Safe Exploration in Dose Finding Clinical Trials with Heterogeneous Participants

Isabel Chien [1] **Wessel Bruinsma** [2] **Javier Gonzalez Hernandez** [3] **Richard E. Turner** [1]

Abstract

In drug development, early phase dose-finding clinical trials are carried out to identify an optimal dose to administer to patients in larger confirmatory clinical trials. Standard trial procedures do not optimize for participant benefit and do not consider participant heterogeneity, despite consequences to participants' health and downstream impacts to under-represented population subgroups. Many novel drugs also do not obey parametric modelling assumptions made in common dose-finding procedures. We present Safe Allocation for Exploration of Treatments (SAFE-T), a procedure for adaptive dose-finding that adheres to safety constraints, improves utility for heterogeneous participants, and works well with small sample sizes. SAFE-T flexibly learns non-parametric multi-output Gaussian process models for dose toxicity and efficacy, using Bayesian optimization, and provides accurate final dose recommendations. We provide theoretical guarantees for the satisfaction of safety constraints. Using a comprehensive set of realistic synthetic scenarios, we demonstrate empirically that SAFE-T generally outperforms comparable methods and maintains performance across variations in sample size and subgroup distribution. Finally, we extend SAFE-T to a new adaptive setting, demonstrating its potential to improve traditional clinical trial procedures.

1. Introduction

New drugs and treatments are typically first investigated in early phase dose-finding studies, which aim to assess safety and recommended dosing for future study. In early phase trials, participant safety is prioritized through strict safeguards against adverse outcomes (toxicity); however, participant benefit (efficacy) is not typically considered. Researchers are increasingly adopting adaptive trial methods, where trial parameters may change based on ongoing outcomes, potentially improving both the efficiency of a trial and the toxicity/efficacy outcomes experienced by trial participants (Villar et al., 2015; Riviere et al., 2018).

Dose-finding studies commonly use the rule-based 3+3 method, which allocates increasing doses to cohorts of size 3, terminating if a threshold of toxicity is exceeded (Kurzrock et al., 2021), or the adaptive continual reassessment method (CRM), which allocates doses based on a continually updated parametric model for toxicity (Wheeler et al., 2019). Both methods suffer from drawbacks: neither method optimizes over efficacy for participants, the 3+3 method has been widely criticized for its inefficiency due to unnecessary early trial termination (Love et al., 2017; Kurzrock et al., 2021), and CRM requires a parametric model and strict prior assumptions, leaving room for misspecification errors (Cheung & Chappell, 2002). Both methods are increasingly unsuitable for modeling dose-response; newly developed drugs may not adhere to requisite prior assumptions, particularly that dose-toxicity and dose-efficacy are monotonically increasing. Innovative therapies, such as immune-oncology agents, may have plateauing or parabolic efficacy profiles (Wages et al., 2018; Zhang et al., 2006).

Furthermore, current dose-finding practices have given rise to ethical concerns and have potentially contributed to societal health inequalities. Clinical trials have often been criticized for *therapeutic misconception*, where participants incorrectly believe that the primary goal of a trial is to improve their own health outcomes, while in practice, scientific validity is prioritized over patient health and trials do not optimize for participant benefit (Chien et al., 2022). Standard methods also assume that the trial population is homogeneous, not accounting for possible variations in toxicity and efficacy due to patient heterogeneity. Due to these assumptions and persisting inequalities in subject selection for clinical trials (Steinberg et al., 2021), reported optimal drug doses are often not generalizable to those who are under-represented in early-phase trials, such as women and under-studied racial/ethnic groups (Özdemir et al., 2022).

*Equal contribution [1]University of Cambridge, Cambridge, UK [2]Microsoft Research AI for Science [3]Microsoft Research. Correspondence to: Isabel Chien <ic390@cam.ac.uk>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Research has shown that women experience a far greater risk of adverse drug effects across all drug classes as compared to men (Zucker & Prendergast, 2020; Unger et al., 2022) and that race/ethnicity can also impact drug response (Ramamoorthy et al., 2021; Dickmann & Schutzman, 2018). This necessitates trial methods that can optimize for participant benefit and address heterogeneity, while still ensuring safety and learning accurate dose recommendations.

Our contributions. While we cannot address the many complex societal and methodological factors that surround inequality in clinical trials, we make a small contribution with our adaptive dose-finding procedure, Safe Allocation for Exploration of Treatments (SAFE-T). SAFE-T models toxicity and efficacy using multi-output Gaussian process classification, which can capture variations between heterogeneous participant subgroups and can flexibly handle different dose-response profiles. Using these models, SAFE-T allocates doses to trial participants while balancing safety and exploration, resulting in improved safety, participant utility, and final dose recommendation accuracy. SAFE-T can be extended to a novel trial setting where dose recommendations are provided over a continuous range. We provide theoretical results for SAFE-T, guaranteeing safety with high probability. In extensive experiments, we compare the performance of SAFE-T to other methods with respect to safety, participant utility, and final dose recommendation accuracy, across varying subgroup ratios and sample sizes.

Simultaneously, we seek to address a concerning health inequity: certain population subgroups may experience greater risks of adverse drug effects due to flaws in the current methodology of dose-finding trials. Standard trial methods assume that the trial population is homogeneous. This, in conjunction with existing inequalities in trial subject selection, may lead to disparities in patient utility during the trial (from sub-optimal dose allocation) and after the trial (from non-representative dose recommendations). We tackle inequity through proper consideration of heterogeneous patient populations, rather than standard fairness constraints. We also examine subgroup disparities in careful experimental evaluation, finding that SAFE-T improves performance (in safety, accuracy, utility, Sec. 5.1) for patients and reduces disparities between subgroups (Sec. 5.2).

Related work. Recent works in machine learning for dose-finding trials have mostly concentrated on multi-armed bandit methods (Aziz et al., 2020; Riviere et al., 2018; Garivier et al., 2017), with some including explicit safety constraints (Lee et al., 2020; Shen et al., 2020; Wang et al., 2021). These works consider dose-finding scenarios with monotonically increasing toxicity and efficacy curves, with Aziz et al. (2020); Shen et al. (2020) also providing separate algorithms for dose selection for plateauing efficacy curves. These works typically assume a logistic parametric formO'Quigley et al. (1990) of the dose-toxicity model (Aziz et al., 2020; Lee et al., 2020; Shen et al., 2020; Riviere et al., 2018), or else do not learn models of dose-response relationships, providing point-estimates of treatment arm performance (Wang et al., 2021; Garivier et al., 2017). Lee et al. (2020), investigates a setting most similar to ours, proposing a dose-finding method that addresses heterogeneous populations. However, they use an inflexible parametric model and rely on an unsafe burn-in period.

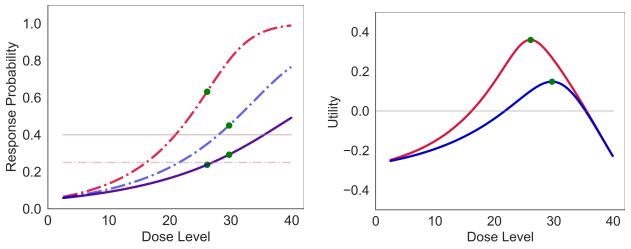
While we focus on a dose-finding setting, our problem is general: given a true underlying safety function and reward function, we desire an algorithm that maximizes reward without violating safety constraints, while accurately learning decision boundaries associated with a small, heterogeneous data set. Safe exploration has been addressed in the context of bandits (Kazerouni et al., 2016) and active learning with Gaussian processes (GPs) (Sui et al., 2015; 2018; Schreiter et al., 2015; Turchetta et al., 2019; Bottero et al., 2022; Berkenkamp et al., 2016; Turchetta et al., 2016). Of these, our setting is most similar to STAGEOPT (Sui et al., 2018), which optimizes a reward function over a safe range determined by a separate safety function. In contrast to our binary outcome setting, these works mostly consider continuous outcomes under a Gaussian additive noise model, allowing use of existing theoretical bounds for GP confidence intervals (Srinivas et al., 2009; Chowdhury & Gopalan, 2017). The exception is Schreiter et al. (2015), which classifies a safe region based on binary safety indicators, but does not consider a separate reward function.

2. Dose-Finding Problem Statement

Problem set-up. Fig. 1a illustrates the problem setting, which reflects standard practice (Wheeler et al., 2019). N trial participants are considered sequentially; each participant is allocated one of K discrete doses, indexed by $k \in \mathbb{K} = \{1, ..., K\}$, where $d_k \in \mathbb{D}$ represents dosage values. At each time n, a new participant of known subgroup $s_n \in \mathbb{S} = \{1, ..., S\}$[1], with arrival probabilities π_s, is allocated to dose $m_n \in \mathbb{K}$ based on a *selection rule*. For each pair (d, s), $q_s(d)$ defines true toxicity probabilities and $p_s(d)$ defines true efficacy probabilities. We observe the binary toxicity outcome $Y_n \sim \text{Ber}(q_s(d_{m_n}))$, $Y_n = +1$ indicating an adverse reaction; and the binary efficacy outcome $X_n \sim \text{Ber}(p_s(d_{m_n}))$, $X_n = +1$ indicating effective treatment.

At the outset of a trial, based on prior clinical knowledge, investigators specify adverse events that represent toxic outcomes, responses that qualify as effective outcomes, a maximum toxicity threshold, τ_T, and a minimum efficacy threshold, τ_E (Wheeler et al., 2019). A dose d for a patient in subgroup s is in an acceptable safe and effective range when

[1]Throughout, we drop the subscript n from s_n for simplicity.



(a) (b)

Figure 1: *(a)* Each subgroup has a corresponding probability of toxicity (—), probability of efficacy (—·), and optimal dose (●). Here, both subgroups 0 (—) and 1 (—) experience the same dose-toxicity responses. Safe doses fall within the range of the maximum toxicity threshold (—) and minimum efficacy threshold (—·) are also depicted. *(b)* The *Thall utility* measure trades off toxicity and efficacy.

$q_s(d) \leq \tau_T$ and $p_s(d) \geq \tau_E$. Trials using a model-based methodology estimate the dose-toxicity function $\hat{q}_s(d)$ and sometimes the dose-efficacy function $\hat{p}_s(d)$. At trial conclusion, a *final recommendation rule* recommends doses for further examination in downstream trials. There is no underlying truth for optimality; standard dose-finding methods recommend the highest safe dose, under the assumption that efficacy is monotonically increasing. However, in novel treatments, dose-efficacy may plateau or be parabolic, where the highest safe dose is not optimal (Wages et al., 2018).

Importance of safety. Because dose-finding trials investigate treatments where adverse effects are largely unknown, the safety of participants is of primary importance. Any dose allocation procedure must incorporate practical safety constraints alongside theoretical safety guarantees (which may rest on untestable assumptions). For example, methods that require an unsafe burn-in period (Lee et al., 2020; Shen et al., 2020), where a bandit algorithm initializes by selecting each arm in succession, cannot be used in practice.

Patient heterogeneity. The trial population may be heterogeneous, where different patient covariate values may correspond to different dose-response profiles, which may not be known *a priori* (Thomas et al., 2018). Patient heterogeneity is typically not investigated in current practice, leading to results that may reflect the (unequal) distribution of the recruited population (Wages et al., 2015); consequently, women and ethnic minorities may experience disparate impacts for the same drugs (Unger et al., 2022; Ramamoorthy et al., 2021). We are concerned with mitigating disparities caused by standard methods, particularly across *protected subgroups*, often defined by demographic characteristics (Barocas et al., 2023; Mehrabi et al., 2019). Our work thus focuses on *pre-defined* heterogeneous subgroups, such as biological sex, which has been shown to impact toxicity and efficacy outcomes (Unger et al., 2022; Ramamoorthy et al., 2021). This setting has received some attention in the clinical trials literature (Chapple & Thall, 2018; Ivanova & Wang, 2006; Salter et al., 2015; Morita et al., 2017) and in ML (Aziz et al., 2020; Lee et al., 2020), under similar motivations for health equity.

However, relying on pre-defined subgroups is limiting, especially when considering intersectionality; we encounter the issue of *infinite regress*, where computation is challenged by the presence of 1) too many subgroups due to many considered attributes and 2) small size of resultant subgroups (Kong, 2022). Kong (2022) recommends instead to "proactively challenge oppression and make society fairer," particularly through collaboration with domain experts. While imperfect, our work is a first step towards addressing the pervasive issue of inequity in dose-finding trials caused by unrecognized patient heterogeneity; some major sources of heterogeneity are already known (such as biological sex) and we can begin by addressing these. Future work may focus, for example, on individualized dose-finding that does not require pre-defined subdivision of patients.

Additional considerations. We also take into account two key considerations: sample size and misspecification. As most early phase trials include fewer than 100 participants (Can, 2022; Huang et al., 2015), dose-response models should perform well with small sample sizes. It is also not possible to know true underlying dose-response functions; models should be robust to misspecification.

3. The SAFE-T Algorithm

3.1. Overview

SAFE-T consists of three components: (1) a safe exploration stage, which encourages allocation to unexplored doses (2) an efficacy optimization stage, and (3) a final dose recommendation rule. The first two stages define selection rules for dose allocation that balance safety, learning, and utility. SAFE-T is initialized with a set of N_0 participants assigned the lowest dose, d_0, which is presumed to be safe. SAFE-T runs the first safe exploration stage for participants $n \in (N_0, N_0 + N_1]$ and the second efficacy optimization stage over the remaining participants. SAFE-T terminates early if the estimated safe set of doses is empty for N_{stop} participants during the trial, as a safeguard in case d_0 is not safe. At trial conclusion, SAFE-T uses *Thall utility* (Thall & Cook, 2004) to select final recommended doses by subgroup for further clinical study. We first discuss important components of SAFE-T and then detail the algorithm in Section 3.2, with pseudocode in Algorithm 1.

GP classification. We model the dose-response with GP classification (Rasmussen & Williams, 2003), where a latent discriminative function $h\colon \mathbb{X} \to \mathbb{R}$ is modelled with a standard GP prior with mean $\mu(\cdot)$ and covariance function $k(\cdot, \cdot)$. The latent $h(x)$ values are mapped through a link function,

in our case the Gaussian CDF Φ, to determine class probabilities over the binary outcome Y such that $Y \mid h(x) \sim \text{Ber}(\Phi(h(x)))$. Dose-toxicity, $q_{1:S}(d)$, and dose-efficacy, $p_{1:S}(d)$, can be viewed as multi-valued functions, where s corresponds to a subgroup. SAFE-T models both functions with two independent multi-output Gaussian processes (MOGPs), using the linear model of co-regionalization (LMC) (Journel & Huijbregts, 1976). We let $f_{1:S}(d)$ and $g_{1:S}(d)$ be two MOGPs such that $\Phi(f_s(d)) = p_s(d)$ and $\Phi(g_s(d)) = q_s(d)$. We then define dose-toxicity estimates as $\hat{q}_s(d) = \mathbb{E}[\Phi(\hat{g}_s(d))]$ and dose-efficacy estimates as $\hat{p}_s(d) = \mathbb{E}[\Phi(\hat{f}_s(d))]$ where are $\hat{g}_s(d)$ and $\hat{f}_s(d)$ are random variables distributed according to the posterior distributions of of $g_s(d)$ and $f_s(d)$ given all observations so far. The hyperparameters of the MOGPs are chosen using heuristics, as described in App. D.1. GPs allow for flexibility in modelling dose-response profiles, unlike commonly used parametric models (Wheeler et al., 2019; Lee et al., 2020), which may be severely misspecified with respect to the ground truth.

Early stopping. We assign the first N_0 participants to a known safe dose d_0; in standard practice, the lowest dose (presumed to be safe) is administered to an initial trial cohort (Wheeler et al., 2019). A guaranteed initial safe set is also a common assumption in the safe exploration literature (Sui et al., 2015; 2018; Schreiter et al., 2015; Bottero et al., 2022), to prevent unnecessary early stopping due to an empty safe set. In Theorem 4.1, we provide a lower bound for N_0 to guarantee success of SAFE-T. However, as with any theoretical bound, we rely on assumptions that may be violated in practice: that we have knowledge of a safe d_0. We thus also provide a practical early stopping criterion: after N_{stop} safety violations, the trial is terminated and recommendations are made with the existing information. N_{stop} can be set according to clinical guidelines or participant preference; analogous parameters are used in the standard 3+3 and CRM trial methods (Wheeler et al., 2019).

Safety constraint. For the remaining participants, SAFE-T guarantees, with probability at least $1 - \delta_T$, that all doses $d_{N_0+1:n}$ are safe, such that the safety constraint $\Phi(g(d)) \leq \tau_T$ is satisfied. We achieve this guarantee by only considering doses d that satisfy the safety constraint:

$$\mu_{\hat{g}_s}(d) + \nu_T \sigma_{\hat{g}_s}(d) \leq \rho_T \tag{1}$$

where $\mu_{\hat{g}_s}(d)$ and $\sigma^2_{\hat{g}_s}(d)$ are the mean and variance of $\hat{g}_s(d)$. Set $\rho_T = \Phi^{-1}(\tau_T)$ and $\nu_T = \Phi^{-1}(1 - \frac{\delta_T}{N - N_0}) > 0$, where $\delta_T \in (0, \frac{1}{2})$. Further details in Theorem 4.2.

Utility measure. We adopt a notion of utility for dose-finding trial participants as proposed by Thall & Cook (2004) and further used in (Koopmeiners & Modiano, 2014), depicted in Figure 1b. *Thall utility* is a weighted L_p norm that codifies a trade-off between toxicity and efficacy:

$$U(p_E, p_T) = 1 - [(p_T/\tau_T)^p + ((1 - p_E)/(1 - \tau_E))^p]^{1/p} \tag{2}$$

Algorithm 1 SAFE-T Algorithm

Input: subgroups $\mathbb{S}$, dose indices $\mathbb{K}$, num. patients N, num. initial safe set N_0, safe expansion timesteps N_1, toxicity threshold τ_T, safety parameter ν_T, GP prior for toxicity func. $g_{1:S}(d) \sim$ MOGP, GP prior for efficacy func. $f_{1:S}(d) \sim$ MOGP, utility func. $U(p_E, p_T)$
Initialize: $n = 1$, $\mathbb{B}_{s,0} = \varnothing$ for all $s \in \mathbb{S}$, $\hat{g}_{1:S}(d) \sim p(g_{1:S}(d))$, and $\hat{f}_{1:S}(d) \sim p(f_{1:S}(d))$
while $n \leq N$ **do**
  if $n \leq N_0$ **then** *# Select lowest dose for N_0 iterations.*
    $m_n = 0$
  else *# Initialization done. Safe dynamic dose selection.*
    $\mathbb{A}_{s,n} = \mathbb{K} \cap (\mathbb{B}_{s,n-1} \cup \{\max(\mathbb{B}_{s,n-1}) + 1\})$
    $\mathbb{M}_{s,n} = \{k \in \mathbb{A}_{s,n} | \mu_{\hat{g}_s(d_k)} + \nu_T \sigma_{\hat{g}_s(d_k)} \leq \Phi^{-1}(\tau_T)\}$
    if $N_0 < n \leq N_1$ **then** *# Safe exploration stage*
      if $\max(\mathbb{M}_{s,n}) \notin \mathbb{B}_{s,n-1}$ **then**
        $m_n = \max(\mathbb{M}_{s,n})$ *# Increase to highest safe dose.*
      else
        $m_n = \arg\max_{k \in \mathbb{M}_{s,n}} c_n(d_k)$ *# Reduce uncertainty.*
      end if
    end if
    if $N_1 < n \leq N$ **then** *# Safe optimization stage*
      $m_n = \arg\max_{k \in \mathbb{M}_{s,n}} \text{ACQUIS.}_{\hat{f}_s}(d_k)$
    end if
  end if
  Administer dose d_{m_n}
  Observe toxicity Y_n, efficacy X_n
  Update posteriors $\hat{g}_{1:S}(d)$, $\hat{f}_{1:S}(d)$
  $\mathbb{B}_{s,n} = \mathbb{B}_{s,n-1} \cup \{m_n\}$
  $n \leftarrow n + 1$
end while
for all $s \in \mathbb{S}$ **do**
  $\mathbb{M}_{s,\text{final}} = \{k \in \mathbb{B}_{s,N} | \mu_{\hat{g}_s(d_k)} + \nu_T \sigma_{\hat{g}_s(d_k)} \leq \Phi^{-1}(\tau_T)\}$
  $m_{s,\text{final}} = \arg\max_{k \in \mathbb{M}_{s,\text{final}}} U(\hat{p}_s(d_k), \hat{q}_s(d_k))$
end for
Output: for every subgroup $s \in \mathbb{S}$, recommend dose $d_{m_{s,\text{final}}}$

where p_T is the probability of toxicity, p_E the probability of an effective treatment, and p is elicited from trial practitioners. Details on the selection of p appear in App. A.

3.2. Algorithm description

During trial: Safe exploration stage. For each subgroup s, at each timestep n, SAFE-T maintains a set of previously sampled doses, $\mathbb{P}_{s,n}$, which is empty at the beginning of a trial. Throughout, dose sets are composed of dose indices. The available set of doses, $\mathbb{A}_{s,n} = \mathbb{K} \cap (\mathbb{P}_{s,n} \cup \{\max(\mathbb{P}_{s,n}) + 1\})$, includes all doses that have been previously sampled and the next highest dose (unless the highest dose has been sampled already or the next highest dose falls outside $\mathbb{K}$). The safe set of doses are those that satisfy Eqn. 1: $\mathbb{M}_{s,n} = \{k \in \mathbb{A}_{s,n} | \mu_{\hat{g}_s}(d_k) + \nu_T \sigma_{\hat{g}_s}(d_k) \leq \rho_T\}$. SAFE-T allocates the highest safe dose available, if this dose has not been previously sampled by subgroup s. If all safe doses have been sampled, SAFE-T selects the dose with the largest confidence interval on $\hat{g}_s(d)$. The priority is exploration to new doses and reduction of uncertainty is secondary. Confidence interval widths are $c_n(d_k) = 2\nu_T \sigma_{\hat{g}_s(d_k)}$. We allow a

slight relaxation of this constraint to prevent premature stopping: if $\mathbb{M}_{s,n}$ is empty, participants are assigned the lowest dose d_0. This constraint violation may occur up to N_{stop} times. Thus, unless $\mathbb{M}_{s,n} = \varnothing$, for dose m_n, we use the selection rule: $m_n = \max(\mathbb{M}_{h_n,n})$ if $\max(\mathbb{M}_{h_n,n}) \notin \mathbb{B}_{h_n,n}$, and $m_n = \arg\max_{k \in \mathbb{M}_{h_n,n}} c_n(d_k)$ otherwise.

During trial: Safe optimization stage. In the second stage, SAFE-T prioritizes optimization of participant efficacy, while still ensuring safety. The safe set $\mathbb{M}_{s,n}$ is constructed as previously defined. From $\mathbb{M}_{s,n}$, SAFE-T selects a dose based on optimization of an acquisition function $\text{ACQUIS.}_{\hat{f}_s}(d)$ on the dose-efficacy model. For example, we may adopt the upper confidence bound (UCB) paradigm (Auer, 2003) for optimistic exploration, where $\text{ACQUIS.}_{\hat{f}_s}(d) = \mu_{\hat{f}_s}(d) + \nu_E \sigma_{\hat{f}_s}(d)$ for some hyperparameter ν_E. We may also use the expected improvement (EI) criterion, where $\text{ACQUIS.}_{\hat{f}_s}(d) = \text{EI}_{\hat{f}_s}(d)$ (as defined in Jones et al. (1998)). We also propose optimization directly over participant utility, where the dose with the maximum estimated Thall utility given the UCB on toxicity (conservative with respect to safety) and the UCB on efficacy (optimistic with respect to efficacy) is selected: $\text{ACQUIS.}_{\hat{f}_s}(d) = U(\mu_{\hat{f}_s}(d) + \nu_E \sigma_{\hat{f}_s}(d), \mu_{\hat{g}_s}(d) + \nu_T \sigma_{\hat{g}_s}(d))$.

After trial: Final dose recommendation. At trial conclusion, SAFE-T recommends a dose for each subgroup using the final posteriors on the GP toxicity and efficacy functions. The safe dose set $\mathbb{M}_{s,\text{final}}$ includes all doses allocated during the trial that satisfy the safety constraint, Eqn. 1. The final recommended dose for each subgroup has the maximum utility out of the safe dose set: $m_{s,\text{final}} = \arg\max_{k \in \mathbb{M}_{s,\text{final}}} U(\hat{p}_s(d_k), \hat{g}_s(d_k))$.

4. Theoretical Results

Theorem 4.1 defines a lower bound on the number of initial safe doses N_0 to ensure success of SAFE-T, and Theorem 4.2 shows that SAFE-T guarantees safety with high probability. Full proofs are shown in Appendix B.2.

Theorem 4.1. *Let φ be the standard normal probability density function, σ_g^2 the maximum variance of the GP prior on any of the latent toxicity functions g_s, $\nu_T > 0$, and $\rho_T = \Phi^{-1}(\tau_T)$. Assume that the GP prior is stationary. Let $\hat{g}_s(d)$ be the posterior conditioned on N_0 initial safe dosages and possibly more dosages. Approximate this posterior $\hat{g}_s(d)$ with the Laplace approximation. Under this approximation, if there exists a dose d that satisfies safety constraint in Equation 1, then N_0 is lower bounded as follows:*

$$N_0 \geq \left(2\varphi\left(\frac{1}{2}(1 + \rho_T - \sqrt{(\rho_T - 1)^2 + 4\nu_T\sigma_g})\right)\right)^{-1} \times \min\left(\frac{\nu_T}{\sqrt{3}\sigma_g} - \frac{\rho_T}{\sigma_g^2}, \frac{1}{\sigma_g^2}\sqrt{\frac{\rho_T^2}{4} + \frac{\sigma_g \nu_T}{\sqrt{3}}} - \frac{\rho_T}{2\sigma_g^2}\right).$$

We provide a lower bound for N_0, which means that SAFE-T requires initialization with *at least* N_0 safe samples. However, providing N_0 safe samples does not guarantee that the safe set will be non-empty. The bound is useful as a guide for implementation; we determine the minimum number of trial participants to assign to the lowest dose and conduct simulation studies to tune this hyperparameter in practice.

Theorem 4.2. *Suppose that the true latent toxicity functions $g_{1:S}$ are indeed drawn from the prior model assumed in* SAFE-T. *In* SAFE-T, *assume that the posteriors $\hat{g}_s$ can be accurately approximated with Gaussian distributions. If* SAFE-T *does not early stop, then all allocated doses are approximately safe with probability at least $1 - \delta_T$:*

$$\Pr(\Phi(g_{s_1}(d_{m_1})) \leq \tau_T, \ldots, \Phi(g_{s_N}(d_{m_N})) \leq \tau_T) \gtrapprox 1 - \delta_T.$$

The approximation quality in the inequality depends on the quality of the Gaussian approximation of the posteriors $\hat{g}_s$.

Under Theorem 4.2, if SAFE-T is initialized with a non-empty safe set of doses, all participants are assigned a safe dose with high probability. This is safeguarded by an early stopping constraint, which ends the trial if the safety constraint is violated more than N_{stop} times.

Discussion. Assumptions made in both theorems are standard in the literature (Schreiter et al., 2015; Sui et al., 2015; 2018; Srinivas et al., 2009; Chowdhury & Gopalan, 2017). See App. B.1 for a detailed discussion. GP-UCB (Srinivas et al., 2009; Chowdhury & Gopalan, 2017) and EI (Wang & de Freitas, 2014), which can be used as Bayesian acquisition functions in SAFE-T, have proven finite-time regret bounds and confidence interval guarantees that have been used for regret bounds in related works (Sui et al., 2015; 2018). For both methods, theoretical results are predicated on assumptions that are not applicable in our problem setting: (1) the true reward obeys an additive noise model, and (2) the posteriors are Gaussian processes (roughly). Future work may explore a regret bound with respect to GP classification given Bernoulli-distributed outcomes.

5. Experimental Results

We evaluate SAFE-T on a thorough set of synthetic scenarios (Sec. 5.1), differing subgroup distributions (Sec. 5.2), and sample size variations (Sec. 5.3). In Sec. 5.4, we apply SAFE-T to a new adaptive setting, demonstrating possible future extensions. The same GP hyperparameters are used in SAFE-T toxicity and efficacy MOGPs across all scenarios and experiments, demonstrating that SAFE-T is performant with minimal prior assumptions and in the presence of misspecification (scenarios are not generated from GPs). In the experiments, we use the EI acquisition function, but provide a comparison with UCB and utility optimization, finding negligible performance differences (Appendix E). Implementation details in Appendix D.1.

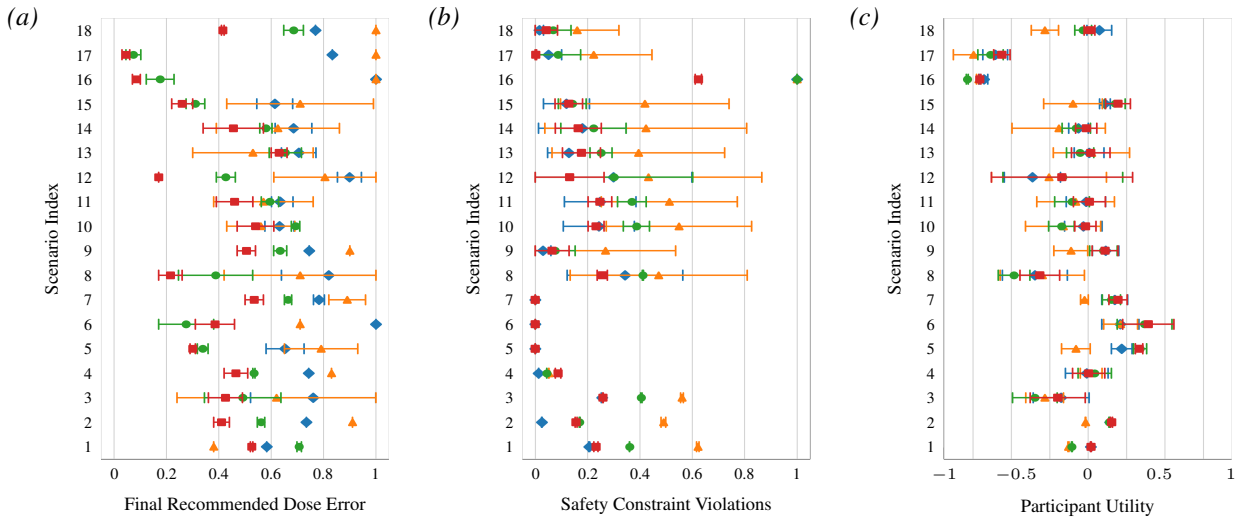


Figure 2: SAFE-T (■) consistently outperforms 3+3 (♦), CRM (▲), and C3T (●) with *(a)* lowest or comparable dose error in $16/18$ scenarios, *(b)* fewest or comparable safety constraint violations in $16/18$ scenarios, and *(c)* highest or comparable participant utility in $17/18$ scenarios. Shapes represent the metric mean across the 2 subgroups, while line caps represent the metric for each subgroup. Wide intervals indicate a large disparity in subgroup performance, indicating inequity.

SAFE-T involves several distinct components; in order to establish the necessity of each component, we conduct experiments with variations of SAFE-T where isolated components are not included. Our results (Appendix E) show that all aspects of SAFE-T (the multi-output GP models, expansion stage, safety constraint, final dose recommendation using Thall utility) are necessary for the highest performing result.

Synthetic scenarios. It is not possible to know ground truth dose-response mechanisms. These can only be hypothesized from clinical trials findings; however, publicly accessible clinical trials data are reported as summary statistics, with no individual-level data on treatments, outcomes, health characteristics, and demographics. Trials that use model-based methods report parametric models trained with patient outcomes (Wheeler et al., 2019) and therefore rely on the *untestable* assumption that the ground truth is well-specified by the selected parametric model. Additionally, the assumptions that justify commonly used parametric models (such as monotonicity) are likely violated by new classes of drugs (Wages et al., 2018; Zhang et al., 2006).

Therefore, we evaluate SAFE-T on a comprehensive set of 18 synthetic scenarios that capture differing dose-response profiles, realistic variations across subgroups, different toxicity and efficacy thresholds, and possible edge cases (such as a scenario where all doses are unsafe). These are constructed based on careful examination of the relevant clinical trials literature on possible dose-response models (Wages et al., 2018), hypotheses regarding newly innovated treatments (Kurzrock et al., 2021; Chiuzan et al., 2017), and reported differences in the impact of drugs on population subgroups (Özdemir et al., 2022; Zucker & Prendergast, 2020; Harkin et al., 2001; Unger et al., 2022; Ramamoorthy et al., 2021). See Appendix C for detailed descriptions.

The use of synthetic experiments is standard in the dose-finding literature, both in the clinical trials field (Ivanova & Wang, 2006; Salter et al., 2015; Morita et al., 2017; Chapple & Thall, 2018) and ML field (Aziz et al., 2020; Lee et al., 2020; Shen et al., 2020; Wang et al., 2021). We find that the extensive and informed nature of our synthetic experiments surpasses the completeness of those in the related works, which all investigate fewer than 5 synthetic examples.

Baselines and evaluation metrics. We evaluate performance on three metrics: final recommended dose error; safety constraint violations, which occur when an allocated dose has a true toxicity probability greater than the toxicity threshold; and participant utility, assessed post-hoc with the Thall utility of the allocated dose. All metrics are reported averaged over 100 experimental trials, with safety violations and utility also averaged over the number of participants. We compare SAFE-T to standard dose-finding procedures: the rule-based 3+3 method, and the Bayesian continual reassessment method (CRM). These methods do not explicitly consider participant heterogeneity (nor do they claim to), but are standard practice. We also compare SAFE-T to the C3T algorithm proposed by Lee et al. (2020), which is the only algorithm we were able to identify in the ML literature that aims to allocate doses optimally with respect to heterogeneous participants. We were not able to identify algorithms in the GP safe exploration space that matched our problem setting (Bernoulli-distributed safety and reward outcomes). CRM and C3T both use a logistic model (O'Quigley et al., 1990) for dose-toxicity. See App. D.2 for further details.

5.1. Performance across synthetic scenarios

For all scenarios, we define a population of $N = 51$, with $N_0 = 3$ and $N_1 = 15$, with 2 subgroups arriving at probabilities $\boldsymbol{\pi} = [0.5, 0.5]$. As seen in Figure 2, **SAFE-T**(■) consistently outperforms baselines for all metrics, exhibiting either lower or comparable final recommended dose error across scenarios. The performance of **CRM** (▲) is quite poor, particularly in final dose recommendation accuracy, with large disparities between subgroups. SAFE-T results in comparatively fewer safety constraint violations, with the rule-based **3+3** (♦) method (App. D.2) outperforming in some scenarios. 3+3 is strict with respect to safety, but can consequently be inefficient due to unnecessary early stopping; we see that it performs generally poorly in accuracy and utility. On balance, SAFE-T is preferable, as safety is still maintained very well, while accuracy and utility are vastly improved. We also highlight the result in scenario 16, which defines an edge case where no doses are considered safe. The impact of our early stopping safeguard is seen here; because SAFE-T terminates, the remaining participants do not receive the unsafe doses, resulting in fewer safety violations. We see the smallest differences in utility across algorithms; however, SAFE-T maintains slightly higher overall utility.

5.2. Impact of subgroup ratios

We desire a dose allocation procedure that performs effectively with heterogeneous participants, particularly when subgroups are unevenly distributed. In Fig. 3, we report performance on safety constraint violations given variations in the subgroup distribution for scenario 11. In scenario 11, dose-toxicity is monotonically increasing, where subgroup 0 experiences higher probability of toxicity at each dose in comparison to subgroup 1. Dose-efficacy, which is the same for both subgroups, is monotonically increasing and plateauing. This scenario exhibits one key difference between subgroups: toxicity. As such, we expect wider variations in the safety constraint violation metric across subgroups for more skewed subgroup distributions. In this setting, we have $N = 201$ participants and 2 subgroups, we vary the arrival probability of subgroups such that the arrival probability of subgroup 0, π_0, ranges from $[0.15, 0.9]$ with step size 0.5, where $\pi_0 + \pi_1 = 1$.

SAFE-T demonstrates a consistently low rate of safety violations across varying subgroup ratios. It effectively maintains a uniformly low level of safety while ensuring comparable safety outcomes across subgroups. This demonstrates the method's relative fairness and its efficacy in addressing participant heterogeneity. In contrast, both 3+3 and CRM exhibit large differences in safety between subgroups, where both result in fewer safety violations for subgroup 1, which experiences a lower rate of toxicity across dose, but a comparatively high proportion of safety violations for subgroup 0, which experiences a higher rate of toxicity. This demonstrates the problematic nature of these standard methods; subgroup variations would not be known before a trial and these methods result in detrimental impacts on subgroups that may experience higher toxicity for the same dosage levels, such as women (Unger et al., 2022). C3T maintains safety with slightly better performance; however, C3T allocates notably more unsafe doses to subgroup 0 when there are few participants from subgroup 0, which would exacerbate health disparities. In App. E, we report the final dose error and utility results over subgroup ratios, as well as results for the same experiment with smaller N (Fig. 11).

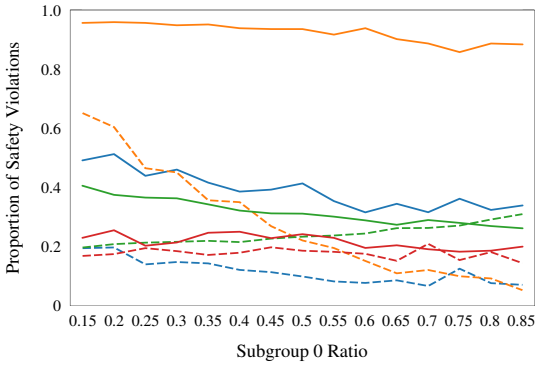


Figure 3: The x-axis shows the proportion of the population from subgroup 0. Solid lines (—) are subgroup 0, dashed (- -) subgroup 1. **SAFE-T** (red) maintains a consistently low rate of safety violations across subgroups. **3+3** (blue) and **CRM** (orange) exhibit large disparities. **C3T** (green) shows larger disparity in safety when subgroup 0 ratio is small.

5.3. Impact of sample size

We examine performance on scenario 9 over sample sizes from $N = 51$ to $N = 537$, with 100 trials of each method every 9 steps. We select scn. 9 as it is a simple scenario with one main complication: plateauing efficacy. Subgroup 0 experiences slightly higher toxicity across doses, efficacy is the same across subgroups. We expect methods to improve on final dose error with increasing sample sizes, as model accuracy improves. However, 3+3 and C3T do not improve with increasing sample sizes, highlighting their ineffectiveness in the presence of misspecification. 3+3 selects the largest safe dose, but when efficacy plateaus, the highest safe dose is not the most effective. C3T selects the estimated most effective safe dose, without considering a utility tradeoff; it does not recognize that a very slightly higher efficacy (as efficacy plateaus) is not worth the higher toxicity of a larger dose. In contrast, SAFE-T shows consistently decreasing dose error with increasing sample size; if practitioners wish to improve final dose accuracy, they may recruit a larger trial cohort. We do not include the time-intensive CRM in this experiment due to lack of sufficient

resources, but note its overall poor performance. In App. E, we report safety and utility results over sample sizes.

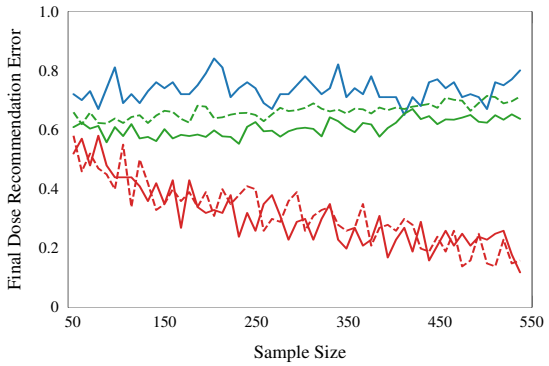


Figure 4: **SAFE-T** (red) improves final dose error over increasing sample sizes for subgroups 0 (—) and 1 (- -) in contrast with **3+3** (blue) and **C3T** (green).

5.4. Learning efficient dose allocations

In a standard dose-finding trial, investigators typically specify a discrete set of 2 to 12 doses (Wheeler et al., 2019). This set remains unchanged, even in adaptive settings (with the exception of unsafe doses being removed). However, the true optimal dose may exist outside of this discrete set. SAFE-T can be extended to recommend doses on a *continuous* scale. Here, SAFE-T expands the available range of doses based on a specified dosage increase interval. A safe dose range is identified based on the same safety constraint. During the efficacy optimization phase, the optimal point is determined across the safe dose *range*, rather than a discrete *set*, allowing for more complete exploration. As in Sec. 5.1, $N = 51$, $N_0 = 3$, and $N_1 = 15$, with 2 subgroups arriving at $\boldsymbol{\pi} = [0.5, 0.5]$. The dosage increase interval is 3.0.

In Table 1, we report the performance of SAFE-T on 4 continuous scenarios (App. C). We report the error between the recommend dose and the true optimal (based on Thall utility) given the true underlying toxicity and efficacy curves), finding that SAFE-T performs extraordinarily well, with objectively small differences in final dose error in Scns. C.1, C.3, and C.4, where the algorithm tends to slightly underestimate the optimal dose (prioritizing safety). We notice slightly worse error in Scn. C.2, where the efficacy curves are strictly monotonically increasing and the optimal dose is relatively high in the provided dose range. This error results from the algorithm behaving conservatively (note that the signed error value is negative), a desirable trait.

Table 1: Optimized continuous dose allocations results

Scn.	Dose Range	Subgroup 0	Subgroup 1
		Safety Violations	
C.1	$2.5 - 15$	0.24 ± 0.00	0.00 ± 0.19
C.2	$2.5 - 40$	0.00 ± 0.00	0.00 ± 0.00
C.3	$2.5 - 20$	0.06 ± 0.02	0.02 ± 0.04
C.4	$0.05 - 15$	0.21 ± 0.18	0.26 ± 0.22
		Final Dose Error	
C.1	$2.5 - 15$	-0.19 ± 2.60	1.34 ± 4.23
C.2	$2.5 - 40$	-4.48 ± 8.29	-10.75 ± 12.11
C.3	$2.5 - 20$	-0.08 ± 2.77	-0.02 ± 2.66
C.4	$0.05 - 15$	-0.22 ± 3.96	-0.31 ± 2.29

6. Conclusion

We propose SAFE-T, an adaptive dose-finding procedure that optimizes for participant benefit and addresses response heterogeneity, while also ensuring participant safety and dose recommendation accuracy. SAFE-T performs effectively within the realistic context of a dose-finding trial, working well with small sample sizes and under misspecification. We present theoretical guarantees for the probability of safety over trial participants. Through extensive synthetic experiments, we demonstrate that SAFE-T outperforms baseline methods in final dose recommendation accuracy, participant safety, and participant utility. SAFE-T also maintains consistent performance across variations in subgroup distribution and improves steadily with increasing sample size. We demonstrate SAFE-T's high performance in a novel continuous setting where the algorithm is not restricted to a discrete set of doses, a promising venue for future work.

There are several limitations to our work. SAFE-T is tailored to the standard setup of an early-phase dose-finding trials, where binary outcomes for toxicity and efficacy are considered. However, crucial information is lost in this setting; future work may focus on dose-finding methods that consider more complex scenarios, such as continuous outcomes, delayed outcomes, multiple outcomes, and missing data (i.e. due to patient dropout). As clinical trials practitioners are hesitant to adopt unproven methods, new research should emphasize safety and robustness of findings. SAFE-T is applicable when participant heterogeneity can be hypothesized pre-trial, making it useful for exploring demographic inequities. However, heterogeneous with respect to latent characteristics that are not easily identifiable at the outset of a trial; future work may incorporate learned subgroups into the dose-finding procedure (Thomas et al., 2018). From a practical standpoint, GPs are more complex than standard statistical methods, which would be a barrier to uptake in practice. This setting has also raised interesting questions with respect to GP classification. We have not been able to identify literature that provides theoretical guarantees on confidence bounds and regret for GP classification, likely due to the analytically intractable GP posterior, making this an interesting venue for fundamental research. While we make a first step towards mitigating inequity in dose-finding trials, further research that actively challenges structural sources of inequality is required.

Acknowledgements

We thank Niki Kilbertus, Jiri Hron, and Adrian Weller for helpful discussion during the early stages of this work. Isabel Chien and Richard E. Turner are supported by an EPSRC Prosperity Partnership EP/T005386/1 between Microsoft Research and the University of Cambridge.

Impact Statement

In this work, we seek to address health inequities that arise from flaws in commonly used methodology in early phase dose-finding clinical trials. While we discuss several concerns, design choices, and limitations of our work in the body of the paper, there are additional ethical issues we can elaborate on: (1) potential sources of algorithmic bias and (2) our approach to fairness considerations. In addition, we discuss possible future societal consequences.

(1) Potential sources of algorithmic bias: Researchers in ethical machine learning for healthcare are typically concerned with algorithms that may unintentionally codify existing structural and individual biases through biased training data. Data generated from clinical practice can reflect harmful biases (such as the issue of physician biases impacting treatment of pain for Black patients). However, we note that SAFE-T is not an algorithm for use in clinical practice; it is specifically for dose-finding clinical trials, where doses are allocated in accordance with strictly pre-defined procedures and patient outcomes are evaluated against strictly defined guidelines (typically involving objective health measurements). While such clinical procedures were developed to prevent confounding in clinical findings, by design they help mitigate biases that may stem from human evaluation of patient outcomes. SAFE-T is only trained with the patient outcomes data collected during a trial. As such, our primary concern would be to ensure that all trial protocols are properly followed, particularly any blinding practices that would mitigate practitioner biases. For success of SAFE-T (as well as any dose-finding trial method), standardized patient outcome assessment is thus particularly important.

(2) Our approach to fairness considerations: In our work, we do not utilize explicit fairness constraints or objectives in the pursuit of health equity. Instead, we identify a clear source of unfairness (unaddressed patient heterogeneity) and develop our method specifically to address this concern. We also ensure that our design prioritize participant safety and utility and also take care to provide thorough empirical experiments on realistic scenarios to properly assess robustness and flexibility.

We value fairness considerations very highly in our work; however, fairness considerations are context-dependent. In many healthcare contexts, fairness considerations would not be resolved with standard fairness constraints (such as demographic parity). In particular, the purpose of our work is to address a concerning health inequity: that certain population subgroups may experience greater risks of adverse drug effects due to flaws in the current methodology of dose-finding trials. As discussed in our paper, standard dose-finding trial methods assume that the patient population is homogeneous, not accounting for possible variations in drug toxicity and efficacy due to patient heterogeneity. This issue, in conjunction with existing inequalities in subject selection for clinical trials, may lead to both disparities in patient utility during the trial (from sub-optimal dose allocation) and after the trial (from non-representative dose recommendations). The way that we tackle inequity in this case is through proper consideration of heterogeneous patient populations, rather than standard fairness constraints. In addition, we examine subgroup disparities in careful experimental evaluation, finding that SAFE-T improves both performance (safety, accuracy, utility, Section 5.1) for patients and also reduces disparities between subgroups (even when subgroup ratios vary, Section 5.2).

As an example: We may have a dose-finding trial examining a drug that is more toxic for women than men at the same dose (which is likely: (Zucker & Prendergast, 2020; Unger et al., 2022)). Women are historically under-represented in clinical trials, particularly in early-phase dose-finding trials (Chien et al., 2022; Steinberg et al., 2021). If dose allocation decisions during the trial are made based on (1) data from the trial, which includes mostly men, and (2) assumptions that the trial population is homogeneous, women are more likely to be allocated unsafe doses. On top of that, the 'optimal' recommended dose based on trial findings may be optimal for men, but not for women. To mitigate this issue (and similar issues), we require a method that can account for patient heterogeneity.

Societal consequences. Our algorithm is a small contribution towards mitigation of health disparities caused by widely-used dose-finding trial methods. However, we acknowledge that the field is complex and subject to many issues that are not simply fixable with new methods. For example, trial practitioners must be willing to adopt new methods and also require the expertise to use any new methods. New methods may also be met with skepticism during trial approval processes, although adaptive designs are increasingly being adopted (Wheeler et al., 2019). In addition, as stated in the Conclusion, the dose-finding problem setting we investigate is restricted to a basic form, where binary toxicity and efficacy outcomes are considered. While this is the structure that standard trials assume, it is simplified from reality; this simplification leads to information loss and does not consider more complex circumstances, such as missing data or multiple target outcomes. We hope that this work brings attention to an interesting problem area that can benefit from greater, careful attention from the ML

community.

References

Phases of clinical trials, Oct 2022. URL https://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/what-clinical-trials-are/phases-of-clinical-trials.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3:397–422, 2003.

Aziz, M., Kaufmann, E., and Riviere, M.-K. On multi-armed bandit designs for dose-finding trials. *J. Mach. Learn. Res.*, 22:14:1–14:38, 2020.

Barocas, S., Hardt, M., and Narayanan, A. *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2023.

Berkenkamp, F., Krause, A., and Schoellig, A. P. Bayesian optimization with safety constraints: Safe and automatic parameter tuning in robotics. *ArXiv*, abs/1602.04450, 2016.

Bottero, A. G., Luis, C. E., Vinogradska, J., Berkenkamp, F., and Peters, J. Information-theoretic safe exploration with gaussian processes. *ArXiv*, abs/2212.04914, 2022.

Boyd, A. Inequalities for mills' ratio. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs*, 6:44–46, 1959.

Brøgger-Mikkelsen, M., Ali, Z. S., Zibert, J. R., Andersen, A. D., and Thomsen, S. F. Online patient recruitment in clinical trials: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 22, 2020.

Chapple, A. G. and Thall, P. F. Subgroup-specific dose finding in phase i clinical trials based on time to toxicity allowing adaptive subgroup combination. *Pharmaceutical Statistics*, 17:734 – 749, 2018. URL https://api.semanticscholar.org/CorpusID:52013376.

Cheung, Y. K. and Chappell, R. A simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics*, 58, 2002.

Chien, I., Deliu, N., Turner, R. E., Weller, A., Villar, S. S., and Kilbertus, N. Multi-disciplinary fairness considerations in machine learning for clinical trials. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022.

Chiuzan, C., Shtaynberger, J., Manji, G. A., Duong, J. K., Schwartz, G. K., Ivanova, A., and Lee, S. M. Dose-finding designs for trials of molecularly targeted agents and immunotherapies. *Journal of Biopharmaceutical Statistics*, 27:477 – 494, 2017.

Chowdhury, S. R. and Gopalan, A. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, 2017.

Dickmann, L. J. and Schutzman, J. L. Racial and ethnic composition of cancer clinical drug trials: How diverse are we? *The oncologist*, 23 2:243–246, 2018.

Gardner, J. R., Pleiss, G., Bindel, D. S., Weinberger, K. Q., and Wilson, A. G. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Neural Information Processing Systems*, 2018.

Garivier, A., M'enard, P., Rossi, L., and Ménard, P. Thresholding bandit for dose-ranging: The impact of monotonicity. *arXiv: Statistics Theory*, 2017.

Harkin, T. S., Snowe, O. J., Mikulski, B. A., and Waxman, H. A. Drug safety : Most drugs withdrawn in recent years had greater health risks for women. 2001.

Hensman, J., de G. Matthews, A. G., and Ghahramani, Z. Scalable variational gaussian process classification. In *International Conference on Artificial Intelligence and Statistics*, 2014.

Hoffman, M. D. and Gelman, A. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *ArXiv*, abs/1111.4246, 2011.

Houlsby, N., Huszár, F., Ghahramani, Z., and Lengyel, M. Bayesian active learning for classification and preference learning. *ArXiv*, abs/1112.5745, 2011.

Huang, J., Su, Q., Yang, J., hua Lv, Y., He, Y., Chen, J., Xu, L., Wang, K., and shan Zheng, Q. Sample sizes in dosage investigational clinical trials: a systematic evaluation. *Drug Design, Development and Therapy*, 9:305 – 312, 2015.

Ivanova, A. and Wang, K. Bivariate isotonic design for dose-finding with ordered groups. *Statistics in Medicine*, 25, 2006. URL https://api.semanticscholar.org/CorpusID:20679378.

Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.

Journel, A. G. and Huijbregts, C. J. Mining geostatistics. 1976.

Kazerouni, A., Ghavamzadeh, M., Abbasi, Y., and Roy, B. V. Conservative contextual linear bandits. In *NIPS*, 2016.

Kong, Y. Are "intersectionally fair" ai algorithms really fair to women of color? a philosophical analysis. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022. URL https://api.semanticscholar.org/CorpusID:249872618.

Koopmeiners, J. S. and Modiano, J. F. A bayesian adaptive phase i–ii clinical trial for evaluating efficacy and toxicity with delayed outcomes. *Clinical Trials*, 11:38 – 48, 2014.

Kurzrock, R., Lin, C., Wu, T.-C., Hobbs, B. P., Pestana, R. C., and Hong, D. S. Moving beyond 3+3: The future of clinical trial design. *American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting*, 41:e133–e144, 2021.

Lee, H.-S., Shen, C., Jordon, J., and van der Schaar, M. Contextual constrained learning for dose-finding clinical trials. *ArXiv*, abs/2001.02463, 2020.

Love, S. B., Brown, S., Weir, C. J., Harbron, C., Yap, C., Gaschler-Markefski, B., Matcham, J., Caffrey, L., McKevitt, C., Clive, S., Craddock, C. F., Spicer, J. F., and Cornelius, V. R. Embracing model-based designs for dose-finding trials. *British Journal of Cancer*, 117:332 – 339, 2017.

Mehrabi, N., Morstatter, F., Saxena, N. A., Lerman, K., and Galstyan, A. G. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54:1 – 35, 2019. URL https://api.semanticscholar.org/CorpusID:201666566.

Morita, S., Thall, P. F., and Takeda, K. A simulation study of methods for selecting subgroup-specific doses in phase 1 trials. *Pharmaceutical Statistics*, 16:143 – 156, 2017. URL https://api.semanticscholar.org/CorpusID:4909935.

Nickisch, H. and Rasmussen, C. E. Approximations for binary gaussian process classification. *Journal of Machine Learning Research*, 9:2035–2078, 2008.

Özdemir, B. C., Gerard, C. L., and da Silva, C. E. Sex and gender differences in anticancer treatment toxicity - a call for revisiting drug dosing in oncology. *Endocrinology*, 2022.

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

Ramamoorthy, A., Kim, H. H., Shah-Williams, E., and Zhang, L. Racial and ethnic differences in drug disposition and response: Review of new molecular entities approved between 2014 and 2019. *The Journal of Clinical Pharmacology*, 62, 2021.

Rasmussen, C. E. and Williams, C. K. I. Gaussian processes for machine learning. In *Adaptive computation and machine learning*, 2003.

Riviere, M.-K., Yuan, Y., Jourdan, J.-H., Dubois, F., and Zohar, S. Phase i/ii dose-finding design for molecularly targeted agent: Plateau determination using adaptive randomization. *Statistical Methods in Medical Research*, 27: 466 – 479, 2018.

Salter, A., Morgan, C. J., and Aban, I. B. Implementation of a two-group likelihood time-to-event continual reassessment method using sas. *Computer methods and programs in biomedicine*, 121 3:189–96, 2015. URL https://api.semanticscholar.org/CorpusID:25315809.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. J. Probabilistic programming in python using pymc3. *PeerJ Prepr.*, 4: e1686, 2016.

Schreiter, J., Nguyen-Tuong, D., Eberts, M., Bischoff, B., Markert, H., and Toussaint, M. Safe exploration for active learning with gaussian processes. In *ECML/PKDD*, 2015.

Shen, C., Wang, Z., Villar, S. S., and van der Schaar, M. Learning for dose allocation in adaptive clinical trials with safety constraints. In *International Conference on Machine Learning*, 2020.

Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W. Gaussian process optimization in the bandit setting: No regret and experimental design. In *International Conference on Machine Learning*, 2009.

Steinberg, J. R., Turner, B. E., Weeks, B. T., Magnani, C. J., Wong, B. O., Rodriguez, F., Yee, L. M., and Cullen, M. R. Analysis of female enrollment and participant sex by burden of disease in us clinical trials between 2000 and 2020. *JAMA Network Open*, 4, 2021.

Sui, Y., Gotovos, A., Burdick, J. W., and Krause, A. Safe exploration for optimization with gaussian processes. In *International Conference on Machine Learning*, 2015.

Sui, Y., Zhuang, V., Burdick, J. W., and Yue, Y. Stagewise safe bayesian optimization with gaussian processes. In *International Conference on Machine Learning*, 2018.

Thall, P. F. and Cook, J. D. Dose-finding based on efficacy–toxicity trade-offs. *Biometrics*, 60, 2004.

Thomas, M., Bornkamp, B., and Seibold, H. Subgroup identification in dose-finding trials via model-based recursive partitioning. *Statistics in Medicine*, 37:1608 – 1624, 2018.

Turchetta, M., Berkenkamp, F., and Krause, A. Safe exploration in finite markov decision processes with gaussian processes. *ArXiv*, abs/1606.04753, 2016.

Turchetta, M., Berkenkamp, F., and Krause, A. Safe exploration for interactive machine learning. *ArXiv*, abs/1910.13726, 2019.

Unger, J. M., Vaidya, R., Albain, K. S., Leblanc, M. L., Minasian, L. M., Gotay, C., Henry, N. L., Fisch, M. J., Lee, S. M., Blanke, C. D., and Hershman, D. L. Sex differences in risk of severe adverse events in patients receiving immunotherapy, targeted therapy, or chemotherapy in cancer clinical trials. *Journal of Clinical Oncology*, 40:1474 – 1486, 2022.

Villar, S. S., Bowden, J., and Wason, J. M. S. 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: 199–215, 2015.

Wages, N. A., Read, P. W., and Petroni, G. R. A phase i/ii adaptive design for heterogeneous groups with application to a stereotactic body radiation therapy trial. *Pharmaceutical Statistics*, 14:302 – 310, 2015. URL https://api.semanticscholar.org/CorpusID:5347921.

Wages, N. A., Chiuzan, C., and Panageas, K. S. Design considerations for early-phase clinical trials of immune-oncology agents. *Journal for Immunotherapy of Cancer*, 6, 2018.

Wang, Z. and de Freitas, N. Theoretical analysis of bayesian optimisation with unknown gaussian process hyper-parameters. *ArXiv*, abs/1406.7758, 2014.

Wang, Z., Wagenmaker, A. J., and Jamieson, K. G. Best arm identification with safety constraints. *ArXiv*, abs/2111.12151, 2021.

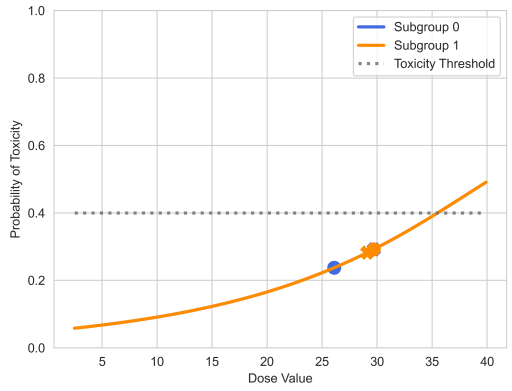
Wheeler, G. M., Mander, A. P., Bedding, A., Brock, K., Cornelius, V. R., Grieve, A. P., Jaki, T. F., Love, S. B., Odondi, L., Weir, C. J., Yap, C., and Bond, S. How to design a dose-finding study using the continual reassessment method. *BMC Medical Research Methodology*, 19, 2019.

Williams, C. K. I. and Barber, D. Bayesian classification with gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20:1342–1351, 1998.

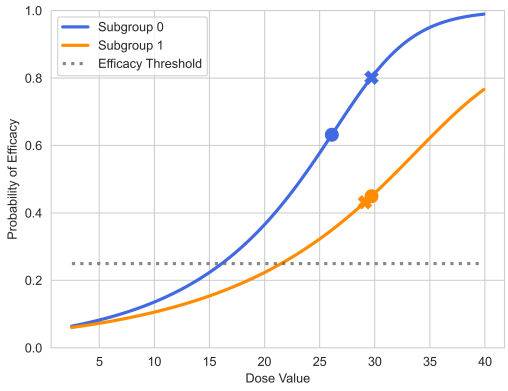
Zhang, W., Sargent, D. J., and Mandrekar, S. J. An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine*, 25, 2006.

Zucker, I. and Prendergast, B. J. Sex differences in pharmacokinetics predict adverse drug reactions in women. *Biology of Sex Differences*, 11, 2020.

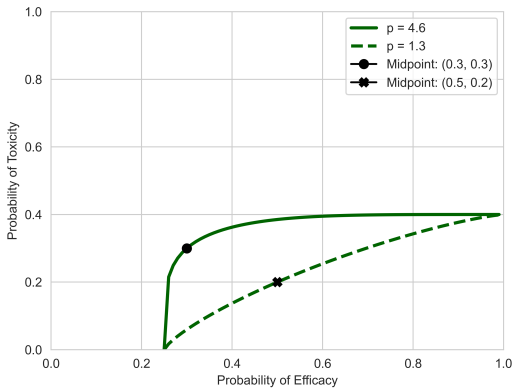
A. Definition of Utility



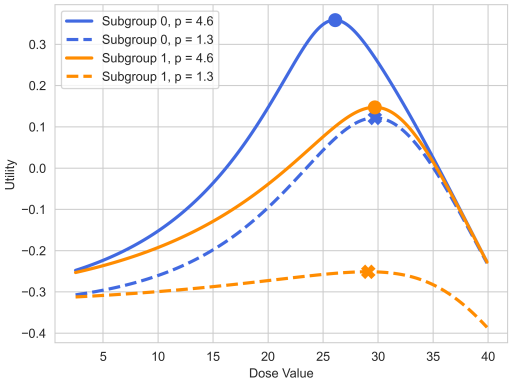
(a) Dose-toxicity



(b) Dose-efficacy



(c) Utility tradeoff curve



(d) Dose-utility

Figure 5: We illustrate the impact of selecting different midpoint values to define the toxicity-efficacy tradeoff curve in a scenario with two subgroups (subgroup 0 in blue, subgroup 1 in orange). For a midpoint of $(0.3, 0.3)$, the utility trade-off *(c)* and subsequent dose-utility values *(d)* are shown in solid lines, and the resultant optimal dose is shown as ●. For a midpoint of $(0.5, 0.2)$, the curves are shown as dashed lines, and the resultant optimal dose is marked as an ✖. The midpoint $(0.3, 0.3)$ results in a steeper curve, indicting that proximity to either the toxicity or efficacy thresholds are heavily penalized. The midpoint $(0.5, 0.2)$ results in a much flatter curve, indicating that incremental increases in probability of toxicity are acceptable given improvements in the probability of efficacy.

The parameter p is determined by setting the utility to 0 and plugging in a midpoint, (p_E^*, p_T^*), elicited from practitioners, that defines the curvature of the contour:

$$1 = \left(\left(\frac{p_T^*}{\tau_T} \right)^p + \left(\frac{1 - p_E^*}{1 - \tau_E} \right)^p \right)^{\frac{1}{p}} \tag{3}$$

Figure 5 shows the impact of the choice of midpoint on the resulting utility trade-off curve and subsequent dose-utility values. Doses that lie on the utility contour (c) are considered to have 0 utility, while doses below the contour have positive utility. A utility of $geq0$ indicates a desirable tradeoff. We notice the manifestation of the defined trade-off in two dimensions: by subgroup (blue or orange) or by choice of midpoint (solid line with ●, or dashed line with ✖). We refer to the midpoint at $(0.3, 0.3)$ as ●, leading to a steeper trade-off curve; we refer to the midpoint at $(0.5, 0.2)$ as ✖, leading to a flatter trade-off curve. When ● is used, we notice that the optimal dose for subgroup 0 (blue) is lower than that of subgroup 1 (orange). This makes intuitive sense: while both subgroups maintain the same dose-toxicity curve, subgroup 0 experiences much higher efficacy at the same doses. Thus, subgroup 0 is capable of experiencing an acceptably high efficacy at a lower toxicity than

subgroup 1. When ✖ is used, the optimal dose for subgroup 0 (blue) increases because proximity to the toxicity threshold is not penalized as much. Therefore, a greater increase in toxicity is acceptable given and increase in efficacy. We note that if ✖ is used to define the utility contour, it is not possible for subgroup 1 (orange) to receive doses that have positive utility. Clinical trial practitioners may select a midpoint consistent with the levels of risk they are willing to condone in the trial.

B. Further theoretical details

B.1. Assumptions

While the assumptions required for Theorems 4.1 and 4.2 can be considered strong, they are also standard in the literature. Theorem 4.1 is predicated on the assumption that we have access to d_0, a known safe dose. The safety bounds in related works by Sui et al. (2015, Theorem 1) and Sui et al. (2018, Theorem 1) rely on the assumption of access to an initial safe set that "contains at least one safe decision" which may be unrealistic in practice. In contrast, while we assume knowledge of a safe d_0, we also include an early stopping constraint N_{stop} in our algorithm as a practical safeguard if the initial set does not include a truly safe point.

Theorem 4.2 is predicated on the assumption that the true latent function is indeed drawn from the given GP prior. Samples of Gaussian processes have a certain level of smoothness, determined by the kernel. For example, see Kolmogorov's continuity theorem: if f is α-Hölder continuous w.r.t. the L^p-norm, then samples are almost surely $(\alpha - \frac{1}{p})$-Hölder continuous. Therefore, assuming that the latent function is distributed according to a GP implies certain assumptions about the smoothness of the underlying function. In our particular implementation, we use the radial basis function (RBF) kernel. This kernel is noted as "probably the most widely-used kernel within the kernel machines field" and is "infinitely differentiable, which means that the GP with this covariance function has mean square derivatives of all orders, and is thus very smooth" (Rasmussen & Williams, 2003).

While this assumption that the latent function is drawn from the GP prior can be considered strong, it is standard in the literature (e.g. Srinivas et al. (2009, Theorem 1)). Such a theoretical result lays the groundwork for future work to generalize to slightly weaker assumptions, namely the "agnostic setting" defined in Srinivas et al. (2009, Theorem 3) and further adopted by Chowdhury & Gopalan (2017, Theorem 3), Sui et al. (2015, Theorem 1), and Sui et al. (2018, Theorem 1). The "agnostic setting" rests on the assumption that the underlying function is distributed according to a GP to the assumption that the underlying function is bounded in the RKHS norm of the selected GP kernel. This generalization can directly be interpreted as a smoothness assumption. The results derived in this agnostic setting rely critically on the observed data being drawn from a model with Gaussian additive noise. In contrast, the data in our setting is Bernoulli-distributed, so the proof techniques behind Srinivas et al. (2009, Theorem 3) and Chowdhury & Gopalan (2017, Theorem 3) cannot be used to generalize our results to the agnostic setting. We plan to generalize our results to the agnostic setting in future work (note that this requires extensive novel theoretical research).

Because the Bernoulli likelihood results in an analytically intractable posterior, we perform a Gaussian approximation over f (Nickisch & Rasmussen, 2008), as in related works (Houlsby et al., 2011; Schreiter et al., 2015; Bottero et al., 2022). We define precisely where the approximation error occurs in the proof of Theorem 4.2 in Appendix B.2.

While we provide theoretical safety bounds to illustrate the correctness of our algorithm, we note that these assumptions may not be exactly satisfied in practice. As such, we took care to highlight these assumptions clearly and define safeguards in our algorithm to minimize implementation risks.

B.2. Proofs

In this section, we provide the full proofs for Theorem 4.1, the lower bound on the number of initial safe samples N_0 that guarantees that SAFE-T succeeds, and Theorem 4.2, a safety guarantee on SAFE-T. While Theorem 1 is a generalization of Theorem 2 by Schreiter et al. (2015) and Theorem 2 is similar to Theorem 3 by Schreiter et al. (2015), there are notable differences due to our differing safety constraint (Equation 1). Our proof strategy follows that by Schreiter et al. (2015), again with notable differences due to the differing safety constraint, and the lemmas that we present are similar to the lemmas presented by Schreiter et al. (2015).

To achieve the result in Theorem 1, we assume that we have N_0 initial safe points satisfying the safety constraint (Equation 1). These safe points correspond to negative labels, as positive labels indicate toxicity. Theorem 1 assumes that the underlying Gaussian processes uses a stationary covariance function, so a repeated sample of the same safe initial dose, d_0, maximally

decreases the mean and variance of the approximated posterior. This allows us to define the mean, μ_{N_0}, and variance, $\sigma^2_{N_0}$, of the posterior after N_0 samples as in Lemma 1. In Lemma 2, we provide an upper bound on μ_{N_0} using the results of Lemma 1. In Lemma 3, we lower bound N_0 with respect to μ_{N_0}. Combining the results of Lemma 2 and Lemma 3, we prove the bound presented in Theorem 1.

To achieve the result in Theorem 2, we first define the probability of selecting an unsafe dose given the safety constraint (1), then use the union bound to determine the probability of safety over all allocated doses.

Lemma B.1 (Lemma 2 by Schreiter et al. (2015))**.** *Assume the setting of Theorem 4.1. Let $\hat{g}_s(d)$ be the posterior latent toxicity function conditioned on N_0 samples of the same safe initial dosage d_0. Under a Laplace approximation of this posterior, the mean and variance at d_0 satisfy*

$$\mu_{N_0} = -N_0\sigma_g^2 q_{N_0} \qquad \textit{and} \qquad \sigma^2_{N_0} = \frac{\sigma_g^2}{1 + N_0\sigma_g^2 w_{N_0}} \tag{4}$$

where

$$q_{N_0} = \frac{\varphi(-\mu_{N_0})}{\Phi(-\mu_{N_0})} \qquad \textit{and} \qquad w_{N_0} = q^2_{N_0} - \mu_{N_0} q_{N_0} \tag{5}$$

and σ_g^2 is the maximum variance of the GP prior on any of the latent toxicity functions g_s.

Proof. The proof follows the proof of Lemma 2 by Schreiter et al. (2015).

Let $\boldsymbol{Y} = -\mathbf{1}$ be a vector of N_0 negative labels. Using the Laplace approximation (Williams & Barber, 1998), we approximate our non-Gaussian posterior $\hat{g}_s(d_0)$ with a Gaussian $\mathcal{N}(\mu_{N_0}, \sigma^2_{N_0})$ where

$$\mu_{N_0} = \arg\max_g \, p(g \mid \boldsymbol{Y}, d_0) \tag{6}$$

and

$$\sigma^2_{N_0} = (w + \sigma_g^{-2})^{-1}, \qquad w = -\left.\frac{\partial^2}{\partial g^2} \log p(\boldsymbol{Y} \mid g, d_0)\right|_{g=\mu_{N_0}}. \tag{7}$$

Here the likelihood of our classification model is given by

$$p(\boldsymbol{Y} \mid g, d_0) = \prod_{n=1}^{N_0} \Phi(Y_n g(d_0)) = \Phi(-g(d_0))^{N_0}. \tag{8}$$

From the Laplace approximation, it follows that

$$\mu_{N_0} = \sigma_g^2 \left.\frac{\partial}{\partial g} \log p(\boldsymbol{Y} \mid g, d_0)\right|_{g=\mu_{N_0}} = -N_0\sigma_g^2 q_{N_0} \tag{9}$$

where q_{N_0} is defined in the lemma statement. In addition, we compute

$$\begin{aligned} -\left.\frac{\partial^2}{\partial g^2} \log p(\boldsymbol{Y} \mid g, d_0)\right|_{g=\mu_{N_0}} &= N_0 \left.\frac{\partial}{\partial g}\frac{\phi(-g)}{\Phi(-g)}\right|_{g=\mu_{N_0}} && (10)\\ &= N_0 \left.\left(-g\frac{\phi(-g)}{\Phi(-g)} + \frac{\phi^2(-g)}{\Phi^2(-g)}\right)\right|_{g=\mu_{N_0}} && (11)\\ &= N_0(q^2_{N_0} - \mu_{N_0} q_{N_0}) = N_0 w_{N_0} && (12) \end{aligned}$$

where w_{N_0} is defined in the lemma statement. Therefore,

$$\sigma^2_{N_0} = \frac{1}{N_0 w_{N_0} + \sigma_g^{-2}} = \frac{\sigma_g^2}{1 + N_0\sigma_g^2 w_{N_0}}. \tag{13}$$

□

Lemma B.2. *Assume the setting of Theorem 4.1. Given N_0 samples of the same safe initial point d_0, if the the safety constraint (1) is satisfied, then the Laplace approximation of the posterior mean of the latent toxicity function must be upper bounded as follows:*

$$\mu_{N_0} \leq \frac{1}{2}(1 + \rho_T - \sqrt{(p-1)^2 + 4\nu_T\sigma_g}) \tag{14}$$

where μ_{N_0} and σ_g are defined as in Lemma B.1, $\nu_T > 0$, and $\rho_T = \Phi^{-1}(\tau_T)$.

Proof. With the definitions of posterior mean and variance from Lemma B.1, we can rewrite the safety constraint (Equation 1) as follows:

$$\rho_T - \mu_{N_0} \geq \frac{\nu_T\sigma_g}{\sqrt{1 - q_{N_0}\mu_{N_0} + \mu_{N_0}^2}}. \tag{15}$$

By bounds on the inverse Mill ratio (Boyd, 1959), we have that $q_{N_0} \in [0, 1]$. Therefore, also using that $\mu_{N_0} < 0$ (see the definition of μ_{N_0} in Lemma 1),

$$1 - q_{N_0}\mu_{N_0} + \mu_{N_0}^2 \leq 1 - 2\mu_{N_0} + \mu_{N_0}^2 = (1 - \mu_{N_0})^2, \tag{16}$$

so, noting that $1 - \mu_{N_0} > 0$,

$$\rho_T - \mu_{N_0} \geq \frac{\nu_T\sigma_g}{1 - \mu_{N_0}}. \tag{17}$$

Then multiply with $1 - \mu_{N_0} > 0$ on both sides to find

$$(\rho_T - \mu_{N_0})(1 - \mu_{N_0}) = \mu_{N_0}^2 - (\rho_T + 1)\mu + \rho_T \geq \nu_T\sigma_g \iff \mu_{N_0}^2 - (\rho_T + 1)\mu + \rho_T - \nu_T\sigma_g \geq 0. \tag{18}$$

The discriminant of this quadratic formula is equal to

$$\text{disc} = (\rho_T + 1)^2 - 4(\rho_T - \nu_T\sigma_g) = (\rho_T - 1)^2 + 4\nu_T\sigma_g > (\rho_T - 1)^2 \geq 0. \tag{19}$$

Since this discriminant is strictly positive and the parabola in (18) is upwards (leading coefficient is positive), we have $\mu_{N_0} \leq \mu_-$ or $\mu_{N_0} \geq \mu_+$ where

$$\mu_\pm = \tfrac{1}{2}((\rho_T + 1) \pm \sqrt{\text{disc}}). \tag{20}$$

The safety constraint requires that $\rho_T - \mu_{N_0} \geq \nu_T\sigma_{N_0} \geq 0$. Since

$$\rho_T - \mu_\pm = \tfrac{1}{2}((\rho_T - 1) \mp \sqrt{\text{disc}}) \tag{21}$$

and $\sqrt{\text{disc}} > |\rho_T - 1|$, we find

$$\rho_T - \mu_+ = \tfrac{1}{2}((\rho_T - 1) - \sqrt{\text{disc}}) < 0. \tag{22}$$

This means that $\mu_{N_0} \geq \mu_+$ implies that $\rho_T - \mu_{N_0} \leq \rho_T - \mu_+ < 0$. Therefore, $\mu_{N_0} \geq \mu_+$ cannot be true, which means that $\mu_{N_0} \leq \mu_-$ must be true. □

Lemma B.3. *Assume the setting of Theorem 4.1. If the safety constraint in Equation 1 is satisfied for some dose d, then N_0 must be lower bounded as follows:*

$$N_0 \geq (2\varphi(\mu_{N_0}))^{-1} \min\Big(\frac{\nu_T}{\sqrt{3}\sigma_g} - \frac{\rho_T}{\sigma_g^2}, \frac{1}{\sigma_g^2}\sqrt{\frac{\rho_T^2}{4} + \frac{\sigma_g\nu_T}{\sqrt{3}}} - \frac{\rho_T}{2\sigma_g^2}\Big). \tag{23}$$

Proof. Like in Lemma B.2, with the definitions of posterior mean and variance from Lemma B.1, we can rewrite the safety constraint (Equation 1) as follows:

$$-\mu_{N_0} = q_{N_0}\sigma_g^2 N_0 \geq -\rho_T + \frac{\nu_T\sigma_g}{\sqrt{1 - q_{N_0}\mu_{N_0} + \mu_{N_0}^2}}. \tag{24}$$

From Lemma B.1, we know that $\mu_{N_0} < 0$. We consider the cases $\mu_{N_0} \in (-\infty, -1]$ and $\mu_{N_0} \in (-1, 0)$ separately.

On the one hand, suppose that $\mu_{N_0} \in (-1, 0)$. Then, recalling that $q_{N_0} \in [0, 1]$ (Boyd, 1959),

$$1 - q_{N_0}\mu_{N_0} + \mu_{N_0}^2 \leq 1 + 1 + 1 = 3, \tag{25}$$

so

$$q_{N_0}\sigma_g^2 N_0 \geq -\rho_T + \frac{\nu_T \sigma_g}{\sqrt{3}} \implies N_0 \geq \frac{1}{q_{N_0}} \left(-\frac{\rho_T}{\sigma_g^2} + \frac{\nu_T}{\sqrt{3}\sigma_g} \right) = \frac{1}{q_{N_0}} \left(\frac{\nu_T}{\sqrt{3}\sigma_g} - \frac{\rho_T}{\sigma_g^2} \right). \tag{26}$$

On the other hand, suppose that $\mu_{N_0} \in (-\infty, -1]$. Then

$$1 - q_{N_0}\mu_{N_0} + \mu_{N_0}^2 \leq \mu_{N_0}^2 + \mu_{N_0}^2 + \mu_{N_0}^2 = 3\mu_{N_0}^2, \tag{27}$$

so

$$q_{N_0}\sigma_g^2 N_0 \geq -\rho_T + \frac{\nu_T \sigma_g}{\sqrt{3}|\mu_{N_0}|} = -\rho_T + \frac{\nu_T}{\sqrt{3 q_{N_0}}\sigma_g N_0}. \tag{28}$$

Rearrange this inequality as follows:

$$q_{N_0}\sigma_g^2 N_0^2 + \rho_T N_0 + \frac{\nu_T}{\sqrt{3 q_{N_0}}\sigma_g} \geq 0. \tag{29}$$

The discriminant of this quadratic formula is equal to

$$\text{disc} = \rho_T^2 + 4 q_{N_0}\sigma_g^2 \frac{\nu_T}{\sqrt{3 q_{N_0}}\sigma_g} = \rho_T^2 + \frac{4\sigma_g \nu_T}{\sqrt{3}} > \rho_T^2 \geq 0. \tag{30}$$

Since this discriminant is strictly positive and the parabola in (29) is upwards (leading coefficient is positive), we have $N_0 \leq N_-$ or $N_0 \geq N_+$ where

$$N_\pm = \frac{1}{2 q_{N_0}\sigma_g^2}(-\rho_T \pm \sqrt{\text{disc}}). \tag{31}$$

Using that N_0 must be positive, we find that

$$N_0 \geq \frac{1}{2 q_{N_0}\sigma_g^2} \left(-\rho_T + \sqrt{\rho_T^2 + \frac{4\sigma_g \nu_T}{\sqrt{3}}} \right) = \frac{1}{q_{N_0}} \left(\frac{1}{\sigma_g^2} \sqrt{\frac{\rho_T^2}{4} + \frac{\sigma_g \nu_T}{\sqrt{3}}} - \frac{\rho_T}{2\sigma_g^2} \right). \tag{32}$$

Putting together the two cases:

$$N_0 \geq \frac{1}{q_{N_0}} \min \left(\frac{\nu_T}{\sqrt{3}\sigma_g} - \frac{\rho_T}{\sigma_g^2}, \frac{1}{\sigma_g^2} \sqrt{\frac{\rho_T^2}{4} + \frac{\sigma_g \nu_T}{\sqrt{3}}} - \frac{\rho_T}{2\sigma_g^2} \right). \tag{33}$$

Finally, using that $-\mu_{N_0} > 0$, so $\Phi(-\mu_{N_0}) \geq \frac{1}{2}$,

$$\frac{1}{q_{N_0}} = \frac{\Phi(-\mu_{N_0})}{\varphi(-\mu_{N_0})} \geq \frac{1}{2\varphi(-\mu_{N_0})}, \tag{34}$$

which gives the equation from the lemma statement. □

These lead us to our result in Theorem 1.

Proof (Theorem 4.1). Using the lower bound of N_0 as defined in Lemma 3 in combination with the upper bound condition on μ_{N_0} given in Lemma 2, we obtain the result of Theorem 1. □

Proof (Theorem 4.2). Recall that SAFE-T sets $\nu_T = \Phi^{-1}(1 - \frac{\delta_T}{N-N_0})$. At each time point n, if we select a dose d_n that fulfills the safety constraint (1), the probability that the dose is unsafe is as follows:

$$\begin{aligned}
\Pr(\Phi(g_{s_n}(d_n)) > \tau_T) &= \Pr(g_{s_n}(d_n) > \Phi^{-1}(\tau_T)) \\
&\leq \Pr(g_{s_n}(d_n) > \mu_{\hat{g}_{s_n}}(d_n) + \nu_T \sigma_{\hat{g}_{s_n}}(d_n)) \\
&= \Pr\left(\frac{g_{s_n}(d_n) - \mu_{\hat{g}_{s_n}}(d_n)}{\sigma_{\hat{g}_{s_n}}(d_n)} > \nu_T\right) \\
&= \mathbb{E}_{Y_{1:(n-1)}}\left[\Pr\left(\frac{g_{s_n}(d_n) - \mu_{\hat{g}_{s_n}}(d_n)}{\sigma_{\hat{g}_{s_n}}(d_n)} > \nu_T \,\middle|\, Y_{1:(n-1)}\right)\right] \\
&\approx \mathbb{E}_{Y_{1:(n-1)}}\left[1 - \Phi^{-1}(\nu_T)\right] \\
&= \frac{\delta_T}{N - N_0}
\end{aligned}$$

where the quality of the approximate equality depends on how well the distribution of $\hat{g}_s(d_n)$ can be approximated with a Gaussian distribution. Using the union bound, we then find that

$$\Pr\Big(\bigcup_{n=N_0+1}^{N} \{\Phi(g_{s_n}(d_n)) > \tau_T\}\Big) \leq \sum_{n=N_0+1}^{N} \Pr(\Phi(g_{s_n}(d_n)) > \tau_T) \approx (N - N_0)\left(\frac{\delta_T}{N - N_0}\right) = \delta_T.$$

Thus, selected doses $d_{N_0:N}$ are approximately safe with at least probability at least $1 - \delta_T$. □

B.3. Distinction from related works

While our theoretical bounds are motivated by Schreiter et al. (2015), there are key differences. Our safety constraint (Equation 1) differs from the one defined in Schreiter et al. (2015, Equation 13). Our constraint (Equation 1) allows a variable practitioner-specified toxicity probability threshold ρ_T, while Schreiter et al. (2015, Equation 13) fixes the constraint ρ_T at 0. Their simplification allows them to approach the derivation of their theoretical bounds Schreiter et al. (2015, Theorems 1 & 2) differently. Our theoretical bounds (Theorems 4.1 and 4.2 (Appendix B.2) can be considered a generalization of this specific case proven in Schreiter et al. (2015), as ρ_T can be set as 0 or whatever a practitioner decides. This difference is critical to the derivation of Theorems 4.1 and 4.2 (Appendix B.2) and involves significant technical contribution in our proofs (Appendix B.2. The proof of Schreiter et al. (2015, Theorem 3) also requires the use of Laplace approximation for Gaussian distributions. We approach our proof of Theorem 4.2 differently such that this requirement is not needed. In addition, we make it explicit that there is approximation error due approximate inference for the posteriors. The proof makes it clear where precisely this approximation comes into play: the fifth equality in the proof of Theorem 4.2 in Appendix B.2.

C. Description of Synthetic Scenarios

For our experiments in Section 5.1 we constructed 18 synthetic scenarios to represent varying dose-response shapes, toxicity and efficacy thresholds, variations between subgroups, and possible edge cases. In Table 3 we record the characteristics captured by each scenario. Figure 12 (displayed at the end of the Appendix) show the true toxicity, efficacy, and Thall utility plots by scenario. The optimal doses by subgroup per scenario are selected using the Thall utility values calculated from the true underlying toxicity and efficacy probabilities (in conjunction with the set toxicity and efficacy thresholds). If no doses satisfy the toxicity and efficacy thresholds, then the optimal dose is no dose. We have also reviewed the optimal doses manually and found that in all scenarios, the intuitively optimal dose matches that selected using Thall utility.

In Table 3, we use a variety of terms to describe the dose-response relationships:

- Increasing: response is generally increasing as dose increases with no obvious pattern
- Increasing (Logarithmic): response is increasing in a logarithmic shape (more slowly) as dose increases that is not quite a plateau

- Increasing (Exponential): response is increasing in an exponential shape as dose increases
- Plateauing: response plateaus, such that after some dose there is no (or very minimal) change in response as dose increases
- Parabolic: response clearly peaks at some dose
- Flat: response does not change with dose
- Vertical transformation: response curve is shifted vertically, retains same shape across subgroups
- Horizontal transformation: response curve is shifted horizontally, retains same shape across subgroups
- Combined transformation: response curve is shifted and dose not retain the same shape across subgroups

Broadly, we investigate cases where there is either no subgroup variation, only efficacy variations, only toxicity variations, or variations in both toxicity and efficacy, with differing shapes and edge cases tested throughout.

For our experiments in Section 5.4, we constructed 4 synthetic continuous scenarios to explore the most common dose-response patterns. In Figure 13 (displayed at end of Appendix) we plot the true toxicity, efficacy, and Thall utility values by scenario. Similarly to the discrete scenarios, the true optimal dose for the continuous scenarios is selected based on the maximum Thall utility value.

D. Experimental Details

D.1. SAFE-T implementation details

In dose-finding trials, patients typically arrive in cohorts of size pre-determined by trial practitioners (often 3), usually based on logistical considerations. Patients in a cohort may belong to different subgroups and all are assumed to be treated at the same time. Models of dose-response relationships are updated after outcomes are observed from all patients in a cohort. We follow this realistic setting in our experiments, first assigning $N_0 = 3$ participants to the lowest dose d_0 and treating all subsequent participants in cohorts of size 3. We note that we avoid the use of an unsafe burn-in period as is done in related works, where each dose is selected in immediate sequence (without regard for toxicity) (Lee et al., 2020; Shen et al., 2020).

SAFE-T models dose-toxicity and dose-efficacy with multi-output Gaussian process, using the linear model of co-regionalization (LMC) (Journel & Huijbregts, 1976). The LMC model assumes that each output dimension is a linear combination of Q learned latent functions, $\boldsymbol{g}(\cdot) = [g^{(1)}(\cdot), \ldots, g^{(Q)}(\cdot)]$:

$$\boldsymbol{f}_s(\boldsymbol{x}) = \sum_{i=1}^{Q} a_s^{(i)} g^{(i)}(\boldsymbol{x}) \tag{35}$$

In our setting, each output dimension (also referred to as a *task*) corresponds to a subgroup. Thus, our Gaussian processes learn subgroup representations that are composed of underlying latent functions, with $s \in \{1, \ldots, S\}$, $S = 2$, and $Q = 3$.

Due to the small sample sizes we expect to work with, we set many of the hyperparameters of the Gaussian processes ahead of time. While we propose hyperparameters that can be defined in order to obtain theoretical probability guarantees, in practice we use those definitions as a starting point for fine-tuning. Our experimental hyperparameters have been manually tuned to work with the standard value ranges of dose-finding trials. In a CRM trial, practitioners are advised to run simulations in order to determine reasonable parameters for the trial setting. In practice, trial practitioners can use prior knowledge and past data to inform hyperparameter tuning with simulation studies (as in our manual tuning procedure) (Wheeler et al., 2019). Practitioners will have knowledge of the toxicity/efficacy thresholds, dose range, dosages of interest, and hypothesized dose-response profiles. With guidance, these values can be used to construct simulated scenarios for hyperparameter tuning: for example, dosages inform the possible kernel length-scale and toxicity/efficacy thresholds inform GP prior means. A constant mean function and the RBF kernel have been shown in our experiments to be compatible with various dose-response profiles and practitioners can stick with these unless they have additional expertise. They may run simulations to semi-manually tune the hyperparameters of the GP (mean value, kernel length scale, LMC coefficients). We believe that introducing a slightly more complex hyperparameter tuning process is worth the tradeoff in performance and flexibility that SAFE-T offers in contrast to existing parametric methods. Oftentimes, practitioners

use software that abstracts the particularities of the modeling procedure (Wheeler et al., 2019). We believe that it is thus reasonable to suggest that practitioners may do the same with a method such as SAFE-T. Early phase trials are likely to have small sample sizes, so fitting hyperparameters could lead to instability.

Hyperparameters remain the same across all 18 test scenarios for comparability, indicating that practitioners need only general prior knowledge to determine suitable hyperparameters. We define $\nu_T = 0.2$ for the safety constraint and $\nu_E = 0.2$ (needed when UCB efficacy optimization is used), while the toxicity threshold (τ_T) and efficacy thresholds (τ_E) are set by the scenario setting. Both the toxicity and efficacy GPs use constant mean functions (we set mean $= -0.3$ for toxicity and mean $= -0.1$ for efficacy) and the stationary radial basis function kernel (RBF kernel) as the covariance function (we set length scale $= 4$ for toxicity and length scale $= 2$ for efficacy). We also set the matrix $\boldsymbol{A}$, with Q rows and S columns, which is composed of the coefficients $a_s^{(i)}$ of the LMC model to $\left(\begin{smallmatrix} 1.0 & 0 \\ 0.2 & 0.2 \\ 0 & 1.0 \end{smallmatrix}\right)$. These hyperparameters were manually tuned and the high performance maintained across all 18 distinct synthetic scenarios suggests that they are applicable across variable dose-finding settings. However, they could be further tuned for improved performance for specific settings. For example, the length scale parameter can be informed by the range of investigated dose values, which will be known ahead of a trial.

While Theorem 4.1 assumes the use of Laplace approximation in order to approximate our GP posteriors as Gaussians, in practice we train our MOGPs using stochastic variational inference (Hensman et al., 2014) due to ease of implementation in GPyTorch (Gardner et al., 2018). We obtain confidence intervals by drawing 1000 samples from the GP posteriors and calculating the 0.025 quantile (lower bound) and 0.975 quantile (upper bound). Future work can explore the impact that different methods of approximate Bayesian inference, such as Laplace approximation, Expectation Propagation, or variational inference, have on performance.

D.2. Baseline algorithms

In our experiments (Section 5.1), we assess the performance of SAFE-T as compared to the standard dose-finding trial procedures of $3 + 3$ (Kurzrock et al., 2021) and the continual reassessment method (CRM) (Wheeler et al., 2019), as well as the C3T algorithm proposed in (Lee et al., 2020). Note that every metric for every experiment for every method is the average over 100 experiment repetitions.

3+3 (Kurzrock et al., 2021) is a rule-based methodology for dose escalation. A cohort c_1 of three participants are allocated to a dose d_1. If one participant in c_1 experiences toxicity, the next three participants at c_2 are allocated the same dose d_1. If more than two out of six participants across c_1 and c_2 experience toxic outcomes, d_1 is considered too toxic. If no participants in c_1 experience a (previously defined) toxic outcome, the trial allocated the next highest dose d_2 to another three participant cohort, c_2. The pattern repeats here; if any of c_2 experiences a toxic outcome, c_3 is also allocated d_2. If more than two participants in c_2 and c_3 experience a toxic outcome, the next cohort c_4 is allocated a lower dose, d_1. In short, the dose where no more than one participant out of six experiences toxicity is considered the maximum tolerated dose (MTD), which is then recommended for future study. This method relies on the assumption that both toxicity and efficacy are monotonically increasing, such that the maximum tolerated dose is also the most effective. However, this assumption may not hold for novel therapies, where efficacy may plateau or decrease with higher dosages (Wages et al., 2018; Zhang et al., 2006).

CRM (Wheeler et al., 2019) is a model-based design for dose-finding trials. In CRM, practitioners learn a model for the dose-toxicity relationship. Typically, a parameterized logistic model is used, such as the model proposed in O'Quigley et al. (1990). This model is used to determine dose allocation for the next patient (or cohort of patients) as well as the MTD to recommend for future study. CRM requires a pre-specified dose skeleton of expected toxicity probabilities. A dose skeleton, as described in (Wheeler et al., 2019), is used to determine dose labels (the input to the parametric dose-toxicity model) that will fit the domain of the parametric dose-toxicity model well. Typically, dose skeletons are defined manually by trial practitioners based on prior knowledge of dose-toxicity. A comprehensive guide to the implementation of CRM for dose-finding trials is available at Wheeler et al. (2019). Note that in standard CRM trials, participant efficacy outcomes are not considered (neither in the modelling process nor the dose allocation decisions) and one model of dose-toxicity is learned for all participants.

In our implementation of CRM, we use the O'Quigley et al. (1990) logistic model, which maintains one parameter. Our model is implemented with the Python probabilistic programming framework, PyMC (Salvatier et al., 2016) and updated via Bayesian inference (with posterior estimated from NUTS sampling (Hoffman & Gelman, 2011)) after observing the outcomes of each cohort. We include common safety and logistic constraints in our implementation of CRM: escalation is

restricted to only the next highest dose (this is also done in SAFE-T) and participants are treated in cohorts of size 3. We simulate dose skeletons by adding random noise to the the true toxicity probabilities of each scenario (mimicking possible human error).

C3T (Lee et al., 2020) is a multi-armed bandit method where a parametric model of dose-toxicity and empirical efficacy estimates are used to allocate doses by subgroup. At each timestep, C3T allocates the dose with the highest empirical efficacy estimate from the set of doses with estimated toxicities (based on the parametric model) lower than the toxicity threshold. C3T treats participants from different subgroups separately; models of dose-toxicity are learned by subgroup and empirical efficacy estimates are maintained by subgroup. At the end of the trial, C3T selects the dose with the highest estimated efficacy from the set of safe doses, by subgroup.

As in CRM, C3T learns the dose-toxicity relationship using a parametric model, namely the logistic model proposed by O'Quigley et al. (1990). Their methods, safety bounds, and regret bounds depend on the use of this particular model. As such, C3T also requires a pre-specified dose skeleton of expected toxicity probabilities. While the experiments in (Lee et al., 2020) use the underlying true toxicity probabilities of their synthetic dose-finding scenarios to determine a dose skeleton, we include random noise in our dose skeleton priors, as prior toxicity estimates are unlikely to be perfectly accurate in practice (as we do in our implementation of CRM experiments).

A notable aspect of C3T is that it allows for patient skipping if the patient budget, B, is less than the maximum number of timesteps, N. We do not include this aspect in our comparison (we set $B = N$), as the difficulties of recruiting for dose-finding trials and ethical considerations in subject selection indicate that skipping patients would be unlikely in a realistic setting (Brøgger-Mikkelsen et al., 2020). To mimic a realistic setting (Wheeler et al., 2019), for our experiments of SAFE-T patients arrive in cohorts of size 3 and models are updated following outcomes observed from all cohort patients. However, the use of cohorts is not discussed in (Lee et al., 2020) so participants are allocated doses one at a time. C3T also utilizes an unsafe burn-in period (as is common for multi-armed bandit methods) where each dose is selected in immediate sequence (without regard for toxicity). Although this would likely not be possible in a practical setting, we have included it in the implementation of the algorithm to best reflect the reported algorithm. Note that no early stopping rule is defined for C3T.

E. Additional Experimental Results

E.1. Comparison of optimization approaches

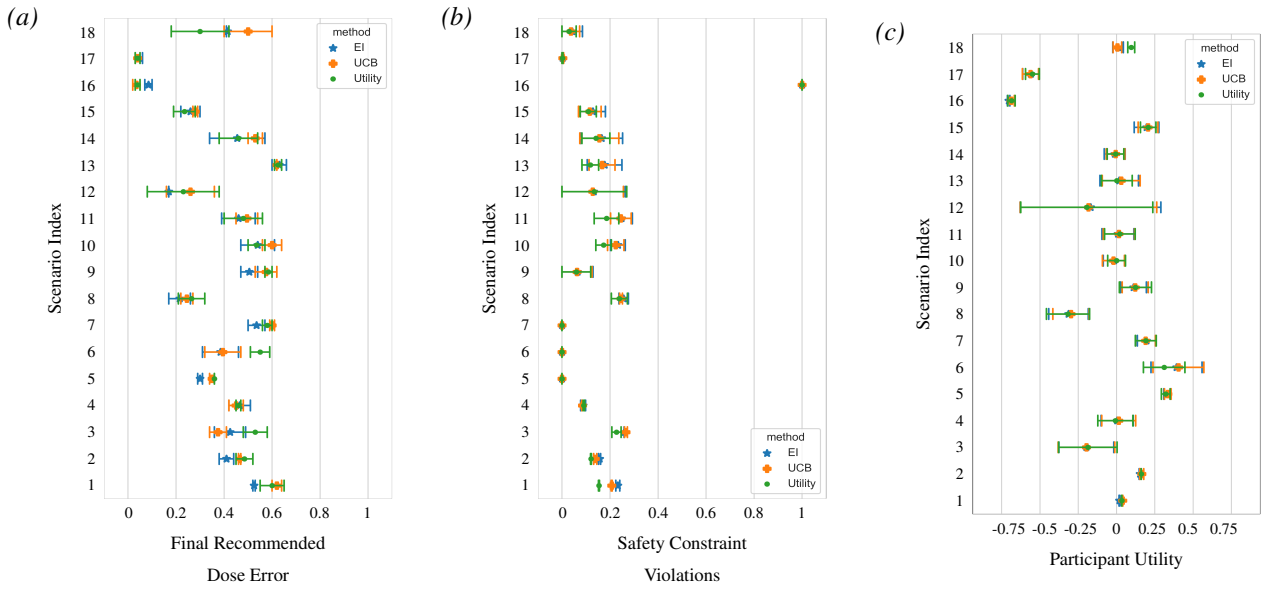


Figure 6: We compare the performance of SAFE-T across scenarios given different optimization techniques: EI (blue star), UCB (orange plus), and utility (green dot).

As discussed in Section 3, the efficacy optimization stage of SAFE-T can employ various optimization approaches. In all other experiments, we report the results of SAFE-T implemented with optimization via expected improvement (EI) on the dose-efficacy function. We also propose the use of the upper confidence bound (UCB) on efficacy or estimated Thall utility as possible acquisition functions.

In Figure 6, we compare the performance of SAFE-T optimized with EI, UCB, or estimated utility on our set of synthetic dose-finding scenarios. Performance is comparable across methods, with very small differences. With respect to final recommended dose error, EI appears to generally perform slightly better and with smaller disparities across subgroups. The utility method results in the best overall participant safety.

E.2. Contribution of SAFE-T components

SAFE-T consists of multiple components that work together to address the complex problem constraints and objectives of dose-finding. Namely, we use multi-output Gaussian processes (MOGP) to address participant heterogeneity in small sample sizes, restrict our dose set based on a safety constraint to improve participant safety, include a safe exploration stage to encourage allocation to new doses while maintaining safety to improve learning, optimize over efficacy to improve participant outcomes, and propose the use of a utility-based final dose recommendation method to improve recommendation accuracy. In Table 2, we summarize the components of SAFE-T alongside comparison methods. In each of these comparison methods, one crucial component of SAFE-T is removed so that we can assess its contribution. For example, the MTD method is SAFE-T without the utility-based final dose recommendation rule, and the No-Exp method is SAFE-T without the safe expansion stage (instead, it immediately begins optimizing over efficacy). We compare other possible optimization methods in Appendix E.1 and so exclude them from this comparison. For clarity, we compare groupings of the comparison algorithms in separate figures.

Table 2: Comparing components of SAFE-T

Name	Subgroup Modelling	Exploration Stage	Optimization Stage	Final Dose Selection
SAFE-T	MOGP	Yes	EI over safe set	Max utility of safe set
MTD	MOGP	Yes	EI over safe set	MTD
No-Exp	MOGP	None	EI over safe set	Max utility of safe set
Unconstrained	MOGP	None	EI	Max utility
Sep	Separate GPs	Yes	EI over safe set	Max utility of safe set
One	One GP	Yes	EI over safe set	Max utility of safe set

In Figure 7, we compare SAFE-T to MTD, where the only change is that the maximum tolerated dose (MTD) is selected as the final recommended dose. This means that we recommend the highest dose that satisfies the safety constraint, in contrast to SAFE-T, where we recommend the dose with the highest Thall utility that satisfies the safety constraint. Performance in safety constraint violations and participant utility are approximately the same as in SAFE-T, which is expected as nothing during the dose allocation process has changed. However, significant differences are seen with respect to final recommended dose error. We expect MTD to perform sub-optimally as a recommendation rule because it does not account for possibilities of plateauing or parabolic dose-efficacy response. And as expected, SAFE-T significantly outperforms MTD in the majority of scenarios. The exceptions are scenario 1, 10, 13 where MTD is slightly more accurate (the largest gap is $\tilde{0}.15$; these are the only scenarios where the optimal dose is the maximum safe dose for both subgroups and the efficacy-response curve is increasing steadily (without flattening or plateauing).

In Figure 8, we compare SAFE-T to No-Exp and Unconstrained. In No-Exp, the safe exploration stage is removed. Final recommended dose error is higher, particularly in scenarios where dose-efficacy response differs by subgroup. No-Exp is slightly safer (resulting in slightly higher utility as well), which makes intuitive sense; the safe exploration stage encourages allocation to higher or uncertain doses in order to improve final dose recommendation accuracy, sacrificing some safety (and thus utility) in the process. In the Unconstrained method, where the safety constraint is completely removed, we see significantly worse performance and larger subgroup disparities across all metrics. Note that in the Unconstrained method, we still maintain the realistic restriction that only the next highest dose can be assigned in the next timestep.

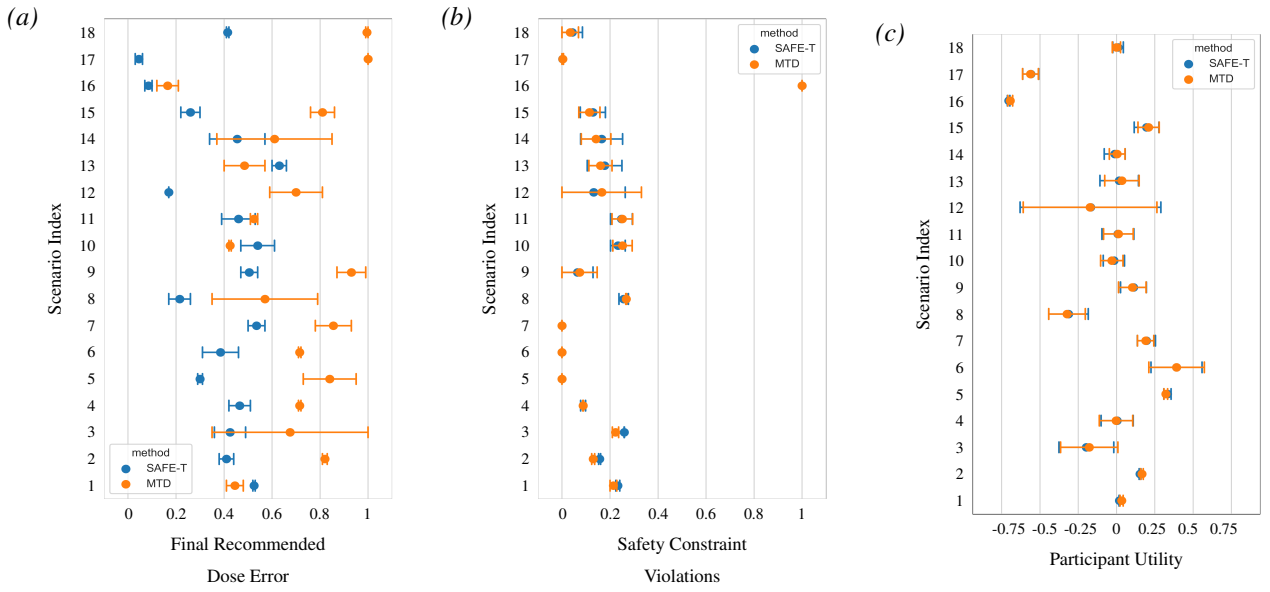


Figure 7: We compare the performance of SAFE-T (blue) and MTD (orange) across scenarios.

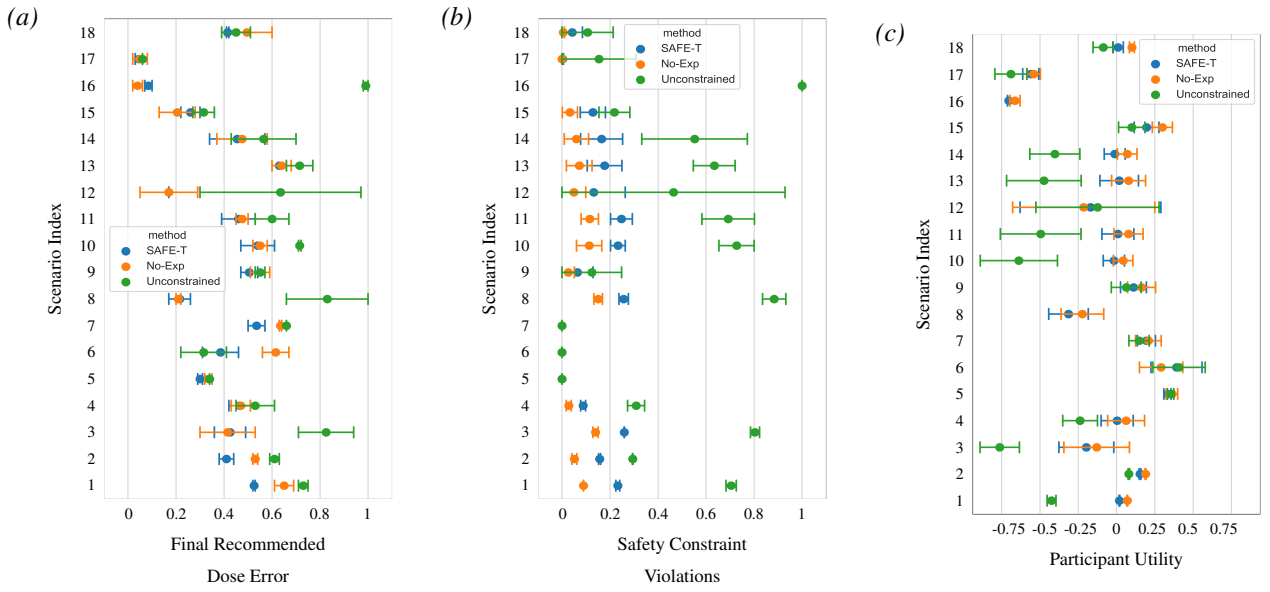


Figure 8: We compare the performance of SAFE-T (blue) across scenarios with No-Exp (orange) and Unconstrained (green).

In Figure 9, we compare SAFE-T to Sep and One, which model dose-response relationships differently. In Sep, we use a separate GPs to model the dose-response relationships of each subgroup. In One, we use one GP to model the dose-response relationships of the entire population, regardless of subgroup. The performance of Sep is largely comparable to the performance of SAFE-T, with SAFE-T resulting in slightly lower final recommended dose error in scenarios where dose-toxicity differs between subgroups. Future work should more thoroughly investigate settings that differentiate the performance of SAFE-T and Sep, identifying where the use of MOGPs is helpful. As expected, the performance of One is mostly inferior across all metrics and also results in larger subgroup disparities, particularly in safety.

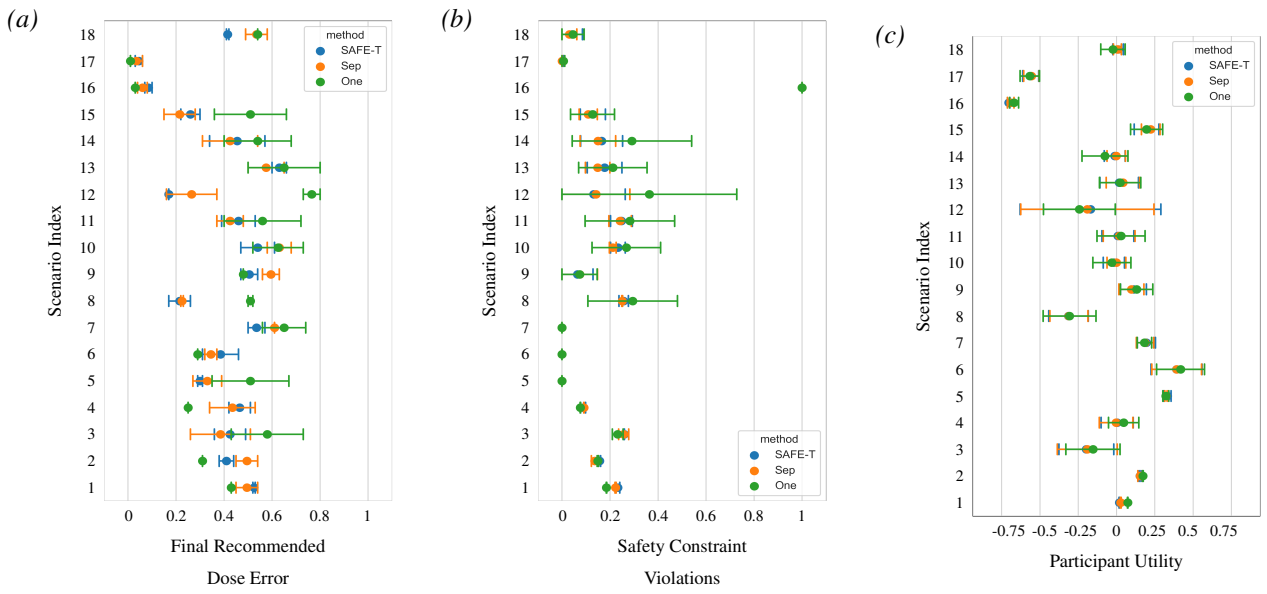
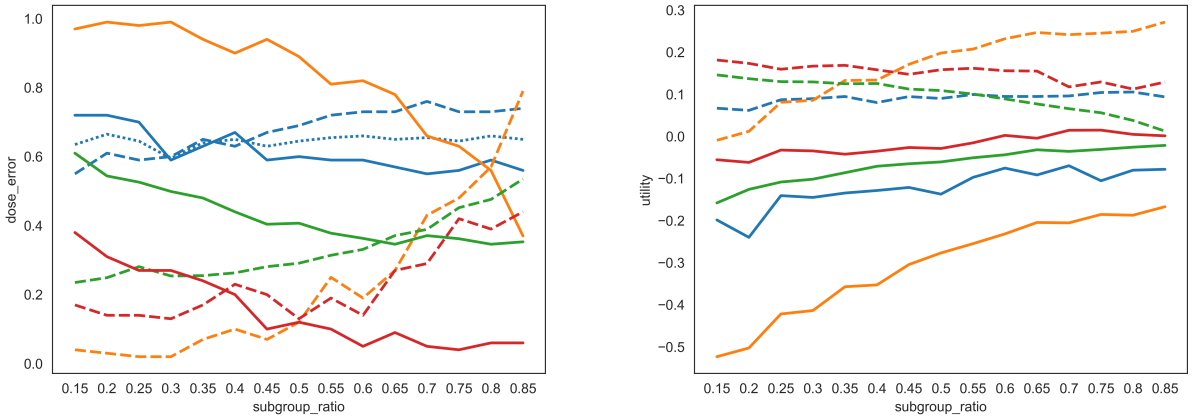


Figure 9: We compare the performance of SAFE-T (blue) with Sep (orange) and One (green).

E.3. Further results on impact of subgroup ratios

In Figure 10, we show the result for the impact of changing subgroup ratios on (a) the final recommended dose error and (b) average participant utility.

This experiment uses scenario 11, where subgroup 0 experiences higher toxicities at each dose and both subgroups experience the same, plateauing efficacy-response. Across all ratios, we see that SAFE-T maintains a comparatively low recommended dose error, although disparities do increase as subgroup become more imbalanced. $3 + 3$ does not fluctuate much, but the dose error is high throughout. CRM exhibits an extremely large disparity with differing ratios, indicating its inability to handle heterogeneity. C3T results in a generally higher final recommended dose error with disparities at imbalanced ratios. In terms of utility, CRM exhibits improvement as the ratio of subgroup 0 increases. Recall that in CRM, we maintain only a dose-toxicity model which is used to allocate doses; because subgroup 0 experiences higher toxicity, when CRM is able to learn a more accurate dose-toxicity model for it, it is able to allocate doses at higher utility. For all other models, utility remains generally consistent, with SAFE-T resulting in consistently higher utility for both subgroups over all ratios.



(a) Final Recommended Dose Error

(b) Utility

Figure 10: Additional results for subgroup ratios experiment (Section 5.2). Algorithms results are differentiated by color: $3+3$ (blue), CRM (orange), C3T (green), SAFE-T (red) and by line style: subgroup 0 (solid) and subgroup 1 (dashed). The ratio shown refers to subgroup 0; for example, at ratio 0.15, subgroup 0 has a 15% probability of arrival.

In Figure 11, we conduct the same experiment with a smaller sample size, where $N = 99$. We do not run the experiment on CRM due to computational intensiveness (we see its poor performance clearly with the larger sample size experiment already as well). We see similar behavior of methods; 3+3 and C3T both exhibit large differences in safety violations, with the performance of C3T varying widely based on subgroup ratio. However, SAFE-T maintains low and consistent level of safety violations across subgroups.

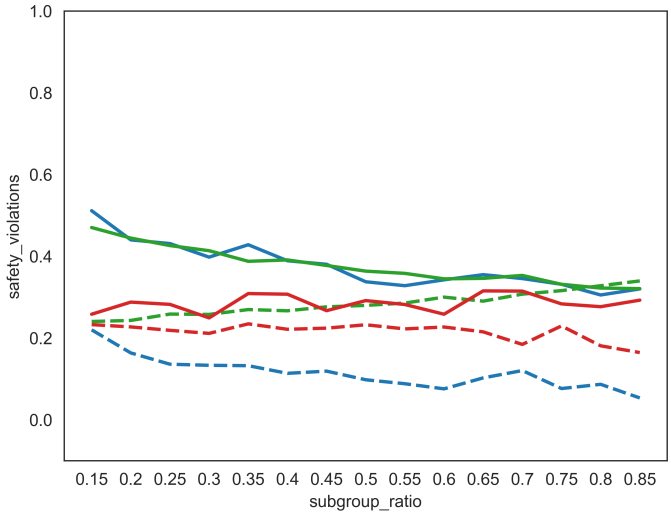
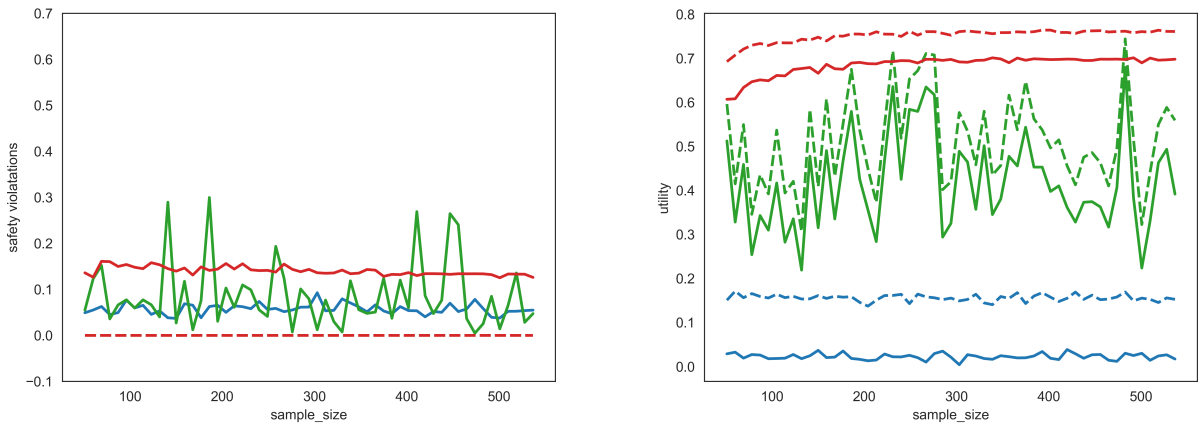


Figure 11: Safety violations in subgroup ratio experiment with smaller sample size ($N = 99$). SAFE-T (red) maintains a consistently low rate of safety violations across subgroups, while 3+3 (blue) and C3T (green) exhibit disparities.

E.4. Further results on impact of sample size

In Figure 12, we show the results for the impact of samples size on (a) rate of safety constraint violations, and (b) average participant utility. The results are relatively uninformative, showing that $3+3$, C3T, and SAFE-T all maintain similar levels

of safety and participant utility regardless of sample size. We do note that while we report the averaged results over 100 trials at each sample size for each algorithm, the performance of C3T fluctuates quite notably.



(a) Safety Constraint Violations

(b) Utility

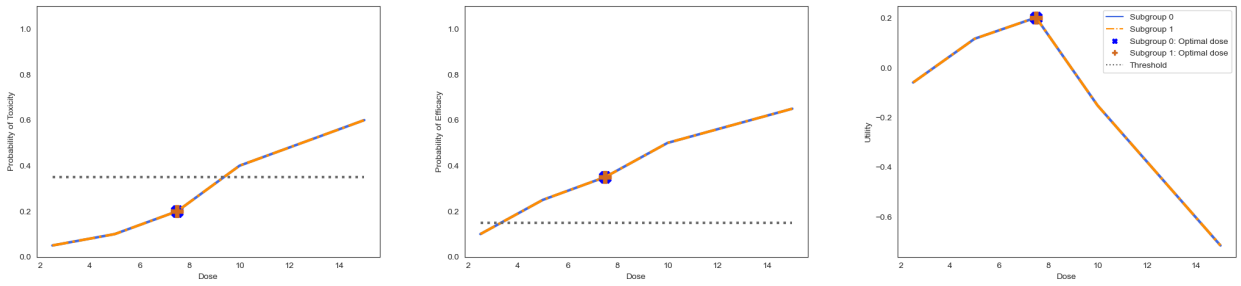
Figure 12: Additional results for sample size experiment (Section 5.3). Algorithms results are differentiated by color: $3 + 3$ (blue), C3T (green), SAFE-T (red) and by line style: subgroup 0 (solid) and subgroup 1 (dashed). All algorithms obtain 0 safety constraint violations for subgroup 1 regardless of sample size, as all doses are safe for subgroup 1.

Table 3: Description of dose-finding scenario characteristics.

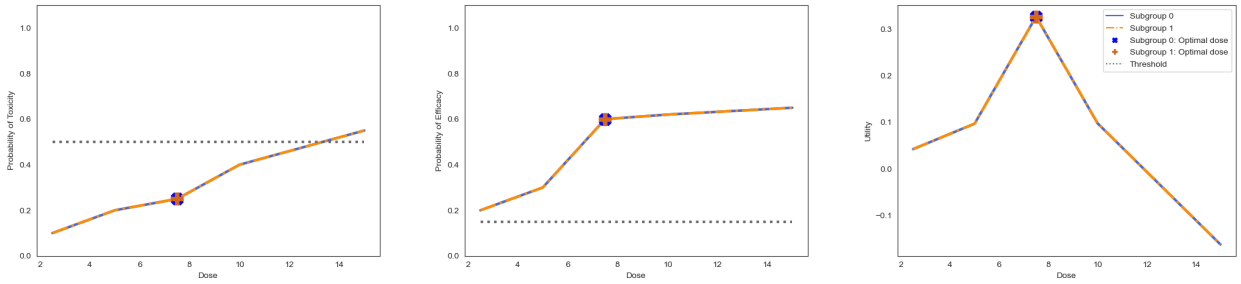
Scn	Toxicity	Efficacy	Subgroup Variations	Case Study
1	Increasing	Increasing	None	Basic case: tox/eff both increasing, subgroups identical.
2	Increasing	Plateauing	None	Basic case: tox increasing, eff plateauing, subgroups identical
3	Increasing (Logarithmic)	Increasing (Logarithmic)	Efficacy: vertical transformation	Subgroup 1 experiences lower efficacy at the same dose, resulting in no optimal dose.
4	Increasing (Exponential)	Plateauing	Efficacy: vertical transformation	Subgroup 0 experiences higher efficacy at the same dose
5	Increasing (Exponential)	Parabolic	Efficacy: horizontal transformation	Differing optimal dose for parabolic efficacy curves
6	Increasing	Increasing	Efficacy: combined transformation	Very low toxicities, optimal dose is highest dose
7	Increasing	Plateauing	Efficacy: horizontal transformation	Subgroup 1 plateaus at lower dose
8	Plateauing	Increasing (Logarithmic)	Toxicity: vertical transformation	Subgroup 1 experiences higher toxicity at same dose, resulting in no optimal dose.
9	Increasing (Exponential)	Plateauing	Toxicity: vertical transformation	Subgroup 0 experiences higher toxicity at same dose
10	Increasing (Logarithmic)	Increasing	Toxicity: combined transformation	Subgroup 0 experiences higher toxicity at same dose
11	Increasing (Logarithmic)	Plateauing	Toxicity: combined transformation	Subgroup 0 experiences higher toxicity at same dose
12	Increasing (Exponential)	Parabolic	Toxicity: vertical transformation	Subgroup 0: all doses are safe, subgroup 1: no doses are safe (no optimal)
13	Increasing (Exponential)	Increasing	Both: combined transformation	Basic case with inconsistent subgroups shifts
14	Increasing	Plateauing	Both: combined transformation	Due to high toxicity, subgroup 1's optimal dose is not its plateau inflection point
15	Increasing	Parabolic	Both: combined transformation	Efficacy peaks at different doses by subgroup
16	Increasing (Logarithmic)	Increasing (Logarithmic)	Both: combined transformation	All doses are toxic - no optimal doses. Early stopping expected.
17	Increasing	Increasing	Both: combined transformation	No doses are effective - no optimal doses.
18	Increasing	Flat	Both: combined transformation	Efficacy curve is completely flat while toxicity increases; therefore the lowest dose is optimal.

Figure 13: 18 synthetic dose-finding scenarios with discrete doses. Leftmost column depicts dose-toxicity, middle columns depicts dose-efficacy, and the rightmost column depicts dose-utility.

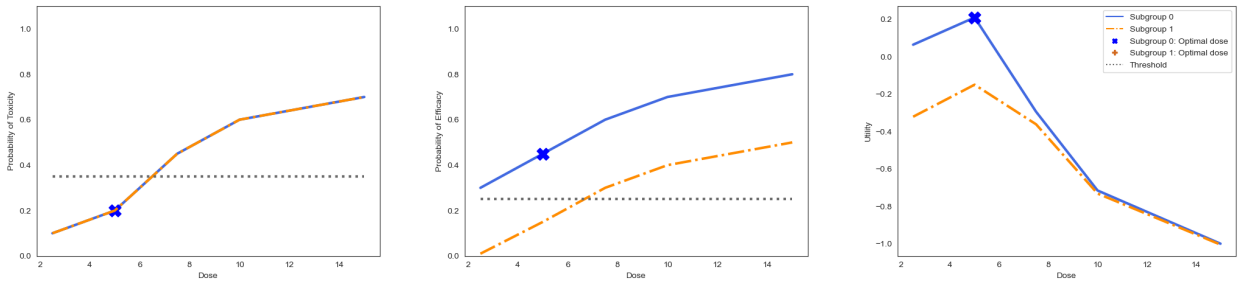
(a) Scenario 1



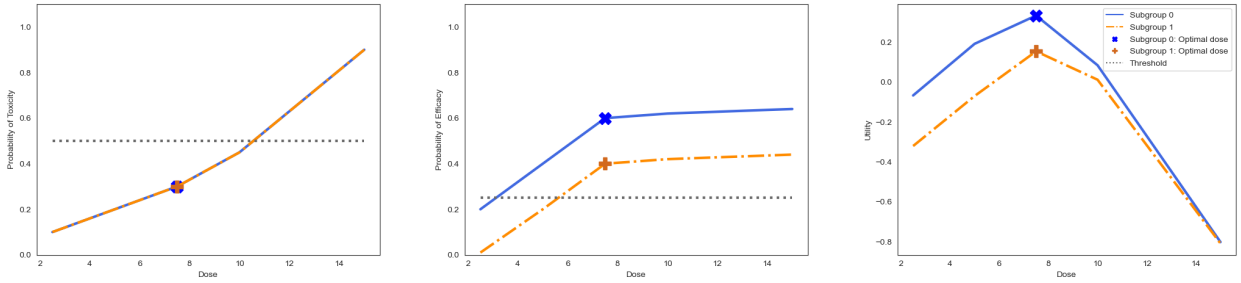
(b) Scenario 2



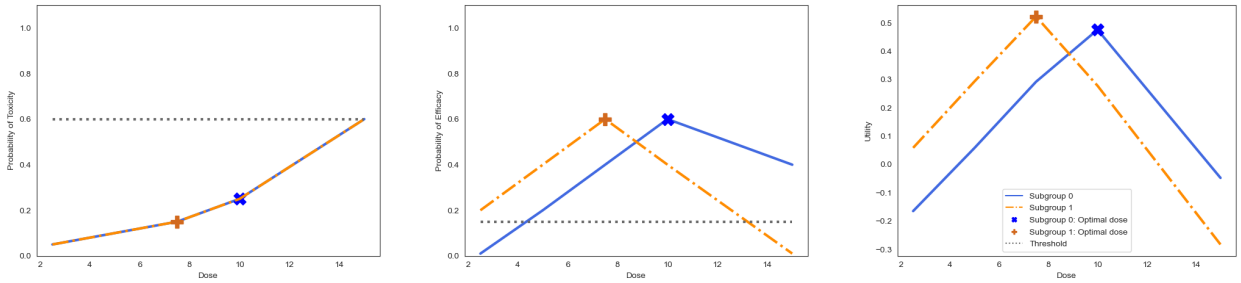
(c) Scenario 3



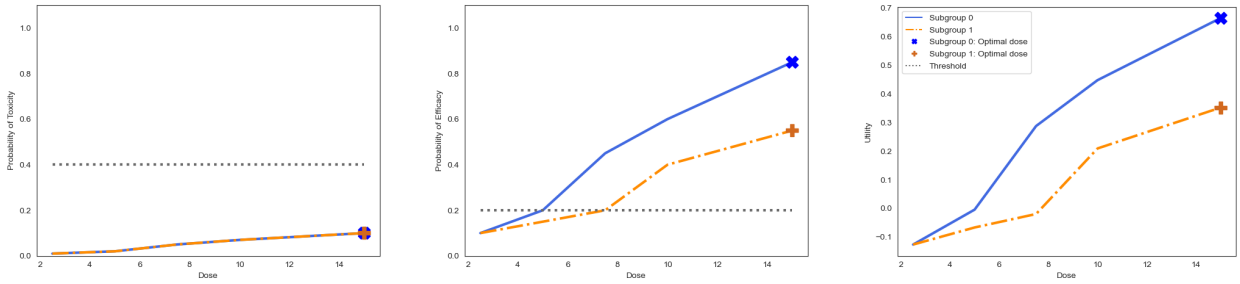
(d) Scenario 4



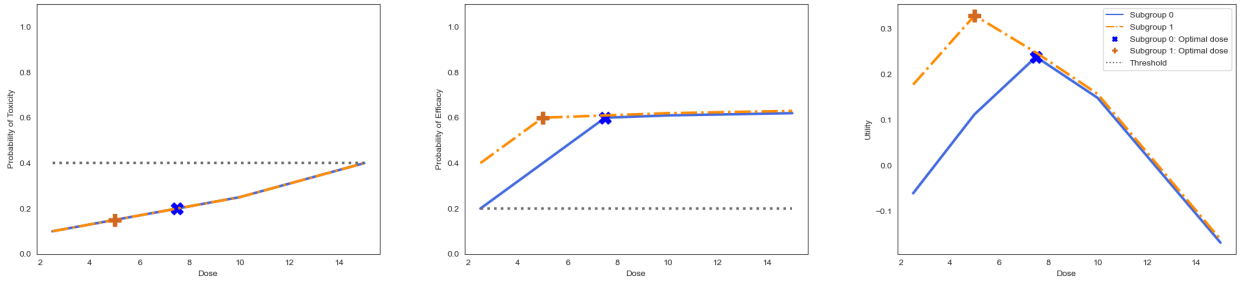
(e) Scenario 5



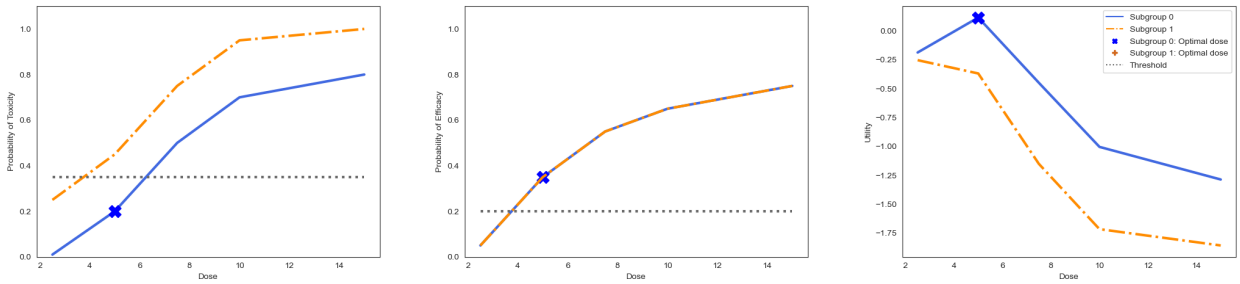
(f) Scenario 6



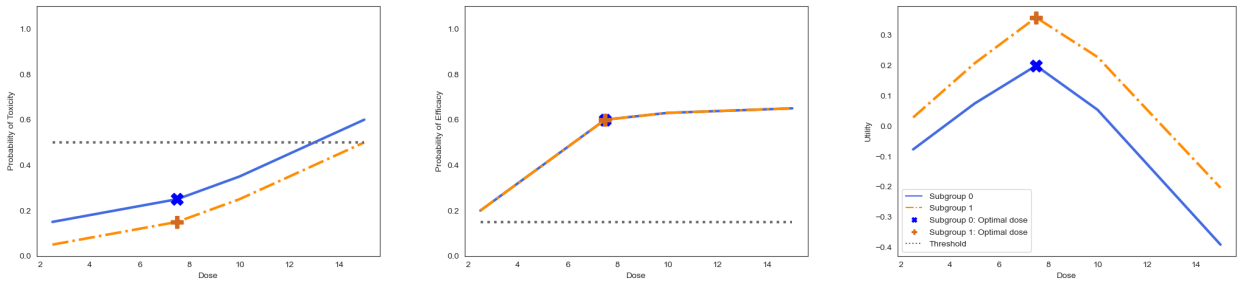
(g) Scenario 7



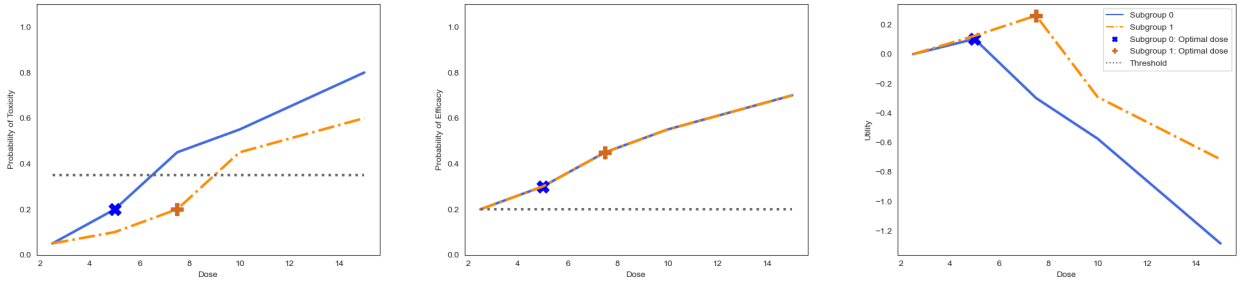
(h) Scenario 8



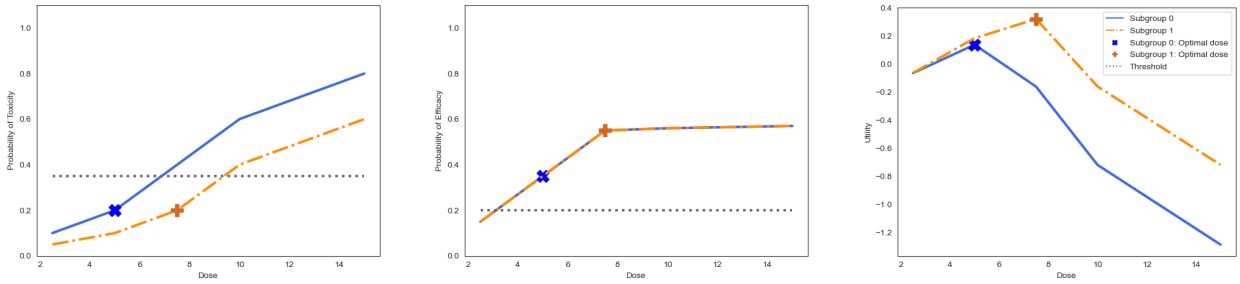
(i) Scenario 9



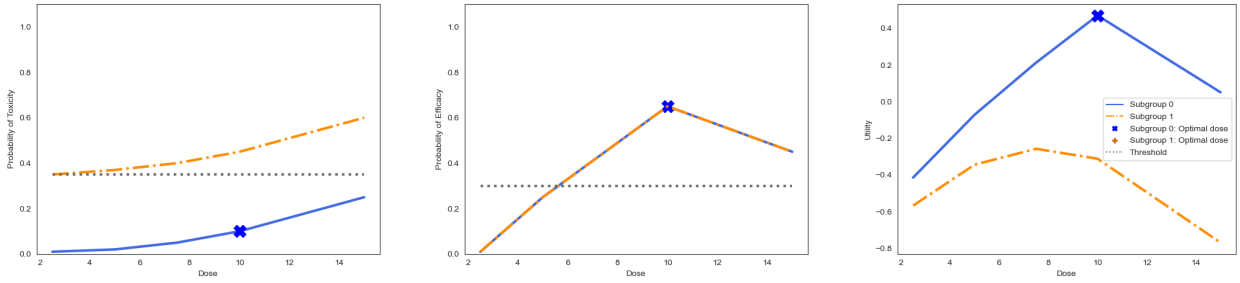
(j) Scenario 10



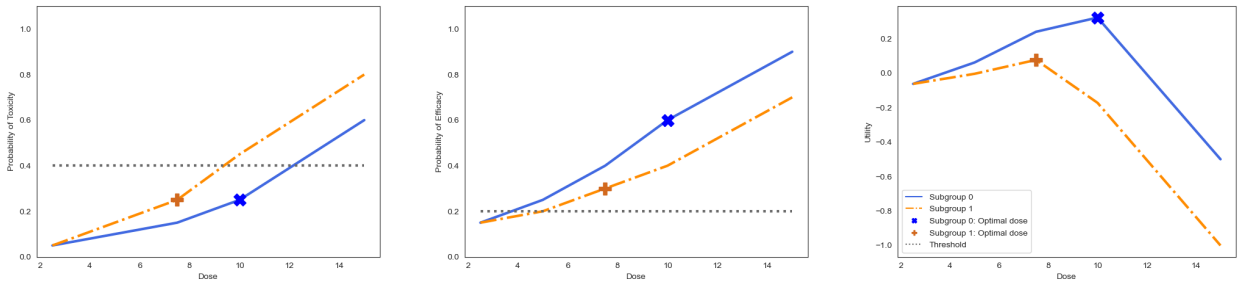
(k) Scenario 11



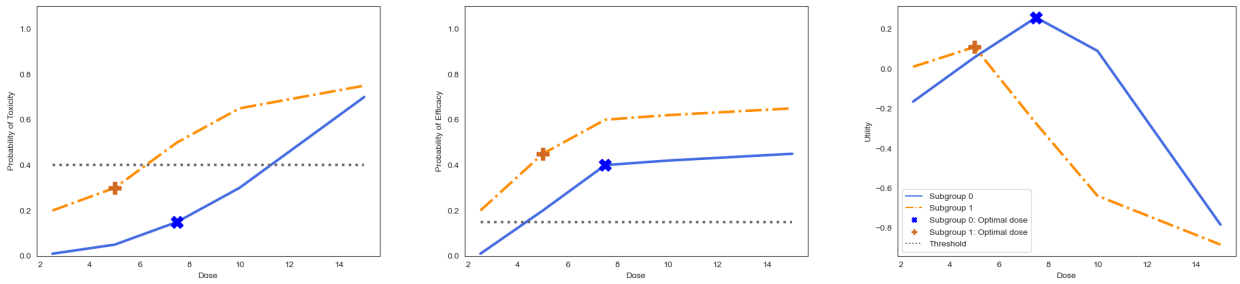
(l) Scenario 12



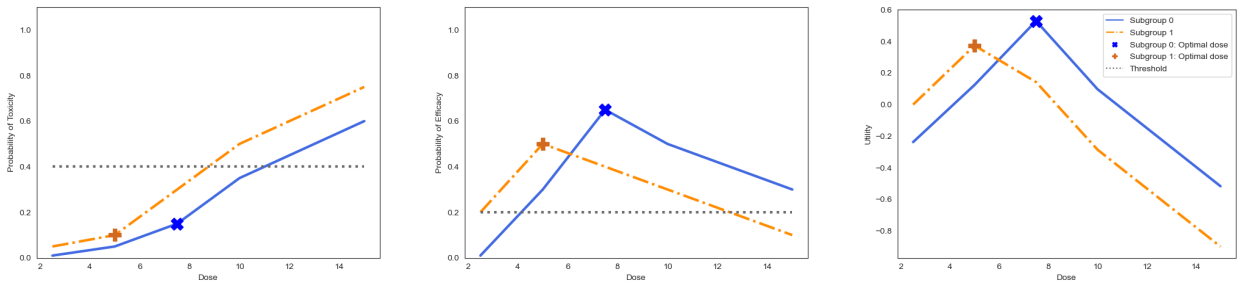
(m) Scenario 13



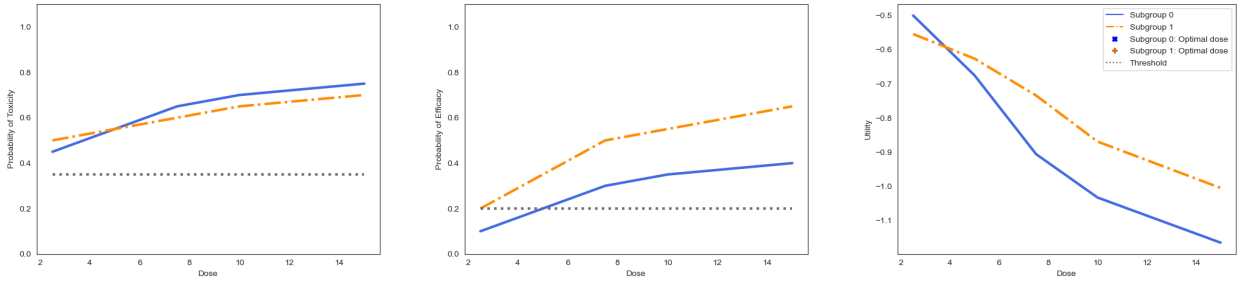
(n) Scenario 14



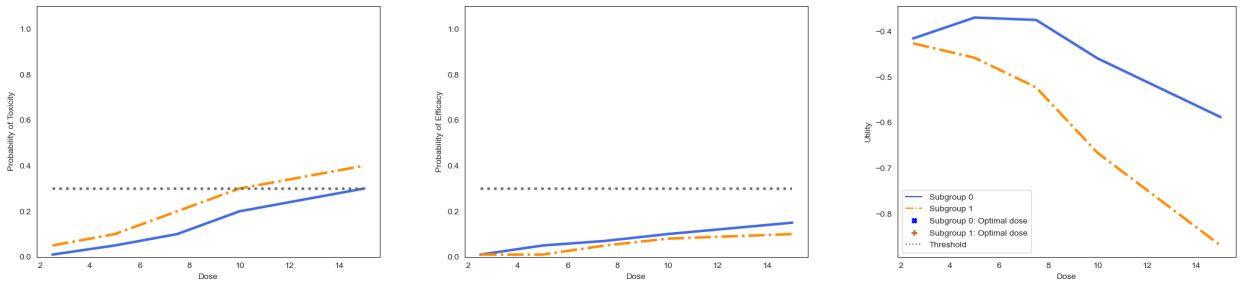
(o) Scenario 15



(p) Scenario 16



(q) Scenario 17



(r) Scenario 18

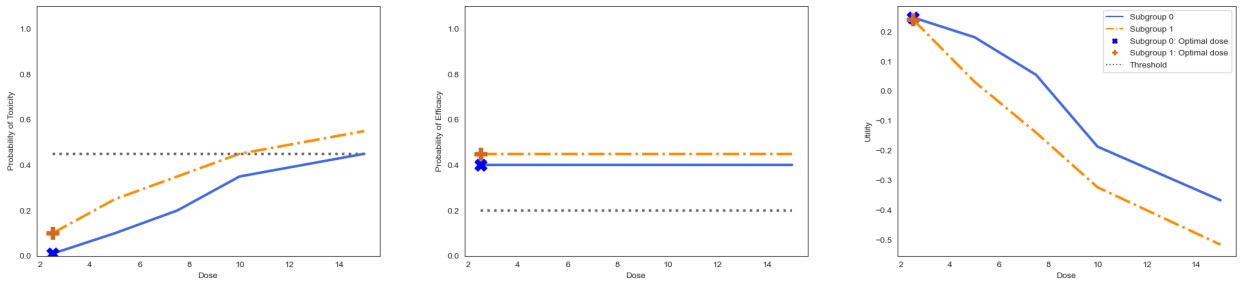
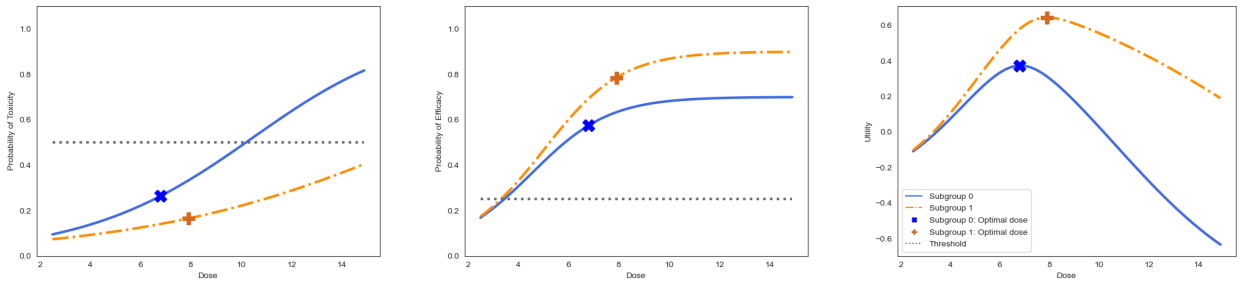
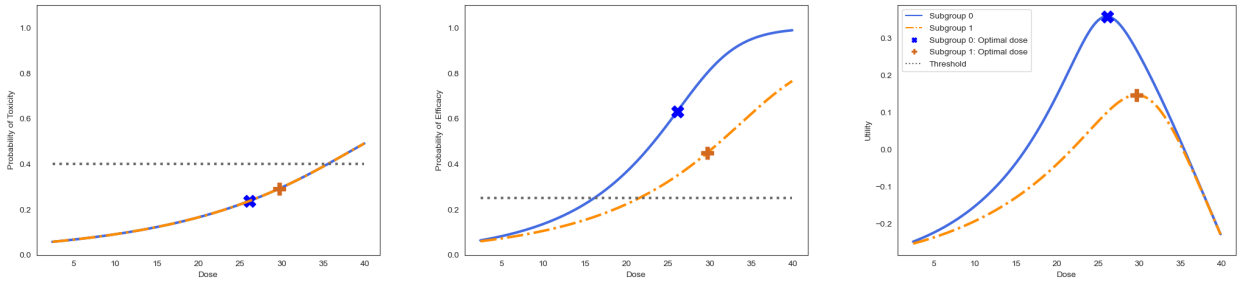


Figure 14: 4 synthetic dose-finding scenarios over a continuous dose range.

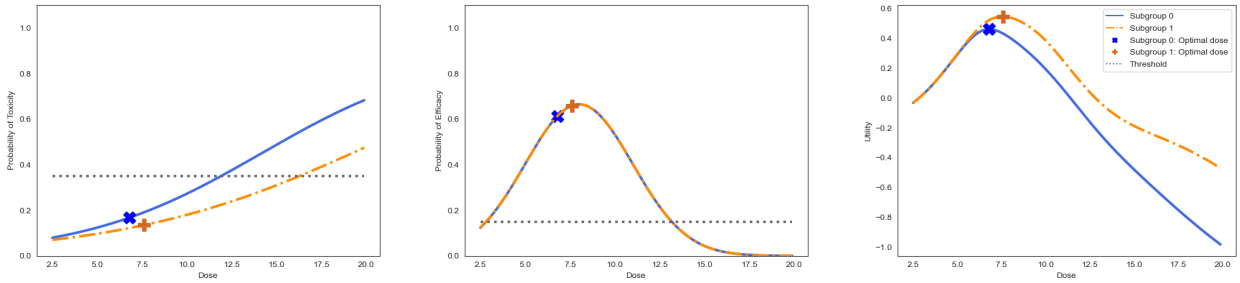
(a) Scenario 1



(b) Scenario 2



(c) Scenario 3



(d) Scenario 4