Observationally Informed Adaptive Causal Experimental Design

Erdun Gao
Responsible AI Research Centre, Adelaide University
Adelaide, Australia
erdun.gao@adelaide.edu.au

Liang Zhang*
Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
liangzhang@hkust-gz.edu.cn

Jake Fawkes
University College London
London, United Kingdom
j.fawkes@ucl.ac.uk

Aoqi Zuo
The University of Sydney
Sydney, Australia
aoqi.zuo@sydney.edu.au

Wenqin Liu
The University of Melbourne
Melbourne, Australia
wenqinl@student.unimelb.edu.au

Haoxuan Li
Peking University
Beijing, China
hxli@stu.pku.edu.cn

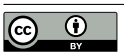
Mingming Gong
The University of Melbourne
Melbourne, Australia
mingming.gong@unimelb.edu.au

Dino Sejdinovic
Responsible AI Research Centre, Adelaide University
Adelaide, Australia
dino.sejdinovic@adelaide.edu.au

Abstract

Randomized Controlled Trials (RCTs) represent the gold standard for causal inference yet remain a scarce resource. While large-scale observational data is often available, it is typically used only for retrospective fusion, and remains discarded in prospective trial design due to bias concerns. We argue that this "tabula rasa" data acquisition strategy is inefficient when observational models contain useful structural information. In this work, we propose *Active Residual Learning*, a new paradigm that leverages the observational model as an informative but biased prior. This shifts the experimental focus from learning target causal quantities from scratch to estimating residual corrections that debias the observational model. To operationalize this, we introduce the R-Design framework. Theoretically, we characterize two key advantages: (1) a conditional structural efficiency gap, showing that estimating lower-complexity residual contrasts can admit faster convergence rates than reconstructing full outcomes; and (2) information efficiency, where we quantify the redundancy in standard parameter-based acquisition, demonstrating that such baselines can waste budget on task-irrelevant nuisance uncertainty. We propose R-EPIG (Residual Expected Predictive Information Gain), a unified criterion that directly targets the downstream causal quantity, reducing residual uncertainty for estimation or clarifying decision boundaries for policy. Experiments on synthetic and semi-synthetic benchmarks show that R-Design significantly outperforms baselines in the intended informative-but-biased regime, supporting the effectiveness of correcting a biased model rather than learning from scratch.

*Corresponding author.



KDD 2026, Jeju Island, Republic of Korea.

ACM ISBN 979-8-4007-2259-2/2026/08
https://doi.org/10.1145/3770855.3817762

CCS Concepts

• **Computing methodologies** → **Machine learning**.

Keywords

Treatment Effect Estimation; Causal Experimental Design; Expected Predictive Information Gain

ACM Reference Format:
Erdun Gao, Liang Zhang, Jake Fawkes, Aoqi Zuo, Wenqin Liu, Haoxuan Li, Mingming Gong, and Dino Sejdinovic. 2026. Observationally Informed Adaptive Causal Experimental Design. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD 2026), August 9–13, 2026, Jeju Island, Republic of Korea.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3770855.3817762

1 Introduction

Precise estimation of individual treatment effects [46] is fundamental to personalized decision-making in domains ranging from economics [28] and recommendation systems [18] to healthcare [16]. While the Average Treatment Effect (ATE) offers a coarse summary, effective policy requires the granularity of the Conditional Average Treatment Effect (CATE), which quantifies how interventions vary across individuals [38]. However, accurately estimating CATE is fundamentally constrained by the data dichotomy. Large-scale observational studies offer population representativeness [4] but suffer from hidden confounding [40], whereas Randomized Controlled Trials (RCTs) guarantee unbiasedness but are crippled by prohibitive costs and limited sample sizes [9, 14, 17]. This trade-off has motivated a growing literature on *post-hoc data fusion* [8, 12, 35, 50]. While these methods demonstrate the benefits of integrating both modalities, they remain fundamentally *retrospective* [7, 44, 49]. They treat experimental data collection as fixed, repairing statistical power after the fact. Consequently, this passive paradigm fails to optimize acquisition, leaving the synergy between observational priors and experimental design largely unexplored.

Motivation. In many real-world application domains, large amounts of observational data are frequently available prior to

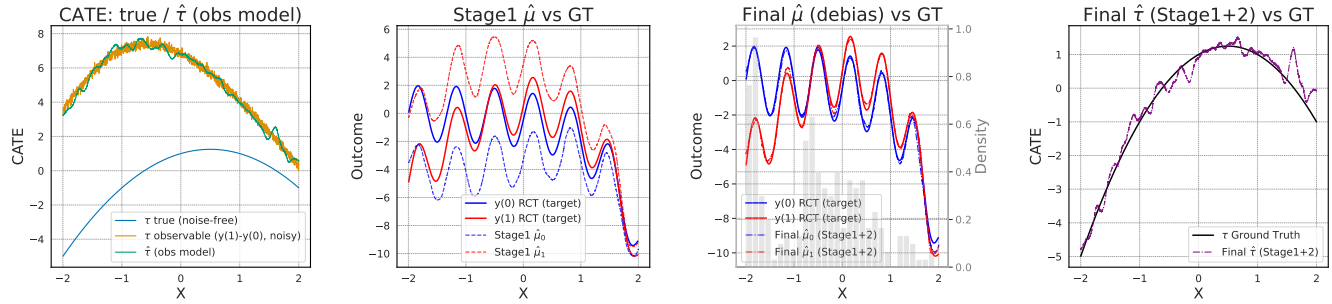


Figure 1: R-Design intuition on 1D synthetic data. In this example, the observational prior is biased (Left) but captures high-frequency structure (Middle-Left). Treating it as a fixed offset, R-Design learns a lower-complexity residual correction (Middle-Right), exploiting this complexity gap to recover the true CATE with few samples (Right), while the naive observational model fails due to uncorrected bias.

experimentation [11, 41]. Yet, standard causal experimental design for CATE estimation typically adopts a tabula rasa approach, ignoring this prior information to learn causal mechanisms from scratch [1, 22]. We argue that this can be statistically inefficient when the observational source is informative despite being biased. First, it overlooks the target population distribution delineated by the large-scale observational data, which characterizes the real-world covariate support where the causal estimand is ultimately evaluated. Second, it discards the observational model itself, which, despite hidden confounding, may capture substantial shared or global structure of the outcomes [27, 35]. This suggests an opportunity to transform experimental design. Rather than treating the confounded observational model as a nuisance to be discarded, it can serve as an informative but biased prior when it captures useful structural information (as illustrated in Fig. 1). This perspective motivates the core research question:

> **Key Question:** How can we leverage the confounded observational prior to guide causal experimental design, shifting the goal from exploring outcomes from scratch to adaptively learning the necessary corrections for the observational bias?

Challenges. Operationalizing this adaptive correction strategy is non-trivial and faces three fundamental challenges. (1) *Disentangling Signal from Bias.* The observational prior intrinsically conflates rich structural information (signal) with hidden confounding (bias). The difficulty lies in designing a mechanism that selectively transfers useful global structure without blindly inheriting confounding errors. This also requires recognizing the regime boundary: if the observational model is too weak or misspecified, the residual may be no easier to learn than the original response. (2) *Objective Misalignment.* There is an inherent discrepancy between identifying the global observational bias and maximizing downstream performance. The regions where the observational model is most biased (high error magnitude) may not overlap with the target population or the decision boundaries. A naive design might waste budget correcting biases in irrelevant regions. (3) *Lack of Residual-Aware Acquisition.* Standard causal experimental design metrics are ill-suited here. While Bayesian strategies can incorporate priors, they typically target the uncertainty of the full treatment effect. Consequently, they fail to focus the budget strictly on learning the residual contrast required for bias correction.

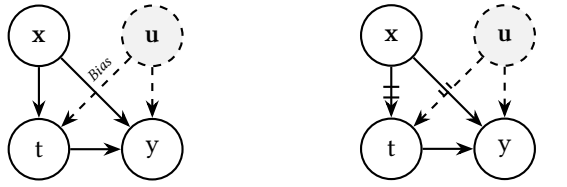
Contributions. To address these gaps, this paper makes the following contributions. (1) *A New Paradigm.* We formalize observationally informed adaptive causal experimental design, where informative but biased observational data shifts the design objective from learning outcomes from scratch to correcting observational bias. (2) *The R-Design Framework.* We propose a unified framework comprising R-EPIG (Sec. 4.1), a task-aware information-theoretic acquisition criterion, and the Two-Stage Residual (TSR) strategy (Sec. 4.2), which decouples base estimation from residual uncertainty modeling. (3) *Theoretical Foundations.* We provide four theoretical results (Sec. 5): a conditional *Structural Efficiency Gap* showing faster rates when residuals have lower effective complexity (Lemma 5.1); *Objective Alignment* with Bayesian CATE risk minimization (Prop. 5.2); *Information Redundancy* of parameter-based baselines (Prop. 5.3); and a TAL-style posterior-uncertainty convergence guarantee for a global residual-target information objective, governed by the residual information capacity (Thm. 5.4). (4) *Extensive Empirical Validation.* We evaluate R-Design on synthetic and semi-synthetic benchmarks (Sec. 6), showing improved sample efficiency over strong baselines in the intended informative-but-biased observational regime.

2 Preliminaries and Problem Setup

In this section, we establish the mathematical foundations of our framework. We first formalize the causal inference problem and the CATE estimand within the potential outcomes framework. Subsequently, we review the necessary background on disparate data integration and the specific active learning setup used in adaptive causal experimental design.

2.1 Causal Estimands and Targets

Potential Outcome Framework. Our analytical framework is grounded in the Neyman-Rubin potential outcomes model [10, 42]. Let $\mathbf{x} \in \mathcal{X}$ denote the covariates, $\mathrm{t} \in \{0, 1\}$ the binary treatment assignment, and $\mathrm{y} \in \mathcal{Y}$ the observed outcome. We use $\boldsymbol{x}$, t, and y to denote realizations of these variables. The two potential outcomes



(a) Observational ($s = o$) (b) Experimental ($s = e$)

Figure 2: Causal diagrams of data-generating processes.

are $\mathrm{y}(0)$ and $\mathrm{y}(1)$, corresponding to the potential outcome under control and treatment, respectively. To accommodate diverse downstream objectives, we define a general target quantity of interest, denoted by $\Phi(\boldsymbol{x})$, evaluated over a target population distribution $p_{\text{tar}}(\boldsymbol{x})$ (typically the observational population $p_o(\boldsymbol{x})$ [8, 35]). We use a hat to denote an estimator of a target quantity, e.g., $\hat{\Phi}$. In this paper, we focus on two primary specifications:

CATE Estimation. When the primary objective is to quantify the heterogeneity of treatment effects across the population, we target the Conditional Average Treatment Effect (CATE). Formally, let $\mu(\boldsymbol{x}, t) \coloneqq \mathbb{E}[\mathrm{y}(t) \mid \mathbf{x} = \boldsymbol{x}]$ denote the expected potential outcome surface for arm t. The CATE is defined as the pointwise difference between the treated and control response surfaces:

$$\tau(\boldsymbol{x}) \coloneqq \mu(\boldsymbol{x}, 1) - \mu(\boldsymbol{x}, 0). \tag{1}$$

This estimand captures how the treatment effect varies as a function of the covariates $\boldsymbol{x}$. The estimation quality is measured by the Precision in Estimation of Heterogeneous Effects (PEHE) [30], which quantifies the expected squared error (or L_2 risk) of the estimator over the target distribution:

$$\epsilon_{\text{PEHE}}(\hat{\tau}) = \mathbb{E}_{\mathbf{x}\sim p_{\text{tar}}(\cdot)}\left[(\hat{\tau}(\mathbf{x}) - \tau(\mathbf{x}))^2\right]. \tag{2}$$

Decision Making. Alternatively, the objective may be policy optimization, where the goal is to determine the optimal treatment assignment for each unit to maximize the total population welfare [13]. This is equivalent to identifying the *sign* of the treatment effect rather than estimating its continuous magnitude. The target estimand is the binary optimal policy $\Phi(\boldsymbol{x}) = \pi(\boldsymbol{x}) = \mathbb{I}(\tau(\boldsymbol{x}) > 0)$, which prescribes the treatment arm yielding the higher expected outcome. Unlike CATE estimation, this task implies that high precision is primarily required near the decision boundary (where $\tau(\boldsymbol{x}) \approx 0$). The performance is measured by the Average Policy Error (APE) [19], defined as the probability of assigning a sub-optimal treatment (misclassification rate) over the target population:

$$\epsilon_{\text{APE}}(\hat{\pi}) = \mathbb{E}_{\mathbf{x}\sim p_{\text{tar}}(\cdot)}\left[\mathbb{I}(\hat{\pi}(\mathbf{x}) \neq \pi(\mathbf{x}))\right]. \tag{3}$$

2.2 Disparate Data Sources

We consider a scenario involving two distinct data sources: a large observational study (OS) and a smaller experimental dataset, as shown in Fig. 2. We denote the source using a random variable $\mathrm{s} \in \{o, e\}$. For each source s, we define the source-specific conditional means as $\mu_s(\boldsymbol{x}, t) \coloneqq \mathbb{E}[\mathrm{y} \mid \mathbf{x} = \boldsymbol{x}, \mathrm{t} = t, \mathrm{s} = s]$. The source-specific contrast for source s is $\tau_s(\boldsymbol{x}) \coloneqq \mu_s(\boldsymbol{x}, 1) - \mu_s(\boldsymbol{x}, 0)$. This quantity measures the observed conditional mean difference. For the OS, the contrast $\tau_o(\boldsymbol{x})$ is generally biased due to hidden confounding [12]. For the experimental source, the controlled assignment mechanism ensures that $\tau_e(\boldsymbol{x})$ identifies the true CATE within the experimental support $\mathcal{X}_e \subseteq \mathcal{X}$:

$$\tau(\boldsymbol{x}) = \tau_e(\boldsymbol{x}), \quad \text{for } \boldsymbol{x} \in \mathcal{X}_e. \tag{4}$$

To integrate these disparate sources, define the arm-level residual $\delta(\boldsymbol{x}, t) \coloneqq \mu_e(\boldsymbol{x}, t) - \mu_o(\boldsymbol{x}, t)$. We characterize the structural relationship between the biased observational contrast $\tau_o(\boldsymbol{x})$ and the true causal effect $\tau(\boldsymbol{x})$ through the following decomposition:

$$\tau(\boldsymbol{x}) = \tau_o(\boldsymbol{x}) + \tau_\delta(\boldsymbol{x}), \quad \text{where } \tau_\delta(\boldsymbol{x}) \coloneqq \delta(\boldsymbol{x}, 1) - \delta(\boldsymbol{x}, 0). \tag{5}$$

Here, $\tau_\delta(\boldsymbol{x})$ represents the residual contrast (debiasing correction) that captures the discrepancy between the observational correlations and the underlying causal mechanism. To ensure that this decomposition yields a valid and identifiable causal target, we impose the following assumptions on the data generation mechanisms.

Assumption 2.1. *Stable Unit Treatment Value Assumption:* $\mathrm{y} = \mathrm{ty}(1) + (1-\mathrm{t})\mathrm{y}(0)$, and no interference between units. *Unconfoundedness by Design:* In the experiment, the treatment assignment is controlled by the design policy and may depend on observed covariates and past history, but not on unobserved potential outcomes, ensuring $\mathrm{t} \perp \{\mathrm{y}(0), \mathrm{y}(1)\} \mid (\mathbf{x}, \mathcal{H}_k, \mathrm{s} = e)$. *Positivity of the Design Space:* For every $\boldsymbol{x}$ in the region of interest, both interventions are feasible experimental actions; that is, $(\boldsymbol{x}, 0)$ and $(\boldsymbol{x}, 1)$ are admissible queries whenever $\boldsymbol{x} \in \mathcal{X}_e$. *Effect Invariance (Transportability):* The underlying causal mechanism is invariant across sources, i.e., $\forall s \in \{o, e\}, \mathbb{E}[\mathrm{y}(1) - \mathrm{y}(0) \mid \mathbf{x} = \boldsymbol{x}, \mathrm{s} = s] = \tau(\boldsymbol{x})$. This rules out unnatural shifts in CATE where the experimental protocol itself, rather than the treatment, interferes with the outcome mechanism. *Support Inclusion:* The support of the target population is contained within the feasible experimental region, $\mathrm{supp}(p_{\text{tar}}) \subseteq \mathcal{X}_e$.

Assump. 2.1 ensures that $\tau_\delta(\boldsymbol{x})$ is a structurally consistent quantity representing pure observational bias. Transportability prevents $\tau_\delta(\boldsymbol{x})$ from mixing confounding correction with cross-source CATE shift, while support inclusion ensures that the target population is experimentally reachable. If either condition fails, the target considered here may no longer be identifiable without additional transport or extrapolation assumptions.

2.3 Problem Setup

Unlike retrospective fusion where data collection is fixed, we address the dynamic allocation of a limited experimental budget n_B. We assume access to a static, labeled observational dataset $\mathcal{D}_O = \{(\boldsymbol{x}_i, t_i, y_i, s_i = o)\}_{i=1}^{n_O}$. We aim to optimize the CATE estimator for a target population, characterized by the marginal covariate distribution $p_{\text{tar}}(\boldsymbol{x})$. Following standard conventions, we assume the observational covariates are representative of this target (i.e., the empirical distribution of $\mathcal{D}_O$ approximates p_{tar}) [8, 35, 47]. For the experimental phase, we have a finite candidate pool of recruitable units $\mathcal{D}_P = \{\boldsymbol{x}_j\}_{j=1}^{n_P}$. The experimental dataset $\mathcal{D}_E$ is initially empty. The design process proceeds in sequential stages $k = 1, \ldots, n_B$. At each stage: ① The learner jointly selects a unit $\boldsymbol{x}_k \in \mathcal{D}_P$ and a treatment $t_k \in \{0, 1\}$; ② The learner observes the experimental outcome y_k; ③ The datasets are updated: $\mathcal{D}_E^{(k)} \leftarrow \mathcal{D}_E^{(k-1)} \cup \{(\boldsymbol{x}_k, t_k, y_k, \mathrm{s}_k = e)\}$ and the unit is removed from the candidate pool $\mathcal{D}_P$.

Design Objective. The central challenge of this task is to design a principled utility function, $U(\cdot \mid \mathcal{D}_O, \mathcal{D}_E^{(k-1)})$, that quantifies the information gain of a specific query $(\boldsymbol{x}, t)$. The acquisition strategy at stage k maximizes this utility over the joint product space:

$$(\boldsymbol{x}_k^*, t_k^*) = \underset{(\boldsymbol{x},t)\in\mathcal{D}_P\times\{0,1\}}{\arg\max} \; U(\boldsymbol{x}, t \mid \mathcal{D}_O, \mathcal{D}_E^{(k-1)}). \tag{6}$$

3 Principles for Observationally Informed Causal Experimental Design

From Outcome Exploration to Residual Correction. Standard experimental design typically adopts a *tabula rasa* approach, expending budget to learn the full outcome surface from scratch. We argue this is structurally inefficient when an informative observational prior is available. Given the decomposition $\tau(\boldsymbol{x}) = \hat{\tau}_o(\boldsymbol{x}) + \tau_\delta(\boldsymbol{x})$ where the observational contrast $\hat{\tau}_o(\boldsymbol{x})$ is estimated prior to experimentation and then frozen, the conditional entropy of the total effect is equivalent to that of the residual contrast, conditional on $\mathcal{D}_O$: $\mathrm{H}(\tau(\boldsymbol{x}) \mid \mathcal{D}_O, \mathcal{D}_E) \equiv \mathrm{H}(\tau_\delta(\boldsymbol{x}) \mid \mathcal{D}_O, \mathcal{D}_E)$. Consequently, acquisition should focus on the uncertainty that remains in $\tau_\delta(\boldsymbol{x})$ rather than re-learning the pre-estimated signal $\hat{\tau}_o(\boldsymbol{x})$. This structural equivalence dictates our first principle:

Principle 3.1 (The Residual Uncertainty Principle)**.** *Experimental design should target the epistemic uncertainty of the residual contrast $\tau_\delta(\boldsymbol{x})$. The acquisition function should treat the estimated observational contrast $\hat{\tau}_o(\boldsymbol{x})$ as a fixed mean function, utilizing the budget to resolve the structural discrepancy between the biased prior and the interventional truth, rather than re-learning the full outcome surface.*

This principle is most beneficial when the residual correction has lower effective complexity than the full response.

From Global Exploration to Estimand Alignment. While Principle 3.1 identifies the *structural target* (the residual), it remains silent on the *spatial allocation* of the budget. Classical active learning criteria typically aim to minimize model variance globally, implicitly assuming a uniform utility over the domain [32]. However, causal inference is intrinsically objective-driven; its validity is defined solely relative to a specific estimand (e.g., CATE or Policy) and a target population $p_{\text{tar}}(\boldsymbol{x})$ [29, 43]. Consequently, reducing residual uncertainty in regions that are irrelevant to the target density or inconsequential for the decision boundary provides negligible value. Pure epistemic uncertainty does not equate to utility.

Principle 3.2 (Estimand-Aligned Acquisition)**.** *Optimal experimental design must be risk-centric rather than uncertainty-centric. The utility of a design point $(\boldsymbol{x}, t)$ is not intrinsic, but is quantified solely by the expected reduction in the target-specific risk. This mandates that the experimental budget be concentrated non-uniformly on regions where the residual uncertainty is both high and consequential for the specified estimand.*

4 The R-Design Framework

In this section, we introduce R-Design, a framework that transforms causal experimental design into active residual learning. Leveraging the decomposition $\tau(\boldsymbol{x}) = \hat{\tau}_o(\boldsymbol{x}) + \tau_\delta(\boldsymbol{x})$, where the observational contrast $\hat{\tau}_o$ acts as a pre-computed offset, R-Design explicitly targets the identification of the residual τ_δ. We first derive the residual-based acquisition criterion, R-EPIG, and then present the Active Learning (AL) strategy that enables its tractable computation.

4.1 The R-Design Criterion: R-EPIG

Our framework operationalizes Principle 3.2 by explicitly targeting the epistemic uncertainty of the specific downstream target $\Phi(\cdot)$. Critically, since the observational base $\hat{\mu}_o(\boldsymbol{x}, t)$ is treated as a fixed mean function, the information content of an experimental query $(\boldsymbol{x}, t)$ resides solely in the *residual outcome*, denoted as the random variable $\mathrm{r} \coloneqq \mathrm{y} - \hat{\mu}_o(\boldsymbol{x}, t)$. We propose the Residual Expected Predictive Information Gain (R-EPIG). Unlike standard criteria operating on raw outcomes [43], R-EPIG selects experiments to maximize the expected information gain regarding the target $\Phi(\cdot)$ specifically via the observed residual r. Let $\mathcal{H}_k$ be the history at step k. The general acquisition objective is:

$$\alpha_{\text{R-EPIG}}(\boldsymbol{x}, t) \coloneqq \mathbb{E}_{\mathbf{x}^*\sim p_{\text{tar}}(\cdot)}\Big[\mathrm{I}\big(\mathrm{r};\, \Phi(\mathbf{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k\big)\Big]. \tag{7}$$

Here, the randomness in the target $\Phi(\cdot)$ is induced solely by the posterior of the residual processes. The outer expectation averages this gain over the target population. As with other posterior-based acquisition rules, R-EPIG relies on usable residual uncertainty; if the residual posterior is poorly calibrated, targeted acquisition can become less reliable.

Remark 4.1. While $\mathrm{I}(\mathrm{r}; \Phi) \equiv \mathrm{I}(\mathrm{y}; \Phi)$, formulating the objective over r aligns with our Residual GP. This structurally decouples inference from $\mathcal{D}_O$, ensuring complexity scales solely with the experimental budget n_B rather than the observational size n_O.

By specifying $\Phi(\cdot)$, this framework yields two task-specific strategies: *R-EPIG-Est* for estimation and *R-EPIG-Policy* for policy.

4.1.1 Design for Estimation (R-EPIG-Est). When the goal is minimizing PEHE, we target the magnitude of the correction terms. We propose two specifications:

Joint-Outcome Residuals (R-EPIG-μ). To recover the full potential outcome surfaces, we define the target as the *residual vector* $\Phi(\boldsymbol{x}) = \boldsymbol{\delta}(\boldsymbol{x}) \coloneqq [\delta(\boldsymbol{x}, 0), \delta(\boldsymbol{x}, 1)]^\top$. Since the observational base $\hat{\mu}_o$ is treated as a fixed offset, reducing uncertainty in $\boldsymbol{\delta}(\boldsymbol{x})$ directly reduces uncertainty in the absolute potential outcomes $\mu_e(\boldsymbol{x}, t)$. The criterion is:

$$\alpha^{\mu}_{\text{R-EPIG}}(\boldsymbol{x}, t) \coloneqq \mathbb{E}_{\mathbf{x}^*\sim p_{\text{tar}}(\cdot)}\Big[\mathrm{I}\big(\mathrm{r};\, \boldsymbol{\delta}(\mathbf{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k\big)\Big]. \tag{8}$$

This strategy is conservative but robust: it identifies regions where the observational prior is unreliable for either arm.

Direct CATE Residuals (R-EPIG-τ). To learn the CATE, we can define the target strictly as the residual contrast $\Phi(\boldsymbol{x}) = \tau_\delta(\boldsymbol{x}) \coloneqq \delta(\boldsymbol{x}, 1) - \delta(\boldsymbol{x}, 0)$. The criterion is:

$$\alpha^{\tau}_{\text{R-EPIG}}(\boldsymbol{x}, t) \coloneqq \mathbb{E}_{\mathbf{x}^*\sim p_{\text{tar}}(\cdot)}\Big[\mathrm{I}\big(\mathrm{r};\, \tau_\delta(\mathbf{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k\big)\Big]. \tag{9}$$

This variant can be more sample-efficient when the residual contrast is lower-complexity. By targeting the difference, it naturally ignores shared baseline errors that cancel out in the subtraction, avoiding budget waste on variations irrelevant to the treatment effect.

4.1.2 Design for Policy (R-EPIG-Policy). When minimizing APE, the estimand collapses to the binary decision boundary. We define the target as the policy indicator $\Phi(\boldsymbol{x}) = \pi(\boldsymbol{x}) \coloneqq \mathbb{I}(\hat{\tau}_o(\boldsymbol{x}) + \tau_\delta(\boldsymbol{x}) > 0)$. The criterion targets the information gain regarding this sign:

$$\alpha^{\pi}_{\text{R-EPIG}}(\boldsymbol{x}, t) \coloneqq \mathbb{E}_{\mathbf{x}^* \sim p_{\text{tar}}(\cdot)}\Big[\mathrm{I}\big(\mathrm{r};\, \pi(\mathbf{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k\big)\Big]. \tag{10}$$

Unlike variance reduction, R-EPIG-Policy is *decision-aware*: it ignores regions where the sign of the total effect is determined with high confidence, concentrating the budget strictly on resolving ambiguities near the decision boundary.

Remark 4.2 (Information Gain vs. Direct Error Reduction). While minimizing APE is the ultimate goal, direct error reduction suffers from sparse signals: for candidates far from the current boundary, the probability of altering the policy is negligible, yielding zero-utility plateaus. In contrast, entropy reduction on π serves as a smooth surrogate, providing informative guidance that drives the learner toward the boundary even before a sign flip is probable.

4.2 Model Architecture: The TSR Strategy

To instantiate the probabilistic terms in R-EPIG, we introduce the Two-Stage Regression (TSR) strategy. TSR structurally decouples the *capacity* requirements from the *uncertainty* requirements.

Stage 1: The Observational Base Learner. We first exploit the abundant observational data $\mathcal{D}_O$ to learn a high-capacity base estimate of the observational conditional outcome surfaces. This stage is model-agnostic, permitting the use of state-of-the-art estimators $\mathcal{M}_O$ (e.g., TabPFN [31], CausalPFN [5]) to capture complex, global structural signals. The resulting observational outcome surfaces $\hat{\mu}_o(\boldsymbol{x}, t)$ are treated as pre-computed functional offsets. This base model acts as a fixed mean function for the subsequent residual model, relieving the active learner from reconstructing the global outcome surface and allowing it to focus its limited capacity on correcting local deviations. This benefit is regime-dependent: if the Stage-1 model is weak or misses too much shared structure, the residual may no longer be easier to learn than raw outcomes, and joint or tabula-rasa experimental models can be preferable.

Stage 2: Bayesian Residual Learning. The sequential phase strictly models the epistemic uncertainty of the residual vector $\boldsymbol{\delta}(\boldsymbol{x})$. Learning $\boldsymbol{\delta}(\boldsymbol{x})$ faces the fundamental counterfactual challenge: for any unit, we observe only the factual residual $r_k = y_k - \hat{\mu}_o(\boldsymbol{x}_k, t_k)$. To address this, we adopt a Multi-task GP [6] as the backend inference engine. We choose MTGP because its kernel structure explicitly encodes cross-arm correlations, allowing information sharing across treatment arms to infer missing counterfactual residuals:

$$\boldsymbol{\delta}(\boldsymbol{x}) \sim \mathcal{GP}(\mathbf{0}, \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}')), \quad r \mid \boldsymbol{x}, t \sim \mathcal{N}(\delta(\boldsymbol{x}, t), \sigma^2). \tag{11}$$

By decomposing the kernel $\boldsymbol{K}$ into task-specific and input-specific components (e.g., via CMGP [3], NSGP [2]), the model captures both the smoothness of the bias and the correlation between potential outcome residuals. Standard GPs scale cubically, $O(n^3)$. TSR resolves this bottleneck: the GP fits *only* the residuals on the small experimental set $\mathcal{D}_E$. The complexity scales with the budget $n_E \ll n_O$, not the massive observational size, ensuring real-time feasibility for sequential acquisition. Although our main implementation uses GP residual learners, R-Design only requires a residual model with usable posterior uncertainty. Recent martingale-posterior constructions can extract uncertainty over loss-defined functionals from tabular foundation models [39]; however, adapting such predictive-rule-based uncertainty to per-candidate residual information gain is non-trivial and left for future work.

Remark 4.3 (The Intuition of Structural Efficiency). The efficiency of TSR relies on a key structural insight: while the full outcome mechanism (e.g., biological response) is often complex and data-hungry, the confounding bias (e.g., clinical guidelines) may vary more smoothly across the population. By offloading the complex base learning to $\mathcal{D}_O$, TSR focuses the expensive experimental budget on the residual correction, which can enable faster convergence when this residual has lower effective complexity.

4.3 Budgeted Acquisition Algorithm

The R-EPIG utility drives our active learning strategy for CATE estimation under a fixed budget (Alg. 1). The process employs a two-stage architecture. First, observational models are trained on $\mathcal{D}_O$ and frozen to serve as functional offsets. Second, in the adaptive residual learning phase, we iteratively select batches of high-utility points. We use softmax-weighted sampling on the utility scores as a lightweight stochastic batching heuristic to avoid sequential re-scoring and encourage batch diversity. Experimental outcomes are queried, immediately transformed into residuals by subtracting the frozen offsets, and used to update the residual model $\hat{\boldsymbol{\delta}}$. This cycle continues until the budget is exhausted, yielding a final estimator that superimposes the learned residual contrast onto the observational baseline.

5 Theoretical Analysis

We justify the R-Design framework through four theoretical results. First, we identify a *Structural Efficiency Gap* (Lemma 5.1), showing that residual learning can admit faster rates when the residual correction has lower effective complexity than the full experimental outcome surface. Second, we establish an *Objective Alignment* result (Prop. 5.2), showing that, under the fixed-offset R-Design decomposition, minimizing the model-based Bayesian PEHE risk reduces to minimizing the integrated posterior variance of the residual contrast. Third, we demonstrate the *Information Efficiency* of target-based acquisition (Prop. 5.3), proving that parameter-based baselines such as BALD can spend information on target-irrelevant nuisance uncertainty. Finally, we provide an *Uncertainty Convergence* result (Thm. 5.4) for the residual-target information objective, whose rate is governed by the residual information capacity.

5.1 Sample Complexity: The Structural Efficiency Gap

We quantify the efficiency of R-Design via nonparametric convergence rates over Hölder-type function classes [2, 48]. For each treatment arm $t \in \{0, 1\}$, decompose the experimental potential-outcome surface into an observational limit and a causal residual:

$$\mu_e(\boldsymbol{x}, t) = \mu_o(\boldsymbol{x}, t) + \delta(\boldsymbol{x}, t). \tag{12}$$

Here, μ_o captures the observational conditional-mean structure, while δ is the correction needed to recover the experimental causal

response. Let $\alpha_{\cdot,t}$ and $d_{\cdot,t}$ denote the smoothness and effective dimension of the corresponding arm-specific function class.

Lemma 5.1 (Structural Efficiency Gap). *Assume that the observational base learner $\hat{\mu}_o$ and the experimental residual learner $\hat{\delta}$ use priors matched to the smoothness of their respective targets. Suppose that the conditional residual target induced by the frozen offset is well approximated by a residual Hölder class with smoothness $\alpha_{\delta,t}$ and effective dimension $d_{\delta,t}$. Assume that the observational arm-specific covariate distributions have sufficient support over the target region, and that the experimental design sequence satisfies the standard arm-specific coverage conditions required for the nonparametric regression rates over $L_2(p_{\text{tar}})$. Let $n_{E,t}$ and $n_{O,t}$ denote the corresponding effective arm-specific sample sizes. Under the identifiability assumptions in Assump. 2.1, the expected PEHE risk $\psi(\hat{\tau})$ satisfies*

$$\mathbb{E}[\psi(\hat{\tau})] \le C_1 \underbrace{\max_{t\in\{0,1\}} n_{E,t}^{-\frac{2\alpha_{\delta,t}}{2\alpha_{\delta,t}+d_{\delta,t}}}}_{\text{Residual convergence}} + C_2 \underbrace{\max_{t\in\{0,1\}} n_{O,t}^{-\frac{2\alpha_{\mu_o,t}}{2\alpha_{\mu_o,t}+d_{\mu_o,t}}}}_{\text{Observational baseline}} + \epsilon_{\text{approx}}, \tag{13}$$

where $C_1, C_2 > 0$ are constants, the expectation is taken over the observational sample, experimental noise, and any acquisition randomness conditional on the stated coverage conditions, and ϵ_{approx} collects approximation and model-misspecification errors. If the effective sample sizes are balanced across arms, $n_{E,t} \asymp n_E$ and $n_{O,t} \asymp n_O$, then Eq. 13 simplifies to

$$\mathbb{E}[\psi(\hat{\tau})] \le C_1 \max_{t\in\{0,1\}} n_E^{-\frac{2\alpha_{\delta,t}}{2\alpha_{\delta,t}+d_{\delta,t}}} + C_2 \max_{t\in\{0,1\}} n_O^{-\frac{2\alpha_{\mu_o,t}}{2\alpha_{\mu_o,t}+d_{\mu_o,t}}} + \epsilon_{\text{approx}}. \tag{14}$$

Interpretation. Lemma 5.1 makes the residual-favorable regime explicit. Since $n_E \ll n_O$, the observational error behaves as a static baseline, while the experimental-budget rate is governed by the residual correction. A tabula-rasa learner that ignores the observational offset must learn the full experimental response surface and is governed by the raw outcome complexity $O\left(n_E^{-\frac{2\alpha_{\mu_e,t}}{2\alpha_{\mu_e,t}+d_{\mu_e,t}}}\right)$. In contrast, R-Design benefits when the residual is smoother or lower-dimensional than the full response, e.g., $\alpha_{\delta,t} > \alpha_{\mu_e,t}$ or $d_{\delta,t} < d_{\mu_e,t}$. If the observational prior is weak or the induced residual is highly irregular, this structural advantage may vanish.

5.2 Objective: Optimality of Residual Learning

Lemma 5.1 characterizes when residual learning can be statistically easier than learning the full response. We next show that targeting residual uncertainty is aligned with one-step model-based Bayesian PEHE minimization. Conditional on $\mathcal{D}_O$, the observational contrast $\hat{\tau}_o(\boldsymbol{x})$ is frozen and deterministic. Therefore, under the R-Design decomposition $\tau(\boldsymbol{x}) = \hat{\tau}_o(\boldsymbol{x}) + \tau_\delta(\boldsymbol{x})$, the posterior uncertainty of τ is exactly the posterior uncertainty of τ_δ.

Proposition 5.2 (Objective Alignment). *Let $\mathcal{H}_k = \mathcal{D}_O \cup \mathcal{D}_E^{(k-1)}$ be the current history. For a candidate query $(\boldsymbol{x}, t)$ and prospective outcome $y \sim p(y \mid \boldsymbol{x}, t, \mathcal{H}_k)$, define the updated history $\mathcal{D}_+ \coloneqq \mathcal{H}_k \cup \{(\boldsymbol{x}, t, y)\}$. Assume that, after observing y, the CATE estimator is the posterior mean $\hat{\tau}_{\mathcal{D}_+}(\boldsymbol{x}^*) \coloneqq \mathbb{E}[\tau(\boldsymbol{x}^*) \mid \mathcal{D}_+]$. For active CATE estimation over p_{tar}, minimizing the expected model-based Bayesian PEHE risk is equivalent, under the R-Design decomposition, to minimizing the expected integrated posterior variance of the residual contrast:*

$$\begin{aligned}(\boldsymbol{x}_k^\dagger, t_k^\dagger) &= \arg\min_{(\boldsymbol{x},t)} \mathbb{E}_{y\sim p(y|\boldsymbol{x},t,\mathcal{H}_k)} \left[\mathbb{E}_{\mathbf{x}^*\sim p_{\text{tar}}} \left[\left(\hat{\tau}_{\mathcal{D}_+}(\mathbf{x}^*) - \tau(\mathbf{x}^*)\right)^2 \,\middle|\, \mathcal{D}_+\right]\right] \\ &\equiv \arg\min_{(\boldsymbol{x},t)} \mathbb{E}_{y\sim p(y|\boldsymbol{x},t,\mathcal{H}_k)} \left[\int \mathbb{V}[\tau_\delta(\boldsymbol{x}^*) \mid \mathcal{D}_+]\, dp_{\text{tar}}(\boldsymbol{x}^*)\right].\end{aligned} \tag{15}$$

Practical Surrogate. Eq. 15 identifies integrated posterior variance reduction as the exact one-step model-based Bayesian objective. Directly optimizing this objective is computationally expensive and lacks the same algorithmic structure as information gain. R-Design therefore uses residual information gain, e.g., $\mathrm{I}(r_{\boldsymbol{x},t}; \tau_\delta(\boldsymbol{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k)$, $r_{\boldsymbol{x},t} \coloneqq \mathrm{y} - \hat{\mu}_o(\boldsymbol{x}, t)$, as a tractable target-aware surrogate. For Gaussian residual models, this corresponds to entropy reduction of the residual target and supports information-theoretic active-learning guarantees.

5.3 Surpassing Parameter-Based Acquisition

Given the residual-learning parameterization in Lemma 5.1, a natural question is whether standard parameter-based active learning methods, such as BALD applied to the residual model parameters θ, are sufficient. We show that parameter-based acquisition optimizes a broader objective that can include nuisance information not required for the residual contrast.

Proposition 5.3 (Information Redundancy). *Fix a candidate query $(\boldsymbol{x}, t)$, current history $\mathcal{H}_k$, and target point $\boldsymbol{x}^*$. Let $r_{\boldsymbol{x},t} \coloneqq \mathrm{y} - \hat{\mu}_o(\boldsymbol{x}, t)$ be the prospective residual observation, and let θ denote the parameters, latent variables, or latent function representation of the residual model. If $\tau_\delta(\boldsymbol{x}^*)$ is determined by θ, then the parameter-oriented residual utility decomposes as*

$$\underbrace{\mathrm{I}(r_{\boldsymbol{x},t}; \theta \mid \boldsymbol{x}, t, \mathcal{H}_k)}_{\text{Residual-BALD}} = \underbrace{\mathrm{I}(r_{\boldsymbol{x},t}; \tau_\delta(\boldsymbol{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k)}_{\text{Target utility}} + \underbrace{\mathrm{I}(r_{\boldsymbol{x},t}; \theta \mid \tau_\delta(\boldsymbol{x}^*), \boldsymbol{x}, t, \mathcal{H}_k)}_{\text{Nuisance redundancy}}. \tag{16}$$

Averaging Eq. 16 over $\boldsymbol{x}^* \sim p_{\text{tar}}$ gives the corresponding population-level decomposition.

The Cost of Over-Modeling. The redundancy term $\Delta(\boldsymbol{x}^*) = \mathrm{I}(r_{\boldsymbol{x},t}; \theta \mid \tau_\delta(\boldsymbol{x}^*), \boldsymbol{x}, t, \mathcal{H}_k)$ is nonnegative and measures information about the residual model representation that remains after the target residual contrast is already known. Such information may correspond to nuisance uncertainty in high-frequency components, redundant parameterizations, or shared baseline variation that cancels in the contrast. Thus, parameter-based methods can spend budget reducing uncertainty that is not directly relevant to the target estimand. In contrast, R-EPIG targets the residual contrast itself over the target population.

5.4 Information-Theoretic Convergence

Finally, we analyze whether residual-targeted acquisition realizes the accelerated uncertainty reduction suggested by Lemma 5.1. The guarantee below is a TAL-style posterior-uncertainty bound for a *global* residual-target information objective. Let $\boldsymbol{X}_{\text{tar}}$ be a finite target set representing p_{tar}, and define the residual target vector $Z_{\text{tar}} \coloneqq \{\tau_\delta(\boldsymbol{x}^*) : \boldsymbol{x}^* \in \boldsymbol{X}_{\text{tar}}\}$. The global residual R-EPIG

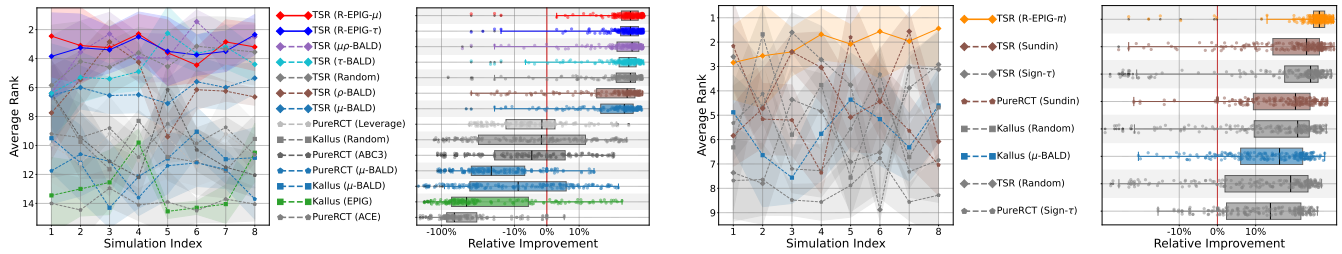


Figure 3: Comparison of acquisition functions using the average rank metric and relative performance improvements for effect estimation and policy learning tasks over eight combinations of base effect functions, bias functions, and CATE functions.

objective selects queries by maximizing $\mathrm{I}(Z_{\mathrm{tar}}; r_{\boldsymbol{x},t} \mid \boldsymbol{x}, t, \mathcal{H}_k)$, rather than the scalable mean-marginal approximation or the stochastic batching heuristic used in the implementation. This distinction is important: the theorem applies directly to the global objective, while the practical mean-marginal R-EPIG is a computational surrogate.

Theorem 5.4 (TAL-style Uncertainty Convergence). *Let $\gamma^{\delta}_{n_B}$ denote the maximum information capacity of the residual target vector Z_{tar} over the feasible experimental action class. Under the residual multitask GP model with fixed kernel and Gaussian observation noise, finite target and candidate sets, and the monotonicity/submodularity or weak-submodularity regularity conditions required by TAL, the global greedy residual R-EPIG strategy satisfies, for every $\boldsymbol{x}^* \in \boldsymbol{X}_{\mathrm{tar}}$,*

$$\mathbb{V}[\tau_\delta(\boldsymbol{x}^*) \mid \mathcal{D}_O, \mathcal{D}_E^{(n_B)}] \le \eta^2(\boldsymbol{x}^*) + \widetilde{O}\left(C_\alpha \frac{\gamma^{\delta}_{n_B}}{\sqrt{n_B}}\right). \tag{17}$$

where $\eta^2(\boldsymbol{x}^*)$ is the irreducible posterior uncertainty associated with the accessible sample class, and C_α is a constant capturing weak-submodularity or feasible-greedy approximation factors. If the feasibility-aware TAL condition holds, then $\eta^2(\boldsymbol{x}^*)$ can be taken as the feasible irreducible uncertainty $\eta^2_{\mathrm{feas}}(\boldsymbol{x}^*)$.

Efficiency Gap. A standard active learner targeting the full CATE or outcome surface satisfies an analogous TAL-style bound governed by the corresponding full-response information capacity. Under the fixed-offset R-Design decomposition, the posterior uncertainty of τ equals that of τ_δ conditional on $\mathcal{D}_O$, but the information capacity can be substantially smaller for the residual target. When the residual has lower effective complexity, as in the regime of Lemma 5.1, $\gamma^{\delta}_{n_B}$ grows more slowly than the capacity of the full target. Thus, the global residual-target strategy can achieve faster posterior-uncertainty reduction in the residual-favorable regime.

6 Experimental Results

In this section, we report empirical results verifying the effectiveness of our framework in acquiring experimental data points across multiple synthetic datasets and two semi-synthetic datasets: the Infant Health and Development Program (IHDP) [30] and the AIDS Clinical Trials Group Study 175 (ACTG-175) [25].

Baseline Acquisition Functions. We benchmark our method against a diverse set of acquisition strategies, ranging from random sampling to advanced baselines: Leverage [1], ACE [44], ABC3 [7], and Causal-BALD [33]. For decision-making settings, we additionally include Sundin [45], DEIG [15], and RFAN [37].

Inference Architectures and Backbones. We compare our proposed R-Design (TSR) framework against two baseline architectures: (1) PureRCT, a standard approach that trains causal estimators solely on experimental data, ignoring observational priors; and (2) Kallus [35], a data fusion method that pools observational and experimental data but corrects for confounding via importance weighting. All frameworks are instantiated using the same set of robust Bayesian learners. For the observational base (Stage 1), we primarily utilize TabPFN-v2.5 [23] due to its strong performance on tabular data, while also evaluating CausalPFN [5]. For the experimental/residual learning phase (Stage 2), we employ established Bayesian estimators capable of providing valid joint posteriors, including CMGP [3], NSGP [2], BART [30], BCF [24], and CMDE [34].

Metrics. We evaluate performance using two primary criteria. First, we report PEHE (Eq. 2), which quantifies the accuracy of estimated treatment effects relative to the ground truth. Second, for decision-making tasks, we report APE (Eq. 3) and Average Regret over the observational population [13, 19]. Unless otherwise noted, results are reported as mean ± standard deviation over 10 independent runs.

6.1 Synthetic Data

We evaluate different acquisition functions using two simulation studies: a univariate and a multivariate setting. The univariate study consists of 8 datasets spanning simple and complex combinations of base functions, bias functions, and CATE functions, and evaluates performance on both CATE estimation and policy learning tasks.

Results. Figs. 4 and 7 summarize the main results on synthetic datasets with dim = 6 covariates. These results separate two effects: the framework gain from using TSR rather than non-residual baselines, and the acquisition gain from using R-EPIG rather than other acquisition rules within TSR.

Effect Estimation Performance. As shown in Fig. 4, our proposed R-EPIG criteria consistently achieve substantial PEHE reductions compared to all baseline methods across the entire acquisition trajectory. Among all 15 methods evaluated, R-EPIG-μ achieves the best average rank of 3.00 and R-EPIG-τ ranks second with an average rank of 3.60 in the CD diagram based on the Nemenyi post-hoc test. The critical difference threshold (CD = 3.03 at $\alpha = 0.05$) confirms that both R-EPIG variants are statistically indistinguishable from each other while being significantly better than all non-TSR

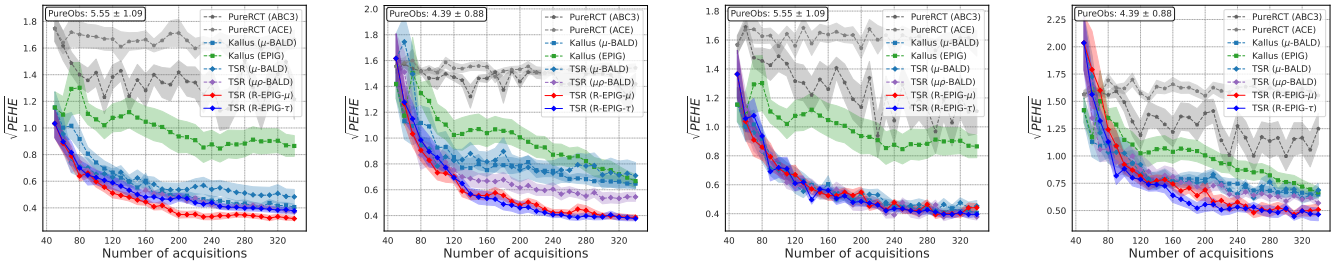


Figure 4: Comparison of $\sqrt{\text{PEHE}}$ on synthetic datasets with two trial estimators (CMGP and NSGP): (1) CMGP, (2) CMGP with heavy covariate shift, (3) NSGP, and (4) NSGP with heavy covariate shift.

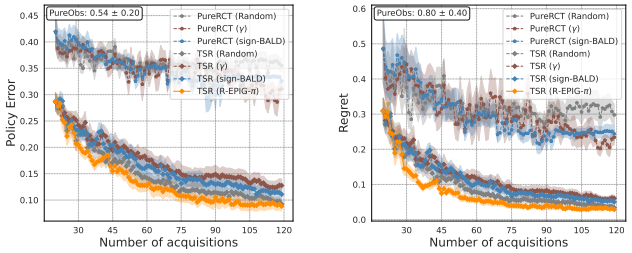


Figure 5: Performance comparison of all methods on the policy learning task: (a) APE and (b) AR.

baselines. Notably, all TSR-based methods (ranks 1–7) substantially outperform Kallus-based methods (ranks 8–10) and PureRCT methods (ranks 11–15), isolating the framework-level benefit of residualizing around an observational prior. Within TSR, the superior ranks of R-EPIG-μ and R-EPIG-τ isolate the additional benefit of estimand-aligned acquisition. Pairwise Wilcoxon signed-rank tests further confirm that R-EPIG-μ significantly outperforms 14 of 15 baselines ($p < 0.05$) with an 84.5% win rate, and R-EPIG-τ outperforms 13 baselines with an 80.5% win rate. Importantly, neither R-EPIG variant loses significantly to any baseline method.

Policy Learning Performance. For treatment policy optimization, we evaluate methods using Average Policy Error (APE) and Average Regret (AR). As illustrated in Fig. 5, R-EPIG-π, our policy-aware acquisition function specifically designed for decision-making, outperforms all baseline acquisition functions on both metrics. This demonstrates the benefit of directly targeting policy-relevant information gain rather than relying on CATE estimation accuracy as a proxy for policy quality.

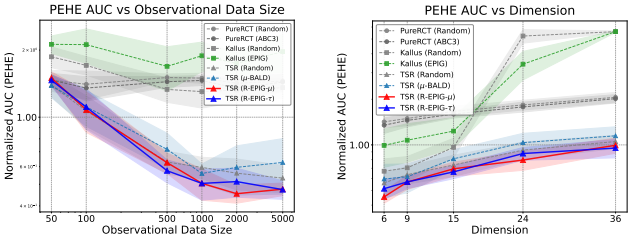


Figure 6: AUC under varying prior sizes and covariate dimensions.

Table 1: Final PEHE of TSR + R-EPIG-μ under varying observational size N_{obs} and experimental budget N_{exp}.

$N_{\text{obs}} \backslash N_{\text{exp}}$	100	200	340
100	1.190	1.045	1.000
500	0.793	0.569	0.509
2000	0.598	0.350	0.319

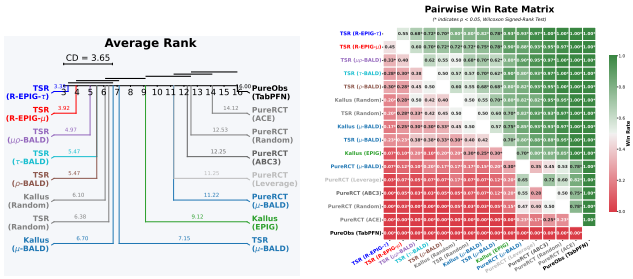


Figure 7: Comparison of acquisition functions on synthetic datasets: (a) critical difference diagram across all settings, and (b) pairwise win rate matrix with statistical significance.

Scalability Analysis. Fig. 6 examines method performance as the covariate dimensionality and observational dataset size vary. As the dimension increases from 6 to 36, the performance gap between TSR methods and baselines widens, with R-EPIG-μ maintaining its top rank (avg. rank 3.00) across all dimensions. This robustness stems from R-EPIG's explicit awareness of the target distribution, which becomes increasingly important under severe distribution shift in higher-dimensional spaces. When varying the observational data size from 50 to 5000 samples, TSR methods demonstrate consistent improvements with more observational data, while the relative ordering of acquisition functions remains stable. Tab. 1 further shows that N_{obs} and N_{exp} are complementary: better observational priors improve the residual baseline, while larger experimental budgets become most useful once that prior is sufficiently informative.

Robustness Across Data Generating Processes. Fig. 3 evaluates the robustness of R-EPIG criteria across 8 diverse combinations of base effect functions, confounding bias patterns, and CATE structures. Across all configurations, R-EPIG-μ and R-EPIG-τ consistently rank among the top performers, demonstrating that their effectiveness is not limited to specific functional forms but generalizes across a broad spectrum of causal data generating processes.

6.2 Semi-synthetic Data

We evaluate our framework on two well-established semi-synthetic benchmarks. The first, IHDP [30] (25 covariates), induces selection bias by removing a non-random subset of treated units from a randomized trial. The second, ACTG-175 [25] (12 covariates), constructs an observational cohort by excluding participants based on enrollment symptoms. Fig. 8 reports relative performance improvements over the PureRCT strategy with random acquisition. On IHDP, TSR with R-EPIG-μ achieves the best performance, followed by R-EPIG-τ. We attribute this to the higher covariate dimensionality of IHDP, where targeting the outcome surfaces provides more robust uncertainty quantification. In contrast, ACTG-175 has a lower-dimensional covariate space, under which R-EPIG-τ ranks first due to its direct focus on treatment effect heterogeneity, while R-EPIG-μ ranks third, demonstrating continued competitiveness across problem settings.

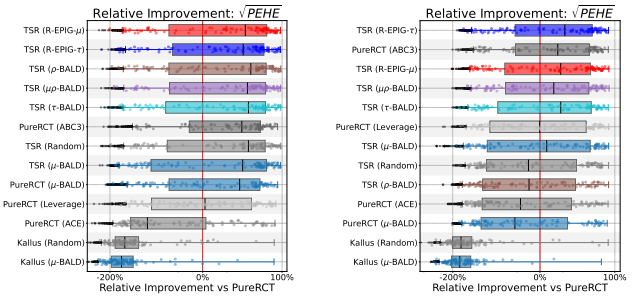


Figure 8: Comparison of acquisition functions on semi-synthetic datasets. Relative performance improvement on the IHDP dataset (left) and the ACTG-175 dataset (right).

7 Related Work

Our framework lies at the intersection of active learning and adaptive experimental design for causal inference. Existing literature can be broadly categorized by the level of control the learner has over the intervention mechanism. One line focuses on active CATE estimation from a fixed observational cohort [20, 21, 33]. In this setting, treatments have already been assigned by an unknown policy, and the learner selectively queries costly outcomes to improve a CATE model. The core challenge here is a structural transduction problem: inferring pairs of unobservable potential outcomes from single factual queries. However, the learner is passive regarding treatment assignment, limiting the ability to explore counterfactuals directly. A second, distinct line concerns adaptive experimental design [1, 49]. Here, the learner has the power to assign treatments. Recent works have explored adaptive randomization to minimize ATE variance [36], covariate balancing in a finite pool [26], and Bayesian experimental design to reduce CATE posterior variance [7, 37]. However, these methods typically operate in a tabula rasa setting, designing experiments from scratch while ignoring existing observational priors.

A related body of work studies the integration of observational and experimental data [8, 12, 35, 50]. These approaches recognize that observational data can improve statistical efficiency, but they are primarily post-hoc: the experimental sample is treated as already collected, and the goal is to debias, reweight, or fuse the two sources retrospectively. More recent work considers using observational data to inform trial design [11, 41], but such designs are typically static or parametric and do not address sequential acquisition of individual treatment assignments under a residual uncertainty model. Our setting is therefore different from both retrospective fusion and conventional experimental design. We use the observational model prospectively, not as a final causal estimator, but as a biased structural offset that changes what the experiment should learn. This leads to a residual active-design problem in which each query is chosen for its expected information about the remaining debiasing correction. We bridge this gap by addressing a hybrid question: given a large but biased observational prior, how should we sequentially assign treatments to efficiently learn the residual correction needed to estimate the target CATE?

8 Discussion and Conclusion

We introduced R-Design to bridge large-scale observational studies and targeted experimental design through active residual learning. Rather than discarding biased observational models, R-Design repurposes them as informative priors and uses scarce experimental samples to learn residual corrections. Our theoretical analysis characterizes when this reparameterization yields a structural efficiency gap: if the observational model captures substantial shared structure and the residual has lower effective complexity, estimating the residual contrast can require fewer experimental samples than reconstructing the full outcome mechanism from scratch. R-EPIG further aligns acquisition with downstream causal objectives by focusing on target-population uncertainty and adapting to estimation or policy learning. Our results show that R-Design improves sample efficiency across synthetic and semi-synthetic benchmarks in the intended informative-but-biased observational regime.

Limitations. The advantage of residual learning depends on Stage-1 prior quality; when the observational model is weak or misspecified, joint or tabula-rasa experimental models may be preferable. Our identifiability assumptions also require effect invariance and support inclusion: mechanism shift across sources or target units outside the feasible experimental region require additional transport or extrapolation assumptions. Finally, our strongest instantiation uses GP-style residual posteriors; scaling to raw image or text covariates will require representation learning with calibrated uncertainty-aware residual heads.

Future work. While R-Design currently focuses on binary interventions with a single prior, future work will extend active residual learning to richer settings, including continuous treatment regimes, multiple observational sources, and adaptive diagnostics for switching between residual and tabula-rasa experimental designs.

Acknowledgments

LZ is supported by the Guangzhou-HKUST(GZ) Joint Funding Program (No.2024A03J0630). HL is supported by the Beijing Major Science and Technology Project (No. Z251100008425006) and the Beijing Natural Science Foundation (No. L257007). EG and DS are supported by the Responsible AI Research Centre (RAIR). MG is supported by the Australian Research Council Discovery Project (DP240102088).

References

[1] Raghavendra Addanki, David Arbour, Tung Mai, Cameron Musco, and Anup Rao. 2022. Sample constrained treatment effect estimation. In *Advances in Neural Information Processing Systems*, Vol. 35. 5417–5430.

[2] Ahmed Alaa and Mihaela Schaar. 2018. Limits of estimating heterogeneous treatment effects: Guidelines for practical algorithm design. In *International Conference on Machine Learning*. PMLR, 129–138.

[3] Ahmed M Alaa and Mihaela Van Der Schaar. 2017. Bayesian inference of individualized treatment effects using multi-task gaussian processes. In *Advances in neural information processing systems*, Vol. 30.

[4] Susan Athey and Stefan Wager. 2019. Estimating treatment effects with causal forests: An application. *Observational studies* 5, 2 (2019), 37–51.

[5] Vahid Balazadeh, Hamidreza Kamkari, Valentin Thomas, Benson Li, Junwei Ma, Jesse C. Cresswell, and Rahul G. Krishnan. 2025. CausalPFN: Amortized Causal Effect Estimation via In-Context Learning. In *Advances in Neural Information Processing Systems*, Vol. 38.

[6] Edwin V Bonilla, Kian Chai, and Christopher Williams. 2007. Multi-task Gaussian process prediction. *Advances in neural information processing systems* 20 (2007).

[7] Taehun Cha and Donghun Lee. 2025. ABC3: Active Bayesian Causal Inference with Cohn Criteria in Randomized Experiments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 26769–26777.

[8] Bénédicte Colnet, Imke Mayer, Guanhua Chen, Awa Dieng, Ruohong Li, Gaël Varoquaux, Jean-Philippe Vert, Julie Josse, and Shu Yang. 2024. Causal inference methods for combining randomized trials and observational studies: a review. *Statistical science* 39, 1 (2024), 165–191.

[9] Irina Degtiar and Sherri Rose. 2023. A review of generalizability and transportability. *Annual Review of Statistics and Its Application* 10, 1 (2023), 501–524.

[10] Peng Ding. 2024. *A first course in causal inference*. CRC Press.

[11] Aristotelis Epanomeritakis and Davide Viviano. 2025. Choosing What to Learn: Experimental Design when Combining Experimental with Observational Evidence. *arXiv preprint arXiv:2510.23434* (2025).

[12] Jake Fawkes, Michael O'Riordan, Athanasios Vlontzos, Oriol Corcoll, and Ciarán Mark Gilligan-Lee. 2025. The Hardness of Validating Observational Studies with Experimental Data. In *International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research, Vol. 258)*. PMLR, 1819–1827.

[13] Carlos Fernández-Loría and Foster Provost. 2022. Causal decision making and causal effect estimation are not the same… and why it matters. *INFORMS Journal on Data Science* 1, 1 (2022), 4–16.

[14] Stefan Feuerriegel, Dennis Frauen, Valentyn Melnychuk, Jonas Schweisthal, Konstantin Hess, Alicia Curth, Stefan Bauer, Niki Kilbertus, Isaac S Kohane, and Mihaela van der Schaar. 2024. Causal machine learning for predicting treatment outcomes. *Nature Medicine* 30, 4 (2024), 958–968.

[15] Louis Filstroff, Iiris Sundin, Petrus Mikkola, Aleksei Tiulpin, Juuso Kylmäoja, and Samuel Kaski. 2024. Targeted Active Learning for Bayesian Decision-Making. *Transactions on Machine Learning Research* (2024).

[16] Jared C Foster, Jeremy MG Taylor, and Stephen J Ruberg. 2011. Subgroup identification from randomized clinical trial data. *Statistics in medicine* 30, 24 (2011), 2867–2880.

[17] Thomas R Frieden. 2017. Evidence for health decision making—beyond randomized, controlled trials. *New England Journal of Medicine* 377, 5 (2017), 465–475.

[18] Chen Gao, Yu Zheng, Wenjie Wang, Fuli Feng, Xiangnan He, and Yong Li. 2024. Causal inference in recommender systems: A survey and future directions. *ACM Transactions on Information Systems* 42, 4 (2024), 1–32.

[19] Erdun Gao, Howard Bondell, Wei Huang, and Mingming Gong. 2024. A variational framework for estimating continuous treatment effects with measurement error. In *The Twelfth International Conference on Learning Representations*.

[20] Erdun Gao, Jake Fawkes, and Dino Sejdinovic. 2025. Causal-EPIG: A Prediction-Oriented Active Learning Framework for CATE Estimation. *arXiv preprint arXiv:2509.21866* (2025).

[21] Erdun Gao and Dino Sejdinovic. 2026. ActiveCQ: Active Estimation of Causal Quantities. In *The Fourteenth International Conference on Learning Representations*.

[22] Mehrdad Ghadiri, David Arbour, Tung Mai, Cameron Musco, and Anup B Rao. 2023. Finite population regression adjustment and non-asymptotic guarantees for treatment effect estimation. *Advances in Neural Information Processing Systems* 36 (2023), 74180–74212.

[23] Léo Grinsztajn, Klemens Flöge, Oscar Key, Felix Birkel, Philipp Jund, Brendan Roof, Benjamin Jäger, Dominik Safaric, Simone Alessi, Adrian Hayler, et al. 2025. TabPFN-2.5: Advancing the State of the Art in Tabular Foundation Models. *arXiv preprint arXiv:2511.08667* (2025).

[24] P Richard Hahn, Jared S Murray, and Carlos M Carvalho. 2020. Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis* 15, 3 (2020), 965–1056.

[25] Scott M Hammer, David A Katzenstein, Michael D Hughes, Holly Gundacker, Robert T Schooley, Richard H Haubrich, W Keith Henry, Michael M Lederman, John P Phair, Manette Niu, et al. 1996. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine* 335, 15 (1996), 1081–1090.

[26] Christopher Harshaw, Fredrik Sävje, Daniel A Spielman, and Peng Zhang. 2024. Balancing covariates in randomized experiments with the gram–schmidt walk design. *J. Amer. Statist. Assoc.* 119, 548 (2024), 2934–2946.

[27] Tobias Hatt, Jeroen Berrevoets, Alicia Curth, Stefan Feuerriegel, and Mihaela van der Schaar. 2022. Combining observational and randomized data for estimating heterogeneous treatment effects. *arXiv preprint arXiv:2202.12891* (2022).

[28] James J Heckman. 2000. Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* 115, 1 (2000), 45–97.

[29] M.A. Hernan and J.M. Robins. [n. d.]. *Causal Inference: What If*. CRC Press.

[30] Jennifer L Hill. 2011. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics* 20, 1 (2011), 217–240.

[31] Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. 2025. Accurate predictions on small data with a tabular foundation model. *Nature* (09 01 2025). doi:10.1038/s41586-024-08328-6

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

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

[34] Ziyang Jiang, Zhuoran Hou, Yiling Liu, Yiman Ren, Keyu Li, and David Carlson. 2023. Estimating causal effects using a multi-task deep ensemble. In *International Conference on Machine Learning*. PMLR, 15023–15040.

[35] Nathan Kallus, Aahlad Manas Puli, and Uri Shalit. 2018. Removing hidden confounding by experimental grounding. *Advances in neural information processing systems* 31 (2018).

[36] Masahiro Kato, Akihiro Oga, Wataru Komatsubara, and Ryo Inokuchi. 2024. Active Adaptive Experimental Design for Treatment Effect Estimation with Covariate Choice. In *International Conference on Machine Learning*. PMLR.

[37] Omer Noy Klein, Alihan Hüyük, Ron Shamir, Uri Shalit, and Mihaela van der Schaar. 2025. Towards Regulatory-Confirmed Adaptive Clinical Trials: Machine Learning Opportunities and Solutions. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 4969–4977.

[38] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences* 116, 10 (2019), 4156–4165.

[39] Kenyon Ng, Edwin Fong, David T Frazier, Jeremias Knoblauch, and Susan Wei. 2025. TabMGP: Martingale Posterior with TabPFN. *arXiv preprint arXiv:2510.25154* (2025).

[40] J Pearl. 2009. *Causality*. Cambridge university press.

[41] Evan TR Rosenman and Art B Owen. 2021. Designing experiments informed by observational studies. *Journal of Causal Inference* 9, 1 (2021), 147–171.

[42] Donald B Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* 100, 469 (2005), 322–331.

[43] Freddie Bickford Smith, Andreas Kirsch, Sebastian Farquhar, Yarin Gal, Adam Foster, and Tom Rainforth. 2023. Prediction-oriented Bayesian active learning. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 7331–7348.

[44] Difan Song, Simon Mak, and Jeff Wu. 2024. ACE: Active Learning for Causal Inference with Expensive Experiments. In *29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining: Workshop on Causal Inference and Machine Learning*. ACM.

[45] Iiris Sundin, Peter Schulam, Eero Siivola, Aki Vehtari, Suchi Saria, and Samuel Kaski. 2019. Active learning for decision-making from imbalanced observational data. In *International conference on machine learning*. PMLR, 6046–6055.

[46] Stefan Wager and Susan Athey. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113, 523 (2018), 1228–1242.

[47] Shu Yang, Siyi Liu, Donglin Zeng, and Xiaofei Wang. 2025. Data fusion methods for the heterogeneity of treatment effect and confounding function. *Bernoulli* 31, 4 (2025), 2987–3012.

[48] Yun Yang and Surya T Tokdar. 2015. Minimax-optimal nonparametric regression in high dimensions. *Annals of Statistics* 43, 2 (2015), 652–674.

[49] Zhiheng Zhang, Haoxiang Wang, Haoxuan Li, and Zhouchen Lin. 2025. Active Treatment Effect Estimation via Limited Samples. In *Forty-second International Conference on Machine Learning*.

[50] Chuan Zhou, Yaxuan Li, Chunyuan Zheng, Haiteng Zhang, Min Zhang, Haoxuan Li, and Mingming Gong. 2025. A Two-Stage Pretraining-Finetuning Framework for Treatment Effect Estimation with Unmeasured Confounding. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2113–2123.

A Appendix

Scope of this appendix. This appendix provides the practical R-Design loop, the acquisition-computation details used in the main experiments, and compact proof sketches for the theoretical results in Sec. 5. The convergence result in Thm. 5.4 applies to the idealized sequential greedy rule for the global residual-target information objective. The mean-marginal R-EPIG utilities and the stochastic batching rule below are scalable computational surrogates.

Extended version. An extended version containing full proofs, additional derivations, extended ablations, and further implementation details is available at

https://arxiv.org/abs/2603.03785.

The source code is available at

https://github.com/ErdunGAO/R-Design.

Algorithm 1 R-Design: Residual-Based Adaptive Causal Experimental Design

Require: Observational data $\mathcal{D}_O$, candidate pool $\mathcal{D}_P = \{\boldsymbol{x}_i\}_{i=1}^{n_P}$, budget n_B, batch size n_b, softmax temperature λ, acquisition target $\Phi \in \{\boldsymbol{\delta}, \tau_\delta, \pi\}$.

Ensure: CATE estimator $\hat{\tau}(\boldsymbol{x}) = \hat{\tau}_o(\boldsymbol{x}) + \hat{\tau}_\delta(\boldsymbol{x})$.

Train observational learners $\hat{\mu}_o(\cdot, 0)$ and $\hat{\mu}_o(\cdot, 1)$ on $\mathcal{D}_O$.
Construct $\hat{\tau}_o(\boldsymbol{x}) \leftarrow \hat{\mu}_o(\boldsymbol{x}, 1) - \hat{\mu}_o(\boldsymbol{x}, 0)$ and freeze $\hat{\mu}_o$.
Initialize experimental data $\mathcal{D}_E \leftarrow \emptyset$, residual data $\mathcal{R}_E \leftarrow \emptyset$, and residual model $\hat{\boldsymbol{\delta}}(\cdot)$.
while $|\mathcal{D}_E| < n_B$ **and** $\mathcal{D}_P \neq \emptyset$ **do**
  Fit/update $\hat{\boldsymbol{\delta}}$ using residual observations in $\mathcal{R}_E$.
  for $\boldsymbol{x} \in \mathcal{D}_P$ **do**
    for $t \in \{0, 1\}$ **do**
      Compute $u_t(\boldsymbol{x}) \leftarrow \mathbb{E}_{\mathbf{x}^* \sim p_{\text{tar}}}[\mathrm{I}(r_{\boldsymbol{x},t}; \Phi(\mathbf{x}^*) \mid \boldsymbol{x}, t, \mathcal{D}_O, \mathcal{R}_E)]$ where $r_{\boldsymbol{x},t} = \mathrm{y} - \hat{\mu}_o(\boldsymbol{x}, t)$.
    end for
    Set $A[\boldsymbol{x}] \leftarrow \arg\max_{t \in \{0,1\}} u_t(\boldsymbol{x})$ and $S[\boldsymbol{x}] \leftarrow \max\{u_0(\boldsymbol{x}), u_1(\boldsymbol{x})\}$.
  end for
  Set $b_k \leftarrow \min\{n_b, n_B - |\mathcal{D}_E|, |\mathcal{D}_P|\}$.
  Sample $\mathcal{B}_x \subseteq \mathcal{D}_P$ of size b_k without replacement using probabilities $\propto \exp(S[\boldsymbol{x}]/\lambda)$.
  Form query batch $\mathcal{B} \leftarrow \{(\boldsymbol{x}, A[\boldsymbol{x}]) : \boldsymbol{x} \in \mathcal{B}_x\}$ and observe outcomes
$$\mathcal{D}_{\text{qry}} \leftarrow \{(\boldsymbol{x}, t, y, s = e) : (\boldsymbol{x}, t) \in \mathcal{B}\}.$$
  Form residual data
$$\mathcal{R}_{\text{new}} \leftarrow \{(\boldsymbol{x}, t, r) : (\boldsymbol{x}, t, y, s = e) \in \mathcal{D}_{\text{qry}},\ r = y - \hat{\mu}_o(\boldsymbol{x}, t)\}.$$
  Update $\mathcal{D}_E \leftarrow \mathcal{D}_E \cup \mathcal{D}_{\text{qry}}$, $\mathcal{R}_E \leftarrow \mathcal{R}_E \cup \mathcal{R}_{\text{new}}$, and $\mathcal{D}_P \leftarrow \mathcal{D}_P \setminus \mathcal{B}_x$.
end while
return $\hat{\tau}(\boldsymbol{x}) = \hat{\tau}_o(\boldsymbol{x}) + \hat{\delta}(\boldsymbol{x}, 1) - \hat{\delta}(\boldsymbol{x}, 0)$.

A.1 Practical Computation of R-EPIG

R-Design treats the observational base learner as a fixed offset. After an experimental query $(\boldsymbol{x}_i, t_i)$ returns outcome y_i, the residual observation is $r_i = y_i - \hat{\mu}_o(\boldsymbol{x}_i, t_i)$. The residual learner models the latent vector $\boldsymbol{\delta}(\boldsymbol{x}) = [\delta(\boldsymbol{x}, 0), \delta(\boldsymbol{x}, 1)]^\top$. In the main implementation, we use a multi-task GP posterior. Equivalently, over the augmented input $a = (\boldsymbol{x}, t)$, we have $\delta(a) \sim \mathcal{GP}(0, k_\delta(a, a'))$, with observations $r_i = \delta(\boldsymbol{x}_i, t_i) + \epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. For separable multi-task kernels, one may write $k_\delta((\boldsymbol{x}, t), (\boldsymbol{x}', t')) = B_{t,t'}\, k_x(\boldsymbol{x}, \boldsymbol{x}')$, where B controls cross-arm covariance and k_x controls smoothness over covariates. The posterior over any finite collection of candidate residuals, residual potential outcomes, and residual contrasts is Gaussian, which yields closed-form acquisition scores for the estimation variants.

Let $\boldsymbol{X}_{\text{tar}} = \{\boldsymbol{x}_j^*\}_{j=1}^m$ be an empirical target set sampled from the observational covariates. For a candidate action $a = (\boldsymbol{x}, t)$, the mean-marginal acquisition used in experiments is $\hat{\alpha}_\Phi(a) = \frac{1}{m}\sum_{j=1}^m \mathrm{I}(r_a; \Phi(\boldsymbol{x}_j^*) \mid a, \mathcal{D}_O, \mathcal{D}_E)$, where r_a is the prospective noisy residual observation at a. The target set $\boldsymbol{X}_{\text{tar}}$ approximates the expectation over p_{tar} in Eq. 7.

R-EPIG-μ. For joint-outcome residual learning, the target is $z_j = \boldsymbol{\delta}(\boldsymbol{x}_j^*) = [\delta(\boldsymbol{x}_j^*, 0), \delta(\boldsymbol{x}_j^*, 1)]^\top$. Under the residual GP posterior, (r_a, z_j) is jointly Gaussian: $[r_a, z_j^\top]^\top \sim \mathcal{N}\big([m_a, m_j^\top]^\top, \left[\begin{smallmatrix} \Sigma_{aa} & \Sigma_{aj} \\ \Sigma_{ja} & \Sigma_{jj} \end{smallmatrix}\right]\big)$. The conditional covariance of z_j after observing r_a is $\Sigma_{j|a} = \Sigma_{jj} - \Sigma_{ja}\Sigma_{aa}^{-1}\Sigma_{aj}$. Thus, $\mathrm{I}(r_a; z_j) = \frac{1}{2}\log\big(\det\Sigma_{jj}/\det\Sigma_{j|a}\big)$. This acquisition reduces posterior uncertainty in both residual potential-outcome arms. It is therefore conservative: it may acquire information about arm-specific residual surfaces even when part of that uncertainty cancels in the CATE contrast.

R-EPIG-τ. For direct CATE residual learning, the target is the scalar contrast $z_j = \tau_\delta(\boldsymbol{x}_j^*) = \delta(\boldsymbol{x}_j^*, 1) - \delta(\boldsymbol{x}_j^*, 0)$. Let $h = [-1, 1]^\top$ and $\boldsymbol{\delta}_j = [\delta(\boldsymbol{x}_j^*, 0), \delta(\boldsymbol{x}_j^*, 1)]^\top$. Then $z_j = h^\top \boldsymbol{\delta}_j$. Since z_j is a linear functional of a Gaussian vector, (r_a, z_j) is jointly Gaussian. Its posterior variance and covariance are $\mathbb{V}[z_j] = h^\top \Sigma_{jj} h$ and $\mathbb{C}[z_j, r_a] = h^\top \Sigma_{ja}$. After observing r_a, the conditional variance is $\mathbb{V}[z_j \mid r_a] = \mathbb{V}[z_j] - \mathbb{C}[z_j, r_a]^2/\mathbb{V}[r_a]$. The mutual information is therefore $\mathrm{I}(r_a; z_j) = \frac{1}{2}\log\big(\mathbb{V}[z_j]/\mathbb{V}[z_j \mid r_a]\big)$. R-EPIG-$\tau$ is the most direct acquisition rule for PEHE because it targets the residual uncertainty that survives the treatment-control subtraction.

R-EPIG-π. For policy learning, the target is binary: $\pi(\boldsymbol{x}_j^*) = \mathbb{I}\big(\hat{\tau}_o(\boldsymbol{x}_j^*) + \tau_\delta(\boldsymbol{x}_j^*) > 0\big)$. Let $q_j = \Pr(\hat{\tau}_o(\boldsymbol{x}_j^*) + \tau_\delta(\boldsymbol{x}_j^*) > 0 \mid \mathcal{D}_O, \mathcal{D}_E)$. The marginal entropy is $H_{\mathrm{B}}(q_j) = -q_j \log q_j - (1 - q_j)\log(1 - q_j)$. For a candidate residual observation r_a, the policy information gain is $\mathrm{I}(r_a; \pi(\boldsymbol{x}_j^*)) = H_{\mathrm{B}}(q_j) - \mathbb{E}_{r_a}[H_{\mathrm{B}}(q_j(r_a))]$, where $q_j(r_a) = \Pr(\hat{\tau}_o(\boldsymbol{x}_j^*) + \tau_\delta(\boldsymbol{x}_j^*) > 0 \mid r_a, \mathcal{D}_O, \mathcal{D}_E)$. We approximate the expectation over r_a with posterior Monte Carlo samples. This criterion prioritizes candidates that are expected to reduce uncertainty about the sign of the total effect, and therefore concentrates acquisition near the decision boundary.

A.2 Batching and Final Estimator

Sequentially recomputing the posterior after every individual query is expensive. We therefore use a softmax batching heuristic. For each unit $\boldsymbol{x}$, we first compute the best arm $A[\boldsymbol{x}] = \arg\max_{t \in \{0,1\}} u_t(\boldsymbol{x})$ and the corresponding unit score $S[\boldsymbol{x}] = \max_{t \in \{0,1\}} u_t(\boldsymbol{x})$. A batch is sampled without replacement with probabilities proportional to $\exp(S[\boldsymbol{x}]/\lambda)$, where $\lambda > 0$ is a temperature parameter. Smaller λ approaches greedy selection, while larger λ increases diversity. This batching rule is used for computational efficiency and is not part of the TAL-style convergence statement.

After the budget is exhausted, the observational learners remain fixed and only the residual model has been updated using experimental data. The final response estimator is $\hat{\mu}_e(\boldsymbol{x}, t) = \hat{\mu}_o(\boldsymbol{x}, t) +$

$\hat{\delta}(\boldsymbol{x}, t)$. The corresponding CATE estimator is $\hat{\tau}(\boldsymbol{x}) = \hat{\mu}_e(\boldsymbol{x}, 1) - \hat{\mu}_e(\boldsymbol{x}, 0) = \hat{\tau}_o(\boldsymbol{x}) + \hat{\tau}_\delta(\boldsymbol{x})$. Thus, R-Design uses observational data to learn a fixed structural offset and uses the experimental budget to learn only the residual correction.

A.3 Proof Sketches and Scope of Guarantees

Structural efficiency. Lemma 5.1 follows by decomposing the experimental response into a frozen observational offset and a residual correction. For each arm t, $\hat{\mu}_{e,t}(\boldsymbol{x}) - \mu_{e,t}(\boldsymbol{x}) = \hat{\delta}_t(\boldsymbol{x}) - \tilde{\delta}_t(\boldsymbol{x})$, where $\tilde{\delta}_t(\boldsymbol{x}) = \mu_{e,t}(\boldsymbol{x}) - \hat{\mu}_{o,t}(\boldsymbol{x})$. Thus, conditional on the Stage-1 offset, the Stage-2 learner solves a nonparametric regression problem for the induced residual target $\tilde{\delta}_t$. The approximation gap between $\tilde{\delta}_t$ and the oracle residual $\delta_t = \mu_{e,t} - \mu_{o,t}$ is controlled by the Stage-1 error: $\|\tilde{\delta}_t - \delta_t\|^2_{L_2(p_{\text{tar}})} = \|\hat{\mu}_{o,t} - \mu_{o,t}\|^2_{L_2(p_{\text{tar}})}$. Combining the arm-specific residual regression rate with the observational baseline rate yields the two-term bound in Lemma 5.1. The bound is conditional on the coverage assumptions stated in the lemma; without sufficient target-region support, a global $L_2(p_{\text{tar}})$ rate need not hold.

Objective alignment. Prop. 5.2 is a model-based Bayesian risk identity. After a candidate outcome is observed and the updated history is $\mathcal{D}_+$, the posterior-mean estimator satisfies $\mathbb{E}\left[(\hat{\tau}_{\mathcal{D}_+}(\boldsymbol{x}^*) - \tau(\boldsymbol{x}^*))^2 \mid \mathcal{D}_+\right] = \mathbb{V}[\tau(\boldsymbol{x}^*) \mid \mathcal{D}_+]$. Under the fixed-offset decomposition $\tau(\boldsymbol{x}) = \hat{\tau}_o(\boldsymbol{x}) + \tau_\delta(\boldsymbol{x})$, the offset $\hat{\tau}_o$ is deterministic conditional on $\mathcal{D}_O$, and therefore $\mathbb{V}[\tau(\boldsymbol{x}^*) \mid \mathcal{D}_+] = \mathbb{V}[\tau_\delta(\boldsymbol{x}^*) \mid \mathcal{D}_+]$. Hence one-step Bayesian PEHE minimization reduces to expected integrated posterior variance minimization for the residual contrast.

Information redundancy. Prop. 5.3 follows from the chain rule of mutual information. If $\tau_\delta(\boldsymbol{x}^*)$ is determined by the residual-model representation θ, then $\mathrm{I}(r_{\boldsymbol{x},t}; \tau_\delta(\boldsymbol{x}^*) \mid \theta, \boldsymbol{x}, t, \mathcal{H}_k) = 0$. Applying the chain rule to $\mathrm{I}(r_{\boldsymbol{x},t}; \theta, \tau_\delta(\boldsymbol{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k)$ gives $\mathrm{I}(r_{\boldsymbol{x},t}; \theta \mid \boldsymbol{x}, t, \mathcal{H}_k) = \mathrm{I}(r_{\boldsymbol{x},t}; \tau_\delta(\boldsymbol{x}^*) \mid \boldsymbol{x}, t, \mathcal{H}_k) + \mathrm{I}(r_{\boldsymbol{x},t}; \theta \mid \tau_\delta(\boldsymbol{x}^*), \boldsymbol{x}, t, \mathcal{H}_k)$. The second term is nonnegative and represents information about the residual model that is not needed once the target residual contrast is known.

TAL-style convergence. For Thm. 5.4, fix a finite target set $\boldsymbol{X}_{\text{tar}}$ and define $Z_{\text{tar}} = \{\tau_\delta(\boldsymbol{x}^*) : \boldsymbol{x}^* \in \boldsymbol{X}_{\text{tar}}\}$. Under the multitask residual GP, the candidate residual observations and Z_{tar} are jointly Gaussian because τ_δ is a linear functional of the residual process. The global residual objective $\mathrm{I}(Z_{\text{tar}}; r_{\boldsymbol{x},t} \mid \boldsymbol{x}, t, \mathcal{H}_k)$ therefore defines a finite Gaussian transductive active-learning instance. Applying the TAL posterior-uncertainty bound to this induced instance yields the rate in Thm. 5.4, governed by the residual information capacity $\gamma^{\delta}_{n_B}$. The statement applies to the global sequential greedy objective; the mean-marginal utilities and stochastic batching used in practice are scalable surrogates.

A.4 Discussion: What R-Design Does and Does Not Claim

R-Design should be understood as an experimental-design principle rather than an observational identification method. The observational model is not assumed to identify the CATE by itself. Instead, it is used as a fixed structural offset that may contain useful information about the global response surface. Identification of the target causal effect still comes from the experimental source under Assump. 2.1. In this sense, R-Design does not remove the need for randomized experimentation; it changes what the scarce experimental budget is asked to learn.

The central advantage is that residualization can alter the effective complexity of the experimental learning problem. A tabula-rasa experimental learner must spend RCT samples learning all components of the response surface, including smooth baseline effects and high-frequency covariate structure that may already be well captured by the observational model. By contrast, TSR asks the experimental data to learn only the correction to the observational model. When this correction is smoother, lower-dimensional, or more localized than the full response, the same experimental budget can produce a more accurate CATE estimator. This is the regime formalized by Lemma 5.1.

The second advantage is acquisition alignment. In standard posterior-based design, high uncertainty is often treated as intrinsically valuable. For causal tasks this is insufficient: uncertainty is useful only insofar as it affects the target estimand and the target population. R-EPIG therefore scores a candidate query by its information about the downstream residual target, such as $\tau_\delta(\boldsymbol{x})$ for CATE estimation or $\pi(\boldsymbol{x})$ for policy learning. This explains why R-EPIG may behave differently from BALD-style rules: parameter uncertainty can include nuisance variation that does not change the residual contrast or the policy boundary.

These advantages are conditional. If the observational model is weak, poorly supported on the target population, or misses important shared structure, the induced residual target $\tilde{\delta}(\boldsymbol{x}, t) = \mu_e(\boldsymbol{x}, t) - \hat{\mu}_o(\boldsymbol{x}, t)$ may be as complex as the original response. In such cases, R-Design may lose its structural efficiency advantage, and a tabula-rasa RCT learner or a joint model may be preferable. Similarly, if the residual posterior is miscalibrated, R-EPIG can be misled because all information-theoretic acquisition scores are computed under that posterior.

R-Design is also sensitive to support and transportability. Randomization protects the experimental contrast from confounding, but it does not justify extrapolation outside the feasible experimental region. If the target population has support outside $\mathcal{X}_e$, then support inclusion fails and the target CATE cannot be recovered without additional extrapolation assumptions. If effect invariance fails, then τ_δ no longer represents only observational confounding bias; it also absorbs source-specific treatment-effect shifts. In such cases, the residual may still be useful predictively, but its interpretation as a pure debiasing correction becomes weaker.

Overall, the main message is not that observational priors should always be trusted. Rather, observational priors should be used when they are informative enough to reduce the complexity of the remaining experimental task, and they should be corrected by randomized data rather than accepted at face value. R-Design provides one way to operationalize this principle: freeze the observational structure, model residual uncertainty, and allocate experiments to the residual aspects that matter for the target causal decision.