Adaptive Experiment Design with Synthetic Controls

Alihan Hüyük
University of Cambridge

Zhaozhi Qian
University of Cambridge

**Mihaela van der Schaar**
University of Cambridge

Abstract

Clinical trials are typically run in order to understand the effects of a new treatment on a given population of patients. However, patients in large populations rarely respond the same way to the same treatment. This heterogeneity in patient responses necessitates trials that investigate effects on *multiple subpopulations*—especially when a treatment has marginal or no benefit for the overall population but might have significant benefit for a particular subpopulation. Motivated by this need, we propose SYNTAX, an exploratory trial design that identifies subpopulations with positive treatment effect among many subpopulations. SYNTAX is sample efficient as it (i) recruits and allocates patients *adaptively* and (ii) estimates treatment effects by forming *synthetic controls* for each subpopulation that combines control samples from other subpopulations. We validate the performance of SYNTAX and provide insights into when it might have an advantage over conventional trial designs through experiments.

1 INTRODUCTION

Randomized controlled trials (RCTs) remain an essential part of evidence-based medicine (Sackett and Rosenberg, 1995) despite their recognized shortcomings (Feinstein and Horwitz, 1997). Many of these shortcomings are starting to be addressed by machine learning (Curth et al., 2023; Hüyük et al., 2023); in this paper, we focus on a specific one: RCTs typically only consider *average* treatment effects across a given population, yet patients in large populations rarely respond the same way to receiving the same treatment. Differences in genetics, environments, and clinical backgrounds among patients lead to differences in the benefit they receive
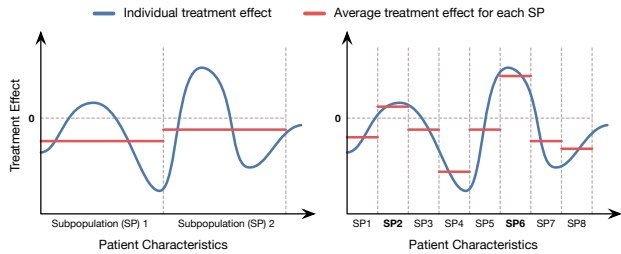
Figure 1: **Tradeoff between individual benefit and cost.** Consider two clinical trials that are both designed to confirm the effectiveness of a new treatment for one or multiple subpopulations. While Trial A investigates only two candidate subpopulations, Trial B investigates eight. As a result, Trial B has the potential to succeed for two subpopulations (SP2 & SP6) while Trial A is likely to fail for all. However, Trial B needs to allocate fewer samples to each subpopulation, which makes confirming positive effects more challenging. We propose SYNTAX as an exploratory pilot study that finds good subpopulations to target (such as SP2 & SP6) ahead of a confirmatory trial.

from a treatment as well. This heterogeneity in patient responses necessitates clinical trial designs that investigate the treatment effect not only for the overall population but also for various *subpopulations* within it (Chiu et al., 2018). Investigating multiple subpopulations at once becomes especially important when a treatment only has marginal benefit for the population as a whole while it might have significant benefit for a particular subpopulation (Moineddin et al., 2008; Lipkovich et al., 2017). Declaring such a treatment to be ineffective after a clinical trial that ignores heterogeneity might result in the treatment being needlessly denied to the subpopulation that would have actually benefited from it.

Running more "complex" trials that investigate larger numbers of subpopulations provides an opportunity to identify more personalized treatments for each subpopulation, rather than identifying one-size-fits-all treatments that are supposed to be beneficial for everyone (Wand et al., 2007; Lazar et al., 2016). However, such trials also come with increased costs as dividing the population into finer subpopulations naturally requires more patients to be recruited into the trial when the

recruitment is done in a randomized manner (Brookes et al., 2004; Ondra et al., 2016). When determining how many subpopulations a trial should investigate, there is a clear tradeoff between the benefit provided to the individual and the cost of running the trial (Figure 1).

Motivated by this issue, we propose SYNTAX (*Synthetic Adaptive eXploration*), an exploratory trial design for subpopulation selection. Exploratory (Phase I & II) trials are commonly run to establish a treatment's safety and to determine the correct dosage for the treatment. In contrast, confirmatory (Phase III) trials aim to validate the efficacy of a treatment with type I error control. SYNTAX, as an exploratory (pre-Phase III, similar to Phase I & II) trial design, aims to find suitable populations for which validating efficacy would be easier in a subsequent confirmatory trial. Specifically, our objective is to identify subpopulations with positive treatment effect among a pre-determined set of candidate subpopulations with a limited budget of samples—that is a limited number of patients that can be recruited into the trial. While a conventional Phase II trial determines the "right" dosage for a Phase III trial, in a similar vein, SYNTAX determines the "right" populations to target in a Phase III trial.

Our trial design has two key characteristics that contribute to its efficiency in identifying subpopulations:

(i) SYNTAX recruits patients and allocates them into control or treatment groups *adaptively*, not randomly. This allows more samples to be allocated into subpopulations that have higher variability in their responses to the treatment.

(ii) When estimating treatment effects, SYNTAX forms *synthetic controls* for each subpopulation, by decomposing them into a linear combination of others, instead of relying on control samples just from the subpopulation itself. This allows information from control samples to be shared between different subpopulations and more samples to be allocated to the treatment group.

An important difference bears emphasis here: Conventionally, synthetic controls replace the control group in a trial outright, they are sourced from *offline datasets* collected ahead of the trial, and no new control samples are observed during the trial itself. In this work, we consider synthetic control in an *online context* and investigate how they can still facilitate more efficient use of the control samples that are being collected as a trial continues. This setting is unique within the synthetic control literature, as it requires an experimenter to decide continually for each subpopulation, whether to (i) collect new real data as controls, or (ii) rely on synthetic controls constructed from previously observed data for other subpopulations.

Contributions   Our contributions are three-fold: *First,* we outline a treatment effect estimator based on decomposing subpopulations into linear combinations of others as in synthetic control. Within this estimator, the unique decision between collecting new real data vs. relying on synthetic controls is reflected by the fact that we allow decompositions of a given subpopulation to include the subpopulation itself.

Second, we provide an upper bound on the variance of our proposed estimator (Proposition 1), which can be tighter than the variance of a naive estimator based on conventional controls (Proposition 2). This upper bound also establishes the rate at which deviations from proper control samples—in favor of synthetic controls—contribute to the error in estimating treatment effects. Identifying this rate leads to an important insight: During an adaptive trial, constructing synthetic controls over collecting real control samples is the most effective when (latent) factors that lead to heterogeneity in patient responses have a stronger effect during the pre-treatment period than the post-treatment period.

Finally, making use of our upper bound, we propose a novel algorithm called SYNTAX that recruits patients from candidate subpopulations and assigns them into control or treatment groups adaptively to identify subpopulations with positive treatment effect in a sample-efficient manner. Although we motivate and formulate SYNTAX from the perspective of clinical trials, it remains generally applicable to any thresholding bandit problem with time-series contexts. Through numerical experiments, we validate the performance of our algorithm and provide insight into when it might have an advantage over conventional trial designs.

2   PROBLEM FORMULATION

We are given a pre-specified set of patient subpopulations $[K] = \{1, 2, \ldots, K\}$. These subpopulations could have been specified according to any arbitrary criteria (for instance, according to biomarkers or geographical location). Each subpopulation $i \in [K]$ has known/observable *features* $\boldsymbol{x}_i \in \mathbb{R}^{D_x}$ and unknown/unobservable *factor loadings* $\boldsymbol{z}_i \in \mathbb{R}^{D_z}$. As an example, the overall cardiovascular health of a subpopulation could be a latent factor that effects heart-related features such as ejection fraction, heart rate, and blood pressure. We let $X = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_K]$ and $Z = [\boldsymbol{z}_1 \cdots \boldsymbol{z}_K]$. Patients from population i, when they receive no treatment, exhibit the *baseline response* with mean

$$\bar{y}_{it} = \delta_t + \boldsymbol{w}_t^\mathsf{T} \boldsymbol{x}_i + \boldsymbol{\mu}_t^\mathsf{T} \boldsymbol{z}_i \quad \text{for} \quad t \in [T] \qquad (1)$$

where $\delta_t \in \mathbb{R}$ determines the constant portion of the baseline response, $\boldsymbol{w}_t \in \mathbb{R}^{D_x}$ are *weights* that determine the portion based on features, and $\boldsymbol{\mu}_t \in \mathbb{R}^{D_z}$ are *factors* that determine the portion based on factor loadings.

Now, suppose a clinical trial is run over episodes where each episode corresponds to one sample or the recruitment of one patient. At each episode $\eta \in [H]$, the experimenter first recruits a patient from some subpopulation $i[\eta] \in [K]$. Then, they observe the baseline response of the patient they have recruited:

$$y_t[\eta] \sim \mathcal{N}_{\sigma^2}(\bar{y}_{i[\eta]t}) \quad \text{for} \quad t \in \{1, \ldots, T-1\} \quad (2)$$

where $\mathcal{N}_{\sigma^2}(\mu)$ is the normal distribution with mean μ and variance σ^2. This is called the *pre-treatment period*. Once the pre-treatment period ends at time $t = T$, the experimenter assigns the recruited patient either to the control group, $\alpha[\eta] = 0$, or to the treatment group, $\alpha[\eta] = 1$. Depending on this assignment, they either continue observing the baseline response or they observe the response affected by the treatment:

$$y_T[\eta] \sim \begin{cases} \mathcal{N}_{\sigma^2}(\bar{y}_{i[\eta]T}) & \text{if} \quad \alpha[\eta] = 0 \\ \mathcal{N}_{\sigma^2}(r_{i[\eta]} + \bar{y}_{i[\eta]T}) & \text{if} \quad \alpha[\eta] = 1 \end{cases} \quad (3)$$

where r_i is the *treatment effect* for subpopulation i.

Objective The experimenter seeks to identify all of the subpopulations with positive treatment effect

$$\mathcal{I}^* = \{i \in [K] : r_i > 0\} \quad (4)$$

at the end of the episode horizon H, without knowing parameters $\{\delta_t, \boldsymbol{w}_t, \boldsymbol{\mu}_t\}$ or observing factor loadings $\{\boldsymbol{z}_i\}$. However, we assume that features $\boldsymbol{x}_i$ are observed ahead of the clinical trial. Denoting with $\hat{\mathcal{I}}^*$ the experimenter's best estimate of $\mathcal{I}^*$, we measure their success through their (i) *true positive rate* (TPR) given by $|\hat{\mathcal{I}}^* \cap \mathcal{I}^*|/|\mathcal{I}^*|$, and their (ii) *false positive rate* (FPR) given by $|\hat{\mathcal{I}}^* \cap [K] \backslash \mathcal{I}^*|/|[K] \backslash \mathcal{I}^*|$.

On Modeling Assumptions We formulated the problem of subpopulation selection with a linear model (cf. Equation 1) and normal error distributions (cf. Equation 2). It should be emphasized that these are not just arbitrary assumptions that we made because they are common in the literature. Instead, they are careful modeling decisions that are particularly suitable for capturing population-level effects as the units of interest in our problem happen to be populations of patients rather than individual patients themselves. Regarding linearity, Shi et al. (2022) explain how linear relationships can emerge on a population level when outcomes are averaged over multiple individuals, even when the generative process for an individual is non-linear—briefly, this is due to the inherent linearity of the expectation operator. Similarly, error distributions for outcomes averaged over multiple individuals can be approximated as normal due to the central limit theorem. We will evaluate the performance of SYNTAX under model mismatch—that is when the linearity in Equation 1 is violated—during our experiments in Section 6.

3 ESTIMATING TREATMENT EFFECTS WITH SYNTHETIC CONTROLS

As it will become more apparent why in Section 4, efficient exploration of the subpopulations and treatment groups rely on forming estimators of the treatment effects that have low variance. In order to find such estimators, we first start by analyzing the case where recruitment decisions $i[\eta], \alpha[\eta]$ are made independently from observed responses $y_t[\eta]$ so that the observed responses remain independent from each other—that is $y_t[\eta] \perp\!\!\!\perp y_{t'}[\eta']$ if $t \neq t'$ or $\eta \neq \eta'$ for all time steps $t, t' \in [T]$ and episodes $\eta, \eta' \in [H]$. For a given episode η, denote with

$$n_i^{(\alpha)} = \sum_{\eta' < \eta} \mathbb{I}\{i[\eta'] = i, \alpha[\eta'] = \alpha\} \quad (5)$$

the number of patients recruited from subpopulation i into treatment group α until that episode, and with

$$\hat{y}_{it}^{(\alpha)} = \sum_{\eta' < \eta} y_t[\eta'] \cdot \frac{\mathbb{I}\{i[\eta'] = i, \alpha[\eta'] = \alpha\}}{n_i^{(\alpha)}} \quad (6)$$

the empirical mean responses of those patients. Similarly, denote with $n_i = n_i^{(0)} + n_i^{(1)}$ the total number of patients recruited from subpopulation i (regardless of their treatment group) and with $\hat{y}_{it} = (\hat{y}_{it}^{(0)} n_i^{(0)} + \hat{y}_{it}^{(1)} n_i^{(1)})/n_i$ the empirical mean responses of those patients. We let $N^{(\alpha)} = \text{diag}(n_1^{(\alpha)}, \ldots, n_K^{(\alpha)})$, $N = \text{diag}(n_1, \ldots, n_K)$, $\hat{\boldsymbol{y}}_T^{(\alpha)} = [\hat{y}_{1T}^{(\alpha)} \cdots \hat{y}_{KT}^{(\alpha)}]^\mathsf{T}$, $\hat{\boldsymbol{y}}_{i \neg T} = [\hat{y}_{i1} \cdots \hat{y}_{i(T-1)}]^\mathsf{T}$, and $\hat{Y}_{\neg T} = [\hat{\boldsymbol{y}}_{1 \neg T} \cdots \hat{\boldsymbol{y}}_{K \neg T}]$.

Ignoring any potential relationship between the responses of different subpopulations, a naive estimate for the treatment effect r_i can be written as

$$\hat{r}_i^{\text{naive}} = \hat{y}_{iT}^{(1)} - \hat{y}_{iT}^{(0)} \quad (7)$$

which is an unbiased estimate of treatment effects, $\mathbb{E}[r_i - \hat{r}_i^{\text{naive}}] = 0$, and has a variance of $\mathbb{V}[r_i - \hat{r}_i^{\text{naive}}] = \sigma^2(1/n_i^{(0)} + 1/n_i^{(1)})$. However, considering true mean responses $\bar{y}_{iT}$ are related to each other as described in (1), it should be possible to form other estimates with tighter variances. Specifically, if the joint feature/factor space has lower dimensionality than the number of subpopulations—that is $D_x + D_z < K$—one can explore the feature/factor space more efficiently than the subpopulations themselves. Moreover, if the pre-treatment responses $\{y_t\}_{t<T}$ are observed for a time period long enough to infer unobserved factor loadings $\{\boldsymbol{z}_i\}$—that is $T > D_z$—such exploration could be feasible. We aim to achieve this via synthetic control.

Synthetic Control Rather than using $\hat{y}_{iT}^{(0)}$ as our control when estimating r_i, we first decompose subpopulation i as a linear combination of other subpopulations with weights $\boldsymbol{\beta} \in \mathbb{R}^K$ such that $\boldsymbol{x}_i = X\boldsymbol{\beta}$,

$z_i \approx Z\boldsymbol{\beta}$, and the elements of $\boldsymbol{\beta}$ sum up to one.[1] Then, (1) would imply that $\bar{y}_{iT} \approx \boldsymbol{\beta}^\mathsf{T}\bar{\boldsymbol{y}}_{\cdot T}$ hence we can use $\boldsymbol{\beta}^\mathsf{T}\hat{\boldsymbol{y}}_{\cdot T}^{(0)}$ as our control instead. This leads to a family of *synthetic estimators*:

$$\hat{r}_i(\boldsymbol{\beta}) = \hat{y}_{iT}^{(1)} - \boldsymbol{\beta}^\mathsf{T}\hat{\boldsymbol{y}}_{\cdot T}^{(0)} \tag{8}$$

In Proposition 1, we show that synthetic estimators become unbiased when $\boldsymbol{\beta}$ satisfies verifiable conditions that do not depend on unknown or unobservable quantities such as factors $\boldsymbol{\mu}_t$ or factor loadings $\boldsymbol{z}_i$. Moreover, we provide an upper bound on their variance. As we will argue next, synthetic estimators include the naive estimate in (7) as well, moreover, the tightness of the variance bound in Proposition 1 makes it possible to find synthetic estimates with even lower variances than the naive estimate.

Proposition 1. *Assuming $M_{\neg T} = [\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_{T-1}]$ has full rank and $T > D_z$, we have $\mathbb{E}[r_i - \hat{r}_i(\boldsymbol{\beta})] = 0$ and*

$$
\begin{aligned}
&\mathbb{V}[r_i - \hat{r}_i(\boldsymbol{\beta})] \\
&\quad \leq V_i(\boldsymbol{\beta}; N^{(0)}, N^{(1)}) \\
&\quad \doteq \sigma^2 \big(\underbrace{1/n_i^{(1)} + \|\boldsymbol{\beta}\|_{(N^{(0)})^{-1}}^2}_{\text{epistemic uncertainty}} + \underbrace{\lambda\|\boldsymbol{\beta} - \mathbf{1}_i\|_{N^{-1}}^2}_{\text{representation error}} \big)
\end{aligned} \tag{9}
$$

for $\lambda = \|M_{\neg T}^\mathsf{T}(M_{\neg T}M_{\neg T}^\mathsf{T})^{-1}\boldsymbol{\mu}_T\|^2$ when $\boldsymbol{\beta}$ is such that $\boldsymbol{x}_i = X\boldsymbol{\beta}$, $\hat{\boldsymbol{y}}_{\neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}$, and $\mathbf{1}^\mathsf{T}\boldsymbol{\beta} = 1$.

Proof. All proofs can be found in the appendix. □

3.1 Interpreting Proposition 1

To gain a more intuitive understanding of the bound in Proposition 1, first consider the trivial decomposition of subpopulation i as itself—that is $\boldsymbol{\beta} = \mathbf{1}_i$ where $(\mathbf{1}_i)_j = \mathbb{I}\{i = j\}$. Normally, having $\hat{\boldsymbol{y}}_{i\neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}$ would not guarantee $\boldsymbol{z}_i = Z\boldsymbol{\beta}$ due to the observation noise. However, for the trivial decomposition uniquely, we know for certain that $\boldsymbol{z}_i = Z\boldsymbol{\beta}$ holds. In other words, $\boldsymbol{\beta} = \mathbf{1}_i$ is the only known representation of subpopulation i that is *perfectly* accurate even in the latent factor space. The *representation error* $\lambda\|\boldsymbol{\beta} - \mathbf{1}_i\|_{N^{-1}}^2$ takes deviations from this perfect representation into account when quantifying the variance of synthetic estimators. Synthetic control relies on matching observed responses $\hat{y}$ between a subpopulation i and its representation $\boldsymbol{\beta}$ such that $\hat{\boldsymbol{y}}_{i\neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}$ as it leads to a near match in terms of unobservable factor loadings $\boldsymbol{z}_i \approx Z\boldsymbol{\beta}$ as well,

and notably, the deviations from the perfect representation are weighted by N^{-1} which corresponds to the uncertainty of observed responses $\hat{Y}_{\neg T}$.

The remaining terms capture the *epistemic uncertainty* of $\hat{y}_{iT}^{(1)}$ and $\hat{\boldsymbol{y}}_{\cdot T}^{(0)}$ which are composed together to form the final estimate $\hat{r}_i(\boldsymbol{\beta})$. Even when $\boldsymbol{\beta}$ is a perfect representation, the accuracy of $\hat{r}_i(\boldsymbol{\beta})$ is limited by how close $\hat{y}_{iT}^{(1)}$ and $\hat{\boldsymbol{y}}_{\cdot T}^{(0)}$ are to their ground-truth values.

Importantly, the trivial representation $\boldsymbol{\beta} = \mathbf{1}_i$ (of subpopulation i as itself) recovers the naive estimate in (7):

$$\hat{r}_i(\mathbf{1}_i) = \hat{r}_i^{\text{naive}} \tag{10}$$

Moreover, when $\boldsymbol{\beta} = \mathbf{1}_i$, the representation error in $V_i(\boldsymbol{\beta})$ vanishes and the epistemic uncertainty becomes identical to the naive estimate's variance—that is $V_i(\mathbf{1}_i) = \mathbb{V}[r_i - \hat{r}_i^{\text{naive}}]$. This means that the family of estimators $\hat{r}_i(\boldsymbol{\beta})$ not only include the naive estimate but also the variance bound in Proposition 1 is optimally tight for the naive estimate.

Having made this important observation, the key idea behind SYNTAX is to search for $\boldsymbol{\beta}$ that leads to even tighter variance bounds $V_i(\boldsymbol{\beta})$. This can easily be achieved by solving the quadratic program

$$\boldsymbol{\beta}_i^* = \operatorname*{argmin}_{\boldsymbol{\beta}: \boldsymbol{x}_i = X\boldsymbol{\beta}, \hat{\boldsymbol{y}}_{i\neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}, \mathbf{1}^\mathsf{T}\boldsymbol{\beta} = 1} V_i(\boldsymbol{\beta}) \tag{11}$$

Due to Proposition 1, it is guaranteed that $\hat{r}_i(\boldsymbol{\beta}_i^*)$ has variance at least as small as that of the naive estimator.

Proposition 2. $\mathbb{V}[r_i - \hat{r}_i(\boldsymbol{\beta}_i^*)] \leq \mathbb{V}[r_i - \hat{r}_i^{\text{naive}}]$.

3.2 When is synthetic control the most effective in estimating treatment effects?

Proposition 1 provides an intuition as to when synthetic control might be the most effective in estimating treatment effects. Notice that, in Proposition 1, the representation error contributes to the total variance proportionally to constant $\lambda = \|M_{\neg T}^\mathsf{T}(M_{\neg T}M_{\neg T}^\mathsf{T})^{-1}\boldsymbol{\mu}_T\|^2$. Earlier, we have discussed how representation error is caused mainly due to a mismatch between a subpopulation's factor loadings $\boldsymbol{z}_i$ and the synthetic composition $Z^\mathsf{T}\boldsymbol{\beta}$. As such, it is no surprise that λ exclusively depends on factors $\boldsymbol{\mu}_t$. Since

$$\lambda = \|M_{\neg T}^\mathsf{T}(M_{\neg T}M_{\neg T}^\mathsf{T})^{-1}\boldsymbol{\mu}_T\|^2 \leq (\|\boldsymbol{\mu}_T\|/\|M_{\neg T}\|)^2 \tag{12}$$

we also expect that λ—hence the contribution of representation error to the total variance—is smaller when post-treatment factors are small in magnitude (i.e. $\|\boldsymbol{\mu}_T\|$ is small) and pre-treatment factors are large in magnitude (i.e. $\|M_{\neg T}\|$).

This makes intuitive sense as well: With synthetic control, we are essentially trying to infer factor loadings $\boldsymbol{z}_i$ from pre-treatment responses $\{y_t\}_{t<T}$ in order to

---

[1]It is intentional here that $\boldsymbol{x}_i = X\boldsymbol{\beta}$, $\boldsymbol{z}_i \approx Z\boldsymbol{\beta}$ and *not* $\boldsymbol{x}_i = X_{\neg i}\boldsymbol{\beta}$, $\boldsymbol{z}_i \approx Z_{\neg i}\boldsymbol{\beta}$. This is part of our contribution: Unlike previous work on synthetic control, we allow weights $\boldsymbol{\beta}$ to include a subpopulation itself such that $\boldsymbol{x}_i = \boldsymbol{x}_i\beta_i + X_{\neg i}\boldsymbol{\beta}_{\neg i}$, $\boldsymbol{z}_i \approx \boldsymbol{z}_i\beta_i + Z_{\neg i}\boldsymbol{\beta}_{\neg i}$. Of course, one trivial $\boldsymbol{\beta}$ that satisfies this condition would be $\beta_i = 1$, $\boldsymbol{\beta}_{\neg i} = \mathbf{0}$ (i.e. $\boldsymbol{\beta} = \mathbf{1}_i$). However, as we will discuss next, alternative weights can lead to lower-variance estimates.

estimate the post-treatment responses y_T. During the pre-treatment period, if the contribution of factors to responses is stronger relative to the noise level—that is when $\|M_{\neg T}\|/\sigma$ is large—we expect our inference of factor loadings to be more accurate. Later, when we estimate the post-treatment response y_T, if the contribution of factors to the final response is now weaker relative to the noise level—that is when $\|\boldsymbol{\mu}_T\|/\sigma$ is small—we expect to gain more from our earlier inference of the factor loadings compared with a naive estimate (which now has to detect a weaker signal without any additional information from the previous time steps $t < T$ during the pre-treatment period). We will confirm this intuition empirically with simulations in Section 6.

4   SYNTAX

Making use of the synthetic estimator $\hat{r}_i(\boldsymbol{\beta}^*)$ given by (11) and the upper bound on its variance given in Proposition 1, we propose SYNTAX described in Algorithm 1. It is an adaptation of the algorithm of Locatelli et al. (2016) for solving thresholding bandit problems. At each episode $\eta \in [H]$:

(i) First, a *sensitivity index* is computed for each sub-population $i \in [K]$:

$$S_i = \frac{|\hat{r}_i(\boldsymbol{\beta}_i^*)|}{\sqrt{V_i(\boldsymbol{\beta}_i^*)}} \leq \frac{|\hat{r}_i(\boldsymbol{\beta}_i^*)|}{\sqrt{\mathbb{V}[r_i - \hat{r}_i(\boldsymbol{\beta}_i^*)]}} \qquad (13)$$

Intuitively, the lower the sensitivity index of a subpopulation is, the harder it is to determine whether its treatment effect is positive or not.

(ii) Then, the subpopulation $i^* = \arg\min S_i$ with the lowest sensitivity index is determined and the patient that is expected to cause the largest increase in S_{i^*} is recruited.

Algorithm 1 SYNTAX

Parameters: Horizon H, factor effect parameter λ
Output: Subpopulations $\hat{\mathcal{I}}^*$ with positive treat. effect

---

1: $n_i^{(\alpha)} \leftarrow 0$, $\hat{y}_{iT}^{(\alpha)} \leftarrow 0$, $\hat{\boldsymbol{y}}_{i\neg T} \leftarrow \mathbf{0}$, $\forall i \in [K], \alpha \in \{0,1\}$
2: **for** $\eta \in \{1, 2, \ldots, H\}$ **do**
3: $\quad i^* \leftarrow \arg\min_{i \in [K]} S_i$
4: $\quad i, \alpha \leftarrow \arg\min_{i \in [K], \alpha \in \{0,1\}}$
$\qquad \min_{\boldsymbol{\beta}: \, \boldsymbol{x}_{i^*} = X\boldsymbol{\beta}, \, \hat{\boldsymbol{y}}_{i^*\neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}, \, \mathbf{1}^\mathsf{T}\boldsymbol{\beta}=1}$
$\qquad V_{i^*}(\boldsymbol{\beta}; N^{(0)} + (1-\alpha)\mathbf{1}_i\mathbf{1}_i^\mathsf{T}, N^{(1)} + \alpha\mathbf{1}_i\mathbf{1}_i^\mathsf{T})$
5: $\quad$ Recruit from population i and treatment group α
6: $\quad$ Observe baseline outcomes $\boldsymbol{y}_{\neg T} = [y_1 \cdots y_{T-1}]^\mathsf{T}$
$\qquad\qquad\qquad\qquad$ and the final outcome y_T
7: $\quad n_i^{(\alpha)} \leftarrow n_i^{(\alpha)} + 1$
8: $\quad \hat{\boldsymbol{y}}_{i\neg T} \leftarrow \hat{\boldsymbol{y}}_{i\neg T} + (\boldsymbol{y}_{\neg T} - \hat{\boldsymbol{y}}_{i\neg T})/n_i$
9: $\quad \hat{y}_{iT}^{(\alpha)} \leftarrow \hat{y}_{iT}^{(\alpha)} + (y_T - \hat{y}_{iT}^{(\alpha)})/n_i^{(\alpha)}$
10: **end for**
11: $\hat{\mathcal{I}}^* \leftarrow \{i \in [K] : \hat{r}_i(\boldsymbol{\beta}_i^*) > 0\}$

Finally, when the experiment ends, SYNTAX simply returns all subpopulations that have a positive treatment effect estimate: $\hat{\mathcal{I}}^* = \{i \in [K] : \hat{r}_i(\boldsymbol{\beta}_i^*)\}$. Note that, computing $V_i(\boldsymbol{\beta})$ hence sensitivity indices S_i requires specifying λ. The ideal λ given Proposition 1 cannot be computed in practice as it depends on unknown factors $\boldsymbol{\mu}_t$, and we call it the *factor effect parameter* due to this dependence.

5   RELATED WORK

Our objective in this paper is to identify subpopulations with positive treatment effect given a fixed budget of samples. A trial designed to do so would consist of (i) an *inference strategy* that dictates which subpopulation is identified at the end of the trial, and (ii) a *sampling strategy* that dictates how the samples are allocated between different subpopulations as well as control and treatment groups. Our trial design, SYNTAX, happen to combine inference techniques from the *synthetic control* literature and sampling principals originally developed for solving *thresholding bandit* problems. Table 1 summarizes alternative trial designs one might consider, which differ from SYNTAX in terms of their inference and sampling strategies. We give an overview of these alternative strategies in this section.

Synthetic Control   It is not always practical to perform large randomized experiments to understand the effects of an intervention (Bica et al., 2020). For instance, investigating the effects of new policies targeting large geographic areas, often at the level of individual countries, is challenging. As an inference technique, synthetic control (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015) was first introduced to address this challenge. In typical synthetic control studies, the outcome observed for a single treated unit is compared against a control unit that is synthetically generated as a linear combination of other untreated units.

More recently, Doudchenko et al. (2021) introduced a trial design based on synthetic control, where treated and untreated units are not fixed but rather chosen by the designer before the experiment is run. Notably though, their trial design does not adapt to the outcomes observed sequentially once the experiment starts. In contrast, SYNTAX recruits and assigns patients—which is akin to choosing which units will be treated—in a fully adaptive way. Farias et al. (2022) study the adaptive use of synthetic controls but in a setting orthogonal to ours: They consider treatment assignments over time steps t for a single unit/patient whereas we consider the treatment assignment at a final time point T for multiple units/patients over multiple episodes η.

Thresholding Bandits   As an online algorithm, SYNTAX is closely related to the multi-armed bandit lit-

Table 1: **Comparison of related trial designs.** SYNTAX makes inferences via synthetic control and recruits adaptively.

Approach	Inference Strategy	Sampling Strategy	Related Work
Conventional studies	Simple statistics	Randomized	–
Thresholding bandits	$\hat{r}_i = \hat{y}_{iT}^{(1)} - \hat{y}_{iT}^{(0)}$	Adaptive	Locatelli et al. (2016)
Synthetic studies	Synthetic control	Randomized	Abadie et al. (2015)
Synthetic design	$\hat{r}_i = \hat{y}_{iT}^{(1)} - (\boldsymbol{\beta}_i^* = \mathrm{argmin}_{\boldsymbol{\beta}} \, V_i(\boldsymbol{\beta}))^{\mathsf{T}} \hat{\boldsymbol{y}}_{\cdot T}^{(0)}$	Pre-planned	Doudchenko et al. (2021)
SYNTAX		Adaptive	**(Ours)**

erature (e.g. Auer et al., 2002; Hüyük and Tekin, 2019, 2020, 2021), and our objective is most similar to that of the thresholding bandit problem (Locatelli et al., 2016; Zhong et al., 2017; Mukherjee et al., 2017; Tao et al., 2019) and the good arm identification problem (Kano et al., 2019; Katz-Samuels and Jamieson, 2020), which similarly aim to identify arms with rewards higher than a given threshold among a set of candidate arms. Subpopulations in our work can be thought of as arms in a thresholding bandit problem with one important caveat: While in thresholding bandits, each arm has only one reward distribution associated with it, in our setting, the treatment effect of each subpopulation depends on two response distributions that cannot be sampled from simultaneously. Complicating matters further, while there are underlying relationships between the responses of different subpopulations for the control group as described in (1), no such relationship exists for the treatment group.

Another related problem in the multi-armed bandit literature is the best arm identification (BAI) problem (Bubeck et al., 2009; Audibert et al., 2010; Gabillon et al., 2011, 2012; Soare et al., 2014; Xu et al., 2018; Alieva et al., 2021; Degenne et al., 2020). In our context, BAI would correspond to finding the subpopulation with the largest treatment effect. This objective is clinically less suitable as it is important to provide the treatment to all patients who would benefit from it, regardless of how little that benefit might be compared to other patients. Aside from "pure-exploration" problems such as thresholding bandits or BAI, there are numerous other work on multi-armed bandits that focus on maximizing the cumulative reward instead. In our context, this would correspond to maximizing the benefit received by all participant of a trial whereas clinical trials are usually exploratory in nature.

Adaptive Clinical Trials   Finally, it should be mentioned that adaptive trial designs that target objectives other than ours exist. Lewis and Bessen (1990); Whitehead (1997) consider when to terminate a trial, Hu and Rosenberger (1997); Berry (2006); Villar et al. (2015) maximizes the benefit for the recruited patients, Bhatt and Mehta (2006) consider when to stop further investigating one of two subpopulations, O'Quigley et al. (1990); Riviere et al. (2014); Wages et al. (2015); Yan et al. (2017); Shen et al. (2020); Lee et al. (2020, 2021)

determine safe but effective dosage for treatments, and Hüyük et al. (2023) consider the portfolio-level management of trials. Notably, Atan et al. (2019); Curth et al. (2023) also aim to identify subpopulations with positive treatment effects but not in the setting where pre-treatment responses are available for all participants. Adaptive enrichment designs (Stallard et al., 2014) perform interim analyses to select subpopulations. However there, the focus is to seamlessly adapt the inclusion criteria of a confirmatory trial without compromising its type I error rate. In contrast, we focus purely on selecting subpopulations in an exploratory trial that can then be targeted in a confirmatory trial.

6   EXPERIMENTS

Environments   To confirm our earlier intuition in Section 3.2 about when synthetic control might be the most effective, we consider two types of environments:

(i) *Diminishing Factor Effects:* In these environments, the contributions of factors $\boldsymbol{\mu}_t$ to the baseline responses gets weaker over time. Formally, we set $\boldsymbol{\mu}_t = (2 - 10^{t-T})\boldsymbol{\mu}_t'$, where $\boldsymbol{\mu}_t'$ is sampled uniformly at random from the unit ball in $\mathbb{R}^{D_z}$.

(ii) *Increasing Factor Effects:* In contrast to *Diminishing Factor Effects*, in these environments, the contributions of factors $\boldsymbol{\mu}_t$ to the baseline responses gets stronger over time. Formally, we set $\boldsymbol{\mu}_t = 10^{t-T}\boldsymbol{\mu}_t'$, where $\boldsymbol{\mu}_t'$ is sampled uniformly at random from the unit ball in $\mathbb{R}^{D_z}$.

Remember that we expect our gain from synthetic control to be larger for *Diminishing Factor Effects* since we infer factor loadings based on earlier responses—for which we would prefer factors to provide a stronger signal—and we aim to estimate later responses—for which we would gain the most from our earlier inference if factors now provide a weaker signal that is harder to pick up with naive estimates. We set $N = 25$, $T = 5$, $D_x = D_z = 2$. Notably, $D_x + D_z < N$ and $T > D_z$. We sample the remaining parameters of the environment so that $x_{ij} \sim \mathcal{N}_1(0)$, $z_{ij} \sim \mathcal{N}_1(0)$, $\delta_t \sim \mathcal{N}_1(0)$, $r_i \sim \mathcal{N}_1(0)$, and $\boldsymbol{w}_t$ is picked uniformly at random from the unit ball in $\mathbb{R}^{D_x}$. We repeat all our experiments ten times to obtain error bars. During each repetition, we sample a new environment and we average FPR and TPR of all our benchmark algorithms over 1000 runs.

Alihan Hüyük, Zhaozhi Qian, Mihaela van der Schaar

Table 2: **Performance comparison of SYNTAX and benchmarking algorithms.** Inline with our intuition from Section 3.2, SYNTAX performs the best for *Diminishing Factor Effects*. Again inline with our intuition, it does not provide any benefit for *Increasing Factor Effects* but still performs on par with *Thresholding Bandits* since synthetic estimators have variances at least as low as the naive estimate (as stated in Proposition 2).

| | Diminishing Factor Effects | | | | Increasing Factor Effects | | | |
| | $H = 200$ | | $H = 400$ | | $H = 200$ | | $H = 400$ | |
Algorithm	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR
Conventional study	19.5% (0.2%)	80.7% (0.3%)	14.9% (0.3%)	85.4% (0.3%)	19.5% (0.2%)	80.7% (0.3%)	14.9% (0.3%)	85.4% (0.3%)
Thresholding bandits	17.6% (0.4%)	82.6% (0.4%)	13.7% (0.4%)	86.4% (0.2%)	**17.6% (0.4%)**	**82.6% (0.4%)**	**13.7% (0.4%)**	**86.4% (0.2%)**
Synthetic study	16.7% (0.3%)	83.4% (0.3%)	12.5% (0.3%)	87.7% (0.2%)	19.5% (0.2%)	80.7% (0.3%)	14.9% (0.3%)	85.4% (0.3%)
Synthetic design	16.4% (0.4%)	83.8% (0.4%)	12.1% (0.4%)	88.2% (0.3%)	19.7% (0.4%)	80.5% (0.3%)	14.9% (0.3%)	85.4% (0.4%)
SYNTAX	**14.6% (0.4%)**	**85.6% (0.3%)**	**11.0% (0.3%)**	**89.1% (0.2%)**	17.5% (0.4%)	82.6% (0.3%)	13.7% (0.4%)	86.4% (0.3%)

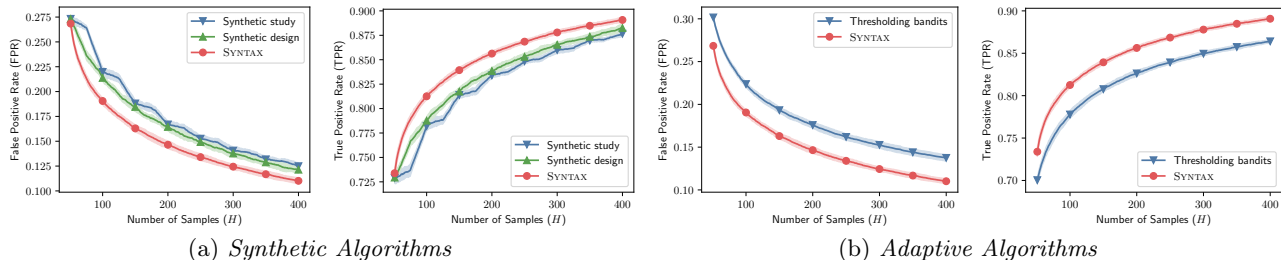(a) *Synthetic Algorithms*   (b) *Adaptive Algorithms*

Figure 2: **Comparison of (a) synthetic algorithms and (b) adaptive algorithms.** Switching to a pre-planned sampling strategy from an adaptive one or switching to a naive inference strategy from a synthetic one both cause FPR to increase and TPR to decrease at comparable scales.

Benchmarks We adapt the alternative trial designs summarized in Table 1 to our problem setting as benchmarks. Adapting *Thresholding Bandits*, when there are two response distributions associated with each subpopulation rather than a single reward distribution, involves forming naive estimates as in (7). *Synthetic Design* normally requires knowing the mean responses for the pre-treatment period before forming a fixed sampling plan. When adapting it, we allowed the sampling plan to be updated using the latest empirical mean responses $\hat{y}_{i \rightarrow T}^{(\alpha)}$ at each episode. However, unlike SYNTAX, *Synthetic Design* still does not take the final responses y_T into consideration when allocating samples. Details regarding each benchmark can be found in the appendix.

Main Results Table 2 compares SYNTAX and other benchmarks in terms of their performance. We consider two metrics: (i) *false positive rate* (FPR), which is the proportion of subpopulations incorrectly identified as having positive treatment effect among all subpopulations with negative treatment effect, and (ii) *true positive rate* (TPR), which is the proportion of subpopulations correctly identifies as having positive treatment effect among all subpopulations with positive treatment effect. We see that our intuition from Section 3.2 holds: SYNTAX performs the best for *Diminishing Factor Effects*. Although it does not provide any additional benefit for *Increasing Factor Effects*, it still performs on par with *Thresholding Bandits*, which relies on the naive estimate in (7). This is because synthetic estimators given by (11) have variances at least as low as the naive estimate (i.e. they are as informative) as stated in Proposition 2.

To understand how much impact each design aspect of SYNTAX has on its performance, we compare all three synthetic algorithms in Figure 2a and the two adaptive algorithms in Figure 2b (for *Diminishing Factor Effects*). These figures show how the FPR/TPR of different algorithms improve as the sample budget H gets larger. We see that switching to a naive inference strategy or a pre-planned sampling strategy both result in similar performance drops—indicating that synthetic and adaptive aspects of SYNTAX are equally important.

Significance of the Main Results Looking at Table 2, the performance difference between SYNTAX and a conventional RCT might not seem like much at a first glance. After all, it is merely (!) a ~5% difference in terms of FPR/TPR for *Diminishing Factor Effects* with $H = 200$. However, the implications of this seemingly small difference are significant in the larger context of clinical trials. We highlight two points: (i) how hard it actually is to gain this ~5% difference and (ii) what tangible benefits it would offer if the same gain were to be realized in practice:

(i) Notice that the performance of a traditional RCT in terms of FPR/TPR also improves only by ~5% when the sample size is increased to $H = 400$. This means that achieving the same performance gain as SYNTAX with a traditional RCT requires doubling the size of the RCT—meaning just by allocating control samples smartly, SYNTAX is able to match the performance of a trial that has double the size.

(ii) A typical RCT costs 12-35 million USD, which primarily scales with the number of patients in-
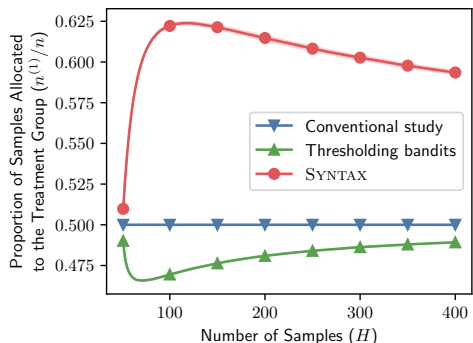
Figure 3: **Proportion of samples allocated to the treatment group over the control group.** By sharing information between the control samples of different subpopulations, SYNTAX is able allocate more of its samples to the treatment group compared with alternative designs.

volved (Moore et al., 2018), and takes 1-2 years, where as much as 86% of trials get delayed due to failures to reach recruitment targets (Huang et al., 2018). Having the required number of samples effectively halved by designs like SYNTAX would save upwards of 17.5 million USD per trial—note that more than 9000 trials are launched each year in the US alone (World Health Organization)—and help deliver much faster results.

Similarly, the 1-2% difference between SYNTAX and the next best-performing benchmark, *Synthetic Design*, for *Diminishing Factor Effects*, translates into a significant difference in terms of the sample size: SYNTAX with $H = 150$ achieves an FPR of 16.3% (0.4%) and TPR of 83.9% (0.2%)—the same performance as synthetic design with $H = 200$ but using 25% fewer samples.

Sample Allocation Characteristics We mentioned how sharing information between the control samples of different subpopulations would allow more samples to be allocated to the treatment group. Figure 3 shows the proportion of samples allocated to the treatment group by different algorithms (for *Diminishing Factor Effects*). Indeed, SYNTAX allocates more of its samples to the treatment group. Proposition 1 might provide insight into how this is achieved: The uncertainty of the treated response of each subpopulation i contributes directly to the variance of synthetic estimators through the term $1/n_i^{(1)}$. Whereas, it is possible to mitigate the uncertainty of baseline responses by distributing its contribution among multiple different subpopulations—more precisely, $\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_{(N^{(0)})^{-1}} \leq \|\mathbf{1}_i\|_{(N^{(0)})^{-1}} = 1/n_i^{(0)}$. In contrast to SYNTAX, a conventional study allocates all samples uniformly at random and *Thresholding Bandits* distributes samples adaptively among subpopulations but tends to not differentiate between different treatment groups.

Performance under Model Mismatch We also evaluate the performance of SYNTAX under model

Table 3: **Performance comparison of SYNTAX and benchmarking algorithms under model mismatch—** when outcomes are non-linear with respect to features.

| Algorithm | $H = 200$ | | $H = 400$ | |
	FPR	TPR	FPR	TPR
Conventional study	19.5% (0.2%)	80.7% (0.3%)	14.9% (0.3%)	85.4% (0.3%)
Thresholding bandits	17.6% (0.4%)	82.6% (0.4%)	13.7% (0.4%)	86.4% (0.2%)
Synthetic study	18.1% (0.4%)	81.8% (0.3%)	14.2% (0.4%)	85.9% (0.3%)
Synthetic design	18.3% (0.5%)	82.0% (0.4%)	14.2% (0.4%)	85.8% (0.3%)
SYNTAX	**16.2% (0.4%)**	**84.0% (0.4%)**	**12.9% (0.4%)**	**87.3% (0.3%)**

Table 4: **Performance comparison of SYNTAX and benchmarking algorithms under model mismatch—** when the number of latent factors $D_z = T$.

| Algorithm | $H = 200$ | | $H = 400$ | |
	FPR	TPR	FPR	TPR
Conventional study	19.5% (0.2%)	80.7% (0.3%)	14.9% (0.3%)	85.4% (0.3%)
Thresholding bandits	17.6% (0.4%)	**82.6% (0.4%)**	13.7% (0.4%)	**86.4% (0.2%)**
Synthetic study	15.8% (0.3%)	71.2% (0.7%)	12.3% (0.3%)	74.9% (0.9%)
Synthetic design	15.7% (0.3%)	71.5% (0.8%)	12.1% (0.3%)	75.0% (0.9%)
SYNTAX	**14.1% (0.3%)**	73.1% (1.0%)	**11.1% (0.2%)**	76.1% (1.0%)

mismatch—specifically when the linearity assumption in (1) is violated (although it is a natural assumption to make in our setting, see "On Modeling Assumptions" in Section 2). This can only happen in two ways: (i) Outcomes y_{it} might not be linear with respect to features $\boldsymbol{x}_i$. (ii) Outcomes y_{it} can always be expressed as a linear function of *some* latent factors $\boldsymbol{z}_i$, however the minimum number of factors needed to do so might be equal to the number of total time steps (i.e. $D_z = T$).

For the first scenario, we consider the same simulation setup as *Diminishing Factor Effects* but generate mean outcomes according to equation

$$\bar{y}_{it} = \delta_t + \boldsymbol{w}_t^\mathsf{T}(\boldsymbol{x}_i^2) + \boldsymbol{\mu}_t^\mathsf{T}\boldsymbol{z}_i \quad \text{for} \quad t \in [T] \qquad (14)$$

instead, where $\boldsymbol{x}_i^2$ denotes the element-wise square of $\boldsymbol{x}_i$. The results are given in Table 3. We see that all methods based on synthetic control, including SYNTAX, naturally lose performance after this change. However, SYNTAX still outperforms all benchmarks.

For the second scenario, we again consider the same simulation setup as *Diminishing Factor Effects* but this time set $D_z = T$. The results are given in Table 4. We see that TPR of SYNTAX is more sensitive to this change than its FPR. This observation is directly related to a balance between FPR-TPR. In all of our experiments, we fixed the threshold for declaring a positive effect at zero (for instance, see line 11 in Algorithm 1). Adjusting this threshold would have tuned the balance between FPR nad TPR (higher thresholds would achieve better FPR but worse TPR). Fixing the threshold at zero is a natural choice when all of our benchmarks are based on unbiased estimates. Under model mismatch however, synthetic inference becomes biased, the effect of which is equivalent to changing the classification threshold. Since the FPR performance stays better relative to the TPR performance, we can tell that the bias introduced happens to be negative (equivalent to a higher classification threshold).

(a) *Diminishing Factor Effects*, $H = 200$

(b) *Diminishing Factor Effects*, $H = 400$
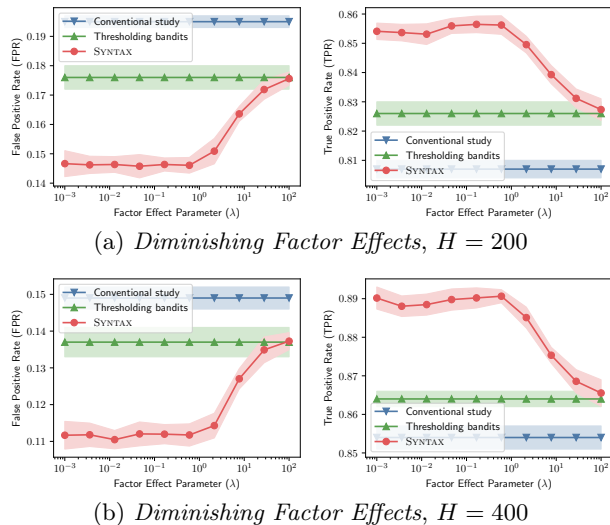
Figure 4: **Sensitivity of SYNTAX to parameter λ.**

Sensitivity to the Factor Effect Parameter Tuning hyper-parameters is a general challenge that affects all online algorithms, including SYNTAX, since in an online setting, no a-priori data would be available to perform cross-validation. We run additional experiments to evaluate the sensitivity of SYNTAX to its hyper-parameter λ for *Diminishing Factor Effects* by varying λ logarithmically from 0.001 to 100 (the ideal λ according to Proposition 1 varies between 0.1 and 1). The results are given in Figure 4. We see that SYNTAX outperforms all of the benchmarks without the hyper-parameter λ—namely *Conventional Study* and *Thresholding Bandits*—for almost all configurations. Notably, the performance of SYNTAX degrades only when λ is exceedingly large and converges to that of *Thresholding Bandits*. This is because λ punishes the errors in synthetic representations ($\|\boldsymbol{\beta} - \mathbf{1}_i\|$). As λ gets larger, SYNTAX reverts back to representing each subpopulation only as itself ($\boldsymbol{\beta} = \mathbf{1}_i$) and hence becomes equivalent to *Thresholding Bandits*.

7 CONCLUSION

We introduced SYNTAX, a clinical trial design that recruits patients adaptively to identify subpopulations with positive treatment effect. As we already argued for in the introduction, running trials that take multiple populations into consideration is absolutely essential. By reducing the length of such trials through designs like SYNTAX that seek to make more efficient use of sample would not only make multi-population trials more feasible but also help treatment get to the market sooner, benefiting the patients. Although we presented the problem mainly from a clinical trial perspective, SYNTAX can generally be applied to any thresholding bandit problem where time-series context (such as baseline responses) are available.

On Clinical Equipoise From the perspective of clinical equipoise, the use of adaptive trials, including SYNTAX, is understood to be situational (Palmer and Rosenberger, 1999; Fillion, 2019). We emphasize that running a SYNTAX-based clinical trial should only be considered if there is a genuine lack of information to determine the target population for a promising new treatment, especially when responses of different subpopulations to the treatment are expected to be highly heterogenous. More specifically, the principle of clinical equipoise requires two conditions to be met: (i) there should be uncertainty among clinicians regarding the effectiveness of a treatment, and (ii) the results of a clinical trial should be convincing enough to resolve this uncertainty (Freedman, 1987; Miller and Brody, 2007). We examine SYNTAX on these tow axises:

(i) SYNTAX is an exploratory trial design rather than a confirmatory trial design. This means that, even at the end of a SYNTAX-based trial, there would still be genuine uncertainty regarding the effectiveness of the treatment for all subpopulations (hence the need for a subsequent confirmatory trial). At no point during the trial, SYNTAX assigns a patient to a treatment that can readily be confirmed to be ineffective for them; and the first condition is maintained through the trial. Moreover, the ultimate goal of SYNTAX is to identify subpopulations as the potential targets of a subsequent Phase III trial. As such, it would be in contradiction with the first condition to use SYNTAX when there is already a clear candidate to target in such a trial. As we already mentioned above, the use of SYNTAX should be reserved to the cases where there is disagreement in terms of which subpopulations would be the most suitable targets of a Phase III trial.

(ii) Being an exploratory trial design, SYNTAX cannot resolve all uncertainties regarding the effectiveness of a treatment. But, it can determine the appropriate target population for a subsequent confirmatory trial, and thereby help satisfy the second condition. As pointed out by Chiu et al. (2018), a confirmatory trial is more likely to be inconclusive regarding subpopulation effects when the potential heterogeneity in patient responses is not taken into account.

On Real-data Validation A real-data validation of SYNTAX would essentially require running a new clinical trial. This would be infeasible (and potentially unethical) for a method development paper as ours. Hence, the standard approach to evaluating adaptive clinical trial designs is only to use simulated data (e.g. Atan et al., 2019; Curth et al., 2023; Hüyük et al., 2023), even in the biostatistics literature (e.g. Friede et al., 2012; Magnusson and Turnbull, 2013; Stallard et al., 2014; Henning and Westfall, 2015; Rosenblum et al., 2016).

References

Abadie, A. and Gardeazabal, J., "The economic costs of conflict: A case study of the basque country," *Amer. Econ. Rev.*, vol. 93, no. 1, pp. 113–132, 2003.

Abadie, A., Diamond, A., and Hainmueller, J., "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *J. Amer. Statist. Assoc.*, vol. 105, no. 490, pp. 493–505, 2010.

Abadie, A., Diamond, A., and Hainmueler, J., "Comparative politics and the synthetic control method," *Amer. J. Political Sci.*, vol. 59, no. 2, pp. 495–510, 2015.

Alieva, A., Cutkosky, A., and Das, A., "Robust pure exploration in linear bandits with limited budget," in *Proc. Int. Conf. Mach. Learn.*, 2021.

Atan, O., Zame, W. R., and van der Schaar, M., "Sequential patient recruitment and allocation for adaptive clinical trials," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2019.

Audibert, J.-Y., Bubeck, S., and Munos, R., "Best arm identification in multi-armed bandits," in *Proc. Annu. Conf. Learn. Theory*, 2010.

Auer, P., Cesa-Bianchi, N., and Fischer, P., "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, pp. 235–256, 2002.

Berry, D. A., "Bayesian clinical trials," *Nature Rev. Drug Discovery*, vol. 5, no. 1, pp. 27–36, 2006.

Bhatt, D. L. and Mehta, C., "Adaptive designs for clinical trials," *New England J Med.*, vol. 375, pp. 65–74, 2006.

Bica, I., Alaa, A. M., and van der Schaar, M., "Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders," in *Proc. Int. Conf. Mach. Learn.*, 2020.

Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., and Peters, T. J., "Subgroup analyses in randomized trials: Risks of subgroup-specific analyses; power and sample size for the interaction test," *J. Clin. Epidemiology*, vol. 57, pp. 229–236, 2004.

Bubeck, S., Munos, R., and Stoltz, G., "Pure exploration in multi-armed bandit problems," in *Proc. Int. Conf. Algorithmic Learn. Theory*, 2009.

Chiu, Y.-D., Koenig, F., Posch, M., and Jaki, T., "Design and estimation in clinical trials with subpopulation selection," *Statist. Med.*, vol. 37, pp. 4335–4352, 2018.

Curth, A., Hüyük, A., and van der Schaar, M., "Adaptively identifying patient populations with treatment benefit in clinical trials," in *Proc. Int. Conf. Mach. Learn.*, 2023.

Degenne, R., Menard, P., Shang, X., and Valko, M., "Gamification of pure exploration for linear bandits," in *Proc. Int. Conf. Mach. Learn.*, 2020.

Doudchenko, N., Khosravi, K., Pouget-Abadie, J., Lahaie, S., Lubin, M., Mirrokni, V., Spiess, J., and Imbens, G., "Synthetic design: an optimization approach to experimental design with synthetic controls," in *Proc. Conf. Neural Inf. Process. Syst.*, 2021.

Farias, V., Moallemi, C., Peng, T., and Zheng, A., "Synthetically controlled bandits," *arXiv preprint arXiv:2202.07079*, 2022.

Feinstein, A. R. and Horwitz, R. I., "Problems in the "evidence" of "evidence-based medicine"," *Amer. J. Med.*, vol. 103, no. 6, pp. 529–535, 1997.

Fillion, N., "Clinical equipoise and adaptive clinical trials," *Topoi*, vol. 38, pp. 457–467, 2019.

Freedman, B., "Equipoise and the ethics of clinical research," *New England J. Med.*, vol. 317, pp. 141–145, 1987.

Friede, T., Parsons, N., and Stallard, N., "A conditional error function approach for subgroup selection in adaptive clinical trials," *Statist. Med.*, vol. 31, no. 30, pp. 4309–4320, 2012.

Gabillon, V., Ghavamzadeh, M., Lazaric, A., and Bubeck, S., "Multi-armed bandit best arm identification," in *Proc. Neural Inf. Process. Syst.*, 2011.

Gabillon, V., Ghavamzadeh, M., and Lazaric, A., "Best arm identification: A unified approach for fixed budget and fixed confidence," in *Proc. Neural Inf. Process. Syst.*, 2012.

Henning, K. S. and Westfall, P. H., "Closed testing in pharmaceutical research: Historical and recent developments," *Statist Biopharmaceutical Res.*, vol. 7, no. 2, pp. 126–147, 2015.

Hu, F. and Rosenberger, W. F., *The theory of response-adaptive randomization in clinical trials.* John Wiley & Sons, 1997.

Huang, G. D., Bull, J., McKee, K. J., Mahon, E., Harper, B., Roberts, J. N., Team, C. R. P. *et al.*, "Clinical trials recruitment planning: a proposed framework from the clinical trials transformation initiative," *Contemporary Clin. Trials*, vol. 66, pp. 74–79, 2018.

Hüyük, A. and Tekin, C., "Analysis of Thompson sampling for combinatorial multi-armed bandit with probabilistically triggered arms," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2019.

Hüyük, A. and Tekin, C., "Thompson sampling for combinatorial network optimization in unknown en-

vironments," *IEEE/ACM Trans. Netw.*, vol. 28, no. 6, pp. 2836–2849, 2020.

Hüyük, A. and Tekin, C., "Multi-objective multi-armed bandit with lexicographically ordered and satisficing objectives," *Mach. Learn.*, vol. 110, no. 6, pp. 1233–1266, 2021.

Hüyük, A., Qian, Z., and van der Schaar, M., "When to make and break commitments?" in *Proc. Int. Conf. Learn. Representations*, 2023.

Kano, H., Honda, J., Sakamaki, K., Matsuura, K., Nakamura, A., and Sugiyama, M., "Good arm identification via bandit feedback," *Mach. Learn.*, vol. 108, no. 5, pp. 721–745, 2019.

Katz-Samuels, J. and Jamieson, K., "The true sample complexity of identifying good arms," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020.

Lazar, A. A., Bonetti, M., Cole, B. F., Yip, W., and Gelber, R. D., "Identifying treatment effect heterogeneity in clinical trials using subpopulations of events: STEPP," *Clin. Trials*, vol. 13, no. 2, pp. 169–179, 2016.

Lee, H.-S., Shen, C., Jordon, J., and van der Schaar, M., "Contextual constrained learning for dose-fnding clinical trials," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020.

Lee, H.-S., Shen, C., Zame, W. R., Lee, J. W., and van der Schaar, M., "SDF-Bayes: Cautious optimism in safe dose-finding clinical trials with drug combinations and heterogenous patient groups," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021.

Lewis, R. J. and Bessen, H. A., "Sequential clinical trials in emergency medicine," *Ann. Emergency Med.*, vol. 19, no. 9, pp. 1047–1053, 1990.

Lipkovich, I., Dmitrienko, A., and D'Agostino Sr., R. B., "Tutorial in biostatistics: Data-driven subgroup identification and analysis in clinical trials," *Statist. Med.*, vol. 36, pp. 136–196, 2017.

Locatelli, A., Gutzeit, M., and Carpentier, A., "An optimal algorithm for the thresholding bandit problem," in *Proc. Int. Conf. Mach. Learn.*, 2016.

Magnusson, B. P. and Turnbull, B. W., "Group sequential enrichment design incorporating subgroup selection," *Statist. Med.*, vol. 32, no. 16, pp. 2695–2714, 2013.

Miller, F. G. and Brody, H., "Clinical equipoise and the incoherence of research ethics," *J. Med. Philos.*, vol. 32, pp. 151–165, 2007.

Moineddin, R., Butt, D. A., Tomlinson, G., and Beyene, J., "Identifying subpopulations for subgroup analysis in a longitudinal clinical trial," *Contemporary Clin. Trials*, vol. 29, pp. 817–822, 2008.

Moore, T. J., Zhang, H., Anderson, G., and Alexander, G. C., "Estimated costs of pivotal trials for novel therapeutic agents approved by the us food and drug administration, 2015-2016," *JAMA internal medicine*, vol. 178, no. 11, pp. 1451–1457, 2018.

Mukherjee, S., Naveen, K. P., Sudarsanam, N., and Ravindran, B., "Thresholding bandits with augmented UCB," *arXiv preprint arXiv:1704.02281*, 2017.

Ondra, T., Dmitrienko, A., Friede, T., Graf, A., Miller, F., Stallard, N., and Posch, M., "Methods for identification and confirmation of targeted subgroups in clinical trials: A systematic review," *J. Biopharmaceutical Statist.*, vol. 26, no. 1, pp. 99–119, 2016.

O'Quigley, J., Pepe, M., and Fisher, L., "Continual reassessment method: A practical design for phase 1 clinical trials in cancer," *Biometrics*, vol. 46, pp. 33–48, 1990.

Palmer, C. R. and Rosenberger, W. F., "Ethics and practice: Alternative designs for phase III randomized clinical trials," *Controlled Clin. Trials*, vol. 20, pp. 172–186, 1999.

Riviere, M.-K., Yuan, Y., Dubois, F., and Zohar, S., "A Bayesian dose-finding design for drug combination clinical trials based on the logistic model," *Pharmaceutical Statist.*, vol. 13, no. 4, pp. 247–257, 2014.

Rosenblum, M., Luber, B., Thompson, R. E., and Hanley, D., "Group sequential designs with prospectively planned rules for subpopulation enrichment," *Statist. Med.*, vol. 35, no. 21, pp. 3776–3791, 2016.

Sackett, D. L. and Rosenberg, W. M. C., "On the need for evidence-based medicine," *J. Public Health*, vol. 17, no. 3, pp. 330–334, 1995.

Shen, C., Villar, S., Wang, Z., and van der Schaar, M., "Learning for dose allocation in adaptive clinical trials with safety constraints," in *Proc. Int. Conf. Mach. Learn.*, 2020.

Shi, C., Sridhar, D., Misra, V., and Blei, D. M., "On the assumptions of synthetic control methods," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2022.

Soare, M., Lazaric, A., and Munos, R., "Best-arm identification in linear bandits," in *Proc. Neural Inf. Process. Syst.*, 2014.

Stallard, N., Hamborg, T., Parsons, N., and Friede, T., "Adaptive designs for confirmatory clinical trials with subgroup selection," *J. Biopharmaceutical Statist.*, vol. 24, no. 1, pp. 168–187, 2014.

Tao, C., Blanco, S., Peng, J., and Zhou, Y., "Thresholding bandit with optimal aggregate regret," in *Proc. Conf. Neural Inf. Process. Syst.*, 2019.

Villar, S. S., Bowden, J., and Wason, J., "Multi-armed bandit models for the optial design of clinical trials:

Benefits and challenges," *Statist. Sci.: Rev. J. Inst. Math. Statist.*, vol. 30, no. 2, p. 199, 2015.

Wages, N. A., Read, P. W., and Petroni, G. R., "A phase I/II adaptive design for heterogenous groups with application to a stereotactic body radiation therapy trial," *Pharmaceutical Statist.*, vol. 14, no. 4, pp. 302–310, 2015.

Wand, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., and Drazen, J. M., "Statistics in medicine—reporting of subgroup analyses in clinical trials," *New England J. Med.*, vol. 357, no. 21, pp. 2189–2194, 2007.

Whitehead, J., *The design and analysis of sequential clinical trials.* John Wiley & Sons, 1997.

World Health Organization. Number of clinical trials by year, country, WHO region and income group (1999-2022). Accessed on Oct. 5, 2023. [Online]. Available: https://www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/number-of-clinical-trials-by-year-country-who-region-and-income-group.

Xu, L., Honda, J., and Sugiyama, M., "A fully adaptive algorithm for pure exploration in linear bandits," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2018.

Yan, F., Mandrekar, S. J., and Yuan, Y., "Keyboard: A novel Bayesian toxicity probability interval design for phase I clinical trials," *Clin. Cancer Res.*, vol. 23, no. 15, pp. 3994–4003, 2017.

Zhong, J., Huang, Y., and Liu, J., "Asynchronous parallel empirical variance guided algorithms for the thresholding bandit problem," *arXiv preprint arXiv:1704.04567*, 2017.

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 2 for the mathematical setting and assumptions, see (1) and (2) in Section 2 for our model, and see Algorithm 1 in Section 4 for our algorithm.

   (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. [Yes] See Section 2 for the full set of assumptions.

   (b) Complete proofs of all theoretical results. Yes, proofs are given in Appendix A.

   (c) Clear explanations of any assumptions. [Yes] See "On Modeling Assumptions" in Section 2.

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] See "Environments" and "Benchmarks" in Section 6 and the benchmark algorithms in Appendix B.

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

   (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 "Environments" in Section 6.

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

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. [Not Applicable]

   (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. [Not Applicable]

   (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]

A PROOFS OF PROPOSITIONS

A.1 Proof of Proposition 1

Denote with $e_{iT}^{(0)} = \hat{y}_{iT}^{(0)} - \bar{y}_{iT}$, $e_{iT}^{(1)} = \hat{y}_{iT}^{(1)} - \bar{y}_{iT} - r_i$, and $e_{it} = \hat{y}_{it} - \bar{y}_{it}$ the observation noises. Similar to Abadie et al. (2010), we start by showing that unobservable factor loadings $\boldsymbol{z}_i$ can be inferred through observable responses $\hat{y}$ if responses are observed for a long enough pre-treatment period. More specifically, a good match in terms of responses $\hat{\boldsymbol{y}}_{i\neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}$ leads to a good match in terms of factor loadings $\boldsymbol{z}_i \approx Z\boldsymbol{\beta}$. We have

$$0 = \hat{\boldsymbol{y}}_{i\neg T} - \hat{Y}_{\neg T}\boldsymbol{\beta} \tag{15}$$
$$= \boldsymbol{e}_{i\neg T} - E_{\neg T}\boldsymbol{\beta} + \bar{\boldsymbol{y}}_{i\neg T} - \bar{Y}_{\neg T}\boldsymbol{\beta} \tag{16}$$
$$= \boldsymbol{e}_{i\neg T} - E_{\neg T}\boldsymbol{\beta} + \boldsymbol{\delta}_{\neg T} + W_{\neg T}^{\mathsf{T}}\boldsymbol{x}_i + M_{\neg T}^{\mathsf{T}}\boldsymbol{z}_i - (\boldsymbol{\delta}_{\neg T}\mathbf{1}^{\mathsf{T}} + W_{\neg T}^{\mathsf{T}}X + M_{\neg T}^{\mathsf{T}}Z)\boldsymbol{\beta} \tag{17}$$
$$= \boldsymbol{e}_{i\neg T} - E_{\neg T}\boldsymbol{\beta} + M_{\neg T}^{\mathsf{T}}(\boldsymbol{z}_i - Z\boldsymbol{\beta}) \tag{18}$$

where (16) is due to (2), (17) is due to (1), and (18) holds if $\boldsymbol{x}_i = X\boldsymbol{\beta}$ and $\mathbf{1}^{\mathsf{T}}\boldsymbol{\beta} = 1$. Hence,

$$\boldsymbol{z}_i - Z\boldsymbol{\beta} = (M_{\neg T}^+)^{\mathsf{T}}E_{\neg T}(\boldsymbol{\beta} - \mathbf{1}_i) \tag{19}$$

where $M_{\neg T}^+ = M_{\neg T}^{\mathsf{T}}(M_{\neg T}M_{\neg T}^{\mathsf{T}})^{-1}$ is the right inverse of $M_{\neg T}$. Notably, this right inverse exists only if $M_{\neg T}$ has full rank and $T > D_z$ (i.e. when the pre-treatment period is long enough). It is already possible to spot that the term in (19) is the source of representation error in Proposition 1. This provides the insight that representation error is essentially caused by the imperfect estimation of factor loadings $\boldsymbol{z}_i$ through noisy observations of responses $\hat{y}$.

With this result, we are now ready to characterize the estimation error $r_i - \hat{r}_i(\boldsymbol{\beta})$ in terms of observation noises $e_{it}, e_{iT}^{(0)}, e_{iT}^{(1)}$. We have

$$r_i - \hat{r}_i(\boldsymbol{\beta})$$
$$= r_i - \hat{y}_{iT}^{(1)} + \boldsymbol{\beta}^{\mathsf{T}}\hat{\boldsymbol{y}}_{\cdot T}^{(0)} \tag{20}$$
$$= r_i - (r_i + \bar{y}_{iT} + e_{iT}^{(1)}) + \boldsymbol{\beta}^{\mathsf{T}}(\bar{\boldsymbol{y}}_{\cdot T} + \boldsymbol{e}_{\cdot T}^{(0)}) \tag{21}$$
$$= -e_{iT}^{(1)} + \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{e}_{\cdot T}^{(0)} - \delta_T - \boldsymbol{w}_T^{\mathsf{T}}\boldsymbol{x}_i - \boldsymbol{\mu}_T^{\mathsf{T}}\boldsymbol{z}_i + \boldsymbol{\beta}^{\mathsf{T}}(\delta_T\mathbf{1} + X^{\mathsf{T}}\boldsymbol{w}_T + Z^{\mathsf{T}}\boldsymbol{\mu}_T) \tag{22}$$
$$= -e_{iT}^{(1)} + \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{e}_{\cdot T}^{(0)} + \boldsymbol{\mu}_T^{\mathsf{T}}(\boldsymbol{z}_i - Z\boldsymbol{\beta}) \tag{23}$$
$$= -e_{iT}^{(1)} + \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{e}_{\cdot T}^{(0)} + \boldsymbol{\mu}_T^{\mathsf{T}}(M_{\neg T}^+)^{\mathsf{T}}E_{\neg T}(\boldsymbol{\beta} - \mathbf{1}_i) \tag{24}$$

where (21) is due to (3), (22) is due to (1), and (23) holds if $\boldsymbol{x}_i = X\boldsymbol{\beta}$ and $\mathbf{1}^{\mathsf{T}}\boldsymbol{\beta} = 1$. Since $e_{iT}^{(0)}$, $e_{iT}^{(1)}$, and $\{e_{it}\}_{t<T}$ all have zero mean and are independent from each other, $\mathbb{E}[r_i - \hat{r}_i(\boldsymbol{\beta})] = 0$ and

$$\mathbb{V}[r_i - \hat{r}_i(\boldsymbol{\beta})] = \mathbb{E}[(r_i - \hat{r}_i(\boldsymbol{\beta}))^2] \tag{25}$$
$$= \mathbb{E}[(e_{iT}^{(1)})^2] + \mathbb{E}[\|\boldsymbol{\beta}\|_{\boldsymbol{e}_{\cdot T}^{(0)}(\boldsymbol{e}_{\cdot T}^{(0)})^{\mathsf{T}}}^2] + \mathbb{E}[\|\boldsymbol{\mu}_T^{\mathsf{T}}(M_{\neg T}^+)^{\mathsf{T}}E_{\neg T}(\boldsymbol{\beta} - \mathbf{1}_i)\|^2] \tag{26}$$
$$\leq \mathbb{E}[(e_{iT}^{(1)})^2] + \mathbb{E}[\|\boldsymbol{\beta}\|_{\boldsymbol{e}_{\cdot T}^{(0)}(\boldsymbol{e}_{\cdot T}^{(0)})^{\mathsf{T}}}^2] + \|M_{\neg T}^+\boldsymbol{\mu}_T\|^2\mathbb{E}[\|\boldsymbol{\beta} - \mathbf{1}_i\|_{E_{\neg T}^{\mathsf{T}}E_{\neg T}}^2] \tag{27}$$
$$= \sigma^2/n_i^{(1)} + \sigma^2\|\boldsymbol{\beta}\|_{(N^{(0)})^{-1}}^2 + \lambda\sigma^2\|\boldsymbol{\beta} - \mathbf{1}_i\|_{N^{-1}}^2 \tag{28}$$

A.2 Proof of Proposition 2

Proposition 2 is a corollary of Proposition 1. We simply have

$$\mathbb{V}[r_i - \hat{r}_i(\boldsymbol{\beta}_i^*)] \leq V_i(\boldsymbol{\beta}_i^*) \tag{29}$$
$$\leq V_i(\mathbf{1}_i) \tag{30}$$
$$= \mathbb{V}[r_i - \hat{r}_i^{\text{naive}}] \tag{31}$$

where (29) is due to Proposition 1 and (30) is by definition of $\boldsymbol{\beta}_i^*$ in (11).

B BENCHMARK ALGORITHMS

Algorithm 2 summarizes all benchmarks (except SYNTAX, which is given in Algorithm 1 instead). We set the factor effect parameter λ to its optimal value given in Proposition 1 (except for the sensitivity experiments in Figure 4). This means the results we present are for perfectly tuned versions of each algorithm. All experiments are run on a personal computer with an Intel i9 processor. Finally, the code to reproduce our experimental results can be found at `https://github.com/alihanhyk/syntax` and `https://github.com/vanderschaarlab/syntax`.

Algorithm 2 Benchmarks

Parameters: Episode horizon H, factor effect parameter λ
Output: Subpopulations $\hat{\mathcal{I}}^*$ with positive treatment effect

---

1: $n_i^{(\alpha)} \leftarrow 0, \ \hat{y}_{iT}^{(\alpha)} \leftarrow 0, \ \hat{\boldsymbol{y}}_{i \neg T} \leftarrow \mathbf{0}, \ \forall i \in [K], \alpha \in \{0, 1\}$
2: **for** $\eta \in \{1, 2, \ldots, H\}$ **do**
3:     **if** *Conventional Study* **or** *Synthetic Study* **then**
4:         Sample i, α uniformly at random from $[K] \times \{0, 1\}$
5:     **else if** *Thresholding Bandits* **then**
6:         $i \leftarrow \operatorname{argmin}_{i \in [K]} S_i^{\text{naive}} = |\hat{r}_i^{\text{naive}}| / (1/n_i^{(0)} + 1/n_i^{(1)})^{1/2}$
7:         $\alpha \leftarrow \operatorname{argmin}_{\alpha \in \{0, 1\}} n_i^{(\alpha)}$
8:     **else if** *Synthetic Design* **then**
9:         $i, \alpha \leftarrow \operatorname{argmin}_{i \in [K], \alpha \in \{0, 1\}} \max_{i^* \in [K]} \min_{\boldsymbol{\beta}: \ \boldsymbol{x}_{i^*} = X\boldsymbol{\beta}, \ \hat{\boldsymbol{y}}_{i^* \neg T} = \hat{Y}_{\neg T}\boldsymbol{\beta}, \ \mathbf{1}^\mathsf{T}\boldsymbol{\beta} = 1}$
$$V_{i^*}(\boldsymbol{\beta}; N^{(0)} + (1 - \alpha)\mathbf{1}_i \mathbf{1}_i^\mathsf{T}, N^{(1)} + \alpha \mathbf{1}_i \mathbf{1}_i^\mathsf{T})$$
10:     **end if**
11:     Recruit from population i and treatment group α
12:     Observe pre-treatment outcomes $\boldsymbol{y}_{\neg T} = [y_1 \cdots y_{T-1}]^\mathsf{T}$ and the final outcome y_T
13:     $n_i^{(\alpha)} \leftarrow n_i^{(\alpha)} + 1$
14:     $\hat{y}_{iT}^{(\alpha)} \leftarrow \hat{y}_{iT}^{(\alpha)} + (y_T - \hat{y}_{iT}^{(\alpha)})/n_i^{(\alpha)}$
15:     $\hat{\boldsymbol{y}}_{i \neg T} \leftarrow \hat{\boldsymbol{y}}_{i \neg T} + (\boldsymbol{y}_{\neg T} - \hat{\boldsymbol{y}}_{i \neg T})/n_i$
16: **end for**
17: **if** *Conventional Study* **or** *Thresholding Bandits* **then**
18:     $\hat{\mathcal{I}}^* \leftarrow \{i \in [K] : \hat{r}_i^{\text{naive}} > 0\}$
19: **else if** *Synthetic Study* **or** *Synthetic Design* **then**
20:     $\hat{\mathcal{I}}^* \leftarrow \{i \in [K] : \hat{r}_i(\boldsymbol{\beta}_i^*) > 0\}$
21: **end if**