Towards Regulatory-Confirmed Adaptive Clinical Trials: Machine Learning Opportunities and Solutions

Omer Noy Klein
Tel-Aviv University

Alihan Hüyük
University of Cambridge

Ron Shamir
Tel-Aviv University

Uri Shalit
Technion
Tel-Aviv University

Mihaela van der Schaar
University of Cambridge

Abstract

Randomized Controlled Trials (RCTs) are the gold standard for evaluating the effect of new medical treatments. Treatments must pass stringent regulatory conditions in order to be approved for widespread use, yet even after the regulatory barriers are crossed, real-world challenges might arise: Who should get the treatment? What is its true clinical utility? Are there discrepancies in the treatment effectiveness across diverse and under-served populations? We introduce two new objectives for future clinical trials that integrate regulatory constraints and treatment *policy value* for both the entire population and under-served populations, thus answering some of the questions above in advance. Designed to meet these objectives, we formulate RANDOMIZE FIRST AUGMENT NEXT (RFAN), a new framework for designing Phase III clinical trials. Our framework consists of a standard randomized component followed by an adaptive one, jointly meant to efficiently and safely acquire and assign patients into treatment arms during the trial. Then, we propose strategies for implementing RFAN based on causal, deep Bayesian active learning. Finally, we empirically evaluate the performance of our framework using synthetic and real-world semi-synthetic datasets.

1 INTRODUCTION

How will a patient respond to a new treatment? For decades, randomized controlled trials (RCTs) have been used for evaluating the effect of new treatments, and are considered the gold standard for clinical evidence of efficacy and safety. Specifically, Phase III (pre-marketing) trials are typically RCTs designed to evaluate the effectiveness of the new treatment compared to a placebo or to a standard treatment, and are the last and crucial step before gaining regulatory approval for wide dissemination. After approval, Phase IV studies (post-marketing surveillance) provide evidence of the treatment's effectiveness and safety on a large scale in the actual target populations and under the actual treatment *policy*[1]. However, waiting until Phase IV to learn the full real-world implications of a treatment might be too late, as by this point many patients will have been affected. A special concern arises from the under-representation in RCTs of populations such as women, ethnic groups, elderly patients, etc., collectively referred to as *under-served* groups. The discrepancy between trial population and the population who ultimately receives the treatment can significantly affect the generalizability of the Phase III trial findings and the effectiveness of care delivery for these populations. A major reason for the existence of these gaps is the fact that currently trials are often designed so as to maximize the chances of obtaining regulatory approval, and recruiting practices that are often deficient in terms of demographic diversity (Food et al., 2020; Schwartz et al., 2023). To facilitate reading, we provide a glossary of key terms for clinical trials in Table A.1.

As an illustration of these challenges, consider the case of warfarin, a widely used oral anticoagulant (Con-

---

[1]Policy here means which patients receive the treatment and which do not.

sortium, 2009; Huynh, 2017). Despite being used for almost 70 years, warfarin dosing remains a significant challenge for practitioners, primarily due to the drug's narrow therapeutic range and substantial variability in individual responses. An incorrect dose might lead to serious adverse outcomes. Although several dose-optimization algorithms have been developed and tested in RCTs, these studies were conducted primarily in White and Asian populations (Asiimwe et al., 2021; Asiimwe and Pirmohamed, 2022). Importantly, there are key genetic differences that influence warfarin response, making studies conducted in Whites/Asians less applicable to many other populations. Yet clinical practice guidelines for warfarin dosing have largely been based on these RCTs. Recent years have seen similar pressing challenges, with the emergence of the COVID-19 pandemic (Zame et al., 2020). For instance, Paxlovid (nirmatrelvir/ritonavir), an antiviral medication for COVID-19, still faces limited distribution due to budget constraints (Pepperrell et al., 2022). Given the variability in patient response (Najjar-Debbiny et al., 2023), identifying who will most benefit from the treatment remains a critical question. Further real-world examples can be found in Table A.2.

This paper seeks to formulate a new pre-marketing trial design that bridges some of the gap between the requirements of regulatory bodies and real-world deployment concerns. The fundamental challenges we address in this work are: (1) How can we conduct a representative and fair yet regulatory-confirmed clinical trial? (2) Can we anticipate the actual clinical utility of a new treatment in a way that allows us to deliver a better treatment policy immediately upon deployment? We give an illustration of the challenges in Figure 1.

We propose to address these challenges by introducing POST-TRIAL MEAN BENEFIT (PTMB) and POST-TRIAL FAIRNESS (PTF) as two new objectives for clinical trials, incorporating both the conventional regulatory constraints and the estimated real-world performance of a new treatment with a corresponding policy. Unlike the typical Phase III goal that focuses on regulatory approval, *PTMB also considers the performance of the new treatment policy once deployed in the target population*, for example taking into account the possibility that the new treatment is nominally effective but will have low overall impact. The PTF objective considers the policy value for under-served groups.

To meet these objectives we propose RANDOMIZE FIRST AUGMENT NEXT (RFAN) as a new framework for designing Phase III clinical trials. The framework consists of two stages: A randomized (*passive*) stage and an augmented (*active*) stage. The randomized stage ensures compliance with regulatory standards. Then, in the augmented stage, the information learned so far
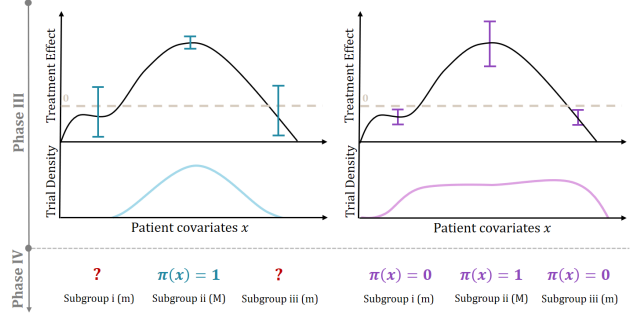

Figure 1: Two Phase III trials with different sample distributions, their impact on the resulting treatment effect estimates (top), and on treatment policy at Phase IV (bottom). Left: Confident treatment effect estimates for the majority (M) subgroup with high uncertainty for the minority groups (m), leading to uncertain treatment policy at Phase IV. Right: While the majority group still can safely benefit from the treatment, better treatment estimates for minority groups are estimated, leading to a beneficial treatment policy.

is iteratively leveraged for active patient recruitment and treatment assignment. RFAN *adaptively* guides *who to recruit* and *which treatment arm to assign them to*, aiming at prioritizing the exploration of heterogeneous subgroups. Building on ideas from Causal-BALD (Bayesian Active Learning by Disagreement) (Jesson et al., 2021), we further propose strategies for practical implementation. Finally, we empirically evaluate realizations of our framework on synthetic data and two semi-synthetic real-world datasets of warfarin and COVID-19 patients, showing its potential impact.

Contributions: (1) We formulate **two new objectives**, PTMB and PTF, that we believe align more closely with the goals of post-marketing effectiveness and fairness while incorporating regulatory compliance. (2) We introduce **RFAN, a new trial design framework** designed to lead to effective and fair treatment policies for the target population while complying with regulatory standards. We present a practical implementation of RFAN trials by adapting novel methods for uncertainty estimation in ML. (3) We empirically evaluate RFAN on synthetic and semi-synthetic real-world datasets and demonstrate its potential impact.

2   RELATED WORK

Adaptive designs The biostatistics community has developed a large body of work on what is known as *adaptive designs* (Stallard et al., 2014; Thall, 2021). In contrast to the static design of RCTs, an adaptive clinical trial allows dynamic modification of different aspects of its future trajectory, based on intermediate

Omer Noy Klein, Alihan Hüyük, Ron Shamir, Uri Shalit, Mihaela van der Schaar

analyses, for better efficiency and ethics. This includes modifications in the trial population, assignment rules, and treatment options as the trial goes on. Notably, major regulatory agencies in Europe and the US have issued detailed guidelines on adaptive designs (Pallmann et al., 2018). Consequently, adaptive designs have received increased attention in recent years. Dominant approaches are *adaptive enrichment* (Mehta et al., 2009; Magnusson and Turnbull, 2013; Simon and Simon, 2013; Wang and Hung, 2013; Simon and Simon, 2018; Ondra et al., 2019; Thall, 2021; Stallard, 2023) and *adaptive signature designs* (Freidlin and Simon, 2005; Freidlin et al., 2010; Mi, 2017; Zhang et al., 2017; Bhattacharyya and Rai, 2019). However, these designs are limited in their adaptivity and flexibility, as they are restricted to a few pre-specified analysis time points. Adaptive enrichment designs are also limited by the number of pre-specified subgroups they can select between, which is typically two. Table A.3 highlights the distinct problem setting of our work compared to adaptive trial designs.

Multi-Armed Bandits The setting of a clinical trial is closely related to *pure exploration* problems in the bandit literature. In pure exploration problems the rewards of played arms are not the primary concern, but rather the reward of the singular arm identified at the end (Bubeck et al., 2009, 2011; Chen et al., 2014; Degenne and Koolen, 2019). This approach is studied in Best Arm Identification problems, which aim to identify the arm with the highest mean reward (e.g., Audibert et al. (2010)). In particular, an RCT can be viewed as a purely-exploratory bandit problem aiming to find the best arm for the entire population. Purely-exploratory bandits seek to identify the best arm, but they do not take into account patient-specific information, and thus cannot lead to patient-specific policies including policies that account for how under-served populations respond to treatment. While best-arm identification for contextual bandits has been studied recently, work so far only optimizes for overall reward marginalized over the entire distribution (Kato et al., 2022), not considering sub-populations or regulatory constraints. See Appendix A.2 for details.

3   PROBLEM SETTING

Here we formalize the problem setting and present our newly proposed objectives for clinical trials.

Patients, Treatments & Outcomes At each time point $t \in \{1, \ldots, T\}$ of the experiment (with finite T for practical feasibility), a pool of patients $\tilde{\mathcal{X}}_t \subseteq \mathcal{X}$ with each patient described by covariates $x \in \mathcal{X}$ becomes eligible for enrollment in the experiment (for the time being, we make no assumptions regarding how $\tilde{\mathcal{X}}_t$

evolves). Next, a particular patient $x_t \in \tilde{\mathcal{X}}_t$ among this pool is selected and enrolled in the experiment, and they are assigned to one of two treatment arms $w_t \in \mathcal{W} = \{0, 1\}$. Then, an outcome $y_t \in \mathcal{Y} \in \mathbb{R}$ is observed immediately. Following the Neyman-Rubin potential outcomes framework (Rubin, 1974), we assume that outcomes are determined by two potential outcome distributions $Y^0, Y^1 \in \Delta(\mathcal{Y})^{\mathcal{X}}$, and we assume causal consistency such that we observe $y_t \sim Y^{w_t}$. We let $\tau(x) = \mathbb{E}[Y^1 - Y^0 | x]$ be the conditional average treatment effect (CATE) for patient $x \in \mathcal{X}$, and given a patient distribution $X \in \Delta(\mathcal{X})$, let $\bar{\tau} = \mathbb{E}[Y^1 - Y^0]$ be the average treatment effect (ATE). Let $\pi \in \mathcal{W}^{\mathcal{X}}$ be a treatment policy that maps patients' covariates to treatment assignments. The expected outcome under a particular policy is known as the *policy value*, defined as: $V^\pi := \mathbb{E}_{x \sim X, w \sim \pi(x)}[Y^w]$. We also denote with $\mathcal{D}_t = \{(x_{t'}, w_{t'}, y_{t'})\}_{t'=1}^t$ the dataset generated by the entire enrollment and assignment process.

Experiment Designs The proposed trial design consists of three functions: (i) an acquisition function α, (ii) a hypothesis test η, and (iii) a planning strategy ρ. The acquisition function decides which patients are enrolled in the experiment and which treatment arms they are assigned to: $(x_t, w_t) \sim \alpha(\mathcal{D}_t, \tilde{\mathcal{X}}_t)$. The hypothesis test decides whether a treatment effect was detected, i.e. $\eta(\mathcal{D}_T) = 1$, or not detected, i.e. $\eta(\mathcal{D}_T) = 0$. The hypothesis test is typically strongly regulated and must be passed for the treatment to obtain approval. Finally, the planning strategy describes how to obtain a treatment policy π for assigning patients to treatments given the final experimental dataset: $\pi = \rho(\mathcal{D}_T)$. We assume that the resulting policy can only be deployed if a valid hypothesis test with type-I error control confirms the efficacy of the new treatment.

3.1   Objectives

Successful Conventional Trial The goal of a conventional RCT can be phrased as:

$$\max_{\alpha \in \mathcal{A}_{\text{non-ad}}} \quad \mathbb{P}_{\mathcal{D}_T \sim \alpha}\{\eta(\mathcal{D}_T) = 1 | \bar{\tau} > 0\}$$
$$\text{subject to} \quad \mathbb{P}_{\mathcal{D}_T \sim \alpha}\{\eta(\mathcal{D}_T) = 1 | \bar{\tau} \leq 0\} \leq \varepsilon,$$

where ε is a predefined error rate (significance level) and $\mathcal{A}_{\text{non-ad}}$ is the set of non-adaptive acquisition functions used in standard RCTs. The degrees of freedom these usually afford are defining apriori inclusion and exclusion criteria (e.g. age bracket, presence or absence of specific comorbidities), and the randomization ratio which is fixed but not always equal for each arm. Conventional clinical trials primarily aim at assessing the treatment's efficacy and safety during the trial, detecting a positive effect, and gaining regulatory approval.

We now introduce two new trial objectives. Our goal

is to develop experiment designs α that attain strong performance on these objectives.

Objective 1: Post-Trial Mean Benefit (PTMB)
We first consider the performance of a treatment policy upon deployment. Denote $V^\pi = \mathbb{E}_{x \sim X, w \sim \pi(x)}[Y^w]$ the deployment-time value of a treatment policy $\pi = \rho(\mathcal{D}_T)$, we have for acquisition function α:

$$\text{PTMB}(\alpha) =$$
$$\max_\rho \quad \mathbb{E}_{\mathcal{D}_T \sim \alpha} \left[\eta(\mathcal{D}_T) \cdot V^{\rho(\mathcal{D}_T)} + (1 - \eta(\mathcal{D}_T)) V^\emptyset \right]$$
$$\text{subject to} \quad \mathbb{P}_{\mathcal{D}_T \sim \alpha} \{ \eta(\mathcal{D}_T) = 1 | \bar{\tau} < 0 \} \leq \varepsilon. \quad (1)$$

We denote by $V^\emptyset = \mathbb{E}[Y^0]$ the value of the *control policy*, i.e. what we expect the patient outcomes to be had the trial failed and the new treatment is not deployed. Intuitively, for a potentially valuable treatment, PTMB maximizes the estimated overall clinical benefit of the new treatment compared to a standard treatment within the target population $(V^\pi - V^\emptyset)$, had the trial passed $(\eta(\mathcal{D}_T) = 1)$. For warfarin, this benefit could mean a reduced risk of adverse events, such as stroke or mortality.

Objective 2: Post-Trial Fairness (PTF) We further modify Objective 1 to explicitly capture a notion of fairness. In this work, by fairness we mean that a model does not disproportionately harm a pre-specified group of individuals defined by sensitive attributes such as sex or race. Let $s \in \mathcal{S}$ be sensitive attributes. Denote the conditional policy value $V_s^{\rho(\mathcal{D}_T)} = \mathbb{E}_{w \sim \rho(\mathcal{D}_t)}[Y^w | s]$ and similarly $V_s^\emptyset = \mathbb{E}[Y^0 | s]$. Here, the goal will be to find experiment designs leading to policies π with good max-min fairness (Frauen et al., 2023):

$$\text{PTF}(\alpha) =$$
$$\max_\rho \min_{s \in \mathcal{S}} \mathbb{E}_{\mathcal{D}_T \sim \alpha} \left[\eta(\mathcal{D}_T) \cdot V_s^{\rho(\mathcal{D}_T)} + (1 - \eta(\mathcal{D}_T)) V_s^\emptyset \right]$$
$$\text{subject to} \quad \mathbb{P}_{\mathcal{D}_T \sim \alpha} \{ \eta(\mathcal{D}_T) = 1 \mid \bar{\tau} \leq 0 \} \leq \varepsilon. \quad (2)$$

In words, Objective 2 aims to maximize the worst-case conditional policy value across sub-populations (e.g., across ethnic subgroups in the case of warfarin).

The novelty of our objectives lies in explicitly integrating both trial success and post-trial benefit into a single, forward-looking formulation. In contrast, existing trial designs focus primarily on maximizing trial success (e.g., Type I error control) or, in some cases, consider treatment effect-related measures separately. Table A.3 compares the objectives of the existing adaptive trial designs with ours. Globally optimizing (1) and (2) is intractable due to the dependence of the policy on the data obtained during the trial. Moreover, directly optimizing would be opaque to regulatory bodies, forming a formidable barrier to adoption. Instead, we next propose a two-stage design striking a balance between

regulatory needs and achieving good performance on the objectives.

4 RANDOMIZE FIRST AUGMENT NEXT (RFAN)

We now introduce RFAN, a framework for designing Phase III clinical trials specifically tailored to deliver robust results on our proposed PTMB and PTF objectives (Section 3.1). The trial procedure is outlined in Algorithm 1 and illustrated in Figure 2. Our approach has the following form:

$$\alpha(\tilde{\mathcal{X}}_t, \mathcal{D}_t, \mathcal{M}_t) =$$
$$\begin{cases} \text{Uniform}(\tilde{\mathcal{X}}_t \times \mathcal{W}) & t \in \{1, \dots, t^*\} \\ \alpha'(\tilde{\mathcal{X}}_t, \mathcal{D}_t, \mathcal{M}_t) & t \in \{t^* + 1, \dots, T\}, \end{cases}$$
$$\eta(\mathcal{D}_T) = \eta(\mathcal{D}_{t^*})$$

where $\mathcal{M}_t$ is a model fitted to $\mathcal{D}_t$ at each time step t. This includes fitting a CATE estimator $\hat{\tau}$ to $\mathcal{D}_t$ and deriving a treatment policy $\pi_t = \rho(\mathcal{D}_t)$. RFAN estimates two key quantities of interest: (i) the ATE, estimated in the randomized stage $(t \leq t^*)$, addressing regulatory and practical needs, and (ii) the CATE function or the resulting treatment policy for the target population, estimated in the augmented stage $(t^* < t)$, depending on the exact setup we use (detailed in Section 4.1). We note that the CATE and policy are functions, not scalar quantities.

RFAN's randomized stage is necessary to prove an error bound, as codified in regulatory procedures. This can be seen as a restriction of a conventional RCT to the first t^* rounds, with a potential loss of statistical power, to hopefully improve our objectives. The randomized stage retains the strong correctness guarantees of a standard RCT, including an upper bound on its error rate (ϵ). This structured separation ensures that any regulatory-mandated guarantees remain intact, while the augmented stage serves as an empirical extension that optimizes treatment policies under real-world considerations. For example, running an RFAN trial for warfarin could yield significant benefits by providing improved treatment policy for the population (Ma et al., 2022), and giving practical insights on subpopulations that might benefit more than others.

We next propose possible practical realizations of RFAN through the selection of the acquisition function α' (Section 4.1) and the switching point t^* (Section 4.2).

4.1 Selection of Acquisition Function

We propose using active learning to enroll patients for whom we have the greatest uncertainty into the
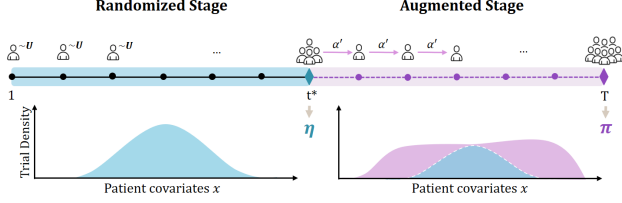
Omer Noy Klein, Alihan Hüyük, Ron Shamir, Uri Shalit, Mihaela van der Schaar

Figure 2: RFAN timeline. t^* and α' are free parameters. The randomized stage establishes a standard regulatory objective (η), and the augmented stage actively refines a treatment policy π, jointly addressing our objectives.

Algorithm 1 *Randomize First Augment Next (RFAN)*

Require: Acquisition function α', switching point t^* (predefined or adaptive, see Section 4.2), acquisition batch size b, experimental steps T, hypothesis test η, planning strategy ρ

1: Initialise: dataset $\mathcal{D}_0 = \emptyset$
2: **I) Randomized (Passive) Stage**
3: **for** $t \in \{1, ..., t^*\}$ **do**
4: Acquire a batch of patients $\{x_i\}_{i=1}^b$ from pool $\tilde{\mathcal{X}}_t$ and assign them uniformly at random to treatment arms $\{w_i\}_{i=1}^b$
5: Observe patients' outcomes $\{y_i\}_{i=1}^b$
6: Update: $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(x_i, w_i, y_i)\}_{i=1}^b$
7: **end for**
8: **II) Augmented (Active) Stage**
9: **for** $t \in \{t^* + 1, ..., T\}$ **do**
10: Fit model $\mathcal{M}_t$ to $\mathcal{D}_{t-1}$.
11: Estimate $\hat{\tau}_{\mathcal{M}_t}(x; \mathcal{D}_{t-1})$ and construct treatment policy $\pi_t = \rho(\mathcal{D}_{t-1})$
12: Use $\alpha'(\tilde{\mathcal{X}}_t, \mathcal{D}_{t-1}, \mathcal{M}_t)$ to acquire a batch of patients $\{x'_i\}_{i=1}^b$ and to assign each to treatment arm $\{w'_i\}_{i=1}^b$
13: Observe patients' outcomes $\{y'_i\}_{i=1}^b$
14: Update: $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(x'_i, w'_i, y'_i)\}_{i=1}^b$
15: **end for**
16: Conduct hypothesis test $\eta(\mathcal{D}_{t^*})$
17: Fit model $\mathcal{M}_T$ to $\mathcal{D}_T$
18: Use $\mathcal{M}_T$ to estimate $\hat{\tau}_{\mathcal{M}_T}(x; \mathcal{D}_T)$ and obtain final treatment policy $\pi_T = \rho(\mathcal{D}_T)$

---

trial and to assign them to treatment arms. In active learning for treatment effect estimation, a model is trained using existing labeled data that contains patient covariates, assigned treatments, and observed outcomes. Then, an acquisition function is used to assess the informativeness of new samples from data comprised of only covariates and treatments. After acquiring the most informative samples, their outcomes are observed, subsequently leading to model retraining and evaluation. We use Bayesian Active Learning by Disagreement (BALD) (Houlsby et al., 2011), which has been shown to be effective when deep learning methods are used. In the Bayesian active learning setup we are given an unlabeled dataset $\mathcal{D}_{\text{pool}} = \{x_i\}_{i=1}^{n_{\text{pool}}}$, a labeled training set $\mathcal{D}_{\text{train}} = \{x_i, y_i\}_{i=1}^{n_{\text{train}}}$, a Bayesian model $\mathcal{M}$ with parameters $\omega \sim p(\omega|D_{\text{train}})$ and predictions $p(y|x, \omega, \mathcal{D}_{\text{train}})$. An oracle provides us with the outcomes y for any data point in D_{pool}. At each step a batch of data $\{x_i^*\}_{i=1}^b$ is chosen from D_{pool} using an acquisition function α and the model is retrained. The choice of a Bayesian adaptive approach aligns with extensive existing literature on adaptive clinical trials, as it enables iterative learning under uncertainty and dynamic trial adaptations (Berry et al., 2010; Williamson et al., 2017; Thall, 2021; Ryan et al., 2022). Bayesian methods are particularly well-suited for such trials as they model uncertainty of the model parameter estimates (e.g., treatment effect) and are able to iteratively update them as new data becomes available. We note that RFAN is not restricted to the acquisition functions proposed below; alternative acquisition functions could be employed depending on the trial context and objectives.

BALD (Houlsby et al., 2011) uses an acquisition function based on estimating the Mutual Information (MI) between the model predictions and the model parameters $I(y; \omega|x, \mathcal{D}_{\text{train}}) = H(y|x, \mathcal{D}_{\text{train}}) - \mathbb{E}_{p(\omega|\mathcal{D}_{\text{train}})}[H(y|x, \omega, \mathcal{D}_{\text{train}})]$. The first term is the entropy of the model's prediction, indicating higher values when the model's predictions are uncertain. The second term is an expectation of the entropy

of the model prediction over the posterior of the model parameters. In words, BALD scores a data point x based on how well its label y would inform us about the true model parameters. The acquisition function is: $\alpha_{\text{BALD}}(\mathcal{D}_{\text{pool}}, p(\omega|D_{\text{train}})) = \text{argmax}_{\{x_i\}_{i=1}^b} I(y; \omega|x, \mathcal{D}_{\text{train}})$.

Causal-BALD Jesson et al. (2021) introduced new acquisition functions for active learning of individual-level causal-treatment effects from high dimensional observational data, based on BALD. In their setting, they assume the pool dataset contains treatment assignments $\mathcal{D}_{\text{pool}} = \{x_i, w_i\}_{i=1}^{n_{\text{pool}}}$, and the labels training set is $\mathcal{D}_{\text{pool}} = \{x_i, w_i, y_i\}_{i=1}^{n_{\text{pool}}}$. Jesson et al. (2021) have proposed $\alpha_{\mu\text{BALD}}$, among other causally-tailored acquisition functions: $\alpha_{\mu\text{BALD}}(\mathcal{D}_{\text{pool}}, p(\omega|D_{\text{train}})) = \text{argmax}_{\{x_i, w_i\}_{i=1}^b} I(Y^w; \omega|\{x_i, w_i\}, \mathcal{D}_{\text{train}})$.
$I(Y^w; \omega|\{x_i, w_i\})$ is a measure of the information gain for the model parameters ω if we obtain a label for the observed potential outcome Y^w given a data point (x, w) and $\mathcal{D}_{\text{train}}$. They show that $I(Y^w; \omega|\{x_i, w_i\}) \approx \text{Var}_{\omega \sim p(\omega|\mathcal{D}_{\text{train}})}(\hat{\mu}_\omega(x, w))$ where $\hat{\mu}_\omega(x, w)$ is a parametric estimator of $\mathbb{E}[Y|W = w, X = x]$.

Developing Acquisition Functions for Clinical Trials Our proposal for α' is built upon $\alpha_{\mu BALD}$ (Jesson et al., 2021). Departing from the causal-BALD setup, our approach guides not only *whom to recruit* but also *which treatment arm to assign them to*. In practice, we only have access to a pool of patients with covariates x, without treatment assignments. We let $\hat{\tau}_t(x) = \hat{\mu}_\omega(x, 1) - \hat{\mu}_\omega(x, 0)$ be the CATE estimator at time step t, based on the model $\mathcal{M}_t$ fit to dataset $\mathcal{D}_{t-1}$. $\hat{\tau}_t(x)$ is then used to inform a treatment policy $\pi_t(x)$. A standard approach for binary treatment is to use the sign of a CATE estimator to construct a policy: $\pi_t(x) = \rho(\mathcal{D}_{t-1}) = \mathbb{1}\{\hat{\tau}(x; \mathcal{D}_{t-1}) > 0\}$. At each time step t, our framework *acquires* a batch of patients $\{x_i\}_{i=1}^b$ and *assigns* each of them into treatment arms $\{w_i\}_{i=1}^b$, aiming to minimize uncertainty in the model parameters to predict the treatment effect. We propose various acquisition functions $\alpha'(\tilde{\mathcal{X}}_t, \mathcal{D}_t)$ for clinical trials, presented in Table A.4. Below we present two of the proposed functions:

(i) $\alpha_{\mu_\pi - Unf}$ acquires patients that maximally reduce uncertainty in model parameters to predict the potential outcome, given the treatment recommended by the current policy $\pi_t(x)$. Patients are assigned to treatment arms uniformly at random: $\alpha_{\mu_\pi - Unf}(\tilde{\mathcal{X}}_t, \mathcal{D}_t, \mathcal{M}_t) = (\{x_i^*\}_{i=1}^b, \{w_i^*\}_{i=1}^b)$, where $\{x_i^*\}_{i=1}^b = \arg\max_{\{x_i\}_{i=1}^b} I(Y^{\pi_t(x)}; \omega | x, \pi_t(x), \mathcal{D}_t)$ and $w_i^* \sim \text{Uniform}(\mathcal{W})$.

(ii) $\alpha_{\mathbf{sign}(\tau) - \pi}$ acquires patients that maximally reduce uncertainty in model parameters predicting the treatment policy $\text{sign}(\tau) = \mathbb{1}\{\tau(x) > 0\}$. Patients are assigned treatment according to the policy:

$$\alpha_{\text{sign}(\tau) - \pi}(\tilde{\mathcal{X}}_t, \mathcal{D}_t, \mathcal{M}_t) = (\{x_i^*\}_{i=1}^b, \{w_i^*\}_{i=1}^b)$$
$$\{x_i^*\}_{i=1}^b = \underset{\{x_i\}_{i=1}^b}{\arg\max} \, I(\text{sign}(Y^1 - Y^0); \omega | x, \mathcal{D}_t)$$
$$\{w_i^*\}_{i=1}^b = \{\pi_t(x_i^*)\}_{i=1}^b.$$

In real-world settings, assigning the correct treatment to a patient is often more clinically important than precisely estimating the potential outcomes. $\alpha_{\text{sign}(\tau) - \pi}$ is built on this principle.

Ideally, our acquisition functions (Table A.4) would lead to a better policy value V^π. We point out that the acquisition functions implicitly account for fairness. First, by actively acquiring patients that maximally reduce uncertainty, they prioritize exploration of underrepresented and/or heterogeneous subpopulations. Second, incorporating the current policy at each time point (e.g., in $\alpha_{\mu_\pi - Unf}$) may lead to a more robust policy at the end of the trial. Thus, they can be viewed as approximately optimizing our PTMB and PTF objectives without the regulatory constraints. We demonstrate these phenomena empirically in Section 5.

4.2 Selection of Switching Point t^*

For the selection of t^* we propose using early stopping strategies on top of our framework. Specifically, we conduct sequential hypothesis testing at predefined intermediate time points to determine whether to switch to the Augmented Stage or continue in the Randomized Stage. *Alpha spending* functions (Demets and Lan, 1994), which are well-established and validated in adaptive clinical trials (O'Brien and Fleming, 1979; Zhang and Saju, 2023), can be used to responsibly establish adjusted alpha values. We use the O'Brien and Fleming (1979) alpha spending function $\varepsilon(f) = 2 - 2\Phi\left(\frac{Z_{\varepsilon/2}}{\sqrt{f}}\right)$, where ε is the overall significance level, $f \in [0, 1]$ is the information fraction at the interim analysis, $Z_{\varepsilon/2}$ is the upper quantile of the standard normal distribution at $\varepsilon/2$ and Φ is the normal cumulative distribution function. The O'Brien-Fleming boundries are widely used in clinical trials Kumar and Chakraborty (2016); Liu and Mehta (2023), offering a conservative approach by reducing the risk of falsely concluding efficacy too early. See Appendix D.4 for more details.

5 EXPERIMENTS

In this section, we aim to evaluate RFAN and investigate its ability to target our newly established objectives. Evaluation of adaptive designs is a major challenge (Appendix B). Since ground truth of treatment effects is unavailable in practice (no access to counterfactual outcomes), a standard approach is using simulated data (Friede et al., 2012; Magnusson and Turnbull, 2013; Stallard et al., 2014; Henning and Westfall, 2015; Rosenblum et al., 2016), as often done in the literature for treatment effect estimation. We emulate RFAN on synthetic and two semi-synthetic real-world datasets: **PharmGKB** (Limdi et al., 2010), a dataset of warfarin patients, and **SIVEP-Gripe** (2020), a Brazilian COVID-19 hospitalization repository. For details on the experimental setup and implementation[2], please refer to Appendix D and Appendix E.

5.1 Datastets

Synthetic Data We modify the synthetic dataset presented in Kallus et al. (2019); Jesson et al. (2021):

$$\begin{aligned} X &\sim \mathcal{N}(0, 1) \\ Y^w &= (2w - 1)x + (2w - 1) - 2\sin((2w - 2)x) + \\ &\quad 2(1 + 0.5x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, 1). \end{aligned}$$

---

[2]The code for all experiments can be found at https://github.com/noyomer/rfan-trial or at https://github.com/Shamir-Lab/rfan-trial.

Omer Noy Klein, Alihan Hüyük, Ron Shamir, Uri Shalit, Mihaela van der Schaar

Table 1: Peformance comparison of realizations of RFAN with different acquisition functions (Table A.4) against Causal-BALD and RCT (Mean ± SEM). Error rate (ϵ) is set to 0.05. WC Policy Val.: the worst-case policy value conditioned on sensitive subgroups. ES: Early stopping. Arrows indicate whether higher or lower is better.

(a) Synthetic Data (T=30, N=300)

Design	Policy Val. ↑	WC Policy Val. ↑	% Succ. ↑	$\sqrt{\epsilon_{PEHE}}$ ↓	PTMB (Obj. 1) ↑	PTF (Obj. 2) ↑
RFAN $\alpha_{\text{sign}(\tau)-\pi}$ (t^*=7)	3.17 ± 0.02	1.23 ± 0.03	0.9 ± 0.1	1.24 ± 0.12	2.95 ± 0.21	1.22 ± 0.03
RFAN $\alpha_{\mu_{\pi}-max}$ (t^*=7)	3.17 ± 0.02	1.23 ± 0.03	0.9 ± 0.1	0.55 ± 0.09	2.95 ± 0.21	1.22 ± 0.03
RFAN $\alpha_{\mu_{max}}$ (t^*=7)	3.17 ± 0.02	1.23 ± 0.03	0.9 ± 0.1	0.55 ± 0.09	2.95 ± 0.21	1.22 ± 0.03
RFAN $\alpha_{\mu_{\pi}-Unf}$ (t^*=7)	3.17 ± 0.02	1.21 ± 0.04	0.9 ± 0.1	0.66 ± 0.11	2.95 ± 0.21	1.2 ± 0.04
RFAN $\alpha_{\mu_{\pi}}$ (t^*=7)	3.16 ± 0.02	1.17 ± 0.03	0.9 ± 0.1	1.12 ± 0.09	2.95 ± 0.21	1.16 ± 0.03
RFAN $\alpha_{\text{sign}\tau-\pi}$ (ES)	3.17 ± 0.02	1.24 ± 0.03	1.0 ± 0.0	0.87 ± 0.12	3.17 ± 0.02	1.24 ± 0.03
RFAN $\alpha_{\mu_{\pi}-max}$ (ES)	3.17 ± 0.02	1.24 ± 0.03	1.0 ± 0.0	0.69 ± 0.12	3.17 ± 0.02	1.24 ± 0.03
RFAN $\alpha_{\mu-max}$ (ES)	3.17 ± 0.02	1.24 ± 0.03	1.0 ± 0.0	0.69 ± 0.12	3.17 ± 0.02	1.24 ± 0.03
RFAN $\alpha_{\mu_{\pi}-Unf}$ (ES)	3.17 ± 0.02	1.23 ± 0.04	1.0 ± 0.0	0.49 ± 0.09	3.17 ± 0.02	1.23 ± 0.04
RFAN $\alpha_{\mu_{\pi}}$ (ES)	3.17 ± 0.02	1.21 ± 0.04	1.0 ± 0.0	0.82 ± 0.09	3.17 ± 0.02	1.21 ± 0.04
Causal-BALD	3.17 ± 0.02	1.22 ± 0.03	0.2 ± 0.13	0.54 ± 0.06	1.43 ± 0.3	1.13 ± 0.05
RCT	3.09 ± 0.04	0.55 ± 0.26	1.0 ± 0.0	1.13 ± 0.19	3.09 ± 0.04	0.55 ± 0.26

(b) Warfarin Data (T=40, N=400)

Design	Policy Val. ↑	WC Policy Val. ↑	%Succ. ↑	$\sqrt{\epsilon_{PEHE}}$ ↓	PTMB (Obj. 1) ↑	PTF (Obj. 2) ↑
RFAN $\alpha_{\text{sign}(\tau)-\pi}$ ($t^* = 20$)	**0.76 ± 0.01**	**0.57 ± 0.03**	1.0 ± 0.0	**0.82 ± 0.01**	**0.76 ± 0.01**	**0.57 ± 0.03**
RFAN $\alpha_{\mu_{\pi}-max}$ ($t^* = 20$)	0.74 ± 0.01	0.52 ± 0.04	1.0 ± 0.0	0.85 ± 0.02	0.74 ± 0.01	0.52 ± 0.04
RFAN $\alpha_{\mu_{max}}$ ($t^* = 20$)	0.75 ± 0.01	0.54 ± 0.06	1.0 ± 0.0	**0.82 ± 0.03**	0.75 ± 0.01	0.54 ± 0.06
RFAN $\alpha_{\mu_{\pi}-Unf}$ ($t^* = 20$)	0.75 ± 0.01	0.56 ± 0.03	1.0 ± 0.0	**0.82 ± 0.01**	0.75 ± 0.01	0.56 ± 0.03
RFAN $\alpha_{\mu_{\pi}}$ ($t^* = 20$)	0.73 ± 0.02	0.52 ± 0.04	1.0 ± 0.0	0.85 ± 0.03	0.73 ± 0.02	0.52 ± 0.04
Causal-BALD	0.75 ± 0.01	0.57 ± 0.03	0.2 ± 0.2	0.83 ± 0.02	0.7 ± 0.0	0.46 ± 0.01
RCT	0.72 ± 0.02	0.46 ± 0.03	1.0 ± 0.0	0.88 ± 0.03	0.72 ± 0.02	0.46 ± 0.03

We define two sensitive subgroups: $s_1 = \{x|x < -1.2\}$ and $s_2 = \{x|x \geq 1.3\}$, each about 10% of the population. Additional details can be found in Appendix D.

Warfarin Data We use the PharmGKB dataset (Limdi et al., 2010) of 5,700 patients treated with warfarin. The dataset contains clinical and genetic factors that have previously been associated with warfarin, including demographics (age, sex, ethnic group, weight, and smoking status), treatment reason, comorbidities, medications, and genetic factors (presence of genotype variants of CYP2C9 and VKORC1). Importantly, for about 4900 patients, the dataset includes the true patient-specific stable therapeutic doses of warfarin, which are initially unknown but were adjusted and determined by physicians over a few weeks, until the patient international normalised ratio (INR) is within a target range (the desired levels of anticoagulation).

We bucket the dosage into two treatment arms: low dosage (< 35 mg/week) and high dosage (≥ 35 mg/week), as in Kallus et al. (2022). We consider binary outcomes that assess the therapeutics' stability and assume for simulation purposes that a warfarin dosage categorized differently from the reported stable one is unstable. Accordingly, for each patient, we set the outcome to $Y = 1$ if they were assigned to the arm corresponding to the patient's true optimal dose, and $Y = 0$ otherwise. We define race and sex as sensitive attributes. Details can be found in Appendix D.2.

COVID-19 Data We analyze COVID-19 hospitalization data using SIVEP-Gripe (Sistema de Informação de Vigilância Epidemiológica da Gripe) (2020), a publicly available dataset in the Brazilian Ministry of Health database. Our cohort contains N=11,321 patients who were tested positive for SARS-CoV-2 and hospitalized. We simulate an experiment that assesses the effect of antiviral medications on patient survival. Additional details can be found at Appendix D.3.

5.2 Model

We use a deep-kernel GP model, which has been shown to perform well on high-dimensional inputs. Deep-kernel GP model uses a deep feature encoder (e.g., Neural network) to transform the inputs and defines a GP kernel over the extracted feature representation to make predictions. We note that, in contrast to Jesson et al. (2021), we do not assume access to a validation set, and therefore our experiments maintain active acquisition of a validation set during the trial. Once the trial is done, the model is tuned and retrained using the acquired unseen validation set. We describe model architecture and hyper-parameter space in Appendix E.

5.3 Baselines and Metrics

While various adaptive clinical trials exist, almost all of them either (i) target different design dimensions to adapt, (ii) operate under different constraints, or
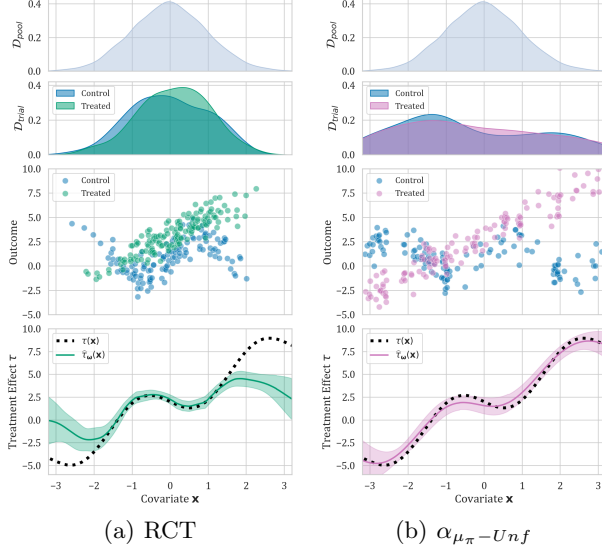
(a) RCT   (b) $\alpha_{\mu_\pi - Unf}$

Figure 3: Distribution of actively acquired synthetic trial population and resulting CATE function.

Figure 4: Performance over varying sample sizes (N) on the synthetic data, over 40 random seeds

(iii) aim to achieve different goals than our trial design. Hence, they are not always suitable benchmarks against RFAN (see Appendix B.2). We highlight the distinct problem setting of our work in Table A.3. We evaluate RFAN with our proposed acquisition functions α' (Section 4.1 and Table A.4) for various values $1 < t* < T$. We compare these strategies against: **(i) RCT:** Uniform patient acquisition and random assignment ($t^* = T$). **(ii) Causal-BALD:** As a reminder, Causal-BALD (Jesson et al. (2021)) assumes that $\mathcal{D}_{\text{pool}} = \{x_i, w_i\}_{i=1}^{n_{\text{pool}}}$, i.e., there is no active treatment assignment. We consider $\alpha_{\mu - max}$ (Table A.4) as an equivalent version of Causal-BALD with active treatment assignment. To closely emulate active learning only, we set $t^* = 1$ for this baseline. Note that $t^* = 0$ is infeasible as there is a minimal required number of data points to randomly acquire at the start for model initialization. Although the objective of Causal-BALD is not regulatory constrained, statistical testing is conducted to simulate its use in clinical trials.

For each dataset, we evaluate the trial designs using different performance metrics: test-set policy value $V^{\rho(\mathcal{D}_T)}$, test-set worst-case policy value over sensitive subgroups $\min_{s \in \mathcal{S}} V_s^{\rho(\mathcal{D}_T)}$, success rate $\%\eta(\mathcal{D}_T) = 1$, square root of precision in estimation of heterogeneous effect (PEHE) and our objectives (Section 3.1). We discuss measures of adaptive trial designs in Appendix B.3.

5.4 Experimental Results

Figure 3 illustrates the effect of each trial design on the distribution of the synthetic trial population (2nd top panel) and on the resulting estimated CATE function.
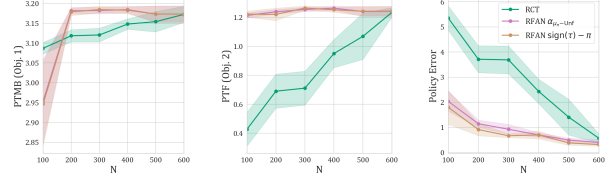
Comparing RCT and RFAN with $\alpha_{\mu_\pi - Unf}$, we can see that the population enrolled in the RCT (Figure 3(a)) is an unbiased sample of the pool data (top panel). However, RFAN with $\alpha_{\mu_\pi - Unf}$ (Figure 3(b)) demonstrates larger coverage of the target population, and in particular prioritization of enrollment of minority groups. The RFAN CATE estimator is more accurate overall, with better estimation for the sensitive subgroups and slightly worse in the central area.

For the synthetic dataset, we use $T = 30$ ($N = 300$) and investigate the performance using two predefined switching points $t^* = 7$ (Table 1a) and $t^* = 15$ (Table A.7). The differences between $t^* = 7$ and $t^* = 15$ illustrate the discussed tradeoff between the success rate and patient benefit. First, we observe that for $t^* = 7$, RFAN, using each of the acquisition functions, outperforms RCT in all metrics except success rate and PTMB (Objective 1). Increasing t^* to 15, the performance of RFAN with any acquisition function outperforms a conventional RCT in all metrics, with a consistent success rate of 100%. We further investigate the performance of RFAN using early stopping (ES) for selecting t^* (Table 1a), showing the RFAN with any acquisition function outperforms Causal-BALD and RCT, without requiring prior selection of t^*. Importantly, while the performance of our framework is comparably good for the Post-Trial Mean Benefit objective, it demonstrates significant improvement in the Post-Trial Fairness (maximizing worst-case policy value conditioned on sensitive subgroups), emphasizing our ability to inform a fair policy without harming, and even improving, the general policy. Table 1b demonstrates similar results for the warfarin dataset using a trial sample size of $N = 400$ ($T = 40$).

In an additional experiment, we investigate the performance of RFAN with $\alpha_{\mu_\pi - Unf}$ and $\alpha_{\text{sign}(\tau) - \pi}$ against RCT, over varying sample sizes. As can be seen from Figure 4, RFAN decreases the policy error rate for all sample sizes from 100 to 600 patients (reasonable sample sizes of clinical trials).

Supplementary Results Additional experiments, including those on COVID-19 data, are provided in Appendix F.

6 DISCUSSION

Significance of Main Results Looking at the synthetic benchmark (Table 1), RFAN with $\alpha_{\text{sign}(\tau)-\pi}$ achieves a two-fold increase in the PTF objective, compared to an RCT, and comparable results in PTMB. The warfarin experiment demonstrates 24% improvement. As shown in Figure 3, RFAN enables the detection of true treatment effects - both positive and negative - that a conventional RCT might overlook. These differences may lead to significant practical consequences, highlighting RFAN's potential to shape more informed future treatment policies and enhance care for under-served subgroups, without compromising the general population benefit. Note that unlike merely oversampling known underserved groups, our uncertainty-aware acquisition approach enables the recruitment of subgroups with diverse, heterogeneous traits that may not be known in advance or even readily interpretable by the trial designer. This approach aims to address the uncertainty in identifying which patients are likely to benefit from or be harmed by the treatment. Importantly, all standard RCT guarantees hold in RFAN's first stage, ensuring regulatory compliance.

Let us demonstrate RFAN's potential for improving warfarin dosing. An incorrect dose can lead to serious adverse outcomes, including thrombotic events and fatal bleeding. Between 2007 and 2009 warfarin was the leading cause of emergency hospitalizations for adverse drug events in older US adults (Budnitz et al., 2011). According to Sombat et al. (2023), the incidence rate of warfarin therapy complications was 4.91 events per 100 person-years. In our context, **the increase in policy value and reduction in error rate indicate a safer warfarin regimen for sensitive subpopulations**. In turn, this implies fewer hospitalizations and adverse events due to incorrect dosing. This could result in substantial improvements in patient outcomes and significant reductions in healthcare costs. Our results (Figure 4) further show a way for more efficient trials, achieving strong performance with fewer patients, thus preventing unnecessary risk and costs.

RFAN holds significant potential value to key healthcare stakeholders. After treatment approval, **payers** are particularly interested in *how the treatment benefits specific subpopulations*. RFAN could provide valuable evidence for better resource allocation across diverse populations, enhancing the drug's marketability. In turn, **caregivers** focus on *how to tailor therapies to individuals* and now could deliver more precise care for heterogeneous populations, which leads **patients** to benefit from improved clinical outcomes. We discuss ethical and regulatory considerations in Appendix G and Appendix H.

Conclusion We investigated how to improve patient welfare and fairness in clinical trials within the real-world regulatory constraints under which clinical trials are held. We examined the existing goals of clinical trials and proposed new objectives that go beyond passing regulatory barriers, bridging the gap between trial success and policy performance in real-world settings. We introduced RFAN, a framework for future designs of clinical trials, designed for achieving our objectives. Although RFAN demonstrates one way to optimize these objectives, we note that there could be other suitable frameworks that require future research. Finally, we present methodological strategies for implementing RFAN based on Causal-BALD (Jesson et al. (2021)). In synthetic and semi-simulated experiments with real-world patient data, we show that RFAN with uncertainty-aware acquisition functions can lead to significantly better outcomes for under-represented populations without harming the overall benefit of the population and with little or no compromise in the success rate of passing the crucial regulatory barriers.

References

(2020). Brasil. Ministério da Saúde. Departamento de informática: Dataset tagged SRAG. `https://opendatasus.saude.gov.br/dataset?tags=SRAG`. Accessed: 2024-01-01.

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.

Asiimwe, I. G. and Pirmohamed, M. (2022). Ethnic diversity and warfarin pharmacogenomics. *Frontiers in Pharmacology*, 13:866058.

Asiimwe, I. G., Zhang, E. J., Osanlou, R., Jorgensen, A. L., and Pirmohamed, M. (2021). Warfarin dosing algorithms: A systematic review. *British journal of clinical pharmacology*, 87(4):1717–1729.

Atan, O., Zame, W. R., and Schaar, M. (2019). Sequential patient recruitment and allocation for adaptive clinical trials. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1891–1900. PMLR.

Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best arm identification in multi-armed bandits. In *COLT*, pages 41–53.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256.

Aziz, M., Kaufmann, E., and Riviere, M.-K. (2021). On multi-armed bandit designs for dose-finding clinical trials. *The Journal of Machine Learning Research*, 22(1):686–723.

Baek, J. and Farias, V. (2021). Fair exploration via axiomatic bargaining. *Advances in Neural Information Processing Systems*, 34:22034–22045.

Baqui, P., Bica, I., Marra, V., Ercole, A., and van Der Schaar, M. (2020). Ethnic and regional variations in hospital mortality from covid-19 in brazil: a cross-sectional observational study. *The Lancet Global Health*, 8(8):e1018–e1026.

Bastani, H. and Bayati, M. (2020). Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294.

Bergstra, J., Yamins, D., and Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *International conference on machine learning*, pages 115–123. PMLR.

Berry, S. M., Carlin, B. P., Lee, J. J., and Muller, P. (2010). *Bayesian adaptive methods for clinical trials*. CRC press.

Bhattacharyya, A. and Rai, S. N. (2019). Adaptive signature design-review of the biomarker guided adaptive phase–iii controlled design. *Contemporary clinical trials communications*, 15:100378.

Bubeck, S., Munos, R., and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory: 20th International Conference, ALT 2009, Porto, Portugal, October 3-5, 2009. Proceedings 20*, pages 23–37. Springer.

Bubeck, S., Munos, R., and Stoltz, G. (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852.

Budnitz, D. S., Lovegrove, M. C., Shehab, N., and Richards, C. L. (2011). Emergency hospitalizations for adverse drug events in older americans. *New England Journal of Medicine*, 365(21):2002–2012.

Chen, S., Lin, T., King, I., Lyu, M. R., and Chen, W. (2014). Combinatorial pure exploration of multi-armed bandits. *Advances in neural information processing systems*, 27.

Chien, I., Deliu, N., Turner, R., Weller, A., Villar, S., and Kilbertus, N. (2022). Multi-disciplinary fairness considerations in machine learning for clinical trials. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 906–924.

Chow, S.-C. and Chang, M. (2008). Adaptive design methods in clinical trials–a review. *Orphanet journal of rare diseases*, 3:1–13.

Chow, S.-C. and Chang, M. (2011). *Adaptive Design Methods in Clinical Trials*. Chapman and Hall/CRC, New York, 2nd edition.

Chu, W., Li, L., Reyzin, L., and Schapire, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214. JMLR Workshop and Conference Proceedings.

Chuang-Stein, C., Anderson, K., Gallo, P., and Collins, S. (2006). Sample size reestimation: a review and recommendations. *Drug information journal: DIJ/Drug Information Association*, 40:475–484.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.

Committee, C. S. et al. (1996). A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (caprie). *The Lancet*, 348(9038):1329–1339.

Consortium, I. W. P. (2009). Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8):753–764.

Curth, A., Hüyük, A., and Van Der Schaar, M. (2023). Adaptive identification of populations with treatment benefit in clinical trials: machine learning challenges and solutions. In *International Conference on Machine Learning*, pages 6603–6622. PMLR.

Degenne, R. and Koolen, W. M. (2019). Pure exploration with multiple correct answers. *Advances in Neural Information Processing Systems*, 32.

Demets, D. L. and Lan, K. G. (1994). Interim analysis: the alpha spending function approach. *Statistics in medicine*, 13(13-14):1341–1352.

Fillion, N. (2019). Clinical equipoise and adaptive clinical trials. *Topoi*, 38(2):457–467.

Food, Administration, D., et al. (2020). Enhancing the diversity of clinical trial populations—eligibility criteria. *Enrollment Practices, and Trial Designs Guidance for Industry*.

Frauen, D., Melnychuk, V., and Feuerriegel, S. (2023). Fair off-policy learning from observational data. *arXiv preprint arXiv:2303.08516*.

Freedman, B. (2017). Equipoise and the ethics of clinical research. In *human experimentation and research*, pages 427–431. Routledge.

Freidlin, B., Jiang, W., and Simon, R. (2010). The cross-validated adaptive signature design. *Clinical cancer research*, 16(2):691–698.

Freidlin, B. and Simon, R. (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clinical cancer research*, 11(21):7872–7878.

Friede, T., Parsons, N., and Stallard, N. (2012). A conditional error function approach for subgroup selection in adaptive clinical trials. *Statistics in medicine*, 31(30):4309–4320.

Henning, K. S. and Westfall, P. H. (2015). Closed testing in pharmaceutical research: Historical and recent developments. *Statistics in biopharmaceutical research*, 7(2):126–147.

Houlsby, N., Huszár, F., Ghahramani, Z., and Lengyel, M. (2011). Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*.

Huynh, K. (2017). Milestone 6: Birth of the inr. *Nature reviews. Cardiology*.

Hüyük, A., Qian, Z., and van der Schaar, M. (2022). When to make and break commitments? In *The Eleventh International Conference on Learning Representations*.

Jesson, A., Tigas, P., van Amersfoort, J., Kirsch, A., Shalit, U., and Gal, Y. (2021). Causal-bald: Deep bayesian active learning of outcomes to infer treatment-effects from observational data. *Advances in Neural Information Processing Systems*, 34:30465–30478.

Kallus, N., Mao, X., and Zhou, A. (2019). Interval estimation of individual-level causal effects under unobserved confounding. In *The 22nd international conference on artificial intelligence and statistics*, pages 2281–2290. PMLR.

Kallus, N., Mao, X., and Zhou, A. (2022). Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, 68(3):1959–1981.

Kato, M., Imaizumi, M., Ishihara, T., and Kitagawa, T. (2022). Best arm identification with contextual information under a small gap. *arXiv preprint arXiv:2209.07330*.

Korn, E. and Freidlin, B. (2018). Interim monitoring for non-inferiority trials: minimizing patient exposure to inferior therapies. *Annals of Oncology*, 29(3):573–577.

Kornek, B. (2015). An update on the use of natalizumab in the treatment of multiple sclerosis: appropriate patient selection and special considerations. *Patient preference and adherence*, pages 675–684.

Kumar, A. and Chakraborty, B. S. (2016). Interim analysis: a rational approach of decision making in clinical trial. *Journal of advanced pharmaceutical technology & research*, 7(4):118–122.

Lachin, J. M. (1988). Statistical properties of randomization in clinical trials. *Controlled clinical trials*, 9(4):289–311.

Lee, H.-S., Shen, C., Jordon, J., and Schaar, M. (2020). Contextual constrained learning for dose-finding clinical trials. In *International Conference on Artificial Intelligence and Statistics*, pages 2645–2654. PMLR.

Lee, H.-S., Shen, C., Zame, W., Lee, J.-W., and van der Schaar, M. (2021). Sdf-bayes: Cautious optimism in safe dose-finding clinical trials with drug combinations and heterogeneous patient groups. In *International Conference on Artificial Intelligence and Statistics*, pages 2980–2988. PMLR.

Lehmacher, W. and Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics*, 55(4):1286–1290.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., and Stoica, I. (2018). Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:1807.05118*.

Limdi, N. A., Wadelius, M., Cavallari, L., Eriksson, N., Crawford, D. C., Lee, M.-T. M., Chen, C.-H., Motsinger-Reif, A., Sagreiya, H., Liu, N., et al. (2010). Warfarin pharmacogenetics: a single vkorc1 polymorphism is predictive of dose across 3 racial groups. *Blood, The Journal of the American Society of Hematology*, 115(18):3827–3834.

Liu, L. and Mehta, C. (2023). What types of formal interim analyses does the dmc review? In *Data Monitoring Committees (DMCs) Past, Present, and Future*, pages 71–89. Springer.

Ma, Z., Wang, P., Mahesh, M., Elmi, C. P., Atashpanjeh, S., Khalighi, B., Cheng, G., Krishnamurthy, M., and Khalighi, K. (2022). Warfarin sensitivity is associated with increased hospital mortality in critically ill patients. *Plos one*, 17(5):e0267966.

Maca, J., Bhattacharya, S., Dragalin, V., Gallo, P., and Krams, M. (2006). Adaptive seamless phase ii/iii designs—background, operational aspects, and examples. *Drug information journal*, 40(4):463–473.

Magnusson, B. P. and Turnbull, B. W. (2013). Group sequential enrichment design incorporating subgroup selection. *Statistics in medicine*, 32(16):2695–2714.

Mehta, C., Gao, P., Bhatt, D. L., Harrington, R. A., Skerjanec, S., and Ware, J. H. (2009). Optimizing trial design: sequential, adaptive, and enrichment strategies. *Circulation*, 119(4):597–605.

Mi, G. (2017). Enhancement of the adaptive signature design for learning and confirming in a single pivotal trial. *Pharmaceutical statistics*, 16(5):312–321.

Miller, F. G. and Brody, H. (2007). Clinical equipoise and the incoherence of research ethics. *The Journal of medicine and philosophy*, 32(2):151–165.

Najjar-Debbiny, R., Gronich, N., Weber, G., Khoury, J., Amar, M., Stein, N., Goldstein, L. H., and Saliba, W. (2023). Effectiveness of paxlovid in reducing severe coronavirus disease 2019 and mortality in high-risk patients. *Clinical Infectious Diseases*, 76(3):e342–e349.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556.

Ondra, T., Jobjörnsson, S., Beckman, R. A., Burman, C.-F., König, F., Stallard, N., and Posch, M. (2019). Optimized adaptive enrichment designs. *Statistical methods in medical research*, 28(7):2096–2111.

Pallmann, P., Bedding, A. W., Choodari-Oskooei, B., Dimairo, M., Flight, L., Hampson, L. V., Holmes, J., Mander, A. P., Odondi, L., Sydes, M. R., et al. (2018). Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC medicine*, 16(1):1–15.

Palmer, C. R. and Rosenberger, W. F. (1999). Ethics and practice: alternative designs for phase iii randomized clinical trials. *Controlled clinical trials*, 20(2):172–186.

Patel, M. R., Mahaffey, K. W., Garg, J., Pan, G., Singer, D. E., Hacke, W., Breithardt, G., Halperin, J. L., Hankey, G. J., Piccini, J. P., et al. (2011). Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *New England Journal of Medicine*, 365(10):883–891.

Pepperrell, T., Ellis, L., Wang, J., and Hill, A. (2022). Barriers to worldwide access for paxlovid, a new treatment for covid-19. In *Open forum infectious diseases*, volume 9, page ofac174. Oxford University Press US.

Polman, C. H., O'Connor, P. W., Havrdova, E., Hutchinson, M., Kappos, L., Miller, D. H., Phillips, J. T., Lublin, F. D., Giovannoni, G., Wajgt, A., et al. (2006). A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *New England Journal of Medicine*, 354(9):899–910.

Robertson, D. S., Lee, K. M., López-Kolkovska, B. C., and Villar, S. S. (2023). Response-adaptive randomization in clinical trials: from myths to practical considerations. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 38(2):185.

Rosenblum, M., Luber, B., Thompson, R. E., and Hanley, D. (2016). Group sequential designs with prospectively planned rules for subpopulation enrichment. *Statistics in medicine*, 35(21):3776–3791.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Rudick, R. A., Stuart, W. H., Calabresi, P. A., Confavreux, C., Galetta, S. L., Radue, E.-W., Lublin, F. D., Weinstock-Guttman, B., Wynn, D. R., Lynn, F., et al. (2006). Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *New England Journal of Medicine*, 354(9):911–923.

Ryan, E. G., Couturier, D.-L., and Heritier, S. (2022). Bayesian adaptive clinical trial designs for respiratory medicine. *Respirology*, 27(10):834–843.

Sampson, A. R. and Sill, M. W. (2005). Drop-the-losers design: normal case. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 47(3):257–268.

Schwartz, A. L., Alsan, M., Morris, A. A., and Halpern, S. D. (2023). Why diverse clinical trial participation matters. *New England Journal of Medicine*, 388(14):1252–1254.

Simon, N. and Simon, R. (2013). Adaptive enrichment designs for clinical trials. *Biostatistics*, 14(4):613–625.

Simon, N. and Simon, R. (2018). Using bayesian modeling in frequentist adaptive enrichment designs. *Biostatistics*, 19(1):27–41.

Sombat, B., Tongkaew, S., Nilwaranon, A., Mungthin, M., Jongcherdchootrakul, K., and Lertwanichwattana, T. (2023). Incidence and risk factors of warfarin therapy complications in community hospitals, central and eastern regions, thailand: a retrospective, multicenter, cohort study. *BMC Research Notes*, 16(1):1–8.

Stallard, N. (2023). Adaptive enrichment designs with a continuous biomarker. *Biometrics*, 79(1):9–19.

Stallard, N., Hamborg, T., Parsons, N., and Friede, T. (2014). Adaptive designs for confirmatory clinical trials with subgroup selection. *Journal of biopharmaceutical statistics*, 24(1):168–187.

Thall, P. F. (2021). Adaptive enrichment designs in clinical trials. *Annual review of statistics and its application*, 8:393–411.

Thorlund, K., Haggstrom, J., Park, J. J., and Mills, E. J. (2018). Key design considerations for adaptive clinical trials: a primer for clinicians. *bmj*, 360.

Varatharajah, Y. and Berry, B. (2022). A contextual-bandit-based approach for informed decision-making in clinical trials. *Life*, 12(8):1277.

Villar, S. S., Bowden, J., and Wason, J. (2015a). Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 30(2):199.

Villar, S. S. and Rosenberger, W. F. (2018). Covariate-adjusted response-adaptive randomization for multi-arm clinical trials using a modified forward looking gittins index rule. *Biometrics*, 74(1):49–57.

Villar, S. S., Wason, J., and Bowden, J. (2015b). Response-adaptive randomization for multi-arm clinical trials using the forward looking gittins index rule. *Biometrics*, 71(4):969–978.

Wang, S.-J. and Hung, H. J. (2013). Adaptive enrichment with subpopulation selection at interim: methodologies, applications and design considerations. *Contemporary clinical trials*, 36(2):673–681.

Williams, J. J., Nogas, J., Deliu, N., Shaikh, H., Villar, S. S., Durand, A., and Rafferty, A. (2021). Challenges in statistical analysis of data collected by a bandit algorithm: An empirical exploration in applications to adaptively randomized experiments. *arXiv preprint arXiv:2103.12198*.

Williamson, S. F., Jacko, P., Villar, S. S., and Jaki, T. (2017). A bayesian adaptive design for clinical trials in rare diseases. *Computational statistics & data analysis*, 113:136–153.

Wu, A. H., White, M. J., Oh, S., and Burchard, E. (2015). The hawaii clopidogrel lawsuit: the possible effect on clinical laboratory testing. *Personalized Medicine*, 12(3):179–181.

Zame, W. R., Bica, I., Shen, C., Curth, A., Lee, H.-S., Bailey, S., Weatherall, J., Wright, D., Bretz, F., and van der Schaar, M. (2020). Machine learning for clinical trials in the era of covid-19. *Statistics in biopharmaceutical research*, 12(4):506–517.

Zhang, J. and Saju, C. (2023). A systematic review of randomised controlled trials with adaptive and traditional group sequential designs–applications in cardiovascular clinical trials. *BMC Medical Research Methodology*, 23(1):200.

Zhang, Q., Dimairo, M., Julious, S. A., Lewis, J., and Yu, Z. (2024). Reporting and communication of sample size calculations in adaptive clinical trials: a review of trial protocols and grant applications. *BMC Medical Research Methodology*, 24(1):216.

Zhang, Z., Li, M., Lin, M., Soon, G., Greene, T., and Shen, C. (2017). Subgroup selection in adaptive signature designs of confirmatory clinical trials. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66(2):345–361.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes] See Section 3 for mathematical setting and Section 4 for a description of our framework.

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Not Applicable]

   (b) Complete proofs of all theoretical results. [Not Applicable]

   (c) Clear explanations of any assumptions. [Not Applicable]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] The code is included as part of the supplementary materials and will be publically available upon acceptance. Additionally, see data sharing statements in Appendix D.

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] See Appendix D and Appendix E.

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] See Section 5.

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] See Appendix D.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

(a) Citations of the creator If your work uses existing assets. [Yes]

(b) The license information of the assets, if applicable. [Not Applicable]

(c) New assets either in the supplemental material or as a URL, if applicable. [Yes] See data sharing statements in Appendix D.

(d) Information about consent from data providers/curators. [Not Applicable]

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Towards Regulatory-Confirmed Adaptive Clinical Trials: Machine Learning Opportunities and Solutions Supplementary Materials

Table A.1: Glossary of Key Terms in Clinical Trials

Term	Description
Adverse Event	An unfavorable medical occurrence in a patient after exposure to a treatment (e.g., disease, sign, or symptom). It may or may not be caused by the intervention being studied.
Eligibility Criteria	A set of rules that determine who can or cannot participate in a clinical study. Eligibility criteria include both *inclusion criteria* (requirements that must be met for a person to participate in the study) and *exclusion criteria* (conditions that prevent a person from participating).
Exclusion Criteria	A type of eligibility criteria. Specific factors that disqualify individuals from participating in a clinical trial (e.g., children, pregnant women).
Experimental Arm	A group of participants that receives the treatment being tested in a clinical trial. The group of participants who receive the placebo or the alternative treatment is called the control arm.
Inclusion Criteria	A type of eligibility criteria. Properties or characteristics that qualify individuals for participation in a clinical trial that define the target population (e.g., age, gender, ethnicity, medical conditions, etc.).
Informed Consent	The process by which participants are informed about the risks, benefits, and objectives of a study before agreeing to participate.
Observational Study	A study where participants are observed, without any treatment or intervention assigned by the researchers.
Outcome	The measured result used to assess the effectiveness of a treatment, such as symptom improvement or adverse effects.
Phase-I trials	Early-stage clinical trials assessing the safety and dosage of a treatment, typically involving a small number of participants.
Phase-II trials	Trials focused on evaluating the efficacy and safety of a treatment, often involving a larger group of participants than Phase I.
Phase-III trials	Trials that are designed to assess the effectiveness of the new treatment over a larger number of participants. They are the final step of research and development before regulatory agencies can approve the new treatment. Phase III trials are nearly always RCTs.
Phase-IV trials	Post-approval studies conducted to gather additional information on treatment effects in the target population.
Target Population	The entire population, or group, that a clinical trial is designed to study and draw conclusions about. The target population includes people who meet the eligibility criteria and represent the condition that the intervention/treatment aims to address.

A   ADDITIONAL LITERATURE REVIEW

A.1   Extended Review of RCTs

For decades, RCTs have been used for evaluating the effect of new treatments, becoming a gold standard. RCTs often enroll patients from a target population and assign them uniformly at random to treatment groups. Clinical trials are typically classified into four phases. Phase I and Phase II trials are used to assess safety and efficacy, respectively. Phase III (pre-marketing) trials are nearly always Randomized Controlled Trials (RCTs) and are designed to evaluate the effectiveness of the new treatment compared to a placebo or to standard treatment, and thus gain crucial regulatory approval for wide dissemination. Regulations usually require to control the familywise error rate (FWER), which bounds the probability of a Type 1 error, i.e., the probability of concluding the new treatment is more effective despite not being so. After approval, Phase IV (post-marketing surveillance stage) provides real-world evidence of the treatment's effectiveness and safety on a large scale in the actual target populations and under the actual treatment assignment policy. While Phase III primarily aims at obtaining regulatory approval, observing and learning real-world implications in Phase IV might be too late.

To illustrate the key challenges involved in the clinical development journey, consider the case of Warfarin. First approved for clinical use in 1954, it is one of the most widely used oral anticoagulant agents worldwide (Consortium, 2009; Huynh, 2017). It is used for the prevention and/or treatment of thromboembolic conditions, such as deep vein thrombosis, rheumatic heart disease, and the prevention of strokes in patients with atrial fibrillation. Despite being used for almost 70 years, warfarin dosing remains a significant challenge for practitioners, primarily due to the drug's narrow therapeutic range and substantial variability in individual responses. An incorrect dose result in serious adverse outcomes: an insufficient dose can trigger thrombotic events, while an excessive dose might lead to fatal bleeding. Between 2007 and 2009 warfarin was the leading cause of emergency hospitalizations for adverse drug events in older US adults (Budnitz et al., 2011). Several studies have been conducted to develop dose-optimization algorithms and assess their clinical utility using RCTs. However, most of these studies were conducted in White and Asian populations (Asiimwe et al., 2021; Asiimwe and Pirmohamed, 2022). Importantly, there are key genetic differences that influence warfarin response, making studies conducted in Whites/Asians less applicable to many other underrepresented populations. Yet clinical implementation guidelines for warfarin dosing have been deployed based on these RCTs, even though they are mostly applicable to White populations. Recent years have faced similar pressing challenges, with the emergence of the COVID-19 pandemic (Zame et al., 2020). For instance, Paxlovid (nirmatrelvir/ritonavir), an antiviral medication for COVID-19, still faces limited distribution due to budget constraints (Pepperrell et al., 2022). Given the variability in patient response (Najjar-Debbiny et al., 2023), identifying who will most benefit from the treatment remains a critical question. Further real-world examples can be found in Table A.2.

A.2   Extended Review of Bandit Literature

In recent years, there has been a growing trend in harnessing machine learning to improve the design of clinical trials, with prominent approaches in the multi-armed bandits (MAB) literature. MAB algorithms are particularly useful for clinical trials since they are designed to address the well-studied *exploration-exploitation* trade-off, which can be interpreted as a trade-off between *clinical research* and *clinical practice* in clinical trials. Recent work has focused on the application of MABs to dose-finding (Bastani and Bayati, 2020; Lee et al., 2020, 2021; Aziz et al., 2021; Baek and Farias, 2021) and treatment allocation (Atan et al., 2019; Varatharajah and Berry, 2022). However, while MABs hold the promise of improving patient welfare in clinical trials, there has been a continual hesitance to deploy them in practice (Chien et al., 2022). The perceived limitations of bandit methods include loss of statistical power, and challenges of statistical analysis, leading to practical barriers to regulatory approval and implementation (Villar et al., 2015a; Williams et al., 2021; Robertson et al., 2023). Some recent work has sought to tackle these challenges from an applied perspective. Villar et al. (Villar et al., 2015a,b; Villar and Rosenberger, 2018) proposed a patient allocation strategy for clinical trials, based on the forward-looking Gittins index to improve statistical power. Williamson et al. (2017) introduce a bandit-based design to improve statistical power and reduce treatment effect biases. Williams et al. (2021) highlight the difficulties of hypothesis testing with data from bandit models, and propose adjustments to existing statistical tests.

The majority of bandit work on clinical trials primarily focuses on maximizing the rewards of all arms that are "played" (e.g., Auer et al. (2002)), which would have corresponded to maximizing the benefit of enrolled patients *during* the trial (Phase III). Instead, in our design, Phase III is completely exploratory. In that sense, our domain

Table A.2: RCT Examples in the Literature

Study	Size	Treatment	Primary Outcome	Result	Real-World Implications
Patel et al. (2011)	14,264	Rivaroxaban vs. dose-adjusted warfarin	Stroke or systemic embolism	In patients with atrial fibrillation, rivaroxaban was non-inferior to warfarin	The evidence may be less applicable to underrepresented populations (trial recruited 83% Whites, 13% Asians, 1% Black, and 3% other) (Asiimwe and Pirmohamed, 2022)
Polman et al. (2006); Rudick et al. (2006)	942; 1,171	Natalizumab vs. placebo	Relapses and disability progression in Multiple Sclerosis (MS)	Natalizumab reduced the risk of the sustained progression of disability and the rate of clinical relapse in patients with relapsing MS	After its approval by the US FDA (2004), Natalizumab was temporarily withdrawn from the market due to cases of progressive multifocal leukoencephalopathy (PML). It was reintroduced in 2006 together with a Global Risk Management Plan (Kornek, 2015)
CAPRIE (1996)	19,185	Clopidogrel vs. aspirin	A composite outcome of myocardial infarction, ischemic stroke and vascular death	In patients at risk of ischaemic events, clopidogrel therapy resulted in a relative risk reduction of ischaemic stroke, myocardial infarction, or vascular death, compared with aspirin therapy	Thirteen years after it was approved and marketed, the US FDA issued a warning for poor metabolizers, which are more common in subpopulations other than the majority recruited in the trial (95% Whites). Such individuals, more common in East Asia and Pacific islands, do not recieve the full benefit of the treatment and therefore remain at risk for heart attack, stroke, and cardiovascular death (Wu et al., 2015).

more relates to the concept of *pure exploration* in the bandit literature, where the rewards of played arms are not the primary concern, but rather the reward of the singular arm identified at the end (Bubeck et al., 2009, 2011; Chen et al., 2014; Degenne and Koolen, 2019). This approach is exemplified in Best Arm Identification (BAI) problems, which aim to identify the arm (or the top-K arms) with the highest mean reward (e.g., Audibert et al. (2010)). In particular, an RCT itself can be viewed as a purely-exploratory bandit problem that aims to find the best arm for the entire population. Yet, while the purely-exploratory problems seek to identify the best arm in the trial, they do not explicitly consider the subsequent individual performances of that arm - in our context, the individualized treatment effects in real-world settings (Phase IV). The version of bandits that can deal with patient-specific effects is known s contextual bandits (Li et al., 2010; Chu et al., 2011), where typically the reward is modelled as an inner-product of some unknown arm-specific vector with the instance covariates. An RCT aimed for uncovering a strong policy, without regard to the ongoing reward, could be considered as a best-arm identification contextual bandit problem. So far this has only been studied in the context of learning the best model marginalized over the population, without accounting for sub-populations, fairness, or regulatory constraints (Kato et al., 2022).

B   EVALUATION CONSIDERATIONS

B.1   Evaluation of Clinical Trials

Evaluation of new designs of clinical trials is a major challenge, even in the biostatistics literature. Conducting a clinical trial that evaluates our new framework is infeasible in the scope of our paper, due to complex ethical, legal, and financial considerations. Ethically, ensuring the well-being and informed consent of participants is fundamental, requiring careful planning and approval from ethical review boards. Legally, compliance with strict regulations and guidelines, often varying across jurisdictions, poses a significant barrier. Financially, the substantial costs involved in clinical trials, from patient recruitment to data analysis, are beyond our current

resources. Beyond that, one of the challenges in our case is that our objectives are defined not only on the trial's performance but also based on Phase IV performance (i.e., policy value after deployment), had the trial been passed. This requires long-term control and monitoring of newly treated patients in the general public population. Given these complexities, a prospective validation of our framework is beyond the focus of this paper. Instead, simulated data are generally used to evaluate newly proposed clinical trial designs (Atan et al., 2019; Hüyük et al., 2022; Curth et al., 2023), even in the biostatistics literature (Friede et al., 2012; Magnusson and Turnbull, 2013; Stallard et al., 2014; Henning and Westfall, 2015; Rosenblum et al., 2016). We therefore evaluate our experimental designs on synthetic data, which enables verification with *known* ground truth. Furthermore, we evaluate our framework on **PharmGKB** (Limdi et al., 2010), a dataset of patients treated with warfarin, and **SIVEP-Gripe** (2020), a Brazilian COVID-19 hospitalization repository, for emulating real-world settings. Future work could explore additional evaluation approaches under these limiting considerations.

B.2   Trial Baselines

While numerous approaches exist for running adaptive clinical trials, almost all of them either (i) target different design dimensions to adapt, (ii) operate under different constraints, or (iii) aim to achieve different goals than our trial design. Hence, they are not always suitable benchmarks against RFAN.

For instance, some common types of adaptive trial designs include (Chow and Chang, 2008, 2011; Thorlund et al., 2018): (i) adaptive randomization design, where the ratio of patient in each treatment arms is updated to increase power (e.g., Lachin (1988)), (ii) group sequential design, which allows for early stopping due to futility (e.g., Lehmacher and Wassmer (1999)), (iii) sample size re-estimation designs, where the sample size is updated with interim analysis to maintain trial power (e.g., Chuang-Stein et al. (2006)), (iv) drop-the-losers design, where ineffective treatment arms are dropped mid-trial (e.g., Sampson and Sill (2005)), (v) adaptive enrichment/signature design, where both the inclusion criteria and the target population of the final analysis can be switched mid-trial (e.g., Simon and Simon (2013)), and (vi) adaptive seamless designs, where data from multiple trials are combined in one analysis (e.g., Maca et al. (2006)).

We have summarized the adapted design dimensions and the fixed design dimensions in these trial designs as well as their objectives in Table A.3 to highlight the distinct problem setting of our work. Adaptive enrichment/signature designs are the most closely related to RFAN as they also adapt the inclusion criteria. However, in an enrichment/signature design, this is done to maximize the success probability of the trial, often by adapting the target population of the final analysis along with the inclusion criteria (i.e. by changing the null hypothesis that is tested at the end of the trial). In our work, the null hypothesis is treated as a fixed design dimension.

B.3   Metrics for Evaluating Adaptive Clinical Trials

Adaptive trial designs have been extensively studied for decades, yet there is no widely accepted measure for comparing their performance and this remains an open research question. Adaptive trials are often evaluated based on the probability of trial success, i.e., whether a valid hypothesis test with type-I error control confirms the efficacy of the new treatment (denoted as $\eta(\mathcal{D}_T) = 1$ in our paper). In some cases, depending on the design context, other metrics are used, such as measures of trial resources needed (e.g., required sample size (Zhang et al., 2024)) and measures of ethical considerations (e.g., quantifying patient exposure to inferior treatment (Korn and Freidlin, 2018)). However, to our knowledge, there is no gold standard besides statistical guarantees such as power and error rate. In particular, no measure combines both regulatory requirements and the predicted patient benefit. Our primary contribution to this challenge is the introduction of two innovative objectives for future clinical trials – PTMB and PTF – which we view as steps towards closing the discrepancy between clinical trials and the way treatments are deployed in practice.

Table A.3: Adaptive trial designs comparison

Adaptive Trial Design	Adapted Design Dimensions	Fixed Design Dimensions	Objective
Adaptive Randomization Design	Distribution of patients across treatments	Error rate, Sample size, Target population, Inclusion criteria	Maximizing power
Group Sequential Design	Final sample size (< Target sample size)	Error rate, Target population, Inclusion criteria	Minimizing sample size while maintaining power
Sample Size Re-estimation Design	Target Sample Size	Error rate, Target population, Inclusion criteria	Minimizing sample size while maintaining power
Drop-the-Losers Design	Set of treatments	Error rate, Sample size, Target population, Inclusion criteria	Maximizing sensitivity in identifying effective treatments
Adaptive Enrichment / Signature Design	Inclusion criteria, Target population	Error rate, Sample size	Maximizing power
Adaptive Seamless Design	Sub-trial designs	Sub-trial objectives, Error rate, Target population	Maximizing power
RFAN (Ours)	Inclusion criteria, Distribution of patients across treatments	Error rate, Sample size, Target population	PTMB / PTF

C   RFAN: ACQUISITION FUNCTIONS

Table A.4: Acquisition Functions for Clinical Trials

Name	Formula	Description
α_{μ_π}	$\alpha_{\mu_\pi}(\tilde{\mathcal{X}}_t, \mathcal{D}_t) =$ $\underset{\{x_i\}_{i=1}^b, \{\pi_t(x_i)\}_{i=1}^b}{\operatorname{argmax}} I(Y^{\pi_t(x)}; \omega \mid x, \pi_t(x), \mathcal{D}_t)$	Acquires patients that maximally reduce uncertainty in model parameters to predict the potential outcome, given the treatment recommended by the current policy $\pi_t(x)$. Patients are assigned to treatment arms according to the policy recommendation.
$\alpha_{\mu-max}$	$\alpha_{\mu-max}(\tilde{\mathcal{X}}_t, \mathcal{D}_t) =$ $\underset{\{x_i\}_{i=1}^b, \{w_i\}_{i=1}^b}{\operatorname{argmax}} I(Y^w; \omega \mid x, w, \mathcal{D}_t)$	Acquires and assigns patients that maximally reduce the uncertainty, under any treatment $w \in \mathcal{W}$, regardless of the current policy.
$\alpha_{\mu_\pi-max}$	$\alpha_{\mu_\pi-max}(\tilde{\mathcal{X}}_t, \mathcal{D}_t) =$ $(\{x_i^*\}_{i=1}^b, \underset{\{w_i\}_{i=1}^b}{\operatorname{argmax}} I(Y^w; \omega \mid x^*, w, \mathcal{D}_t))$ Where $\{x_i^*\}_{i=1}^b = \underset{\{x_i\}_{i=1}^b}{\operatorname{argmax}} I(Y^{\pi_t(x)}; \omega \mid x, \pi_t(x), \mathcal{D}_t)$	Acquires patients according to the current policy, as in α_{μ_π}. Here, the treatment assignments are according to the treatment $w \in \mathcal{W}$ that maximally reduces the uncertainty for each patient.
$\alpha_{\mu_\pi-Unf}$	$\alpha_{\mu_\pi-Unf}(\tilde{\mathcal{X}}_t, \mathcal{D}_t) = (\{x_i^*\}_{i=1}^b, \{w_i^*\}_{i=1}^b)$ Where $\{x_i^*\}_{i=1}^b = \underset{\{x_i\}_{i=1}^b}{\operatorname{argmax}} I(Y^{\pi_t(x)}; \omega \mid x, \pi_t(x), \mathcal{D}_t)$ and $w_i^* \sim \operatorname{Uniform}(\mathcal{W})$	Acquires patients according to the current policy, as in α_{μ_π}, and assigns treatments uniformly at random.
$\alpha_{\operatorname{sign}(\tau)-\pi}$	$\alpha_{\operatorname{sign}(\tau)-\pi}(\tilde{\mathcal{X}}_t, \mathcal{D}_t) = (\{x_i^*\}_{i=1}^b, \{w_i^*\}_{i=1}^b)$ $\{x_i^*\}_{i=1}^b = \underset{\{x_i\}_{i=1}^b}{\operatorname{argmax}} I(\operatorname{sign}(Y^1 - Y^0); \omega \mid x, \mathcal{D}_t)$ $\{w_i^*\}_{i=1}^b = \{\pi_t(x_i^*)\}_{i=1}^b$	Acquires patients that maximally reduce uncertainty in model parameters to predict the treatment policy $\operatorname{sign}(\tau) = \mathbb{1}\{\tau(x) > 0\}$. Patients are assigned to treatment arms according to the policy recommendation. In real-world clinical settings, determining a correct individual treatment assignment is often more clinically important than accurately approximating the potential outcomes themselves. $\alpha_{\operatorname{sign}(\tau)-\pi}$ builds on this concept.

Omer Noy Klein, Alihan Hüyük, Ron Shamir, Uri Shalit, Mihaela van der Schaar

D EXPERIMENTAL DETAILS

Below we provide experimental details for the synthetic and semi-synthetic simulations. All experiments are run on an internal cluster consisting of several servers, including those equipped with Intel Xeon processors, AMD Opteron 6140 processors, and AMD EPYC 7702 64-Core processors. Code to implement our framework and reproduce all experiments is available at https://github.com/noyomer/rfan-trial or at https://github.com/Shamir-Lab/rfan-trial.

D.1 Synthetic Simulation

We modify the synthetic dataset presented in Kallus et al. (2019); Jesson et al. (2021):

$$X \sim \quad \mathcal{N}(0, 1)$$
$$Y^w = \quad (2w - 1)x + (2w - 1) - 2\sin((2w - 2)x) + 2(1 + 0.5x) + \epsilon$$

Where $\epsilon \sim \mathcal{N}(0, 1)$. Each random generation of the simulated data generates a pool dataset of size $10,000$ and a test set of size 2000. During each trial, at each acquisition step, the acquired data is randomly partitioned online into training and unseen validation sets (train/validation ratio = 0.9). We define two sensitive subgroups defined as $s_1 = \{x | x < -1.2\}$ and $s_2 = \{x | 1.3 \geq x\}$, each reflects about $10 - 11\%$ of the population. In this experiment, for illustrative purposes, while patients are acquired from D_{pool}, no new patients can join after the trial starts. We initialize D_{pool} to size $10K$.

We investigate two trial sample sizes: $N = 300$ ($T = 30$) and $N = 100$ ($T = 10$). In both cases, the batch size, b, (i.e., number of patients acquired at each acquisition step) is set to 10 patients. The planning strategy for obtaining treatment policy is set to sign$(\tau)=\mathbb{1}\{\tau(x) > 0\}$. We report results over 20 and 10 random realizations for $N = 100$ and $N = 300$, respectively. The seeds are i and $i + 1$ ($i \in [0, ..., 19]$), for the training and test set, respectively (the validation set is actively collected by the acquired data of the training set).

D.2 Warfarin Semi-Synthetic Simulation

PharmGKB Data This section describes the preprocessing of the PharmGKB dataset (Limdi et al., 2010). PharmGKB contains 5,700 patients who were treated with warfarin. The data for each patient include demographics (age, sex, ethnic group, weight, bmi, and smoking status), treatment reason (e.g., stroke), comorbidities (e.g., diabetes), current medications (e.g., aspirin), and genetic factors (presence of genotype variants of CYP2C9 and VKORC1).

We included only patients with available true stable doses of warfarin and its corresponding stable observed INR (N=4,850). We excluded patients with no BMI available (N=3,964).

All of the variables are categorical besides age, weight, and BMI. Categorical variables were encoded using *one-hot encoding*, including missingness indicators (i.e., indicate which values of the variable are originally observed). A missingness indicator was added for missing age as well. Variables with small variance (standard deviation < 0.05) were excluded. Continuous variables were standardized separately in the train and test sets of each trial.

Simulation We bucket the dosage into two treatment arms: low dosage (< 35 mg/week) and high dosage (≥ 35 mg/week) (as done in Kallus et al. (2022)). We consider binary outcomes that assess the therapeutics' stability and assume for simulation purposes that a warfarin dosage categorized differently from the reported stable one is considered unstable. Accordingly, for each patient, we set the outcome to 1 ($Y = 1$) if they were assigned to the arm corresponding to the patient's true optimal dose. Otherwise, the outcome is set to ($Y = 0$). We define race and sex as sensitive attributes.

We investigate sample size of $N = 400$ patients ($T = 40$), with $t^* = 20$. The batch size, b, (i.e., number of patients acquired at each acquisition step) is set to 10 patients. The planning strategy for obtaining treatment policy is set to sign$(\tau)=\mathbb{1}\{\tau(x) > 0\}$. We report results over 5 random seeds. For each seed, the dataset is split into training and test subsets (train/test ratio = 0.8) using the *scikit-learn* function train_test_split(). Then, during each trial, at each acquisition step, the acquired training data is randomly partitioned online into training and unseen validation subsets (train/validation ratio = 0.9).

Data Sharing The data is publicly available at PharmGKB repository.

D.3 COVID-19 Semi-Synthetic Simulation

SIVEP-Gripe Data   We analyze COVID-19 hospitalization data using the SIVEP-Gripe (Sistema de Informação de Vigilância Epidemiológica da Gripe) repository (2020), a publicly available dataset in the Brazilian Ministry of Health database. The full dataset comprised $N = 99,557$ patients. The data contains clinical features such as age, sex, symptoms, comorbidities, antiviral medications, ethnic groups and regions. Patient ethnicity was classified according to five categories: Branco (White), Preto (Black), Amarelo (East Asian), Indígeno (Indigenous), and Pardo (mixed ethnicity).

We included only patients who have tested positive for SARS-CoV-2 ($N = 19,940$). We excluded patients without ethnicity recorded ($N = 12,221$). we then excluded patients who were not hospitalized, resulting in a cohort of $N = 11,321$ patient admissions.

Simulation   We simulate an experiment that assesses the effect of receiving antiviral medications on patient survival. We consider two treatment arms: Receiving antiviral medication on the first day of hospital admission ($w = 1$), or not receiving antiviral medication ($w = 0$). For each patient, we set the outcome to 1 ($Y = 1$) if survived at the end of hospitalization, otherwise $Y = 0$. Following Baqui et al. (2020), we define region and race as sensitive attributes.

We investigate the scenario where the treatment is potentially successful. Therefore, we undersampled non-survivors from the treated group to create a biased sample for which the survival rates of treated patients are generally higher than the non-treated patients. To estimate the counterfactual outcomes, we trained 2 *XGBoost* classifiers: one classifier was trained on the treated group and the other on the control group. The trained models were then used to predict the counterfactual outcomes for each patient.

We investigate sample size of $N = 400$ patients ($T = 40$), with $t^* = 20$. The batch size, b, (i.e., number of patients acquired at each acquisition step) is set to 10 patients. The planning strategy for obtaining treatment policy is set to $\text{sign}(\tau) = \mathbb{1}\{\tau(x) > 0\}$. We report results over 8 random seeds. For each seed, the dataset is split into training and test subsets (train/test ratio = 0.8) using the *scikit-learn* function train_test_split(). Then, during each trial, at each acquisition step, the acquired training data is randomly partitioned online into training and unseen validation subsets (train/validation ratio = 0.9).

Data Sharing   The data is publicly available at SRAG 2020 repository.

D.4 Statistical Testing and Early Stopping

We employed the *scipy.stats* Independent two-sample t-test as the hypothesis test (η), with a significance level (ε) of 0.05. The test was conducted on the randomized portion only, i.e., on the sample recruited in the RFAN's Randomized Stage, or on the full sample of an RCT. We set $\eta = 1$ if a positive effect is detected (p-value $< \varepsilon$).

To select t^* using early stopping, we conduct sequential hypothesis testing. In experiments, we use *alpha spending* to responsibly adjust ε values at intermediate time points, such that the overall Type 1 error rate remains at the desired level. We practically use the O'Brien-Fleming alpha spending function (O'Brien and Fleming, 1979):

$$\varepsilon(t) = 2 - 2\Phi\left(\frac{Z_{\varepsilon/2}}{\sqrt{t}}\right) \tag{3}$$

Where ε is the overall significance level of the test, $t \in [0, 1]$ is the information fraction at the interim analysis, $Z_{\varepsilon/2}$ is the upper quantile of the standard normal distribution at $\varepsilon/2$ and Φ is the normal cumulative distribution function. The information fractions for intermediate evaluation were set apriori to $[0, 0.25, 0.5, 0.75, 1]$.

E   MODEL DETAILS

We use a deep-kernel GP model in our experiments. Deep-kernel GP model uses a deep feature encoder (e.g., Neural network) to transform the inputs and defines a (GP) kernel over the extracted feature representation to make predictions. The treatment variable is appended to the extracted features to form the input which is then fed into the GP.

Architecture   The full model architecture is as follows:

Layer (type:depth-idx)	Param #
DeepKernelGP	–
|-Sequential: 1-1	–
|-NeuralNetwork: 2-1	–
|-Sequential: 3-1	20,400
|-Activation: 2-2	–
|-Sequential: 3-2	–
|-VariationalGP: 1-2	–
|-VariationalStrategy: 2-3	1,515
|-CholeskyVariationalDistribution: 3-3	240
|-ConstantMean: 2-4	1
|-ScaleKernel: 2-5	1
|-RBFKernel: 3-4	1
|-Positive: 3-5	–
Total params: 22,158	

Hyper-parameters During each trial, the acquired data is randomly split into training and validation sets at each acquisition step. Once the trial is done, the model is tuned and retrained using the acquired unseen validation set. We use ray tune (Liaw et al., 2018) with the HyperOpt (Bergstra et al., 2013) optimization algorithm for hyper-parameter tuning. The hyper-parameters search space is given in Table A.5. The hyper-parameters selected for our final model are those found to minimize the mean loss of the defined objective.

Table A.5: Hyper-parameter search space

Hyper-parameter	Search Space
Kernel	[RBF, Matern]
ν (Matern)	[0.5, 1.5, 2.5]
inducing points	[15, 30, 60]
hidden units	[50, 100, 200]
network depth	[2, 3, 4]
negative slope	[ReLU (Agarap, 2018), 0.1, ELU (Clevert et al., 2015)]
dropout rate	[0.1, 0.2, 0.5]
spectral norm	[None, 0.95, 1.5]
batch size	[32, 64, 100, 200]
learning rate	[2e-4, 5e-4, 1e-3]

F SUPPLEMENTARY RESULTS

F.1 COVID-19

Table A.6: Performance comparison on COVID-19 data (N=400, T=40, t^*=20)

Design	Policy Val.	WC Policy Val.	% Succ.	% Policy Err.	$\sqrt{\epsilon_{PEHE}}$	PTMB (Obj. 1)	PTF (Obj. 2)
RFAN $\alpha_{\text{sign}(\tau)-\pi}$ (t^*=20)	0.76 ± 0.0	0.47 ± 0.02	0.88 ± 0.12	7.53 ± 0.86	0.52 ± 0.0	0.73 ± 0.02	0.44 ± 0.02
RFAN $\alpha_{\mu_{\pi-max}}$ (t^*=20)	0.76 ± 0.0	0.47 ± 0.02	0.88 ± 0.12	7.09 ± 0.76	0.52 ± 0.0	0.74 ± 0.02	0.44 ± 0.02
RFAN $\alpha_{\mu_{max}}$ (t^*=20)	0.76 ± 0.0	0.47 ± 0.02	0.88 ± 0.12	7.09 ± 0.76	0.52 ± 0.0	0.74 ± 0.02	0.44 ± 0.02
RFAN $\alpha_{\mu_{\pi-Unf}}$ (t^*=20)	0.75 ± 0.01	0.47 ± 0.02	0.88 ± 0.12	10.8 ± 3.88	0.53 ± 0.0	0.72 ± 0.03	0.44 ± 0.02
RFAN α_{μ_π} (t^*=20)	0.76 ± 0.0	0.47 ± 0.02	0.88 ± 0.12	7.45 ± 0.85	0.52 ± 0.0	0.73 ± 0.02	0.44 ± 0.02
Causal-BALD	0.76 ± 0.01	0.46 ± 0.02	0.12 ± 0.12	7.38 ± 0.46	0.53 ± 0	0.59 ± 0.02	0.36 ± 0.03
RCT	0.76 ± 0.0	0.39 ± 0.06	1.0 ± 0.0	9.97 ± 3.7	0.52 ± 0.01	0.76 ± 0.0	0.39 ± 0.06

F.2 Synthetic

We investigate the performance of RFAN with different acquisition functions, in three settings:

(i) Using a predefined t^*.

(ii) Selection of t^* using early stopping.

(iii) Evaluation over varying t^*.

We run our experiments on two sample sizes of $N = 300$ ($T = 30$) and $N = 100$ ($T = 10$).

F.2.1 N=300

Table A.7: Performance comparison on synthetic data (N=300, T=30, t^*=15)

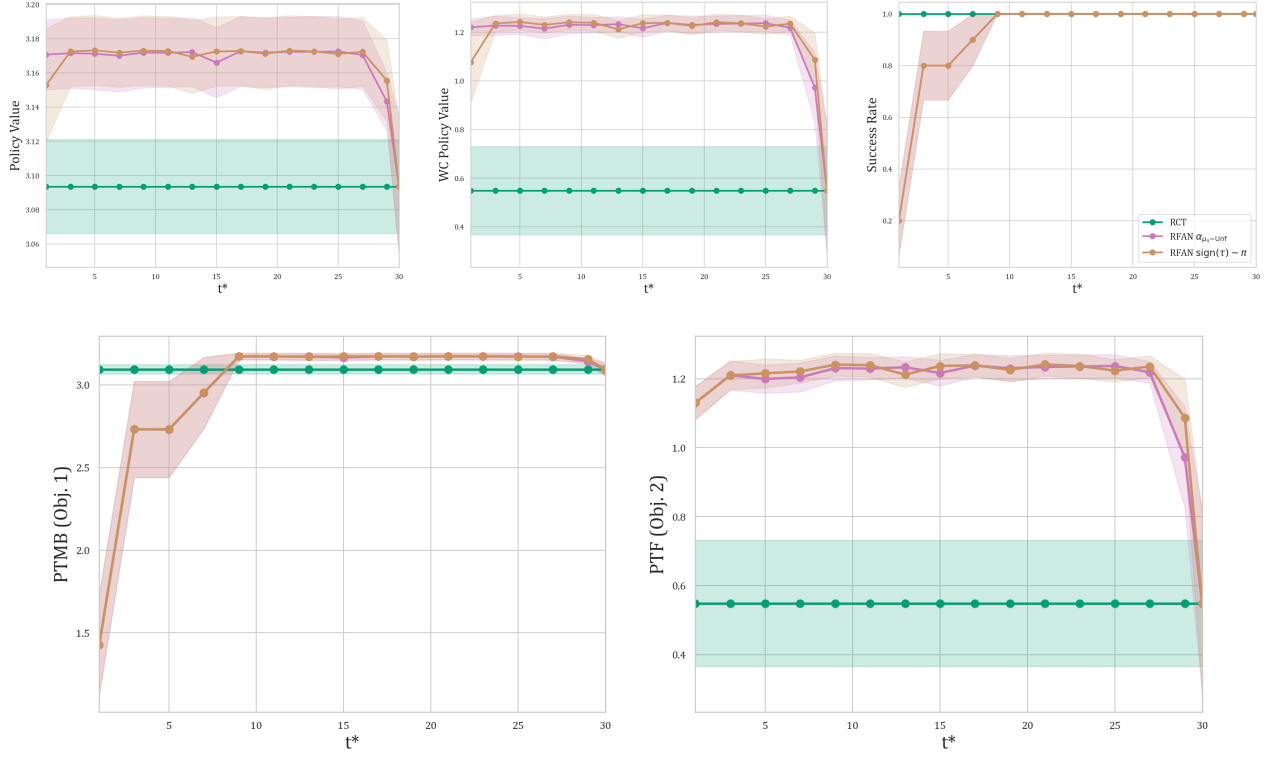| Design	Policy Val.	WC Policy Val.	% Succ.	% Policy Err.	$\sqrt{\epsilon_{PEHE}}$	PTMB (Obj. 1)	PTF (Obj. 2)
RFAN $\alpha_{\text{sign}(\tau)-\pi}$ (t^*=15)	3.17 ± 0.02	1.24 ± 0.04	1.0 ± 0.0	0.56 ± 0.15	0.77 ± 0.1	3.17 ± 0.02	1.24 ± 0.04
RFAN $\alpha_{\mu_{\pi-max}}$ (t^*=15)	3.17 ± 0.02	1.23 ± 0.03	1.0 ± 0.0	0.62 ± 0.17	0.45 ± 0.03	3.17 ± 0.02	1.23 ± 0.03
RFAN $\alpha_{\mu_{max}}$ (t^*=15)	3.17 ± 0.02	1.24 ± 0.03	1.0 ± 0.0	0.54 ± 0.17	0.46 ± 0.03	3.17 ± 0.02	1.24 ± 0.03
RFAN $\alpha_{\mu_{\pi-Unf}}$ (t^*=15)	3.17 ± 0.02	1.22 ± 0.04	1.0 ± 0.0	1.31 ± 0.65	0.54 ± 0.12	3.17 ± 0.02	1.22 ± 0.04
RFAN α_{μ_π} (t^*=15)	3.12 ± 0.02	0.77 ± 0.24	1.0 ± 0.0	2.77 ± 1.08	1.0 ± 0.26	3.12 ± 0.02	0.77 ± 0.24
Causal-BALD	3.17 ± 0.02	1.22 ± 0.03	0.2 ± 0.13	0.99 ± 0.23	0.54 ± 0.06	1.43 ± 0.3	1.13 ± 0.05
RCT	3.09 ± 0.04	0.55 ± 0.26	1.0 ± 0.0	4.17 ± 1.26	1.13 ± 0.19	3.09 ± 0.04	0.55 ± 0.26

Figure A.1: Performance over varying $t*$ on synthetic data (N=300, T=30)

F.2.2   N=100

Table A.8: Performance comparison on synthetic data (N=100, T=10, t^*=5)

Design	Policy Val.	WC Policy Val.	% Succ.	% Policy Err.	$\sqrt{\epsilon_{PEHE}}$	PTMB (Obj. 1)	PTF (Obj. 2)
RFAN $\alpha_{\text{sign}(\tau)-\pi}$ ($t^* = 5$)	3.17 ± 0.01	1.22 ± 0.02	0.85 ± 0.08	1.29 ± 0.27	1.29 ± 0.08	2.84 ± 0.18	1.21 ± 0.02
RFAN $\alpha_{\mu_{\pi}-max}$ ($t^* = 5$)	3.14 ± 0.03	1.13 ± 0.08	0.85 ± 0.08	3.16 ± 1.05	1.19 ± 0.09	2.83 ± 0.18	1.12 ± 0.08
RFAN $\alpha_{\mu_{max}}$ ($t^* = 5$)	3.14 ± 0.03	1.13 ± 0.08	0.85 ± 0.08	3.2 ± 1.04	1.19 ± 0.09	2.83 ± 0.18	1.12 ± 0.08
RFAN $\alpha_{\mu_{\pi}-Unf}$ ($t^* = 5$)	3.17 ± 0.02	1.23 ± 0.02	0.85 ± 0.08	1.68 ± 0.58	0.94 ± 0.07	2.84 ± 0.18	1.22 ± 0.02
RFAN $\alpha_{\mu_{\pi}}$ ($t^* = 5$)	3.16 ± 0.02	1.19 ± 0.02	0.85 ± 0.08	2.8 ± 0.57	1.18 ± 0.07	2.83 ± 0.18	1.18 ± 0.02
RFAN $\alpha_{\text{sign}\tau-\pi}$ (ES)	3.17 ± 0.01	1.23 ± 0.02	0.95 ± 0.05	1.27 ± 0.23	1.04 ± 0.08	3.07 ± 0.11	1.23 ± 0.02
RFAN $\alpha_{\mu_{\pi}-max}$ (ES)	3.17 ± 0.01	1.2 ± 0.02	0.95 ± 0.05	2.24 ± 0.32	1.04 ± 0.08	3.06 ± 0.11	1.2 ± 0.02
RFAN $\alpha_{\mu-max}$ (ES)	3.17 ± 0.01	1.2 ± 0.02	0.95 ± 0.05	2.17 ± 0.3	1.04 ± 0.08	3.06 ± 0.11	1.2 ± 0.02
RFAN $\alpha_{\mu_{\pi}-Unf}$ (ES)	3.17 ± 0.01	1.2 ± 0.03	0.95 ± 0.05	1.99 ± 0.41	1.09 ± 0.08	3.06 ± 0.11	1.2 ± 0.03
RFAN $\alpha_{\mu_{p}i}$ (ES)	3.16 ± 0.02	1.11 ± 0.08	0.95 ± 0.05	2.18 ± 0.43	1.16 ± 0.09	3.05 ± 0.11	1.11 ± 0.08
Causal-BALD	3.0 ± 0.08	1.2 ± 0.03	0.3 ± 0.11	9.12 ± 3.15	1.33 ± 0.11	1.65 ± 0.23	1.16 ± 0.03
RCT	3.07 ± 0.02	0.36 ± 0.17	1.0 ± 0.0	5.35 ± 0.8	1.46 ± 0.13	3.07 ± 0.02	0.36 ± 0.17
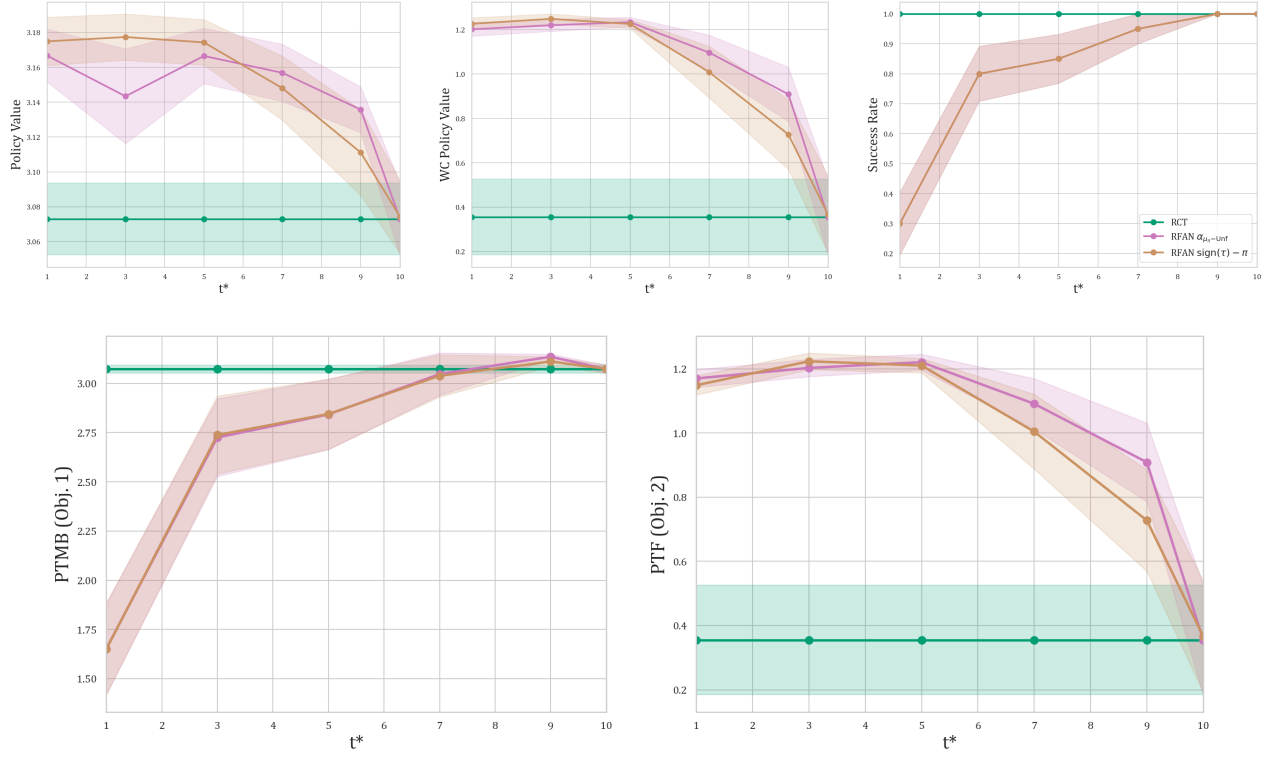
Figure A.2: Performance over varying $t*$ on synthetic data (N=100, T=10)

G ETHICAL CONSIDERATIONS

The ethical justification of clinical trials relies on the principle of clinical equipoise. Clinical equipoise establishes two conditions to be met: (i) There must be uncertainty among domain experts regarding the effectiveness of a treatment, and (ii) the trial should be designed to resolve this uncertainty (Miller and Brody, 2007; Freedman, 2017). As demonstrated in our proposed objectives, our trial is designed to learn and optimize how to treat patients once the treatment is widely deployed. In that sense, we believe that our design is more suited to the clinical equipoise, by reducing uncertainty about the treatment in practice.

The choice of employing adaptive trial designs as opposed to conventional non-adaptive trials is highly situational (Palmer and Rosenberger, 1999; Fillion, 2019). In many scenarios, conventional trial designs are still preferable, e.g., when there are long delays between patient enrollment and the observation of outcomes, or if a homogeneous response of the target population is expected. However, there are various cases in which RFAN is expected to be beneficial. We believe that employing RFAN would be of most value when there is high uncertainty regarding the effectiveness of the new treatment and when the treatment responses are expected to be highly heterogeneous among different subpopulations. We note that it should be only considered when the pool of patients eligible for enrollment shares the same characteristics as the target population. In such cases, we believe that our design is not only ethically permissible but also has the potential to improve patient care with a confirmatory guarantee.

Validating RFAN in real-world settings would essentially entail running a new trial. Conducting such a trial is beyond our scope, as a method development paper (discussed further in Appendix B). We therefore adhere to the standard approach of evaluating adaptive designs using simulated data (e.g., Atan et al. (2019); Hüyük et al. (2022); Curth et al. (2023); Friede et al. (2012); Magnusson and Turnbull (2013); Stallard et al. (2014); Henning and Westfall (2015); Rosenblum et al. (2016).

H REGULATORY CONSIDERATIONS

Regulatory approval is a crucial barrier for any clinical trial design. Our framework, RFAN, is purposely designed to narrow the gap between the regulatory requirements (i.e., trial success) and real-world treatment performance, without requiring a complete overhaul of the regulatory procedures currently in place. The first stage of RFAN follows a conventional randomized design, incorporating only well-established regulatory-compliant components, such as early stopping. This ensures that the initial phase of the trial aligns with existing standards, making approval relatively straightforward. The adaptive stage is designed to enhance fairness and real-world policy impact, compared to current practice, without compromising regulatory rigor. To assess RFAN's applicability and feasibility, we consulted with domain experts in clinical pharmacology and drug development. Their feedback affirmed the practical relevance and validity of our framework, ensuring that it addresses pressing concerns in current clinical trial designs. While regulatory bodies remain a key audience, we also emphasize the potential value of our methodology for other key healthcare stakeholders such as payers, pharmaceutical companies, caregivers, and healthcare providers.