to a control outcome among patients with covariate x, and $\Psi_x > 1/2$ means that treatment is preferred at x. Taking $\mathcal{Y} \subseteq \mathbb{R}$ and $w(y_1 \mid y_0, \tilde{x}) = y_1 - y_0$ recovers the CATE, so the CPTE generalises it to outcomes that can be compared but not substracted. Under the no-ties assumption, the expected preference for control at x is $1 - \Psi_x$. The policy maximising expected preference is therefore $\pi^\star(x) = \mathbf{1}\{\Psi_x > 1/2\}$, and any policy π incurs regret $\mathbb{E}_X\big[\,|2\Psi_X - 1| \cdot \mathbf{1}\{\pi(X) \neq \pi^\star(X)\}\,\big]$: decision errors are weighted by how clear-cut the decision is at X.

3 Active learning problem

In our use case, the rule w is not given: it has to be elicited from clinicians, one pairwise comparison at a time, asked at the query covariate $\tilde{x}$ rather than at the full covariate x.

Definition 3 (Query) *A query is a triplet $c = (\tilde{x}, a, b)$, in which a clinician compares outcome profiles $a, b \in \mathcal{Y}$ presented at query covariate $\tilde{x}$, and returns a binary answer $V_c \in \{0, 1\}$, with $V_c = 1$ if a is preferred.*

Comparisons are collected one at a time: at round t the method selects a query c_{t+1} from a candidate pool $\mathcal{C}_t$, observes a single answer v_{t+1}, and updates its beliefs about w. The pool is finite and built from outcome profiles observed in the cohort, so that the profiles shown to a clinician remain clinically realistic. The procedure stops after a budget of T queries, reflecting available clinician time, and we write $\mathcal{D}_t = \{(c_k, v_k)\}_{k=1}^t$ for the comparisons observed after t rounds. We write $p^\star$ for a distribution over full covariates $x^\star$, describing the population on which treatment decisions will be made. In practice, $p^\star$ is the empirical distribution of the target cohort. This yields the following problem, which our paper addresses.

Problem 1 *Given a budget of T comparisons and the estimated potential-outcome laws $\hat{\mu}_a$ of Section 2.1, choose queries $c_1, \ldots, c_T \in \mathcal{C}_t$ so that the resulting CPTE estimate and the policy it induces are as accurate as possible on $p^\star$.*

Active learning tackles the following question: which query should we ask next? As in Bayesian experimental design, we choose the query with the largest expected information gain about a quantity of interest. Here, this quantity is Ψ_x, the proportion of patients with covariate x for whom treatment is preferred, which alone determines the optimal policy $\pi^\star$ and its regret.

Definition 4 (CPTE-EIG) *For a target covariate x and a candidate query c, the expected information gain (EIG) of c about Ψ_x is the following mutual information (Cover & Thomas, 1991): $\mathrm{EIG}_{\Psi(x)}(c) = \mathbb{I}(\Psi_x; V_c \mid \mathcal{D}_t, c)$. The CPTE-EIG score of c averages this over the target population, and the next query maximises it:*

$$\text{CPTE-EIG}(c) = \mathbb{E}_{x^\star \sim p^\star}\big[\,\mathrm{EIG}_{\Psi(x^\star)}(c)\,\big], \quad c_{t+1} \in \arg\max_{c \in \mathcal{C}_t} \text{CPTE-EIG}(c). \tag{4}$$

Equation 4 is well defined for any probabilistic model of w, but in general does not admit a closed form or cannot be computed efficiently: evaluating it would require sampling w from its posterior and integrating against $\nu_{x^\star}$ for every candidate query and target covariate at every round, by Monte Carlo (Robert, 2011). We next resolve this difficulty by proposing a computationally efficient approach.

4 Efficient CPTE-EIG computation under Gaussian processes

Utility contrast models. While the active learning criterion CPTE-EIG is well-defined for any probabilistic model of w, we specify a model under which we can derive closed form updates.

Following the standard Bradley-Terry model (Bradley & Terry, 1952), we assume preferences come from a latent utility that assigns a single score per outcome profile. One profile is preferred to another whenever its utility is higher. We thus define $u : \tilde{\mathcal{X}} \times \mathcal{Y} \to \mathbb{R}$ and for $r = (r_1, r_0) \in (\tilde{\mathcal{X}} \times \mathcal{Y})^2$, we define the *utility contrast* as:

$$D_r u = u(r_1) - u(r_0). \tag{5}$$

A query $c = (\tilde{x}, a, b)$ corresponds to $r = ((\tilde{x}, a), (\tilde{x}, b))$, which gives the notation $D_c u$. A treated–control pair $z = (y_1, y_0)$ at full covariate x (as in Equation (1)) corresponds to $r = ((\tilde{x}, y_1), (\tilde{x}, y_0))$, giving $D_{x,z}u = u(\tilde{x}, y_1) - u(\tilde{x}, y_0)$, where the index carries the full covariate x, but u itself is evaluated at $\tilde{x} = \gamma(x)$. The clinician's noisy answer and the CPTE both compare an outcome pair through the same probit link (Thurstone, 1927), and we tie them to a single noise scale λ. Formally,

$$V_c \mid u, c \sim \text{Bernoulli}\left(\left(\frac{D_c u}{\lambda}\right)\right), \qquad w_u(x, z) = \Phi\left(\frac{D_{x,z}u}{\lambda}\right), \tag{6}$$

where Φ is the standard normal cumulative distribution function. A large λ drives both towards a coin flip. A small λ makes each nearly deterministic in the sign of its contrast. Because Φ is antisymmetric, $\Phi(t) + \Phi(-t) = 1$, swapping y_1 and y_0 in Equation (6) flips the sign of the contrast and gives $w_u(\tilde{x}, (y_0, y_1)) = 1 - w_u(\tilde{x}, (y_1, y_0))$: the no-ties condition holds for every fixed u.

Utililty prior. We treat u as random and place a Gaussian process prior on it (Rasmussen & Williams, 2005) with marginal variance σ^2, and write U for the resulting random function, reserving u for a fixed realisation. Since Equation (6) is not Gaussian, the exact posterior is unavailable in closed form. As is standard for Gaussian process preference models (Chu & Ghahramani, 2005a), we use a Laplace approximation (Williams & Barber, 1998) which gives $U \mid \mathcal{D}_t \approx q_t = \text{GP}(m_t, k_t)$, and take all probabilities, expectations and covariances involving U under q_t from here on. The contrast operator is linear, so every contrast $D_r U$ is Gaussian under q_t. We write $m_t^\Delta(r)$ for its posterior mean and $k_t^\Delta(r, r')$ for the posterior covariance between $D_r U$ and $D_{r'} U$.

Induced law of the estimand. Substituting Equation (6) into Definition 2, the estimand becomes a functional of the utility,

$$T_x(u) = \int_{\mathcal{Y}^2} \Phi\Big(\frac{D_{x,z}u}{\lambda}\Big)\, \nu_x(dz), \qquad \Psi_x = T_x(u). \tag{7}$$

This holds for a fixed u. Since u is now a draw from the posterior q_t of Section 4, Ψ_x is random too. Let $U \sim q_t$, the *induced estimand* at $x \in \mathcal{X}$ after t queries is the random variable $\Psi_x = T_x(U) \in [0, 1]$, whose law we denote $\Pi_{t,x}$, and it represents our belief about Ψ_x after observing $\mathcal{D}_t$. Because T_x is non-linear, $\Pi_{t,x}$ is not Gaussian and its density has no closed form. Its mean also differs from the plug-in value, $\mathbb{E}_{q_t}[\Psi_x] \neq T_x(m_t)$ in general.

4.1 Linearisation and error control

Definition 4 scores a candidate query c by $\mathbb{I}(\Psi_x; V_c \mid \mathcal{D}_t, c)$. Since V_c depends on U only through the contrast $D_c U$, this score is determined by the joint law of $(\Psi_x, D_c U)$. If this pair were Gaussian, the criterion would reduce to a one-dimensional integral. We therefore approximate Ψ_x by an affine function of U, which makes the pair jointly Gaussian, and we bound the error this introduces. Throughout we assume the posterior kernel k_t is bounded by K_t on its diagonal (Assumption 2, Appendix A), which holds for standard kernels such as RBF and Matérn (Rasmussen & Williams, 2005) and bounds the variance of every posterior contrast by $4K_t$. Consequently, $z \mapsto D_{x,z}h$ lies in $L^2(\nu_x) \subset L^1(\nu_x)$ for q_t-almost every $h = U - m_t$ (Lemma 1, Appendix A), which allows us to differentiate under the integral sign below (Folland, 2013).

Theorem 1 (Linearised estimand and its error) *Under Assumption 2, write $h = U - m_t$ and, for $x \in \mathcal{X}$ and $z \in \mathcal{Y}^2$, the sensitivity weight*

$$\alpha_t(x, z) = \frac{1}{\lambda}\varphi\Big(\frac{m_t^\Delta(x, z)}{\lambda}\Big), \qquad \varphi = \Phi', \tag{8}$$

which is largest on pairs the model finds hardest to rank. Then, the map $\epsilon \mapsto T_x(m_t + \epsilon h)$ is differentiable at 0, and its first-order expansion at $\epsilon = 1$ gives

$$\Psi_x \approx T_x(m_t) + \int \alpha_t(x, z)\big\{D_{x,z}U - m_t^\Delta(x, z)\big\}\, \nu_x(dz), \tag{9}$$

which is affine in U, hence jointly Gaussian with any contrast D_cU. The remainder $R_x(h)$ neglected in Equation (9) *satisfies* $\mathbb{E}_{q_t}\big[|R_x(h)|\big] \;\le\; \frac{\varphi(1)}{2\lambda^2}\int k_t^{\Delta}\big((x,z),(x,z)\big)\,\nu_x(dz) \;=:\; B_t(x)$.

The proof is in Appendix B. From here on, we take Equation (9) as the definition of Ψ_x: a ν_x-average of utility contrasts, each weighted by $\alpha_t(x,z)$, that is, by how much the pair can still move the estimand. The joint law of (Ψ_x, D_cU) is then Gaussian and known in closed form, and Section 4.2 uses it to compute CPTE-EIG. Up to the factor $\varphi(1)/(2\lambda^2)$, the bound $B_t(x)$ is the ν_x-average of the posterior variances of these contrasts. It is small when the model is certain about the contrasts on the target pairs, and loose otherwise. Since it depends only on the current posterior, it can be computed at every round.

4.2 Closed-form CPTE-EIG

We now turn the joint law of (Ψ_x, D_cU) into a computable acquisition score, using the symmetry of mutual information that gives $\mathrm{EIG}_{\Psi(x)}(c) = \mathbb{H}[V_c \mid \mathcal{D}_t, c] - \mathbb{E}_{\Psi_x\sim\Pi_{t,x}}\big[\mathbb{H}[V_c \mid \Psi_x, \mathcal{D}_t, c]\big]$. Expanding on the V_c side, avoids updating the full posterior for every candidate: since V_c is binary, each term is the binary entropy $h(\cdot)$ of a success probability, and the two terms differ only in what that probability conditions on. The first conditions on what we already know before answering, giving the marginal $p_{t,c} = \mathbb{P}(V_c = 1 \mid \mathcal{D}_t, c)$: how likely a "yes" is now. The second additionally conditions on the value of the estimand itself, giving $\eta_{t,x,c}(\psi) = \mathbb{P}(V_c = 1 \mid \Psi_x = \psi, \mathcal{D}_t, c)$: how likely a "yes" would be if $\Psi_x = \psi$ were already known. The gap between $h(p_{t,c})$ and the Ψ_x-averaged $h(\eta_{t,x,c}(\Psi_x))$ is how much knowing Ψ_x would sharpen our prediction of the answer, which also means by symmetry, how much the answer tells us about Ψ_x.

Theorem 2 (Joint law and closed-form CPTE-EIG) *Fix a target covariate $x \in \mathcal{X}$ and a candidate query c, and abbreviate $\alpha_t(z) = \alpha_t(x,z)$ and $k_t^{\Delta}(z,z') = k_t^{\Delta}\big((x,z),(x,z')\big)$. Under q_t, (Ψ_x, D_cU) is Gaussian with mean $\big(T_x(m_t), m_t^{\Delta}(c)\big)$ and covariance $\Sigma = \begin{pmatrix}\sigma_x^2 & \sigma_{x,c}\\ \sigma_{x,c} & \sigma_c^2\end{pmatrix}$, where*

$$\sigma_x^2 = \iint \alpha_t(z)\alpha_t(z')\,k_t^{\Delta}(z,z')\,\nu_x(dz)\,\nu_x(dz'), \tag{10}$$

$$\sigma_{x,c} = \int \alpha_t(z)\,k_t^{\Delta}(z,c)\,\nu_x(dz), \qquad \sigma_c^2 = k_t^{\Delta}(c,c). \tag{11}$$

Writing $\tau_{x,c} = \sigma_{x,c}/\sigma_x$ and $s_{x,c}^2 = \lambda^2 + \sigma_c^2 - \tau_{x,c}^2$, then,

$$\mathrm{EIG}_{\Psi(x)}(c) \approx h\left(\Phi\left(\frac{m_t^{\Delta}(c)}{\sqrt{\lambda^2+\sigma_c^2}}\right)\right) - \sum_{i=1}^{n}\frac{w_i}{\sqrt{\pi}}\,h\left(\Phi\left(\frac{m_t^{\Delta}(c)+\sqrt{2}\,\beta_i\,\tau_{x,c}}{s_{x,c}}\right)\right), \tag{12}$$

where $(\beta_i, w_i)_{i\le n}$ are Gauss–Hermite nodes and weights.

The proof, including the Gauss–Hermite quadrature step (Liu & Pierce, 1994), is in Appendix C. The covariance $\sigma_{x,c}$ links the query to the estimand. If $\sigma_{x,c} = 0$, then $\tau_{x,c} = 0$ and Equation (12) is zero, even when $\sigma_c^2 > 0$: the model is uncertain about the query, but its answer carries no information about Ψ_x. Example 1 in Appendix D gives such a query. Equation (12) requires no posterior sampling and no model refit: all its terms are computed from m_t and k_t through Equations (10) and (11), at a fixed cost per candidate. The CPTE-EIG score then averages $\mathrm{EIG}_{\Psi(x)}(c)$ over $x^\star \sim p^\star$. We approximate this outer expectation and the ν_x-integrals by Monte Carlo (Robert, 2011), as detailed in Algorithm 1 (Appendix E).

5 Experiments

We evaluate CPTE-EIG on the Parkinson's disease example, where dopamine agonists are compared with levodopa, two standard first-line treatments. The main quantities are summarised in Table 1. We consider four follow-up horizons $h \in \{1,3,5,7\}$ years, which give four estimands $\Psi_x^{(h)}$. We compare CPTE-EIG with three baselines: BALD (Houlsby et al., 2011), a widely used criterion for preference learning, the prediction-oriented EPIG (Smith et al., 2023), and random acquisition.

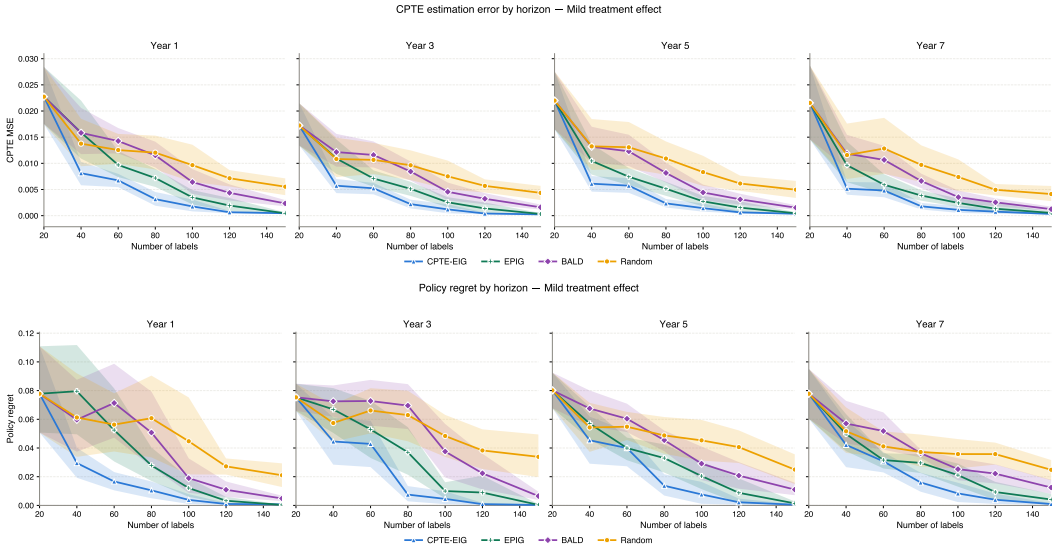


Figure 2: Mild treatment effect, per horizon h. Top: MSE of $T_x(m_t)$ against the oracle CPTE over 100 target patients from $p^\star$. Bottom: regret of $\pi_t(x) = \mathbf{1}\{T_x(m_t) > 1/2\}$. Mean over 15 seeds, 95% bootstrap intervals. Lower is better.

EPIG is given the same targets as CPTE-EIG, namely the treated–control pairs drawn from ν_x for $x \sim p^\star$. The two methods therefore differ only in the quantity they target: the preference label of each pair for EPIG, and their average Ψ_x for CPTE-EIG.

Setup. We build a semi-synthetic cohort of 2,000 patients from PPMI (Marek et al., 2018) with both potential outcomes known, giving an oracle CPTE against which every method is scored, and a separate pool of 1.37M candidate comparisons between real PPMI profiles for elicitation. The clinician is simulated with a known utility that is linear in the outcomes, with covariate-dependent weights, $u(\tilde{x}, y) = -\sum_d w_d(\tilde{x}) \sum_{j \in G_d} y_j$, where G_d groups the MDS-UPDRS items into five domains: neuropsychiatric, sleep and autonomic, motor, motor complications, and quality of life. All weights are positive, so u is decreasing in every symptom. The neuropsychiatric weight increases linearly with age, the motor weight decreases linearly with age, and the motor-complication weights depend on sex. The other weights are constant. Learning u over the whole covariate space therefore requires comparisons across all ages and both sexes. We consider three treatment effects. In the *tiny* effect, the potential outcomes are those of the semi-synthetic cohort, where the two treatments lead to nearly identical outcomes, leading to $\Psi_x^{(h)} \approx 1/2$. In the *mild* and *strong* effects, we keep the original Y_0 and set $Y_1 = \mathrm{clip}(Y_0 + s\,\delta(\tilde{x}, h), 0, 4)$, with $s = 1$ and $s = 2$. Here the agonist improves tremor but worsens cognition by an amount that grows with h and age, so the best treatment depends on the patient, the horizon and the utility. All methods use the same Gaussian process model for the latent utility, the same warm start of 20 random comparisons, and the same candidate sets. The kernel hyperparameters and λ are estimated by marginal likelihood at each iteration. The budget ranges from 20 to 150 comparisons, and each experiment is repeated with 15 independent seeds. Reported intervals are 95% bootstrap intervals for the mean over seeds, computed from 3000 resamples. Further details on the GP and the setup are given in Appendix F.

Results. With a mild treatment effect (Figure 2), CPTE-EIG has the lowest CPTE MSE and policy regret at every budget from 40 comparisons on. With 80 comparisons, its error (2.4×10^{-3}) is already below that of EPIG with 100 comparisons and of BALD with 120, and its regret (0.012) is about a third of EPIG's and a quarter of BALD's. Since the three criteria share the same utility model, the differences in performance come only from the quantity they target: the utility as a whole for BALD, the label of each treated–control pair for EPIG, and the average over these pairs, Ψ_x, for CPTE-EIG. With a strong effect, the gap is larger: at 80 comparisons, the error of CPTE-EIG is five times lower than that of EPIG and fourteen times lower than that of BALD (Table 2), and the interval

Table 2: CPTE MSE ($\times 10^3$) against the number of comparisons, over 100 target patients from $p^\star$ (fixed across seeds and methods) and averaged over horizons: mean over 15 seeds and 95% percentile-bootstrap interval. Star * for lowest mean (ties included); **bold** for methods with interval overlapping that of a * method.

		Budget (number of comparisons)			
Effect	Method	40	80	120	150
Tiny	Random	5.0 [4.0, 6.2]	2.9 [2.3, 3.5]	2.2 [1.8, 2.5]	2.5 [2.1, 3.0]
	BALD	*3.2 [2.7, 3.7]	1.3 [1.1, 1.4]	*0.6 [0.5, 0.7]	*0.3 [0.3, 0.4]
	EPIG	3.3 [2.7, 3.9]	*1.1 [1.0, 1.2]	0.7 [0.6, 0.8]	0.4 [0.3, 0.5]
	CPTE-EIG	*3.2 [2.8, 3.7]	1.3 [1.1, 1.4]	*0.6 [0.5, 0.7]	0.4 [0.3, 0.5]
Mild	Random	12.4 [8.4, 17.0]	10.6 [7.6, 13.7]	6.0 [4.7, 7.3]	4.7 [3.3, 6.2]
	BALD	13.2 [10.2, 16.9]	8.7 [7.0, 10.6]	3.3 [2.3, 4.4]	1.7 [1.2, 2.2]
	EPIG	11.7 [8.5, 15.9]	5.3 [4.0, 6.7]	1.5 [0.7, 2.8]	0.4 [0.3, 0.6]
	CPTE-EIG	*6.3 [4.7, 7.7]	*2.4 [1.5, 3.3]	*0.6 [0.4, 0.9]	*0.3 [0.3, 0.5]
Strong	Random	29.4 [19.3, 41.1]	24.3 [16.6, 32.7]	13.7 [10.8, 16.6]	11.3 [7.6, 15.2]
	BALD	29.4 [23.2, 36.9]	18.3 [14.4, 22.3]	7.0 [4.8, 9.3]	3.4 [2.4, 4.6]
	EPIG	20.5 [13.4, 29.8]	6.6 [4.5, 8.9]	2.5 [1.3, 4.0]	0.6 [0.4, 0.7]
	CPTE-EIG	*13.4 [10.7, 16.1]	*1.3 [1.0, 1.6]	*0.9 [0.7, 1.2]	*0.4 [0.4, 0.5]

of BALD does not overlap that of CPTE-EIG at any budget. With a tiny effect, CPTE-EIG, EPIG and BALD are tied, with overlapping intervals at every budget, and all three have a lower error than random acquisition. The errors are also smaller than with larger effects: at 80 comparisons, the error of random acquisition is 2.9, against 10.6 and 24.3 with mild and strong effects (all $\times 10^{-3}$). Overall, CPTE-EIG is the best-performing method in terms of CPTE accuracy, policy regret and robustness across scenarios: it is as good as the other methods when the effect is tiny, and outperforms them when the effect is mild or strong.

6 Discussion and conclusion

In this paper, we introduced CPTE-EIG, an active learning criterion that selects clinician comparisons by their expected information gain about the CPTE. We linearised the CPTE around the posterior mean utility, which gave a Gaussian approximation of its posterior law, with an error bound. Under this approximation, CPTE-EIG becomes computationally tractable. On a semi-synthetic Parkinson's disease cohort with 13 outcomes, CPTE-EIG was the best-performing method in terms of CPTE estimation accuracy, policy regret and robustness across scenarios.

Our framework has limitations that suggest several extensions. We condition throughout on the estimated potential-outcome laws $\hat{\mu}_a$ and treat the preference rule as the only source of uncertainty. Propagating the uncertainty in $\hat{\mu}_a$ as well would give a more complete account of the uncertainty in the CPTE estimate. We also model a single, average clinician, whereas real clinicians can disagree with one another. Multi-user preference models (Houlsby et al., 2012), for instance with a shared utility and a per-clinician offset, could represent this disagreement. The choice of which clinician to ask would then become a second decision, next to the choice of which pair to show. Our experiments use 13 outcomes, more than the two to four typically considered in the composite-endpoint or multi-criteria decision literature (Freemantle et al., 2003), and CPTE-EIG performs best in this setting. With many more outcomes, the limit may come from elicitation rather than from the model, as profiles with too many dimensions may be hard to compare in a single judgment. Finally, the clinician in our experiments is simulated from a known utility, so that the error can be measured against an oracle CPTE. An important next step is to test the method on real preference data, which will include disagreement between clinicians and inconsistent answers.

When a treatment effect is built from clinician preferences, how these preferences are elicited matters as much as how the effect is estimated. In our Parkinson's disease experiments, targeting the estimand itself made a limited elicitation budget go further, and we expect the same to hold wherever outcomes are compared rather than measured on a common scale.

AI USE STATEMENT

We used large language models (LLMs) to help derive our theoretical analysis, and LLM coding assistants to help implement our method and experiments. We also used LLMs to help with the writing and proofreading of the paper, and to give feedback on drafts, including on the methodology and the experiments. The authors verified all AI-assisted derivations, reviewed and tested all AI-assisted code, and checked and edited all AI-assisted text. We take full responsibility for the final content of this work, including text, claims, proofs and code produced with the help of LLMs.

ETHICS STATEMENT

Our experiments use data from the Parkinson's Progression Markers Initiative (PPMI) (Marek et al., 2018), an observational cohort approved by the ethics committees of the participating sites, in which all participants gave written informed consent. We used the de-identified data under the PPMI data use agreement and do not redistribute them. We collected no new data from patients or clinicians: the clinician who answers the comparisons is simulated.

The experiments are designed to evaluate the methodology, not to answer a clinical question. The evaluation cohort is semi-synthetic, and the utilities that play the role of the clinician were chosen by us rather than elicited from clinicians. The estimated treatment effects and policies therefore have no clinical meaning and should not be used to guide the treatment of patients with Parkinson's disease. More generally, a CPTE is only as valid as its inputs. It requires the causal assumptions of Section 2.1, and it depends on whose preferences are elicited and on which patient characteristics enter the query covariate: our choice of γ, for instance, excludes ethnicity (Section 2.2). Since CPTE-EIG concentrates comparisons on the target population, the learned preference rule may also be less accurate for groups that are rare in this population. Clinical use would require preferences elicited from real clinicians, and ideally from patients, and validation on independent data.

REPRODUCIBILITY STATEMENT

All assumptions are stated in the paper (Assumption 1 in Section 2.1 and Assumption 2 in Appendix A), and complete proofs of Theorems 1 and 2 are given in Appendices A–C. Algorithm 1 (Appendix E) details the query selection procedure. Appendix F describes the construction of the semi-synthetic cohort and of the candidate pool, the simulated clinicians, the compared methods, the budgets, the seeds and the metrics. Anonymized code for the method and the experiments is provided in the supplementary material. The PPMI data cannot be redistributed, but researchers can request access at `https://www.ppmi-info.org`.

REFERENCES

Joshua Angrist and Guido Imbens. Identification and estimation of local average treatment effects, 1995.

Viktor Bengs, Aadirupa Saha, and Eyke Hüllermeier. Stochastic Contextual Dueling Bandits under Linear Stochastic Transitivity Models. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 1764–1786. PMLR, June 2022.

Ralph Allan Bradley and Milton E. Terry. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4):324–345, 1952. ISSN 0006-3444. doi: 10.2307/2334029.

Eric Brochu, Nando Freitas, and Abhijeet Ghosh. Active Preference Learning with Discrete Choice Data. In *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007.

Emily L Butler, Eric B Laber, Sonia M Davis, and Michael R Kosorok. Incorporating patient preferences into estimation of optimal individualized treatment rules. *Biometrics*, 74(1):18–26, 2018.

Marc Buyse. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine*, 29(30):3245–3257, December 2010. ISSN 1097-0258. doi: 10.1002/sim.3923.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical science*, pp. 273–304, 1995.

Siu Lun Chau, Javier Gonzalez, and Dino Sejdinovic. Learning Inconsistent Preferences with Gaussian Processes. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 2266–2281. PMLR, May 2022.

Wei Chu and Zoubin Ghahramani. Preference learning with gaussian processes. In *Proceedings of the 22nd international conference on Machine learning*, pp. 137–144, 2005a.

Wei Chu and Zoubin Ghahramani. Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research*, 6(35):1019–1041, 2005b. ISSN 1533-7928.

T. M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. Wiley, New York, 1991. ISBN 978-0-471-06259-2.

Mathieu Even and Julie Josse. Rethinking the win ratio: A causal framework for hierarchical outcome analysis. *arXiv preprint arXiv:2501.16933*, 2025.

Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley & Sons, June 2013. ISBN 978-1-118-62639-9.

Nick Freemantle, Melanie Calvert, John Wood, Joanne Eastaugh, and Carl Griffin. Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA*, 289(19): 2554–2559, May 2003. ISSN 0098-7484. doi: 10.1001/jama.289.19.2554.

Erdun Gao, Jake Fawkes, and Dino Sejdinovic. Causal-epig: Causally aligned active cate estimation. In *Forty-third International Conference on Machine Learning*, 2026.

Christopher G. Goetz, Barbara C. Tilley, Stephanie R. Shaftman, Glenn T. Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B. Stern, Richard Dodel, Bruno Dubois, Robert Holloway, Joseph Jankovic, Jaime Kulisevsky, Anthony E. Lang, Andrew Lees, Sue Leurgans, Peter A. LeWitt, David Nyenhuis, C. Warren Olanow, Olivier Rascol, Anette Schrag, Jeanne A. Teresi, Jacobus J. van Hilten, Nancy LaPelle, and Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders: Official Journal of the Movement Disorder Society*, 23(15): 2129–2170, November 2008. ISSN 1531-8257. doi: 10.1002/mds.22340.

Pamela Gongora-Salazar, Stephen Rocks, Patrick Fahr, Oliver Rivero-Arias, and Apostolos Tsiachristas. The Use of Multicriteria Decision Analysis to Support Decision Making in Healthcare: An Updated Systematic Literature Review. *Value in Health*, 26(5):780–790, May 2023. ISSN 1098-3015, 1524-4733. doi: 10.1016/j.jval.2022.11.007.

Javier González, Zhenwen Dai, Andreas Damianou, and Neil D. Lawrence. Preferential Bayesian optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1282–1291, 2017.

Miguel A. Hernan and James M. Robins. *Causal Inference: What If*. CRC Press, July 2025. ISBN 978-1-4200-7616-5.

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

Neil Houlsby, Ferenc Huszar, Zoubin Ghahramani, and Jose Hernández-lobato. Collaborative Gaussian Processes for Preference Learning. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

Andrew Jesson, Panagiotis Tigas, Joost van Amersfoort, Andreas Kirsch, Uri Shalit, and Yarin Gal. Causal-BALD: Deep Bayesian active learning of outcomes to infer treatment-effects from observational data. In *Advances in Neural Information Processing Systems*, volume 34, pp. 30465–30478, 2021.

Lorraine V. Kalia and Anthony E. Lang. Parkinson's disease. *Lancet*, 386(9996):896–912, August 2015. ISSN 1474-547X. doi: 10.1016/S0140-6736(14)61393-3.

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

Qing Liu and Donald A. Pierce. A Note on Gauss-Hermite Quadrature. *Biometrika*, 81(3):624–629, 1994. ISSN 0006-3444. doi: 10.2307/2337136.

Kenneth Marek, Sohini Chowdhury, Andrew Siderowf, Shirley Lasch, Christopher S. Coffey, Chelsea Caspell-Garcia, Tanya Simuni, Danna Jennings, Caroline M. Tanner, John Q. Trojanowski, Leslie M. Shaw, John Seibyl, Norbert Schuff, Andrew Singleton, Karl Kieburtz, Arthur W. Toga, Brit Mollenhauer, Doug Galasko, Lana M. Chahine, Daniel Weintraub, Tatiana Foroud, Duygu Tosun-Turgut, Kathleen Poston, Vanessa Arnedo, Mark Frasier, and Todd Sherer. The Parkinson's progression markers initiative (PPMI) – establishing a PD biomarker cohort. *Annals of Clinical and Translational Neurology*, 5(12):1460–1477, October 2018. ISSN 2328-9503. doi: 10.1002/acn3.644.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

Dovid Parnas, Mathieu Even, Julie Josse, and Uri Shalit. Preference-based Conditional Treatment Effects and Policy Learning. In *Proceedings of The 29th International Conference on Artificial Intelligence and Statistics*, pp. 226–234. PMLR, August 2026.

François Petit, Gérard Biau, and Raphaël Porcher. Optimal treatment regimes for the net benefit of a treatment. *arXiv preprint arXiv:2503.22580*, 2025.

Stuart J. Pocock, Cono A. Ariti, Timothy J. Collier, and Duolao Wang. The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal*, 33(2):176–182, January 2012. ISSN 1522-9645. doi: 10.1093/eurheartj/ehr352.

Tom Rainforth, Adam Foster, Desi R Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, November 2005. ISBN 978-0-262-25683-4. doi: 10.7551/mitpress/3206.001.0001.

Christian Robert. Monte Carlo Methods in Statistics. In *International Encyclopedia of Statistical Science*, pp. 854–858. Springer, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_376.

Paul R. Rosenbaum and Donald B. Rubin. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55, 1983. ISSN 0006-3444. doi: 10.2307/2335942.

Binxin Ru, Ahsan Alvi, Vu Nguyen, Michael A Osborne, and Stephen Roberts. Bayesian optimisation over multiple continuous and categorical inputs. In *International Conference on Machine Learning*, pp. 8276–8285. PMLR, 2020.

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

Donald B. Rubin. Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980. ISSN 0162-1459. doi: 10.2307/2287653.

Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom Rainforth. Prediction-oriented bayesian active learning. In *International conference on artificial intelligence and statistics*, pp. 7331–7348. PMLR, 2023.

L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927. ISSN 1939-1471. doi: 10.1037/h0070288.

C.K.I. Williams and D. Barber. Bayesian classification with Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1342–1351, December 1998. ISSN 01628828. doi: 10.1109/34.735807.

Appendices

A Bounded contrasts and integrability **14**

B Proof of Theorem 1 **15**

B.1 The linearisation 16

B.2 The remainder 17

C Proof of Theorem 2 **18**

C.1 Mean and covariance of $(\Psi_x, D_c U)$ 18

C.2 Two properties 19

C.3 Closed-form answer probabilities 20

C.4 Gauss–Hermite quadrature and the final closed form 20

D Targeting the right information **21**

E Query selection algorithm **22**

F Additional experimental details and results **22**

F.1 Data generation 22

F.2 The simulated clinician 23

F.3 Methods and protocol 23

F.4 Metrics 24

F.5 Tiny and strong treatment effects 25

F.6 Learning w is not learning Ψ_x 26

F.7 Gauss–Hermite quadrature versus Monte Carlo 26

A Bounded contrasts and integrability

This appendix states the assumption and regularity result used by Theorem 1 in Section 4.1, proves the variance bound they rely on, and then proves lemma 1.

Assumption 2 (Bounded posterior variance) *The posterior kernel k_t is measurable and bounded on its diagonal: there exists $K_t < \infty$ such that $k_t(r, r) \le K_t$ for all $r \in \tilde{\mathcal{X}} \times \mathcal{Y}$.*

Recall that for two contrast indices $r = (r_1, r_0)$ and $r' = (r'_1, r'_0)$ in $\tilde{\mathcal{X}} \times \mathcal{Y}$,

$$k_t^{\Delta}(r, r') = k_t(r_1, r'_1) - k_t(r_1, r'_0) - k_t(r_0, r'_1) + k_t(r_0, r'_0).$$

Setting $r' = r$ gives the variance, under q_t, of a single contrast:

$$k_t^{\Delta}(r, r) = \mathrm{Var}_{q_t}(D_r U) = k_t(r_1, r_1) + k_t(r_0, r_0) - 2\, k_t(r_1, r_0),$$

using the symmetry $k_t(r_1, r_0) = k_t(r_0, r_1)$ of the kernel.

Assumption 2 bounds the two diagonal terms, $k_t(r_1, r_1) \le K_t$ and $k_t(r_0, r_0) \le K_t$. Since k_t is a covariance kernel it is positive semi-definite, so Cauchy–Schwarz applies to the cross term:

$$|k_t(r_1, r_0)| \le \sqrt{k_t(r_1, r_1)\, k_t(r_0, r_0)} \le K_t.$$

Combining the three bounds,

$$0 \;\le\; k_t^{\Delta}(r,r) \;\le\; K_t + K_t + 2K_t \;=\; 4K_t \;<\; \infty, \tag{13}$$

for every $r \in \tilde{\mathcal{X}} \times \mathcal{Y}$. The same argument applies to k_t^{Δ} itself, which is a valid covariance kernel on contrasts (each $D_r U$ is a linear functional of the Gaussian field U), so for two possibly different contrasts r, r',

$$|k_t^{\Delta}(r,r')| \le \sqrt{k_t^{\Delta}(r,r)\, k_t^{\Delta}(r',r')} \le 4K_t.$$

Every posterior contrast therefore has variance at most $4K_t$, and every pair of contrasts has covariance at most $4K_t$ in absolute value.

Lemma 1 (Regularity of the contrast) *Under Assumption 2, write $h = U - m_t$. For every $x \in \mathcal{X}$ and q_t-almost every realisation of h, the map $z \mapsto D_{x,z}h$ belongs to $L^2(\nu_x) \subset L^1(\nu_x)$.*

Proof 1 *Fix $x \in \mathcal{X}$ and write $h = U - m_t$, a q_t-centered Gaussian process with covariance k_t. For $z = (y_1, y_0) \in \mathcal{Y}^2$, recall that $D_{x,z}h = h(\gamma(x), y_1) - h(\gamma(x), y_0)$, and define the nonnegative random variable*

$$S_x(h) = \int_{\mathcal{Y}^2} \big(D_{x,z}h\big)^2 \, \nu_x(dz).$$

Since k_t is measurable (Assumption 2), the map $(z,\omega) \mapsto \big(D_{x,z}h(\omega)\big)^2$ is jointly measurable and nonnegative, so Tonelli's theorem lets us exchange the expectation under q_t with the integral over ν_x:

$$\mathbb{E}_{q_t}[S_x(h)] = \mathbb{E}_{q_t}\left[\int (D_{x,z}h)^2 \, \nu_x(dz)\right] = \int \mathbb{E}_{q_t}\left[(D_{x,z}h)^2\right] \nu_x(dz).$$

Because h is centered under q_t, so is $D_{x,z}h$, hence

$$\mathbb{E}_{q_t}\big[(D_{x,z}h)^2\big] = \mathrm{Var}_{q_t}(D_{x,z}U) = k_t^{\Delta}\big((x,z),(x,z)\big) \le 4K_t,$$

taking $r = (x,z)$, i.e. $r_1 = (\gamma(x), y_1)$, $r_0 = (\gamma(x), y_0)$, in Equation (13)*. Since ν_x is a probability measure,*

$$\mathbb{E}_{q_t}[S_x(h)] \;\le\; \int_{\mathcal{Y}^2} 4K_t \, \nu_x(dz) \;=\; 4K_t \;<\; \infty.$$

$S_x(h)$ is nonnegative with finite expectation, so it is q_t-almost surely finite: $S_x(h) < \infty$ for q_t-almost every h. By definition of S_x, this says exactly that, for q_t-almost every h,

$$z \mapsto D_{x,z}h \;\in\; L^2(\nu_x).$$

Since ν_x is a probability measure, Cauchy–Schwarz gives, for q_t-almost every h,

$$\int_{\mathcal{Y}^2} |D_{x,z}h| \, \nu_x(dz) \;\le\; \left(\int_{\mathcal{Y}^2} (D_{x,z}h)^2 \, \nu_x(dz)\right)^{1/2} = S_x(h)^{1/2} \;<\; \infty,$$

so $z \mapsto D_{x,z}h$ also belongs to $L^1(\nu_x)$. This holds for every $x \in \mathcal{X}$, which proves the lemma.

B Proof of Theorem 1

Fix a target covariate $x \in \mathcal{X}$. The exact CPTE induced by a posterior draw $U \sim q_t$ is

$$T_x(U) = \int_{\mathcal{Y}^2} \Phi\left(\frac{D_{x,z}U}{\lambda}\right) \nu_x(dz).$$

Because of the nonlinear map Φ, T_x is not a linear functional of U. We linearise T_x around the posterior mean m_t, bound the error this introduces, and record the resulting joint Gaussianity of $(\Psi_x, D_c U)$.

Write $U = m_t + h$, where $h = U - m_t$. By Lemma 1, for q_t-almost every realisation of h,

$$D_{x,\cdot}h \in L^2(\nu_x) \subset L^1(\nu_x).$$

We fix such a realisation of h for the remainder of the proof.

B.1 THE LINEARISATION

Following the standard approach to linearising a nonlinear functional, we study the change of T_x along the one-dimensional path $m_t + \epsilon h$, $\epsilon \in \mathbb{R}$. Define

$$g(\epsilon) = T_x(m_t + \epsilon h).$$

When it exists, $g'(0)$ is exactly the Gateaux derivative of T_x at m_t in the direction h, $g'(0) = \mathrm{d}T_x(m_t)(h)$. We later establish its existence directly, by differentiating under the integral sign, without refering to any general Gateaux-differentiability theorem. Two values of g matter:

$$g(0) = T_x(m_t), \qquad g(1) = T_x(m_t + h) = T_x(U).$$

A first-order Taylor expansion of g around $\epsilon = 0$, evaluated at $\epsilon = 1$, will therefore give the linearisation of $T_x(U)$ that we want.

By linearity of the contrast operator $D_{x,z}$,

$$D_{x,z}(m_t + \epsilon h) = m_t^{\Delta}(x,z) + \epsilon\, D_{x,z}h, \qquad m_t^{\Delta}(x,z) := D_{x,z}m_t,$$

so that

$$g(\epsilon) = \int_{\mathcal{Y}^2} f(\epsilon, z)\, \nu_x(dz), \qquad f(\epsilon, z) = \Phi\left(\frac{m_t^{\Delta}(x,z) + \epsilon\, D_{x,z}h}{\lambda}\right). \tag{14}$$

We justify that $g'(0)$ exists and can be computed by differentiating Equation (14) under the integral sign (Folland, 2013). This standard result requires three conditions to hold for $f(\epsilon, z)$ on some open interval I including 0:

(i) for every $\epsilon \in I$, the map $z \mapsto f(\epsilon, z)$ is ν_x-integrable;

(ii) for ν_x-almost every z, the map $\epsilon \mapsto f(\epsilon, z)$ is differentiable on I;

(iii) there exists $M \in L^1(\nu_x)$ such that, for every $\epsilon \in I$ and ν_x-almost every z, $|\partial_\epsilon f(\epsilon, z)| \leq M(z)$.

If (i)–(iii) hold, then g is differentiable on I and $g'(\epsilon) = \int \partial_\epsilon f(\epsilon, z)\, \nu_x(dz)$.

Condition (i). Immediate, since $0 \leq \Phi \leq 1$, so $f(\epsilon, \cdot)$ is bounded and ν_x is a probability measure.

Condition (ii). Immediate, since Φ is continuously differentiable ($\Phi \in C^1$), with derivative $\varphi = \Phi'$, the standard normal density. Explicitly,

$$\partial_\epsilon f(\epsilon, z) = \frac{D_{x,z}h}{\lambda}\, \varphi\left(\frac{m_t^{\Delta}(x,z) + \epsilon\, D_{x,z}h}{\lambda}\right). \tag{15}$$

Condition (iii). Since φ is bounded, $0 \leq \varphi(u) \leq 1/\sqrt{2\pi}$ for every $u \in \mathbb{R}$, we get, uniformly in ϵ,

$$|\partial_\epsilon f(\epsilon, z)| \leq \frac{|D_{x,z}h|}{\lambda\sqrt{2\pi}} =: M(z).$$

By Lemma 1, $D_{x,\cdot}h \in L^1(\nu_x)$, so $M \in L^1(\nu_x)$: condition (iii) holds.

All three conditions hold, so differentiation under the integral sign applies. Evaluating Equation (15) at $\epsilon = 0$ gives

$$g'(0) = \int_{\mathcal{Y}^2} \frac{D_{x,z}h}{\lambda}\, \varphi\left(\frac{m_t^{\Delta}(x,z)}{\lambda}\right) \nu_x(dz) = \int_{\mathcal{Y}^2} \alpha_t(x,z)\, D_{x,z}h\, \nu_x(dz),$$

where we recognise the sensitivity weight $\alpha_t(x,z)$ of Equation (8).

By Taylor's theorem, $g(1) = g(0) + g'(0) + R_x(h)$, where $R_x(h)$ is the remainder. Substituting $g(0) = T_x(m_t)$, $g(1) = T_x(U)$, and the expression for $g'(0)$ from Step 2,

$$T_x(U) = T_x(m_t) + \int_{\mathcal{Y}^2} \alpha_t(x,z)\, D_{x,z}h\, \nu_x(dz) + R_x(h).$$

The linearised approximation consists in neglecting $R_x(h)$. Since $D_{x,z}h = D_{x,z}U - m_t^{\Delta}(x,z)$, this gives

$$T_x(U) \approx T_x(m_t) + \int_{\mathcal{Y}^2} \alpha_t(x,z)\left\{D_{x,z}U - m_t^{\Delta}(x,z)\right\}\nu_x(dz), \tag{16}$$

which is exactly Equation (9). We adopt the right-hand side of Equation (16) as the definition of Ψ_x from here on.

The right-hand side of Equation (16) is an affine functional of the Gaussian process U, so Ψ_x is Gaussian under q_t. For any candidate query c, the contrast D_cU is also a linear functional of U. Two jointly Gaussian-process-linear functionals of the same Gaussian field are themselves jointly Gaussian, so (Ψ_x, D_cU) is jointly Gaussian, as used in Theorem 2.

B.2 The remainder

We control $R_x(h)$ using Taylor's theorem with integral remainder: if g' is absolutely continuous on $[0,1]$ and $g'' \in L^1([0,1])$, then

$$g(1) = g(0) + g'(0) + R_x(h), \qquad R_x(h) = \int_0^1 (1-\epsilon)\, g''(\epsilon)\, d\epsilon. \tag{17}$$

We verify the hypothesis by differentiating $\partial_\epsilon f$ under the integral sign a second time, checking conditions (i)–(iii) again.

Since $\Phi \in C^2$, the second derivative of the integrand is

$$\partial_\epsilon^2 f(\epsilon, z) = \frac{(D_{x,z}h)^2}{\lambda^2}\,\Phi''\left(\frac{m_t^{\Delta}(x,z) + \epsilon\, D_{x,z}h}{\lambda}\right). \tag{18}$$

(i) For every ϵ, $z \mapsto \partial_\epsilon f(\epsilon, z)$ is ν_x-integrable, since it is bounded in absolute value by a constant times $|D_{x,z}h| \in L^1(\nu_x)$ (already used above).

(ii) For ν_x-almost every z, $\epsilon \mapsto \partial_\epsilon f(\epsilon, z)$ is differentiable, since $\Phi \in C^2$.

(iii) We need a dominating $L^1(\nu_x)$ function for Equation (18). Using $\Phi''(u) = \varphi'(u) = -u\,\varphi(u)$,

$$|\Phi''(u)| \le \sup_{v\in\mathbb{R}} |v\,\varphi(v)| = \varphi(1),$$

so, uniformly in ϵ,

$$|\partial_\epsilon^2 f(\epsilon, z)| \le \frac{\varphi(1)}{\lambda^2}(D_{x,z}h)^2.$$

By Lemma 1, $D_{x,\cdot}h \in L^2(\nu_x)$, so the right-hand side lies in $L^1(\nu_x)$: condition (iii) holds.

All three conditions hold, so g is twice differentiable on $[0,1]$ with

$$g''(\epsilon) = \int_{\mathcal{Y}^2} \frac{(D_{x,z}h)^2}{\lambda^2}\,\Phi''\left(\frac{m_t^{\Delta}(x,z) + \epsilon\, D_{x,z}h}{\lambda}\right)\nu_x(dz),$$

and, using the bound established in condition (iii) above, for every $\epsilon \in [0,1]$,

$$|g''(\epsilon)| \le \frac{\varphi(1)}{\lambda^2}\int_{\mathcal{Y}^2} (D_{x,z}h)^2\,\nu_x(dz).$$

Combining this bound with Equation (17),

$$|R_x(h)| \le \int_0^1 (1-\epsilon)\,|g''(\epsilon)|\,d\epsilon \le \frac{\varphi(1)}{\lambda^2}\left(\int_0^1 (1-\epsilon)\,d\epsilon\right)\int_{\mathcal{Y}^2}(D_{x,z}h)^2\,\nu_x(dz).$$

Since $\int_0^1 (1-\epsilon)\,d\epsilon = 1/2$,

$$|R_x(h)| \le \frac{\varphi(1)}{2\lambda^2}\int_{\mathcal{Y}^2}(D_{x,z}h)^2\,\nu_x(dz). \tag{19}$$

Taking expectation under q_t on both sides of Equation (19), and exchanging expectation and ν_x-integral by Tonelli's theorem exactly as in the proof of Lemma 1 (the integrand is nonnegative), we get

$$\mathbb{E}_{q_t}\big[|R_x(h)|\big] \le \frac{\varphi(1)}{2\lambda^2} \int_{\mathcal{Y}^2} \mathbb{E}_{q_t}\big[(D_{x,z}h)^2\big]\, \nu_x(dz).$$

Since h is centered under q_t, so is $D_{x,z}h$, and hence $\mathbb{E}_{q_t}[(D_{x,z}h)^2] = \mathrm{Var}_{q_t}(D_{x,z}U) = k_t^{\Delta}\big((x,z),(x,z)\big)$. Substituting,

$$\mathbb{E}_{q_t}\big[|R_x(h)|\big] \le \frac{\varphi(1)}{2\lambda^2} \int_{\mathcal{Y}^2} k_t^{\Delta}\big((x,z),(x,z)\big)\, \nu_x(dz) =: B_t(x),$$

which is the bound from the theorem. This completes the proof.

Interpretation. $B_t(x)$ is the average, over $z \sim \nu_x$, of the posterior variance of the contrast $D_{x,z}U$. The linearisation is therefore accurate when the posterior is already confident about how outcome pairs at x compare (small k_t^{Δ} relative to λ^2), and can be loose early in active learning, when the posterior is still wide, but tightens as informative comparisons accumulate.

C Proof of Theorem 2

Throughout, fix a target covariate $x \in \mathcal{X}$ and a candidate query c, and use the shorthand $\alpha_t(z) = \alpha_t(x,z)$ and $k_t^{\Delta}(z,z') = k_t^{\Delta}\big((x,z),(x,z')\big)$ from the theorem statement. The proof has four parts: (1) the mean and covariance of (Ψ_x, D_cU); (2) two standard properties about jointly Gaussian and probit-linked random variables; (3) the closed-form answer probabilities $\eta_{t,x,c}$ and $p_{t,c}$, obtained by applying these properties; and (4) the Gauss–Hermite quadrature that turns the resulting expectation into Equation (12).

C.1 Mean and covariance of (Ψ_x, D_cU)

By Theorem 1,

$$\Psi_x = T_x(m_t) + \int_{\mathcal{Y}^2} \alpha_t(z)\,\big\{D_{x,z}U - m_t^{\Delta}(x,z)\big\}\,\nu_x(dz),$$

an affine functional of U, and D_cU is a linear functional of U. Since U is Gaussian under q_t, so is any finite collection of linear (or affine) functionals of U. In particular (Ψ_x, D_cU) is jointly Gaussian.

Mean Taking expectations, $\mathbb{E}_{q_t}[D_{x,z}U - m_t^{\Delta}(x,z)] = 0$ for every z (by definition of m_t^{Δ} as the posterior mean contrast), so $\mathbb{E}_{q_t}[\Psi_x] = T_x(m_t)$. Similarly $\mathbb{E}_{q_t}[D_cU] = D_cm_t = m_t^{\Delta}(c)$. The mean vector is therefore $\big(T_x(m_t), m_t^{\Delta}(c)\big)$.

Variance σ_x^2 Write $G(z) = D_{x,z}U$ and $\widetilde{G}(z) = G(z) - \mathbb{E}_{q_t}[G(z)]$, so that

$$\Psi_x - \mathbb{E}_{q_t}[\Psi_x] = \int_{\mathcal{Y}^2} \alpha_t(z)\,\widetilde{G}(z)\,\nu_x(dz).$$

Then

$$\begin{aligned}\sigma_x^2 = \mathrm{Var}_{q_t}(\Psi_x) &= \mathbb{E}_{q_t}\left[\left(\int \alpha_t(z)\,\widetilde{G}(z)\,\nu_x(dz)\right)\left(\int \alpha_t(z')\,\widetilde{G}(z')\,\nu_x(dz')\right)\right] \\ &= \mathbb{E}_{q_t}\left[\iint \alpha_t(z)\alpha_t(z')\,\widetilde{G}(z)\widetilde{G}(z')\,\nu_x(dz)\,\nu_x(dz')\right].\end{aligned}$$

We exchange $\mathbb{E}_{q_t}$ and the double ν_x-integral by Fubini's theorem. This is justified because the integrand is absolutely integrable with respect to $q_t \otimes \nu_x \otimes \nu_x$:

- α_t is uniformly bounded, since $\alpha_t(z) = \varphi\big(m_t^{\Delta}(x,z)/\lambda\big)/\lambda \le 1/(\lambda\sqrt{2\pi}) =: \bar{\alpha}$ for every z;

- by Cauchy–Schwarz in $L^2(q_t)$ and the bound Equation (13),

$$\mathbb{E}_{q_t}\big[|\widetilde{G}(z)\widetilde{G}(z')|\big] \le \sqrt{\mathrm{Var}_{q_t}(G(z))\ \mathrm{Var}_{q_t}(G(z'))} = \sqrt{k_t^{\Delta}(z,z)\,k_t^{\Delta}(z',z')} \le 4K_t;$$

- $\nu_x \otimes \nu_x$ is a probability measure on $\mathcal{Y}^2 \times \mathcal{Y}^2$.

The integrand is therefore bounded in absolute value by the constant $\bar{\alpha}^2 \cdot 4K_t$, which is integrable against the finite measure $q_t \otimes \nu_x \otimes \nu_x$. Fubini's theorem applies, so we may exchange the order of integration:

$$\sigma_x^2 = \iint \alpha_t(z)\alpha_t(z')\,\mathbb{E}_{q_t}\big[\widetilde{G}(z)\widetilde{G}(z')\big]\,\nu_x(dz)\,\nu_x(dz') = \iint \alpha_t(z)\alpha_t(z')\ \mathrm{Cov}_{q_t}\big(D_{x,z}U, D_{x,z'}U\big)\,\nu_x(dz)\,\nu_x(dz').$$

By definition of k_t^{Δ} , $\mathrm{Cov}_{q_t}(D_{x,z}U, D_{x,z'}U) = k_t^{\Delta}(z,z')$.

Covariance $\sigma_{x,c}$. By the same argument,

$$\sigma_{x,c} = \mathrm{Cov}_{q_t}(\Psi_x, D_cU) = \mathrm{Cov}_{q_t}\left(\int \alpha_t(z)\,D_{x,z}U\,\nu_x(dz),\ D_cU\right).$$

The exchange of Cov_{q_t} and the ν_x-integral is justified as above, now for a single integral: $|\alpha_t(z)| \le \bar{\alpha}$ and $|\,\mathrm{Cov}_{q_t}(D_{x,z}U, D_cU)| \le \sqrt{k_t^{\Delta}(z,z)\,\sigma_c^2} \le \sqrt{4K_t\sigma_c^2}$ by Cauchy–Schwarz, a constant, and ν_x is a probability measure. Hence

$$\sigma_{x,c} = \int \alpha_t(z)\ \mathrm{Cov}_{q_t}(D_{x,z}U, D_cU)\,\nu_x(dz) = \int \alpha_t(z)\,k_t^{\Delta}(z,c)\,\nu_x(dz),$$

.

Variance σ_c^2 Immediate from the definition of k_t^{Δ}: $\sigma_c^2 = \mathrm{Var}_{q_t}(D_cU) = k_t^{\Delta}(c,c)$.

This establishes the covariance matrix Σ.

C.2 Two properties

Lemma 2 (Gaussian conditioning) *Let (A,B) be jointly Gaussian with means μ_A, μ_B, variances $\sigma_A^2, \sigma_B^2 > 0$, and covariance σ_{AB}. Then*

$$B \mid A = a\ \sim\ \mathcal{N}\left(\mu_B + \frac{\sigma_{AB}}{\sigma_A^2}(a-\mu_A),\ \ \sigma_B^2 - \frac{\sigma_{AB}^2}{\sigma_A^2}\right).$$

Proof 2 *Define $B' = B - \frac{\sigma_{AB}}{\sigma_A^2}(A-\mu_A)$, a linear combination of A and B, hence jointly Gaussian with A. Its covariance with A is*

$$\mathrm{Cov}(A,B') = \sigma_{AB} - \frac{\sigma_{AB}}{\sigma_A^2}\,\mathrm{Var}(A) = \sigma_{AB} - \sigma_{AB} = 0.$$

For jointly Gaussian random variables, zero covariance implies independence, so $B' \perp A$. Its mean and variance are

$$\mathbb{E}[B'] = \mu_B, \qquad \mathrm{Var}(B') = \sigma_B^2 - 2\frac{\sigma_{AB}}{\sigma_A^2}\sigma_{AB} + \frac{\sigma_{AB}^2}{\sigma_A^4}\sigma_A^2 = \sigma_B^2 - \frac{\sigma_{AB}^2}{\sigma_A^2},$$

so $B' \sim \mathcal{N}\big(\mu_B,\ \sigma_B^2 - \sigma_{AB}^2/\sigma_A^2\big)$, independently of A. Since $B = B' + \frac{\sigma_{AB}}{\sigma_A^2}(A-\mu_A)$, conditioning on $A = a$ leaves B''s law unchanged (by independence) and shifts it by the deterministic term $\frac{\sigma_{AB}}{\sigma_A^2}(a-\mu_A)$, giving the stated conditional law.

Lemma 3 (Probit–Gaussian integral identity) *For $W \sim \mathcal{N}(\mu, s^2)$ and $\lambda > 0$,*

$$\mathbb{E}\big[\Phi(W/\lambda)\big] = \Phi\left(\frac{\mu}{\sqrt{\lambda^2 + s^2}}\right).$$

Proof 3 *Let $Y \sim \mathcal{N}(0,1)$ be independent of W. Since Φ is the standard normal cdf, $\Phi(w/\lambda) = \mathbb{P}(Y \le w/\lambda) = \mathbb{P}(\lambda Y \le w)$ for any fixed w, using $\lambda > 0$. Hence, conditioning on W and using the tower property,*

$$\mathbb{E}\big[\Phi(W/\lambda)\big] = \mathbb{E}\big[\mathbb{P}(\lambda Y \le W \mid W)\big] = \mathbb{P}(\lambda Y \le W) = \mathbb{P}(W - \lambda Y \ge 0).$$

Since W and Y are independent Gaussians, $W - \lambda Y \sim \mathcal{N}(\mu, s^2 + \lambda^2)$. Standardising,

$$\mathbb{P}(W - \lambda Y \ge 0) = \mathbb{P}\left(Z \ge \frac{-\mu}{\sqrt{s^2+\lambda^2}}\right) = 1 - \Phi\left(\frac{-\mu}{\sqrt{s^2+\lambda^2}}\right) = \Phi\left(\frac{\mu}{\sqrt{\lambda^2+s^2}}\right),$$

for $Z \sim \mathcal{N}(0,1)$, using the antisymmetry $1 - \Phi(-t) = \Phi(t)$.

C.3 Closed-form answer probabilities

Recall $V_c \mid U \sim \text{Bernoulli}\big(\Phi(D_c U/\lambda)\big)$ from Equation (6), so V_c's conditional law given U depends on U only through $D_c U$. Since Ψ_x is a deterministic (affine) function of U, $\sigma(\Psi_x) \subset \sigma(U)$, and the tower property gives, for any ψ,

$$\eta_{t,x,c}(\psi) = \mathbb{P}(V_c = 1 \mid \Psi_x = \psi) = \mathbb{E}\big[\mathbb{P}(V_c = 1 \mid U) \mid \Psi_x = \psi\big] = \mathbb{E}\big[\Phi(D_c U/\lambda) \mid \Psi_x = \psi\big].$$

Computing $\eta_{t,x,c}(\psi)$ By C, $(\Psi_x, D_c U)$ is jointly Gaussian with the mean and covariance derived above. Applying Lemma 2 with $A = \Psi_x$, $B = D_c U$,

$$D_c U \mid \Psi_x = \psi \;\sim\; \mathcal{N}\left(m_t^{\Delta}(c) + \frac{\sigma_{x,c}}{\sigma_x^2}\big(\psi - T_x(m_t)\big),\;\; \sigma_c^2 - \frac{\sigma_{x,c}^2}{\sigma_x^2}\right).$$

Writing $\tau_{x,c} = \sigma_{x,c}/\sigma_x$, so that $\sigma_{x,c}/\sigma_x^2 = \tau_{x,c}/\sigma_x$ and $\sigma_{x,c}^2/\sigma_x^2 = \tau_{x,c}^2$, this conditional law is $\mathcal{N}\big(\mu_\psi, s^2\big)$ with

$$\mu_\psi = m_t^{\Delta}(c) + \frac{\tau_{x,c}}{\sigma_x}\big(\psi - T_x(m_t)\big), \qquad s^2 = \sigma_c^2 - \tau_{x,c}^2.$$

Applying Lemma 3 conditionally, with $W = D_c U \mid \Psi_x = \psi$,

$$\begin{aligned}\eta_{t,x,c}(\psi) &= \Phi\left(\frac{\mu_\psi}{\sqrt{\lambda^2+s^2}}\right) = \Phi\left(\frac{m_t^{\Delta}(c) + \tau_{x,c}(\psi - T_x(m_t))/\sigma_x}{\sqrt{\lambda^2 + \sigma_c^2 - \tau_{x,c}^2}}\right) \\ &= \Phi\left(\frac{m_t^{\Delta}(c) + \tau_{x,c}(\psi - T_x(m_t))/\sigma_x}{s_{x,c}}\right),\end{aligned} \tag{20}$$

where $s_{x,c}^2 := \lambda^2 + \sigma_c^2 - \tau_{x,c}^2$.

Computing $p_{t,c}$ By C, the marginal law of $D_c U$ is $\mathcal{N}\big(m_t^{\Delta}(c), \sigma_c^2\big)$. Since $V_c \mid U$ depends on U only through $D_c U$,

$$p_{t,c} = \mathbb{P}(V_c = 1) = \mathbb{E}\big[\Phi(D_c U/\lambda)\big].$$

Applying Lemma 3 with $W = D_c U$, $\mu = m_t^{\Delta}(c)$, $s^2 = \sigma_c^2$,

$$p_{t,c} = \Phi\left(\frac{m_t^{\Delta}(c)}{\sqrt{\lambda^2 + \sigma_c^2}}\right). \tag{21}$$

C.4 Gauss–Hermite quadrature and the final closed form

Remind that

$$\text{EIG}_{\Psi(x)}(c) = \mathbb{H}[V_c \mid \mathcal{D}_t, c] - \mathbb{E}_{\Psi_x \sim \Pi_{t,x}}\big[\mathbb{H}[V_c \mid \Psi_x, \mathcal{D}_t, c]\big]. \tag{22}$$

By Equation (22) and the reduction to binary entropies,

$$\text{EIG}_{\Psi(x)}(c) = h(p_{t,c}) - \mathbb{E}_{\Psi_x \sim \Pi_{t,x}}\big[h(\eta_{t,x,c}(\Psi_x))\big].$$

Under the linearisation of Theorem 1, Ψ_x is Gaussian under q_t with mean $T_x(m_t)$ and variance σ_x^2, so $\Pi_{t,x} = \mathcal{N}(T_x(m_t), \sigma_x^2)$ has density

$$f_{\Psi_x}(\psi) = \frac{1}{\sigma_x}\,\varphi\!\left(\frac{\psi - T_x(m_t)}{\sigma_x}\right), \qquad \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2},$$

so that

$$\mathbb{E}_{\Psi_x\sim\Pi_{t,x}}\big[h(\eta_{t,x,c}(\Psi_x))\big] = \int_{-\infty}^{\infty} h(\eta_{t,x,c}(\psi))\,\frac{1}{\sigma_x}\,\varphi\!\left(\frac{\psi - T_x(m_t)}{\sigma_x}\right) d\psi.$$

Substitute $\beta = \frac{1}{\sqrt{2}} \cdot \frac{\psi - T_x(m_t)}{\sigma_x}$, i.e. $\psi = T_x(m_t) + \sqrt{2}\,\beta\,\sigma_x$ and $d\psi = \sqrt{2}\,\sigma_x\,d\beta$. Then $\frac{1}{\sigma_x}\varphi\big(\frac{\psi - T_x(m_t)}{\sigma_x}\big)\,d\psi = \frac{1}{\sigma_x} \cdot \frac{1}{\sqrt{2\pi}} e^{-\beta^2} \cdot \sqrt{2}\,\sigma_x\,d\beta = \frac{1}{\sqrt{\pi}}\,e^{-\beta^2}\,d\beta$, so

$$\mathbb{E}_{\Psi_x\sim\Pi_{t,x}}\big[h(\eta_{t,x,c}(\Psi_x))\big] = \int_{-\infty}^{\infty} h\big(\eta_{t,x,c}(T_x(m_t) + \sqrt{2}\,\beta\,\sigma_x)\big)\,\frac{1}{\sqrt{\pi}}\,e^{-\beta^2}\,d\beta.$$

Gauss–Hermite quadrature approximates integrals of the form $\int_{-\infty}^{\infty} e^{-\beta^2} g(\beta)\,d\beta \approx \sum_{i=1}^{n} w_i\,g(\beta_i)$, where $(\beta_i)_{i\le n}$ are the roots of the degree-n Hermite polynomial and $(w_i)_{i\le n}$ the corresponding weights (Liu & Pierce, 1994). Applying this with $g(\beta) = h\big(\eta_{t,x,c}(T_x(m_t) + \sqrt{2}\,\beta\,\sigma_x)\big)$,

$$\mathbb{E}_{\Psi_x\sim\Pi_{t,x}}\big[h(\eta_{t,x,c}(\Psi_x))\big] \approx \sum_{i=1}^{n} \frac{w_i}{\sqrt{\pi}}\,h\big(\eta_{t,x,c}(T_x(m_t) + \sqrt{2}\,\beta_i\,\sigma_x)\big).$$

From Equation (20), at $\psi = T_x(m_t) + \sqrt{2}\,\beta_i\,\sigma_x$, the term $\big(\psi - T_x(m_t)\big)/\sigma_x$ simplifies to $\sqrt{2}\,\beta_i$, so

$$\eta_{t,x,c}\big(T_x(m_t) + \sqrt{2}\,\beta_i\,\sigma_x\big) = \Phi\!\left(\frac{m_t^{\Delta}(c) + \sqrt{2}\,\beta_i\,\tau_{x,c}}{s_{x,c}}\right).$$

Substituting back,

$$\mathrm{EIG}_{\Psi(x)}(c) \approx h(p_{t,c}) - \sum_{i=1}^{n} \frac{w_i}{\sqrt{\pi}}\,h\!\left(\Phi\!\left(\frac{m_t^{\Delta}(c) + \sqrt{2}\,\beta_i\,\tau_{x,c}}{s_{x,c}}\right)\right),$$

which is Equation (12).

D TARGETING THE RIGHT INFORMATION

By Equation 9, the linearised estimand is $\Psi_x = c_{t,x} + Z_x$, where $c_{t,x}$ is a constant given D_t and $Z_x = \int \alpha_t(x,z)\,D_{x,z}U\,\nu_x(dz)$ is a single linear functional of U, the α_t-weighted average over the target pool of the pairwise contrasts. Since Ψ_x is a deterministic shift of Z_x, $\mathrm{EIG}_{\Psi(x)}(c) = I(Z_x; V_c \mid D_t, c)$ for every c: CPTE-EIG only sees what a query reveals about this one averaged contrast, never about any individual pair entering it. A query can be informative about every pair in the pool taken separately, which is exactly what EPIG rewards, and still be uninformative about their average. The following example gives such a query, and contrasts it with both EPIG and BALD.

Example 1 (Target the right information) *Let $y = (y^{(1)}, y^{(2)})$ be the outcome, where $y^{(1)}$ records mobility and $y^{(2)}$ records pain relief. Model the utility as linear in y, $u(\tilde{x}, y) = \theta_1 y^{(1)} + \theta_2 y^{(2)}$, with $\theta \sim \mathcal{N}(0, \sigma^2 I)$ under q_t: an instance of the Gaussian process model of Section 4, using a linear kernel and a posterior mean of zero. Contrasts are then plain dot products, $\mathrm{Cov}(\theta \cdot d, \theta \cdot d') = \sigma^2\,d \cdot d'$, for any two directions d, d'.*

Fix a target x with $\nu_x = \frac{1}{2}\delta_{z_1} + \frac{1}{2}\delta_{z_2}$, two pairs sharing the query covariate: at z_1 the treated outcome gains on both dimensions, with contrast $d_1 = (1,1)$; at z_2 it trades mobility for pain relief, with contrast $d_2 = (1,-1)$. Consider a query c with contrast $d_c = (0,1)$, turning entirely on pain relief. Since $m_t = 0$, every posterior mean contrast vanishes, and by Equation 8 z_1 and z_2 receive the same sensitivity weight $\alpha_0 = \varphi(0)/\lambda$.

The contrast the target pool averages over is $\frac{1}{2}(d_1 + d_2) = (1, 0)$: the θ_2 component of z_1 and z_2 cancel, since the two pairs trade mobility against pain relief in opposite directions. Hence $\sigma_{x,c} = \alpha_0\sigma^2\,(1, 0) \cdot (0, 1) = 0$, while $\sigma_x^2 = \alpha_0^2\sigma^2 > 0$ and $\sigma_c^2 = \sigma^2 > 0$, so $\tau_{x,c} = \sigma_{x,c}/\sigma_x = 0$. By Equation 20, $\eta_{t,x,c}(\psi) = \Phi(0)$ for every ψ, equal to $p_{t,c}$: the answer to c tells us nothing about Ψ_x, and $\mathrm{EIG}_{\Psi(x)}(c) = h(p_{t,c}) - h(p_{t,c}) = 0$ by Equation 22.

EPIG does not cancel this way, since it scores z_1 and z_2 separately rather than through their average: $\mathrm{Cov}(\theta \cdot d_1, \theta \cdot d_c) = \sigma^2$ and $\mathrm{Cov}(\theta \cdot d_2, \theta \cdot d_c) = -\sigma^2$, both nonzero, so c is informative about the answer at z_1 and at z_2 on their own, and $\mathrm{EPIG}(c) > 0$. BALD is positive as well, since V_c depends on θ_2, which has positive prior variance and is otherwise unconstrained by D_t. CPTE-EIG is the only one of the three that scores c as uninformative: c resolves only the direction θ_2, exactly the direction the target pool's own trade-off cancels.

E Query selection algorithm

Algorithm 1 CPTE-EIG query selection at round t

Input: Potential-outcome laws $\hat{\mu}_0, \hat{\mu}_1$; target distribution $p^\star$; quadrature order n; Monte Carlo sizes M (target draws) and K (outcome pairs per target)
Output: Next query c_{t+1}

Given: comparisons $\mathcal{D}_t$, candidate pool $\mathcal{C}_t$
Fit the Laplace-approximate posterior $q_t = \mathrm{GP}(m_t, k_t)$ from $\mathcal{D}_t$, with the kernel hyperparameters and the noise scale $\hat{\lambda}_t$ maximising the Laplace marginal likelihood (kernel amplitude fixed to 1)
Draw target covariates $x_1^\star, \ldots, x_M^\star \sim p^\star$, and for each m, outcome pairs $z_1^{(m)}, \ldots, z_K^{(m)} \sim \hat{\mu}_1(\cdot \mid x_m^\star) \otimes \hat{\mu}_0(\cdot \mid x_m^\star)$ approximating $\nu_{x_m^\star}$
for $m = 1, \ldots, M$ **do**
    Compute sensitivity weights $\alpha_t(x_m^\star, z_k^{(m)})$, $k = 1, \ldots, K$, via Equation (8), and estimate $\sigma^2_{x_m^\star}$ by Monte Carlo over $z_1^{(m)}, \ldots, z_K^{(m)}$, via Equation (10)
end for
for each candidate query $c \in \mathcal{C}_t$ **do**
    Compute $m_t^\Delta(c)$, $\sigma_c^2 = k_t^\Delta(c, c)$
    for $m = 1, \ldots, M$ **do**
        Estimate $\sigma_{x_m^\star, c}$ by Monte Carlo over $z_1^{(m)}, \ldots, z_K^{(m)}$, via Equation (11)
        Compute $\tau_{x_m^\star, c}$, $s_{x_m^\star, c}$ and, using the n-point Gauss–Hermite rule, $\mathrm{EIG}_{\Psi(x_m^\star)}(c)$ via Equation (12)
    end for
    $\mathrm{CPTE\text{-}EIG}(c) \leftarrow \frac{1}{M}\sum_{m=1}^{M} \mathrm{EIG}_{\Psi(x_m^\star)}(c)$
end for
return $c_{t+1} \in \arg\max_{c \in \mathcal{C}_t} \mathrm{CPTE\text{-}EIG}(c)$

F Additional experimental details and results

F.1 Data generation

Two datasets. Both datasets are built from PPMI (Marek et al., 2018), but they serve different needs. Eliciting preferences requires outcome profiles a clinician could plausibly be shown, so the queries use real PPMI profiles: we take every pair of distinct observed profiles that share the same query covariate $\tilde{x}$, which gives 1,369,284 candidate queries. Of these, 2,000 are held out as a test set and never queried.

Computing an oracle CPTE, on the other hand, requires both potential outcomes of the same patient, which no real cohort provides. We therefore simulate a cohort of 2,000 patients from PPMI. Covariates are drawn from a multivariate Gaussian fitted to the real baseline covariates and clipped to their observed range. Treatment is assigned by a logistic propensity model fitted on the real cohort, so overlap is realistic. $Y^{(h)}(1)$ and $Y^{(h)}(0)$ are obtained by nearest-neighbour matching into the real

treated (dopamine agonist) and control (levodopa) arms. No method ever sees these ground-truth quantities.

Outcome laws. For a target patient x, let $N_k(x)$ be its $k = 30$ nearest cohort patients under a Gower distance on the full baseline covariates, excluding x itself. We set $\hat{\mu}_a(\cdot \mid x)$ to the empirical distribution of $\{Y_j^{(h)}(a)\}_{j \in N_k(x)}$. The law $\nu_x^{(h)} = \hat{\mu}_1 \otimes \hat{\mu}_0$ of Equation (1) is then uniform over the $k^2 = 900$ (treated, control) pairs, so every integral against $\nu_x^{(h)}$ used for evaluation is computed exactly by enumeration. The acquisition rules only sample from $\nu_x^{(h)}$.

Follow-up horizons. Treatment effects in Parkinson's disease are reported at horizons fixed in advance with clinicians, here $h \in \{1, 3, 5, 7\}$ years. For each h, $Y^{(h)}(0)$ and $Y^{(h)}(1)$ are the profiles recorded at that horizon, and a separate estimand is defined from the same covariates. The target distribution $p^\star$ is uniform over (patient, horizon) pairs, so a single run serves all four horizons.

Treatment effects. We use three versions of the cohort, which differ only in $Y(1)$.

- *Tiny effect:* the matched outcomes above. The two arms have similar outcome laws. Over the 100 evaluation patients, the mean oracle CPTE is 0.49–0.56 depending on the horizon.
- *Mild* ($s = 1$) and *strong* ($s = 2$) effects: $Y(0)$ is kept and

$$Y^{(h)}(1) = \mathrm{clip}\big(Y^{(h)}(0) + s\,\delta(\tilde{x}, h),\, 0,\, 4\big),$$

  where δ is zero on every item except two. Tremor (NP2TRMR) improves by one point, $\delta_{\text{tremor}} = -1$. Cognition (NP1COG) worsens by $\delta_{\text{cog}} = (0.5 + 0.15\,h)(0.5 + a)$, with $a = (\text{age group} - 1)/7 \in [0, 1]$ the normalised age. The harm therefore grows with the time on treatment and with age.

The effect is deterministic given $(\tilde{x}, h)$ and additive on the 0–4 item scale. Clipping creates a floor: a patient without tremor cannot improve on it. Both items vary a lot across the query pool, so the utility weights they depend on can be learned from ordinary queries. The magnitudes were chosen from properties of the oracle CPTE only, so that at every horizon some patients benefit and some do not, never from the performance of any method.

F.2 The simulated clinician

Each query is answered as the model of Section 4 assumes, $V_c \sim \mathrm{Bernoulli}\big(\Phi(D_c u^\star / \lambda^\star)\big)$, with $\lambda^\star = 0.1\sqrt{2}$. The utility is

$$u^\star(\tilde{x}, y) = -\sum_d w_d(\tilde{x}) \sum_{j \in G_d} y_j,$$

with five domains G_d. The weights are positive and were drawn once from fixed ranges:

- neuropsychiatric (NP1COG, NP1HALL, NP1DPRS, NP1APAT, NP1DDS): $w = 0.38 + 1.06\,a$, increasing with age;
- motor (NP2TRMR, NP2WALK, NP2FREZ): $w = 0.36 + 1.20\,(1 - a)$, decreasing with age;
- motor complications (NP4DYSKI, NP4FLCTI): $w = (0.40, 0.64) + 1.15\,\text{sex}$, with sex coded $0/1$;
- sleep and autonomic (NP1SLPD, NP1LTHD): $w = (0.49, 0.38)$;
- quality of life: $w = 1.22$.

Since all weights are positive, $u^\star$ is decreasing in every symptom. Learning $u^\star$ everywhere needs comparisons over all ages and both sexes, while Ψ_x only depends on the weights at $\tilde{x}$, because both profiles of a treated–control pair share the same $\tilde{x}$.

F.3 Methods and protocol

Shared model. All methods use the same model. It is a PairwiseGP from *botorch* library with Laplace approximation (Chu & Ghahramani, 2005b), on inputs rescaled to $[0, 1]$. The kernel is a

mixed kernel $k = \theta\,(k_{\mathrm{RBF}} + k_{\mathrm{cat}}) + (1-\theta)\,k_{\mathrm{RBF}}\,k_{\mathrm{cat}}$ (Ru et al., 2020). k_{RBF} is an ARD RBF kernel on age group, disease duration and the 13 outcomes, with lengthscales at least 0.05, and k_{cat} is a categorical kernel on sex. As in Chu & Ghahramani (2005b), the kernel amplitude is fixed, which is needed for λ to be identifiable.

Noise scale. λ and the kernel amplitude are not separately identifiable. We therefore fix the output scale to 1 and treat $\lambda \in [10^{-3}, 10^{3}]$ as a hyperparameter. At every round, the lengthscales, θ and λ are fitted jointly by maximising the Laplace marginal likelihood with L-BFGS, starting from $\lambda = 1$. The estimate $\hat{\lambda}_t$ is then used everywhere: in the likelihood, in each acquisition score, and in the CPTE estimate. The oracle alone uses $\lambda^\star = 0.1\sqrt{2}$.

Failed fits. If the optimiser fails, we refit with λ held at the previous round's estimate, and if that also fails, the round's query is drawn at random from the pool. The same rule applies to every method and to the final evaluation fit. Across all runs, fewer than 1% of fits needed the held-λ refit, and a random query was used 10 times in total, spread across methods.

Acquisition rules. We write V_c for the answer to a query c. All methods start from the same 20 random comparisons, so they coincide at the smallest budget. At each later round, each method scores 2,000 unlabelled comparisons drawn at random from the pool and queries the best one.

- *Random* queries a comparison drawn uniformly from the pool.
- *BALD* (Houlsby et al., 2011) scores the information the answer brings about the utility, $I(V_c; u \mid \mathcal{D}_t)$. We use BoTorch's BALD for pairwise GPs (`PairwiseBayesianActiveLearningByDisagreement`), with the learned $\hat{\lambda}_t$ in place of its fixed $\lambda = 1$, and its Monte Carlo expectation replaced by the deterministic 40-node Gauss–Hermite rule used for CPTE-EIG, which removes sampling noise.
- *EPIG* (Smith et al., 2023) scores the information the answer brings about the answers to target queries. We give it the treated–control queries behind the CPTE. For a target $x \sim p^\star$ and a pair $z = (z_1, z_0) \sim \nu_x$, let $c^\star(x, z)$ compare the treated profile $(\tilde{x}, z_1)$ with the control profile $(\tilde{x}, z_0)$. Its answer $V_{c^\star(x,z)}$ says whether the clinician prefers the treated outcome, and Ψ_x is the average of its probability over $z \sim \nu_x$. Then $\mathrm{EPIG}(c) = \mathbb{E}_{x\sim p^\star,\, z\sim\nu_x}\big[I(V_c; V_{c^\star(x,z)} \mid \mathcal{D}_t)\big]$, $\mathrm{CPTE\text{-}EIG}(c) = \mathbb{E}_{x\sim p^\star}\big[I(V_c; \Psi_x \mid \mathcal{D}_t)\big]$. Both criteria use the same targets, drawn from the same law. They differ only in what they aim to learn about them: each individual answer for EPIG, and their average Ψ_x for CPTE-EIG. EPIG uses 200 target queries per round, each from a fresh draw of x and z, and estimates each mutual information with 512 Monte Carlo samples from the joint GP posterior at c and $c^\star$.
- *CPTE-EIG* follows Algorithm 1, with $M = 20$ targets $x \sim p^\star$ and $K = 20$ pairs $z \sim \nu_x$ per target, drawn at every round, and an $n = 40$-node Gauss–Hermite rule.

Budget and repeats. Budgets range from 20 to 150 comparisons, with 15 seeds per budget. After n comparisons, the GP is refitted on all of them and evaluated.

F.4 Metrics

All CPTE metrics use the same 100 evaluation patients, drawn once from the cohort and shared by all seeds and methods. The estimate is $T_x^{(h)}(m_t) = \frac{1}{k^2}\sum_{i,j\in N_k(x)} \Phi\big(m_t^{\Delta}(x, z_{ij})/\hat{\lambda}_t\big)$, and the oracle $\Psi_x^{(h)}$ is the same average with $u^\star$ and $\lambda^\star$, both over the 900 pairs of $\nu_x^{(h)}$.

- *CPTE MSE:* the mean of $(T_x^{(h)} - \Psi_x^{(h)})^2$ over the 100 patients, per horizon.
- *Policy regret:* $V(\pi^\star) - V(\pi_t)$, where $V(\pi) = \frac{1}{100}\sum_x \big[\pi(x)\Psi_x + (1-\pi(x))(1-\Psi_x)\big]$ and $\pi_t(x) = \mathbf{1}\{T_x(m_t) > 1/2\}$. Each wrong decision is weighted by $|2\Psi_x - 1|$, so errors on clear-cut patients cost more.
- *Held-out prediction:* accuracy, Brier score and negative log-likelihood of the predicted answers on the 2,000 test comparisons, against answers simulated from the same clinician.

Tables and bands report the mean over the 15 seeds with a 95% percentile-bootstrap confidence interval for the mean (3,000 resamples).

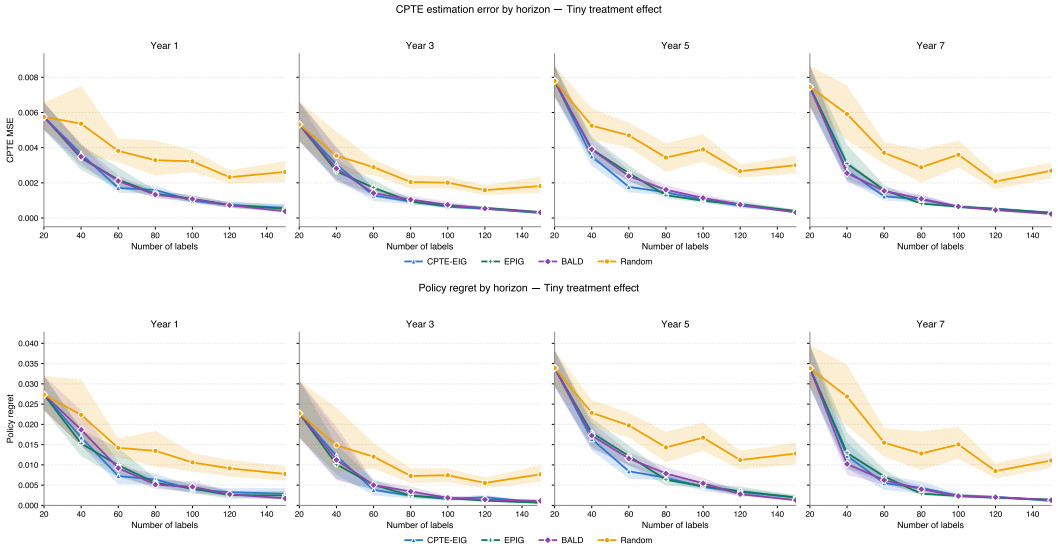


Figure S1: Tiny treatment effect. CPTE MSE (top) and policy regret (bottom) per horizon, as in Figure 2.

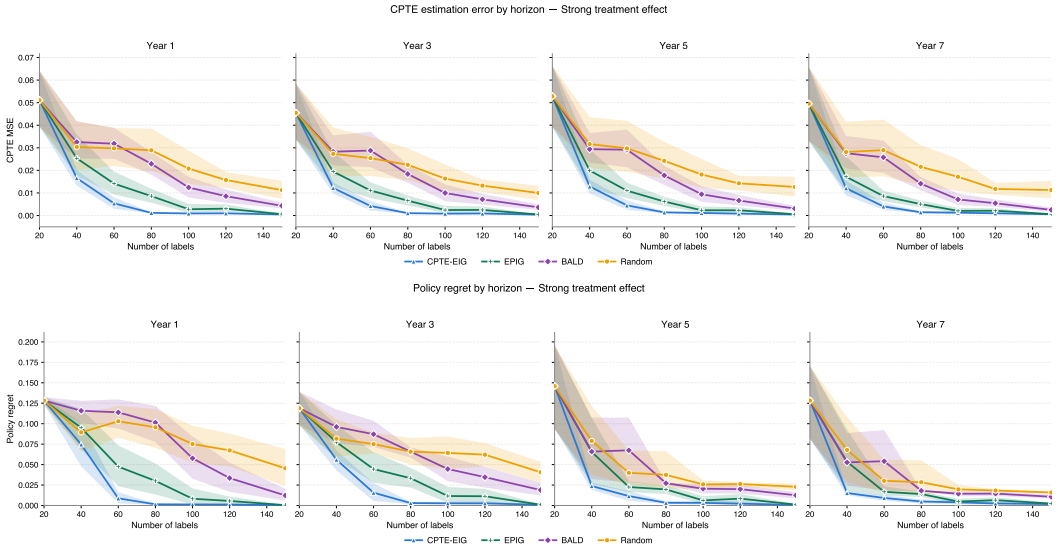


Figure S2: Strong treatment effect. CPTE MSE (top) and policy regret (bottom) per horizon, as in Figure 2.

F.5 Tiny and strong treatment effects

With a strong effect (Figure S2), the ordering of the mild case holds at every horizon and every budget from 40 comparisons on, for both metrics, with wider gaps. At 80 comparisons, CPTE-EIG's regret is 0.003, against 0.024 for EPIG and 0.053 for BALD. On CPTE MSE, it beats BALD on at least 14 of the 15 seeds at every budget from 40 on, and EPIG on 12 of 15 seeds at 80.

With a tiny effect (Figure S1), CPTE-EIG, EPIG and BALD overlap at every horizon. The estimand is easy here: after the 20 initial random comparisons, the CPTE MSE is already 6.6×10^{-3}, against

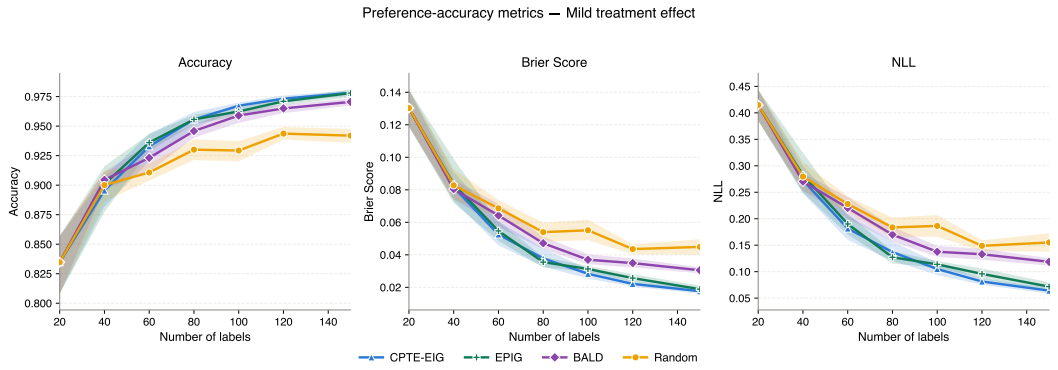


Figure S3: Mild treatment effect. Accuracy (higher is better), Brier score and negative log-likelihood (lower is better) of the predicted answers on the 2,000 held-out comparisons. These measure the fitted preference rule, not the estimand.

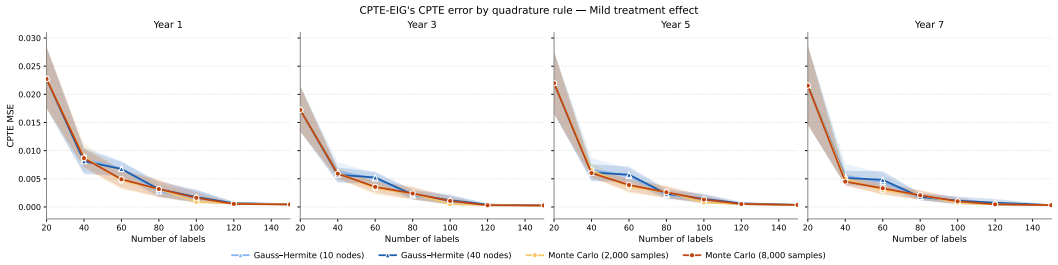


Figure S4: CPTE MSE against the oracle, per horizon, for four rules approximating the expectation in Equation (22)–Equation (12): Gauss–Hermite quadrature with $n = 10$ and $n = 40$ nodes, and Monte Carlo with 2,000 and 8,000 samples of Ψ_x.

2.1×10^{-2} with a mild effect and 5.0×10^{-2} with a strong one. Little is left for any acquisition rule to gain.

F.6 Learning w is not learning Ψ_x

Figure S3 reports held-out prediction under the mild effect. CPTE-EIG and EPIG are within 0.01 accuracy of each other at every budget, and BALD is at most 0.01 behind them from 80 comparisons on. On CPTE MSE, the same methods differ by factors of 2 to 5 between 40 and 120 comparisons (Table 2).

BALD and Random never use the outcomes of the target patients, so their queries and fitted models are identical under the three treatment effects. We checked that their held-out metrics coincide exactly across the three runs. Only their CPTE error changes, because the estimand does. The same fitted utility can thus be good for one causal question and poor for another. What matters is how well it is known along the contrasts Ψ_x depends on.

F.7 Gauss–Hermite quadrature versus Monte Carlo

Equation 12 approximates the remaining expectation over $\Psi_x \sim \Pi_{t,x}$ by an n-point Gauss–Hermite rule rather than by direct Monte Carlo sampling of Ψ_x. We check that this choice does not trade accuracy for speed. We rerun CPTE-EIG with four rules for this one expectation, all else unchanged: Gauss–Hermite with $n = 10$ and $n = 40$ nodes, and plain Monte Carlo with 2,000 and 8,000 samples from $\mathcal{N}(T_x(m_t), \sigma_x^2)$.

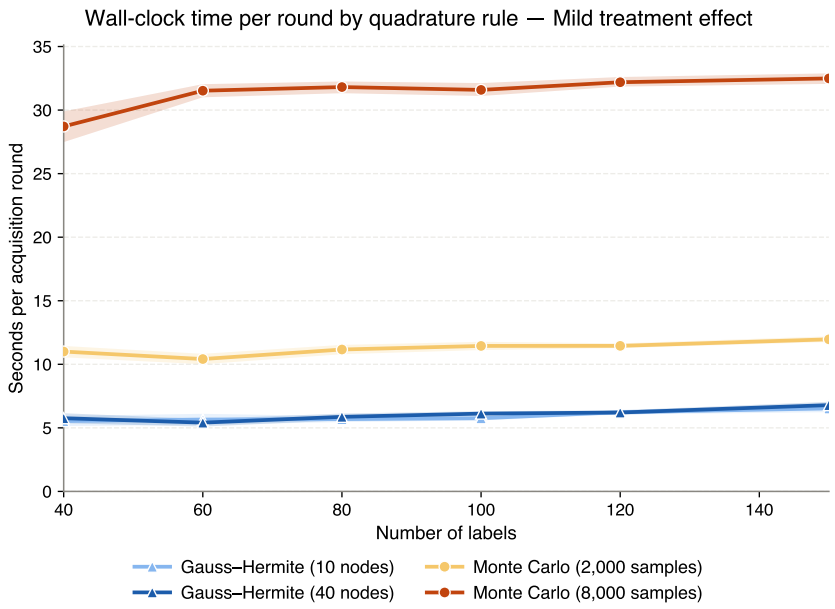


Figure S5: Median wall-clock time per acquisition round (log scale), for the same four rules as Figure S4.

Figure S4 shows the four rules give statistically indistinguishable CPTE MSE at every budget and horizon: even $n = 10$ nodes match 8,000 Monte Carlo samples. This is expected. Under the linearisation of Theorem 1, $\Pi_{t,x}$ is Gaussian, and Gauss–Hermite quadrature is the natural rule for integrating against a Gaussian density, so few nodes already suffice. The two rules differ in cost, however (Figure S5). Gauss–Hermite is several times cheaper than Monte Carlo at every budget. We use $n = 40$ nodes in all other experiments: the same accuracy as Monte Carlo, at a fraction of its cost.