Learning for Dose Allocation in Adaptive Clinical Trials with Safety Constraints

Cong Shen [1]   Zhiyang Wang [2]   Sofía S. Villar [3]   Mihaela van der Schaar [3,4]

Abstract

Phase I dose-finding trials are increasingly challenging as the relationship between efficacy and toxicity of new compounds (or combination of them) becomes more complex. Despite this, most commonly used methods in practice focus on identifying a Maximum Tolerated Dose (MTD) by learning only from toxicity events. We present a novel adaptive clinical trial methodology, called Safe Efficacy Exploration Dose Allocation (SEEDA), that aims at maximizing the cumulative efficacies while satisfying the toxicity safety constraint with high probability. We evaluate performance objectives that have operational meanings in practical clinical trials, including cumulative efficacy, recommendation/allocation success probabilities, toxicity violation probability, and sample efficiency. An extended SEEDA-Plateau algorithm that is tailored for the increase-then-plateau efficacy behavior of molecularly targeted agents (MTA) is also presented. Through numerical experiments using both synthetic and real-world datasets, we show that SEEDA outperforms state-of-the-art clinical trial designs by finding the optimal dose with higher success rate and fewer patients.

1. Introduction

An adaptive clinical trial utilizes the accumulated results to dynamically modify its future trajectory for better efficiency and ethics, while preserving the integrity and validity of the study. Studies such as the phase I trial in Acute Myeloid Leukaemia in (Yap et al., 2013) and Cancer Research UK study CR0720-11 in (Whitehead et al., 2012) have suggested that even some simple forms of adaptive design lead to better usage of resources and require fewer participants. These promising results have spawned the interest in developing adaptive clinical trial methodologies in recent years (Villar et al., 2015a; Pallmann et al., 2018; Atan et al., 2019; Lee et al., 2020), which is of great importance because running an actual clinical trial on human subjects is expensive and ethically sensitive. A well-designed trial methodology with thorough theoretical and simulated investigation is widely acknowledged as a crucial first step.

Traditionally, the goal of phase I clinical trials is to identify the Maximum Tolerated Dose (MTD) of a cytotoxic (CTX) or therapeutic agent, which is then used for subsequent studies (Storer, 1989). However, modern cancer phase I trials test antineoplastic agents in patients with advanced cancer stages, who have often exhausted all other available treatment options (Roberts et al., 2004). These participants usually expect therapeutic benefit from participating in the trial, which has motivated the trial design to *include efficacy as a co-primary end point of phase I dose-finding studies* (Yan et al., 2017; Paoletti & Postel-Vinay, 2018). In addition, the monotonic assumption for the dose-efficacy relationship is widely adopted in state of the art designs, which is reasonable for cytotoxic agents but may not apply to the new molecularly targeted agents (MTA) such as monoclonal antibodies (see (Postel-Vinay et al., 2009) for an exemplary trial that illustrates this issue). Designing adaptive clinical trials that can properly address the intrinsic conflict between learning and treatment effectiveness for general dose-response models has become an important task for phase I clinical trials.

In addition to the well-known 3+3 design (Storer, 1989) and continual reassessment method (CRM) (O'Quigley et al., 1990) (and its many variants), Bayesian approaches such as Thompson Sampling (TS) (Aziz et al., 2019) and Gittins index (Villar et al., 2015a;b) have been proposed in the literature for dose-finding studies. However, these methods were originally designed for simplified models that do not capture some of the unique characteristics of clinical trials, often leading to lack of randomization (Villar et al., 2015b), inefficient use of side information (Villar & Rosenberger, 2018), and reduced power levels and estimation issues. Notably, for cases where the best dose for combination therapies

[1]University of Virginia, USA [2]University of Pennsylvania, USA [3]University of Cambridge, United Kingdom [4]University of California, Los Angeles, USA. Correspondence to: Cong Shen <cong@virginia.edu>.

Table 1: Representative adaptive clinical trial studies

Study	Treatment	Category	Methodology	Evaluation
(Tighiouart et al., 2014)	Veliparib	CTX	EWOC-PH	simulated trial
(Whitehead et al., 2012)	MK-0752	CTX	joint phase I and II design	simulated trial
(Lee et al., 2017)	Erlotinib	MTA	extended TITE-CRM	simulated trial
(Thiessen et al., 2010)	Lapatinib	MTA	escalation to DLT	real-world trial data

is to be found, unknown synergistic/antagonist effects are likely to exist and naive designs will fail to identify them. For MTA, the existence of a plateau of efficacy has been discussed in (Zang et al., 2014) and (Riviere et al., 2018), which indicates that the toxicity constraint must be jointly studied with the dose-efficacy relationship for certain new compounds. This is also confirmed by the real-world trial result; see (Tighiouart et al., 2014). Last but not the least, safety constraints such as minimizing the adverse events (AE) (Petroni et al., 2017) have not been properly evaluated with theoretical guarantees. Table 1 summarizes some representative studies in this direction.

In this paper, we address these challenges by developing new dose-finding methods that explicitly impose safety constraints to the allocation and recommendation of dose levels in a phase I clinical trial. Through the lens of multi-armed bandits (MAB), we propose the *Safe Efficacy Exploration Dose Allocation (SEEDA)* algorithm that adaptively updates the admissible set of dose levels satisfying the safety constraints, thus limiting the exploration of doses with harmful effect. Performance analysis for SEEDA is carried out with respect to several measures that have operational meanings in clinical trials, including the probability of safety constraints violation, the average efficacy for patients, and the recommendation and allocation probabilities. Noting that SEEDA only leverages the dose-toxicity logistic model and makes no assumptions on the efficacy, we then show that, by considering the increasing-then-plateau feature of the dose-efficacy relationship for MTA, *SEEDA-Plateau* leads to improved bandit learning performance over SEEDA by leveraging the unimodal structure. Experiments on simulated datasets as well as clinical trials built from real-world datasets show that the proposed methods are capable of finding the optimal dose with higher success rate and fewer patients in most cases, compared to other state-of-the-art designs.

2. Model and Problem Formulation

2.1. The dose-finding model

In a phase I dose-finding clinical trial, a total of K doses are given where the k-th dose is denoted as $d_k \in \mathcal{D}, k \in \mathcal{K} = \{1, 2, ..., K\}$. The performance is characterized by both *efficacy* and *toxicity*. We model the efficacy X and toxicity Y for dose d_k as Bernoulli random variables with unknown

probabilities q_k and p_k, respectively, where $X = 1$ ($X = 0$) indicates that the dose level is effective (not effective), and $Y = 1$ ($Y = 0$) suggests that the dose is harmful (not harmful) to the patient[1].

We consider adaptive clinical trials where information learned from previous trial patients can be used in allocating doses to subsequent patients (Atan et al., 2019; Villar et al., 2015a; Aziz et al., 2019). For the t-th patient, dose $I(t)$ is selected based on a policy that uses past observations, and administrated to the patient. The efficacy outcome X_t and toxicity response Y_t are realized based on their distributions $X_t \sim Ber(q_{I(t)})$ and $Y_t \sim Ber(p_{I(t)})$, and observed by the trialist.

We adopt a well-known dose-toxicity logistic model proposed by in (O'Quigley et al., 1990) to describe the toxicity probability for different dose levels:

$$p_k(a) = \left(\frac{\tanh d_k + 1}{2} \right)^a, \qquad (1)$$

where a is a global parameter for all the dose levels. It can be verified that Eqn. (1) satisfies the assumption that the toxicity monotonically increases with dose d_k. The unsafe dose levels are defined as those whose toxicity probabilities p_k's are above a pre-determined target toxicity probability θ, which is referred as the MTD threshold. Hence the toxicities of all doses can be written as $p_1 \leq p_2 \leq \cdots \leq p_M < \theta < p_{M+1} \leq \cdots p_K$ where the (unknown) M denotes the number of safe doses. The efficacy-dose relationship is not modeled to allow for the development of a general algorithm. The specific increase-then-plateau efficacy behavior of MTA will be exploited in Section 4.

2.2. Problem formulation

Several objectives are often desired for a successful dose-finding study, which are summarized as follows.

- **Successful recommendation.** At the end of the trial (n patients) a *dose recommendation* $\hat{k}_n$ is made, which is desired to match the optimal dose k^* that is the lowest safe dose that achieves the highest efficacy (Zang et al., 2014): $k^* = \min\{k : q_k = \max_{l:l \in \mathcal{K}, p_l \leq \theta} q_l\}$.

- **Effective treatment.** The cumulative treatment for trial

---

[1]This is typically measured by the presence of absence of a dose-limiting toxicity (DLT) reported in a fixed evaluation window after administrating the drug.

participants $\sum_{i=1}^{n} X_t$ is desired to be maximized.

- **Minimal violation of the safety constraint.** There are different formulations for the safety constraint. One is to minimize $\mathbb{E}[\sum_{k \in \mathcal{K}, p_k > \theta} N_k(n)/n]$ where $N_k(t)$ denotes the number of times dose k is allocated to the first t patients. Another formulation is to minimize the probability that the average toxicity exceeds the MTD threshold.

- **Small sample size.** Most phase I trials have a pre-determined n which is decided as the minimum number of trial participants to achieve a pre-defined confidence level of successful recommendation. It is desirable to have a small n for cost and efficiency considerations.

Proposing a learning model that explicitly guarantees all of the above objectives is elusive and non-constructive in developing the dose-allocation policy. We thus formulate dose-finding clinical trials as *an online efficacy learning problem with explicit safety constraint*, and subsequently provide performance analysis on the metrics of interest. Specifically, we aim at maximizing the cumulative efficacy over a finite number of patients n while simultaneously guaranteeing that the average toxicity observed from the n dose allocations is kept under the probability threshold θ with high probability. This can be written as:

$$\text{maximize} \quad \mathbb{E}\left[\sum_{t=1}^{n} X_t\right]$$

$$\text{subject to} \quad \mathbb{P}\left[\frac{1}{n}\sum_{t=1}^{n} Y_t \le \theta\right] \ge 1 - \delta. \quad (2)$$

Essentially, problem formulation (2) focuses on safe exploration among all the dose levels to maximize cumulative efficacies. Clinical trials using a design based on (2) pursue simultaneously both objectives of toxicity and efficacy.

3. The SEEDA Algorithm

3.1. Algorithm description

The proposed Safe Efficacy Exploration Dose Allocation (SEEDA) design is completely described in Algorithm 1. In particular, $\hat{p}_k(t)$ and $\hat{q}_k(t)$ are the estimated toxicity and efficacy, respectively, after administrating the t-th patient. The principle of dose selection is to first dynamically construct the admissible set $\mathcal{D}_1(t)$ using the Upper Confidence Bound (UCB) principle (Auer et al., 2002), where the confidence interval $\alpha(t)$ is constructed as

$$\alpha(t) = \bar{C}_1 K \left(\frac{\log \frac{2K}{\delta}}{2t}\right)^{\frac{\bar{\gamma}_1}{2}}, \quad (3)$$

where $\bar{C}_1$ and $\bar{\gamma}_1$ are algorithm parameters[2]. Note that the admissible set consists of doses that, with high confidence,

---

[2]See Section B in the supplementary material for a discussion on how to select these algorithm parameters.

satisfy the toxicity constraint.

Then, limiting to those in the admissible set $\mathcal{D}_1(t)$, the algorithm again applies the UCB principle (UCB-1 from (Auer et al., 2002)) to select a dose with the largest $F(p, s, n)$ for the efficacy estimate:

$$F(p, s, n) = p + \sqrt{\frac{c \log(n)}{s}}, \quad (4)$$

with c denoting the UCB-1 coefficient. It should be noted that (4) can be replaced by other UCB principles, e.g., KL-UCB (Garivier & Cappè, 2011).

Algorithm 1 The Safe Efficacy Exploration Dose Allocation (SEEDA) Algorithm

Input: $p_k(a)$ for each $k \in \mathcal{K}$; MTD threshold θ; total number of patients n.

Initialize: $N_k(1) = 0, \hat{p}_k(1) = 0, \hat{q}_k(1) = 0, \forall k \in \mathcal{K}$; Sample each dose once and set: $I(t) = t, \hat{q}_{I(t)}(K) = X_t, \hat{p}_{I(t)}(K) = Y_t, N_{I(t)}(K) = 1$, for $t = 1$ to K; $t = K + 1$.

1: **while** $t \le n$ **do**
2:     Compute the estimated parameter: $\hat{a}(t) = \sum_{k=1}^{K} w_k(t-1)\hat{a}_k(t-1)$;
3:     Set the admissible set: $\mathcal{D}_1(t) = \{d_k \in \mathcal{D} : p_k(\hat{a}(t) + \alpha(t)) \le \theta\}$;
4:     Select dose: $I(t) = \arg\max_{d_k \in \mathcal{D}_1(t)} F(\hat{q}_k(t), N_k(t), t)$,;
5:     Observe the revealed outcomes X_t and Y_t;
6:     Update estimations: $\hat{q}_{I(t)}(t) = \frac{\hat{q}_{I(t)}(t-1)N_{I(t)}(t-1)+X_t}{N_{I(t)}(t-1)+1}$, $\hat{p}_{I(t)}(t) = \frac{\hat{p}_{I(t)}(t-1)N_{I(t)}(t-1)+Y_t}{N_{I(t)}(t-1)+1}, N_{I(t)}(t) = N_{I(t)}(t-1)+1$;
7:     Update parameter estimation: $\hat{a}_{I(t)}(t) = \arg\min_{a \in \mathcal{A}} |p_{I(t)}(a) - \hat{p}_{I(t)}(t)|$;
8:     Update weights: $w_k(t) = N_k(t)/t, \forall d_k \in \mathcal{D}$;
9:     $t = t + 1$.
10: **end while**

Output: $\hat{d}(n) = \arg\max_{d_k : p_k(\hat{a}(n)) \le \theta} p_k(\hat{a}(n))$.

3.2. Performance analysis

The SEEDA algorithm is developed with the aim to solve problem (2). It is thus important to analyze (a) whether the cumulative efficacy is maximized, and (b) how often the toxicity constraint is violated. For metric (a), it can be equivalently formulated as regret minimization, i.e., the cumulative efficacy difference between the oracle policy with full information and that of the learning algorithm. Formally, the efficacy regret is defined as

$$R(n) = q^* n - \mathbb{E}\left[\sum_{t=1}^{n} q_{I(t)}\right], \quad (5)$$

where $q^* = q_{k^*}$ denotes the efficacy associated with the optimal dose defined in Section 2.2, and a^* denotes the true parameter in (1). As for metric (b), we need to evaluate

$$e(n) = \mathbb{P}\left[\frac{1}{n}\sum_{t=1}^{n} p_{I(t)}(a^*) > \theta\right],$$

in conjunction with (5), i.e., whether the proposed *SEEDA* algorithm minimizes $R(n)$ and satisfies $e(n) \leq \delta$ at the same time. In addition, other performance measures such as successful recommendation probability and sample efficiency are of practical interest, and we provide theoretical guarantees for them as well. Due to space limitations, all proofs are provided in the supplementary material.

3.2.1. CUMULATIVE EFFICACY

We start the theoretical analysis by showing that for each patient t in SEEDA, the dose levels whose toxicities are below the MTD threshold are included in the admissible set with high probability. This corresponds to the "false alarm" (type I) error event.

Lemma 1 $\mathbb{P}[p_k(\hat{a}(t) + \alpha(t)) > \theta] \leq \delta, \forall p_k(a^*) \leq \theta.$

Next we prove that with sufficient patients, the dose levels exceeding the toxicity threshold are excluded from the admissible set with high probability. This corresponds to the "miss detection" (type II) error event.

Lemma 2 *If* $t > t_1 = \frac{1}{2}\left(\frac{\bar{C}_1 K}{|\Delta - \epsilon|}\right)^{\frac{2}{\bar{\gamma}_1}} \log \frac{2K}{\delta}$, $\Delta = \min_{k \in \mathcal{K}} \Delta_k$, *where* $\Delta_k = |a^* - p_k^{-1}(\theta)|$ *is the gap between* a^* *and the parameter when the toxicity is at* θ, *then:*

$$\mathbb{P}[p_k(\hat{a}(t) + \alpha(t)) \leq \theta] \leq \exp(-2t\epsilon^2), \forall p_k(a^*) > \theta. \quad (6)$$

Combining Lemmas 1 and 2 leads to the main result on cumulative efficacy regret.

Theorem 1 *With* t_1 *defined in Lemma 2, the regret of SEEDA can be upper bounded as:*

$$R(n) \leq \sum_{d_k:p_k(a^*)\leq\theta} \frac{c \log(n)}{q^* - q_k} + \left(n\delta Q + \frac{1}{2}t_1 + \frac{K-M}{2\epsilon^2}\right) \quad (7)$$

where $Q = \max_{i \in \mathcal{K}} |q_i - q_{k^*}|$ *denotes the maximal single-step regret, and* $\epsilon > 0$ *is a constant. Furthermore, if* $\delta = O(\frac{1}{n})$, *we have that* $R(n) \leq O(\log n)$.

Theorem 1 indicates that the efficacy regret is bounded by $O(\log n)$. A closer look at this scaling reveals that it consists of two parts. The first is due to the structureless model for efficacy – we impose no assumption on the efficacy of different dose levels. The second part, which is reflected through t_1, is determined by the structured model for toxicity, which

affects the admissible set. As will be shown in Section 4, with the increase-then-plateau efficacy assumption, the first $\log n$ component can be further improved.

3.2.2. SAFETY CONSTRAINT VIOLATION

We now move on to analyzing the safety constraint violation. The first result is to verify whether the SEEDA algorithm indeed satisfies the safety constraint in problem (2).

Theorem 2 *For any given* n, *the average toxicity observed from the SEEDA algorithm satisfies*

$$\mathbb{P}\left[\frac{1}{n}\sum_{t=1}^{n} p_{I(t)} - \theta \leq C_2 \epsilon^{\gamma_2}\right] \geq 1 - \delta,$$

for arbitrary $\epsilon > 0$. C_2 *and* γ_2 *are problem-dependent parameters defined in Section A of the supplementary material.*

The safety constraint in problem (2) is formulated based on the average toxicity exceeding the MTD threshold. In practice, we are often interested in minimizing the number of patients that have been exposed to unsafe dose levels, $\mathbb{E}[\sum_{k \in \mathcal{K}, p_k > \theta} N_k(n)/n]$. Corollary 1 analyzes this metric.

Corollary 1 *The number of unsafe dose allocations from SEEDA, i.e., the selected dose levels exceed the MTD threshold, can be bounded as:*

$$\mathbb{E}\left[\sum_{d_k:p_k>\theta} N_k(n)\right] \leq t_1 + \frac{K-M}{2\epsilon^2}.$$

Interestingly, Corollary 1 indicates that unsafe dose allocations in SEEDA are upper bounded by a constant, which is linear in the number of unsafe doses $K - M$ regardless of the number of trial participants n.

3.2.3. RECOMMENDATION ACCURACY

Finally, we analyze the recommendation accuracy of SEEDA at the end of the n-th dose allocation.

Corollary 2 *The probability that SEEDA recommends the MTD satisfies:*

$$\mathbb{P}\left[\hat{d}(n) = \arg\max_{d_k:p_k\leq\theta} p_k\right] \geq 1 - \delta_1, \quad (8)$$

where $\delta_1 = 2K \exp\left(-2\left(\frac{\Delta_M}{C_1 K}\right)^{2\gamma_1} n\right)$.

Corollary 2 guarantees the finding of the MTD with high probability. The recommendation error rate decays *exponentially* with the number of trial participants, which is a nice property. It is worth noting that a lower bound of the

minimal number of trial participants for a given accuracy requirement can be inferred from the upper bound of recommendation error rate (8). This is a practically important result, as sample efficiency directly relates to the cost and ethical constraints of a trial. This is further illustrated in the numerical experiments in Section 5.1.3.

4. Extension to the Increase-Then-Plateau Efficacy Model

Algorithm 2 The SEEDA-Plateau Algorithm

Input: $p_k(a)$ for each $k \in \mathcal{K}$; MTD threshold θ; total number of patients n.

Initialize: $N_k(1) = 0, \hat{p}_k(1) = 0, \hat{q}_k(1) = 0, \forall k \in \mathcal{K}$; $L(1) = K; \eta = 2; l_k = 0, \forall k \in \mathcal{K}$; Sample each dose once and set: $I(t) = t, \hat{q}_{I(t)}(K) = X_t, \hat{p}_{I(t)}(K) = Y_t$, $N_{I(t)}(K) = 1$, for $t = 1$ to $K; t = K + 1$.

1: **while** $t \leq n$ **do**
2: Compute the estimated parameter: $\hat{a}(t) = \sum_{k=1}^{K} w_k(t-1)\hat{a}_k(t-1)$;
3: Set the admissible set: $\mathcal{D}_1(t) = \{d_k \in \mathcal{D} : p_k(\hat{a}(t) + \alpha(t)) \leq \theta\}$;
4: Set $L(t) = \arg\max_{d_k \in \mathcal{D}_1(t)} \hat{q}_k(t)$ and increase $l_{L(t)}$ by 1;
5: If $\frac{l_{L(t)} - 1}{\eta + 1} \in \mathbb{N}$, $I(t) = L(t)$; Otherwise $I(t) = \arg \max_{\substack{\{L(t)-1, L(t), L(t)+1\} \\ \bigcap \mathcal{D}_1(t)}} F(\hat{q}_k(t), N_k(t), t)$;
6: Observe the revealed outcomes X_t and Y_t;
7: Update estimations: $\hat{q}_{I(t)}(t) = \frac{\hat{q}_{I(t)}(t-1)N_{I(t)}(t-1)+X_t}{N_{I(t)}(t-1)+1}$, $\hat{p}_{I(t)}(t) = \frac{\hat{p}_{I(t)}(t-1)N_{I(t)}(t-1)+Y_t}{N_{I(t)}(t-1)+1}$, $N_{I(t)}(t) = N_{I(t)}(t-1)+1$;
8: Update parameter estimation: $\hat{a}_{I(t)}(t) = \arg\min |p_{I(t)}(a) - \hat{p}_{I(t)}(t)|$;
9: Update weights: $w_k(t) = N_k(t)/t, \forall d_k \in \mathcal{D}$;
10: $t = t + 1$.
11: **end while**
12: Estimate the turning point of efficacy as:
$$L_1(n) = \min_{k:d_k \in \mathcal{D}_1(n)} \left\{ m \geq k : |\hat{q}_m(n) - \hat{q}_{m+1}(n)| \right.$$
$$\left. \leq \sqrt{\frac{c \log(n)}{N_m(n)}} + \sqrt{\frac{c \log(n)}{N_{m+1}(n)}}, \hat{q}_m(n) \leq \hat{q}_{m+1}(n) \right\},$$
$$L_2(n) = \arg \max_{d_k:p_k(\hat{a}(n)) \leq \theta} p_k(\hat{a}(n)).$$
Output: $\hat{d}(n) = \min\{L_1(n), L_2(n)\}$.

The proposed SEEDA dose allocation policy is general in the sense that no efficacy model is assumed. In practice, however, efficacy often exhibits certain structure which, if utilized correctly, may further improve the performance. For conventional cytotoxic agents, efficacy monotonically

increases with dose levels. The same is not true for MTAs, for which the dose-efficacy curve increases initially and then plateaus after reaching the level of saturation (Zang et al., 2014; Riviere et al., 2018). In this section, we modify the SEEDA algorithm to handle the increase-then-plateau efficacy model, and analyze its performance.

Formally, we introduce the following increase-then-plateau efficacy assumption, which holds for MTA.

Assumption 1 $q_k, k \in \mathcal{K}$ satisfies $q_1 \leq q_2 \leq q_3 \leq \cdots \leq q_N = q_{N+1} = \cdots = q_K$.

The *SEEDA-Plateau* algorithm is given in Algorithm 2. With Assumption 1, the efficacy has an inherent non-decreasing structure. The key idea is to combine the selection rule of OSUB in (Combes & Proutière, 2014) and reform step 1 in Algorithm 1. Note that step 2 calculates $L(t)$ as the estimated dose level with the optimal efficacy and safe toxicity at t. Algorithm 2 not only selects this dose level frequently enough, but also keeps exploring its neighboring dose levels.

We now analyze the regret of SEEDA-Plateau and present the result in Theorem 3. Compared to Theorem 1 for SEEDA without the increase-then-plateau efficacy model, one can see that the first $\log(n)$ coefficient improves from $c \sum_{d_k:p_k(a^*) \leq \theta} (q^* - q_k)^{-1}$ to $c(q^* - q_{N-1})^{-1}$. This gain comes precisely from the increase-then-plateau efficacy model, as the unimodal structure that exploits this structure leads to $\log(n)$ regret only from the neighboring arm.

Theorem 3 *The regret of SEEDA-Plateau satisfies:*
$$R(n) \leq \frac{c \log(n)}{q^* - q_{N-1}} + O\left(\log\log(n)\right) + \left(n\delta Q + t_1 + \frac{K-M}{2\epsilon^2}\right). \tag{9}$$

Furthermore, if $\delta = O(\frac{1}{n})$, we have that $R(n) \leq O(\log n)$.

The optimal dose level we have defined before can be rewritten as $k^* = \min\{M, N\}$, the recommendation accuracy of SEEDA-Plateau is given in Theorem 4.

Theorem 4 *With c set as $2 < c < \frac{5}{2}$, the probability that SEEDA-Plateau fails to recommend the optimal dose can be bounded as:*
$$\mathbb{P}[\hat{d}(n) \neq k^*] \leq \frac{3}{n^c} + \delta_1. \tag{10}$$

Compared to Corollary 2, the error probability of SEEDA-Plateau is increased by $\frac{3}{n^c}$. This is due to the ambiguity of the efficacy-optimal dose and the toxicity-optimal one, which leads to the two candidate doses $L_1(n)$ and $L_2(n)$. In practice, however, this ambiguity can be eliminated via preliminary experiments.

5. Experiments

5.1. Synthetic dataset

To investigate the operational characteristics and evaluate the performance of the proposed adaptive designs, we present an experimental study with $K = 6$ dose levels and $n = 300$ trial cohorts, with each cohort consists of 3 patients. The estimation is updated after observing all individual outcomes from a cohort. All experiment results are obtained with 1000 trial repetitions. The MTD threshold is set as $\theta = 0.35$.

The trial setup is the same as (Riviere et al., 2018) and (Zang et al., 2014), and we have simulated eight different efficacy and toxicity scenarios[3]. Due to the space limitation, we only report the results of the first scenario, where efficacy reaching the maximal value (the optimal dose) before toxicity hits MTD threshold. Additional results for this setting as well as the other seven scenarios are reported in Section K to M in the supplementary material.

The following baseline designs are used for comparison (whenever appropriate), whose details can be found in the supplementary material: 3+3, CRM, MCRM, Independent TS, KL-UCB, UCB-1, and multi-objective bandits. Note that MTA-RA and other TS variants in (Riviere et al., 2018) are not included because they assume a different truncated efficacy model, which needs to be perfectly known to the algorithm. For algorithms that require prior information of toxicity and efficacy, they are set as $[0.02, 0.06, 0.12, 0.20, 0.30, 0.40]$ and $[0.12, 0.20, 0.30, 0.40, 0.50, 0.59]$, respectively.

5.1.1. RECOMMENDATION AND ALLOCATION ACCURACY

We report the allocation and recommendation percentages of each dose for all considered designs in Table 2. Dose 3 (in bold font) is the optimal biological dose for this scenario. However, we comment that dose 4 also satisfies the optimality condition without violating the safety constraint. Nevertheless, it has a higher toxicity probability (although still below MTD) without increasing efficacy; thus less preferable to Dose 3. We note that for all the evaluated designs, the recommendation rule is $\hat{d}(n) = \arg\max_{k:\hat{p}_k(n) \leq \theta} \hat{q}_k(n)$, where $\hat{q}_k(n)$ and $\hat{p}_k(n)$ are the final estimations of toxicity and efficacy for dose level d_k, respectively. This suggests that safety constraint is considered in recommendation.

We can see from the results that SEEDA almost equally recommends dose 3 and 4 with a total probability of 94.6%.

---

[3]We remark that although no real-world trial data is utilized in the experiment, this approach is commonly accepted in clinical trials as the first-step study for a new methodology; see (Whitehead et al., 2012; Yap et al., 2013; Zang et al., 2014; Riviere et al., 2018).
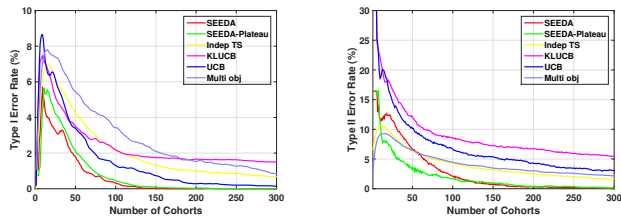

Figure 1: False alarm (left) and miss detection (right) error rates as a function of number of cohorts.

This is because the algorithm cares about maximizing efficacy without violating safety constraint, and both dose 3 and 4 satisfy such conditions. As a result, SEEDA treats both equally as the optimal solution. However, by leveraging the increase-then-plateau model assumption, SEEDA-Plateau can further break the "tie" between dose 3 and 4, and correctly recognize that dose 3 is the optimal biological dose: it chooses dose 3 at 86.6% while dose 4 only 10.4%. We see that the gain of SEEDA-Plateau in terms of recommending the optimal dose is significant over all the other designs (even compared to SEEDA). For a more detailed understanding of the recommendation accuracy, the corresponding false alarm and miss detection error rates (definitions are given in Section J in the supplementary material) are plotted in Fig. 1, and we observe that both SEEDA and SEEDA-Plateau outperform other baseline methods over the range of cohorts.

As for allocation, we observe that both SEEDA and SEEDA-Plateau concentrate at dose 3 and 4, while spending very little budget on both tail ends of the dosage. In particular, SEEDA-Plateau allocates the fewest percentages (1%) of patients to the most toxic dose 6 among all designs.

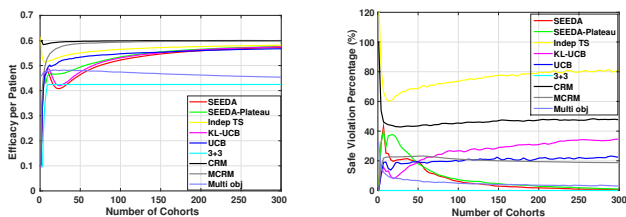### 5.1.2. CONVERGENCE AND SAFETY VIOLATION

Figure 2: Comparison of efficacy per patient (left) and the safety violation percentage (right).

To have a deeper understanding of the tradeoff between efficacy and toxicity, we plot side-by-side the convergence of efficacy and toxicity as t increases in Fig. 2. KL-UCB, UCB and Independent TS have good convergence but suffer from significant safety violation in the process since they do not

Table 2: Recommendation & allocation percentages of different designs. Optimal biological dose is #3. In each cell the first row reports the mean value over 1000 repetitions, and the second row reports the (standard deviation).

	Recommended						Allocated					
Toxicity probabilities	0.01	0.05	**0.15**	0.2	0.45	0.6	0.01	0.05	**0.15**	0.2	0.45	0.6
Efficacy probabilities	0.1	0.35	**0.6**	0.6	0.6	0.6	0.1	0.35	**0.6**	0.6	0.6	0.6
SEEDA	0	1	47.20	47.40	4.40	0	11.18	9.18	30.76	31.71	12.06	5.11
	(0)	(0.71)	(3.40)	(3.41)	(2.46)	(0)	(0.58)	(1.99)	(7.76)	(7.69)	(3.45)	(0.62)
SEEDA-Plateau	0.80	2.20	86.60	10.40	0	0	7.83	8.98	30.12	37.17	14.91	1.00
	(0.32)	(1.96)	(8.58)	(3.65)	(0)	(0)	(1.61)	(4.21)	(6.01)	(7.54)	(3.02)	(0.61)
Independent TS	2.60	9.40	44.60	35.40	6.60	1.40	3.66	7.26	22.22	21.00	22.26	23.60
	(2.47)	(3.86)	(10.25)	(10.32)	(2.96)	(0.69)	(0.97)	(3.85)	(15.47)	(10.44)	(10.43)	(9.22)
KL-UCB	0.20	4.60	48.80	43.60	2.80	0	10.93	7.16	21.33	20.91	21.21	18.46
	(0.13)	(2.71)	(11.68)	(11.36)	(2.64)	(0)	(0.81)	(0.94)	(10.52)	(11.31)	(10.92)	(11.10)
UCB	0	2.40	54.00	40.40	3.20	0	5.45	9.50	22.13	20.93	20.43	24.57
	(0)	(2.15)	(9.92)	(9.05)	(3.09)	(0)	(0.49)	(1.16)	(2.11)	(2.24)	(2.15)	(2.11)
3+3	0	2.40	12.00	17.60	45.20	22.80	16.04	17.82	20.19	18.12	16.81	5.82
	(0)	(0.41)	(4.31)	(5.31)	(7.15)	(4.35)	(5.12)	(4.23)	(10.25)	(9.15)	(8.15)	(4.12)
CRM	0	0	0	33.80	65.80	0.40	0.12	0.35	2.62	33.90	57.69	5.33
	(0)	(0)	(0)	(8.26)	(10.63)	(0.40)	(0.11)	(0.25)	(0.32)	(10.21)	(11.24)	(0.23)
MCRM	0	0	0.20	61.00	38.80	0	1.47	1.18	5.64	55.48	34.63	,1.60
	(0)	(0)	(0.15)	(9.67)	(8.65)	(0)	(0.24)	(0.67)	(3.62)	(8.63)	(7.65)	(0.67)
Multi-obj	0.81	3.23	47.90	41.03	5.88	1.15	18.42	21.92	23.36	18.48	9.92	7.89
	(0.19)	(0.94)	(11.86)	(11.90)	(1.80)	(0.36)	(6.07)	(5.55)	(6.68)	(7.05)	(5.17)	(4.65)

consider the safety constraint during exploration. CRM has higher efficacy at the cost of bad safety constraint violation, while 3+3 performs poorly in efficacy but has the lowest safety probability; this behavior is similarly observed for multi-objective bandits. The SEEDA(-Plateau) algorithm, in comparison, converges to the optimal efficacy at a slower rate, but the exploration process is carefully controlled so that the safety violation is minimized, which is evident from the right subplot of Fig. 2.

5.1.3. SAMPLE EFFICIENCY

Sample efficiency is measured by the minimum number of trial participants to achieve a pre-specified recommendation accuracy (also known as *early stopping* (Montori et al., 2005)). We start the trial with a minimum of 6 patients, and continue recruiting patients until the stopping condition is triggered. Fig. 3 plots the average minimum number of patients to achieve a given a recommendation accuracy for different algorithms[4]. We see that SEEDA-Plateau outperforms all other algorithms by a large margin, thanks to the "double dipping" of the model assumptions which gives the most accurate estimation of the optimal dose. In comparison, SEEDA performs similarly to the baseline algorithms. The reason is that the goal of SEEDA is to recommend the efficacy-maximal dose that satisfies the safety constraint. In this particular setting, both dose 3 and 4 satisfy this condition, and SEEDA does not have the mechanism to further minimize toxicity. This leads to a recommendation error that is similar to other baseline designs.

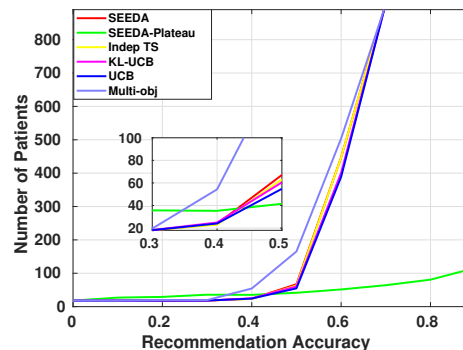[4]3+3, CRM and MCRM are excluded since they only target finding MTD.

Figure 3: The minimum number of trial participants to achieve a given a recommendation accuracy.

The sample efficiency advantage of SEEDA-Plateau is of critical importance in practice, as the significant cost associated with clinical trials is mostly proportional to the number of trial participants. Furthermore, reducing the number of patients while achieving the same level of accuracy minimizes the safety and ethical concern in the trial, which is another important consideration.

5.2. Real-world datasets

We evaluate the SEEDA algorithms in two real-world datasets *neurodeg* and *IBSCovars* based on (Biesheuvel & Hothorn, 2002). We first extract dose and resp variables from the observations reported in the dataset. With these samples, we fit them into a commonly used Emax dose-response model as in (Bornkamp et al., 2011) with an R

Table 3: Recommendation & allocation percentages of the neurodeg dataset. In each cell the first row reports the mean value over 1000 repetitions, and the second row reports the (standard deviation).

	Recommended					Allocated				
Toxicity	0.01	0.08	**0.30**	0.60	0.80	0.01	0.08	**0.30**	0.60	0.80
Efficacy	0.01	0.35	**0.45**	0.52	0.57	0.01	0.35	**0.45**	0.52	0.57
SEEDA	0.60	32.91	66.14	0	0	5.58	34.14	59.60	0.33	0.33
	(0.40)	(10.57)	(10.59)	(0)	(0)	(0.42)	(6.08)	(6.25)	(0.25)	(0.01)
SEEDA-Plateau	0.99	32.66	66.00	0	0	5.09	35.02	59.21	0.33	0.33
	(0.31)	(10.12)	(10.36)	(0)	(0)	(2.05)	(7.78)	(6.64)	(0.02)	(0)
Independent TS	3.39	51.28	44.47	0.38	0.46	0.80	3.50	7.59	23.37	64.68
	(2.56)	(9.92)	(10.34)	(0.33)	(0.37)	(0.60)	(2.70)	(5.21)	(10.19)	(12.51)
KL-UCB	0.07	55.74	28.67	5.43	0.07	98.0	0.42	0.47	0.52	0.55
	(0.06)	(12.38)	(12.76)	(2.10)	(0.05)	(2.62)	(0.23)	(0.28)	(0.49)	(0.04)
UCB	0.81	41.68	57.24	0.23	0.01	6.88	15.09	20.10	25.75	32.16
	(0.74)	(16.07)	(16.07)	(0.17)	(0.01)	(0.29)	(1.60)	(2.10)	(2.55)	(2.87)
3+3	0	2.40	12.00	17.60	45.20	16.04	17.82	20.19	18.12	16.81
	(0)	(1.02)	(2.35)	(3.44)	(10.34)	(5.60)	(9.48)	(1.84)	(1.60)	(4.20)
CRM	0	0	0	100	0	0	0	0	99.66	0.33
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0.01)	(0.01)
MCRM	4.33	26.47	69.18	0	0	4.67	26.40	68.92	0	0
	(0.25)	(1.80)	(1.86)	(0)	(0)	(0.25)	(1.80)	(2.10)	(0)	(0)
Multi-obj	0.24	15.33	17.59	0.12	0.03	24.33	26.12	18.95	16.05	14.52
	(0.13)	(9.65)	(9.71)	(0.05)	(0.03)	(4.28)	(3.51)	(6.11)	(2.93)	(2.61)

package implementation provided by (Yoshida, 2019). The resulting models are as follows.

$$\text{neurodeg:} \quad \text{resp} = 169.94 + \frac{12.95\text{dose}}{1.85 + \text{dose}},$$

$$\text{IBScovars:} \quad \text{resp} = 0.26 + \frac{0.68\text{dose}}{4.01 + \text{dose}}.$$

As for the toxicity event, since it is not reported in the dataset, we resort to simulations with model (1).

The allocation and recommendation percentages of each dose for all the algorithms are shown in Table 3 and Table 4 for both datasets. We have similar observations as in the synthetic experiment that SEEDA and SEEDA-Plateau recommend the correct doses majority of the times, while the suboptimal recommendation is mostly safe in that the doses immediately below MTD are recommended second most. The same is true for allocation.

6. Related Works

This work is concerned with adaptive phase I clinical trials, whose uptake in practice is starting to increase considerably. See (Bretz et al., 2017; Pallmann et al., 2018) for recent comprehensive surveys. The main motivation to use these adaptive designs is to learn as the trial progresses and use this learning to deliver more efficient or more ethically appealing trials. Adaptive clinical trial with sequential patient recruitment is considered in (Atan et al., 2019), but it does

not address the subsequent dose allocation. The 3+3 and the CRM designs or their variations remain the de facto adaptive designs in practice for dose-finding studies (Petroni et al., 2017; Pallmann et al., 2018), although new methodologies that aim at better safety protection are also proposed (Lee et al., 2017). In recent years, there is a growing interest in adaptive trial designs for MTA because of its different dose-response relationships (Zang et al., 2014; Riviere et al., 2018), but these studies do not explicitly enforce the safety constraints during the trial; neither do they provide theoretical guarantees on the trial performance.

Multi-armed bandit has long been considered as an important tool for the clinical trial methodology design, dating back to the early papers of (Thompson, 1933; Robbins, 1952). Developing bandit models and algorithms that better suit the specific requirements of adaptive clinical trials has attracted some attention in recent years. Villar et. al (Villar et al., 2015b; Villar & Rosenberger, 2018) adopted the (modified) forward-looking Gittins index rule for multi-arm clinical trials. The authors of (Wang et al., 2018) propose a regional bandit model that can be applied to learning the drug dosage and patient response relationship. The sample complexity of thresholding bandit is analyzed in (Garivier et al., 2017), which matches MTD identification. Furthermore, dose-finding clinical trials with heterogeneous groups are investigated in (Lee et al., 2020) from a MAB perspective. Probably the closest work to ours is (Aziz et al., 2019), which also considers both toxicity and efficacy. However,

Table 4: Recommendation & allocation percentages of the IBScovars datasets. In each cell the first row reports the mean value over 1000 repetitions, and the second row reports the (standard deviation).

	Recommended					Allocated				
Toxicity probabilities	0.01	0.10	**0.30**	0.70	0.95	0.01	0.10	**0.30**	0.70	0.95
Efficacy probabilities	0.01	0.20	**0.27**	0.33	0.43	0.01	0.20	**0.27**	0.33	0.43
SEEDA	1.14	35.04	63.47	0	0	10.11	34.35	54.86	0.33	0.33
	(1.31)	(7.58)	(7.60)	(0)	(0)	(0.82)	(5.42)	(5.57)	(0.17)	(0.01)
SEEDA-Plateau	2.08	36.51	61.06	0	0	8.97	34.60	55.75	0.33	0.33
	(2.76)	(10.31)	(10.42)	(0)	(0)	(4.00)	(4.34)	(7.83)	(0.02)	(0.01)
Independent TS	7.52	48.47	43.30	0.31	0.39	1.82	3.89	26.66	20.65	23.17
	(7.15)	(9.79)	(9.71)	(0.60)	(0.17)	(0.85)	(3.61)	(10.74)	(13.54)	(10.05)
KL-UCB	28.55	44.90	23.24	3.28	0	98.33	0.37	0.40	0.42	0.45
	(9.21)	(9.95)	(10.06)	(2.60)	(0)	(0.35)	(0.44)	(0.87)	(0.06)	(0.60)
UCB	1.73	45.26	52.81	0.17	0.01	9.41	15.37	18.94	23.22	33.04
	(1.21)	(9.21)	(9.25)	(0.08)	(0.01)	(0.38)	(1.45)	(1.87)	(2.26)	(2.61)
3+3	2.40	12.00	17.60	45.20	22.80	16.04	17.82	20.19	18.12	22.81
	(0.88)	(7.65)	(6.87)	(6.86)	(8.87)	(2.85)	(5.29)	(8.29)	(5.52)	(5.45)
CRM	0	0	0	1.35	98.65	0	0	0	99.66	0.33
	(0)	(0)	(0)	(0.10)	(0.04)	(0)	(0)	(0)	(0.86)	(0.03)
MCRM	4.34	26.91	68.74	0	0	4.67	26.83	68.49	0	0
	(0.26)	(2.15)	(2.20)	(0)	(0)	(0.04)	(0.05)	(0.91)	(0)	(0)
Multi-obj	0.45	16.18	16.56	0.10	0.02	1.23	3.27	5.79	13.11	76.56
	(0.25)	(5.49)	(10.53)	(0.03)	(0)	(1.14)	(3.12)	(5.12)	(6.43)	(6.04)

the safety constraint, which is an essential constraint of real-world phase I trials, has not been explicitly considered in these papers.

On the other hand, the problem of safe exploration has attracted a lot of attention recently, albeit often in control (Koller et al., 2018) and general reinforcement learning (Berkenkamp et al., 2017). The authors in (Sui et al., 2015) propose the SAFEOPT algorithm for safe exploration in Gaussian processes, and (Kazerouni et al., 2017) presents a variant of linear UCB method for the contextual linear bandit problem. A different line of works (Maillard, 2013; Galichet et al., 2013) consider minimizing risk in MAB, but they are mostly casted in the mean-variance framework with respect to the reward distribution.

7. Conclusions

Learning in adaptive clinical trials faces several unique challenges that have not been well addressed, which may have contributed to their lack of adoption in actual clinical trials. In particular, the safety constraints resulting from ethical and societal considerations have been insufficiently researched, which has motivated us to develop the SEEDA algorithm that explicitly imposes safety constraints (in terms of toxicity) while also aiming for maximum patient response in a dose-finding study. Theoretical analysis of SEEDA is carried out and the proposed algorithm is further extended to the increase-then-plateau efficacy model and shown to

have smaller regret thanks to the unimodal structure. The performance advantages over state-of-the-art adaptive clinical trial designs are illustrated with experiments on both synthetic and real-world datasets.

8. Acknowledgements

References

Atan, O., Zame, W. R., and van der Schaar, M. Sequential patient recruitment and allocation for adaptive clinical trials. In *Proceedings of The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1891–1900, Apr 2019.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, May 2002.

Aziz, M., Kaufmann, E., and Riviere, M.-K. On multi-armed bandit designs for phase I clinical trials. *arXiv e-prints*, art. arXiv:1903.07082, March 2019.

Berkenkamp, F., Turchetta, M., Schoellig, A. P., and Krause, A. Safe model-based reinforcement learning with stability guarantees. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 908–919, Long Beach, California, USA, December 2017.

Biesheuvel, E. and Hothorn, L. Many-to-one comparisons in stratified designs. *Biometrical Journal*, 44:101?–116, 2002.

Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. C. Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics*, 5:1611–?1631, 2011.

Bretz, F., Gallo, P., and Maurer, W. Adaptive designs: The swiss army knife among clinical trial designs? *Clinical Trials*, 14(5):417–424, 2017.

Combes, R. and Proutière, A. Unimodal bandits: Regret lower bounds and optimal algorithms. In *Proceedings of the 31th International Conference on Machine Learning*, pp. 521–529, Beijing, China, June 2014.

Galichet, N., Sebag, M., and Teytaud, O. Exploration vs exploitation vs safety: risk-aware multi-armed bandits. In *Proceedings of the 5th Asian Conference on Machine Learning*, pp. 245–260, November 2013.

Garivier, A. and Cappè, O. The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of Conference On Learning Theory (COLT)*, 2011.

Garivier, A., Ménard, P., and Rossi, L. Thresholding bandit for dose-ranging: The impact of monotonicity. *arXiv e-prints*, art. arXiv:1711.04454, November 2017.

Kazerouni, A., Ghavamzadeh, M., Abbasi, Y., and Van Roy, B. Conservative contextual linear bandits. In *Proceedings of Advances in Neural Information Processing Systems*, pp. 3910–3919, 2017.

Koller, T., Berkenkamp, F., Turchetta, M., and Krause, A. Learning-based model predictive control for safe exploration. In *IEEE Conference on Decision and Control (CDC)*, pp. 6059–6066, December 2018.

Lee, H.-S., Shen, C., Jordon, J., and van der Schaar, M. Contextual constrained learning for dose-finding clinical trials. In *Proceedings of The 23rd International Conference on Artificial Intelligence and Statistics*, Aug. 2020.

Lee, S. M., Ursino, M., Cheung, Y. K., and Zohar, S. Dose-finding designs for cumulative toxicities using multiple constraints. *Biostatistics*, 20(1):17–29, Nov. 2017.

Maillard, O.-A. Robust risk-averse stochastic multi-armed bandits. In *Proceedings of the 24th International Conference on Algorithmic Learning Theory*, pp. 218–233, Singapore, 2013.

Montori, V. M. et al. Randomized trials stopped early for benefit: A systematic review. *JAMA*, 294(17):2203–2209, Nov. 2005.

Neuenschwander, B., Branson, M., and Gsponer, T. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine*, 27(13):2420–2439, 2008.

O'Quigley, J., Pepe, M., and Fisher, L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*, 43(1):33–48, 1990.

Pallmann, P. et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine*, 16(1):29, Feb 2018.

Paoletti, X. and Postel-Vinay, S. Phase I–II trial designs: how early should efficacy guide the dose recommendation process? *Annals of Oncology*, 29(3):540–541, Feb. 2018.

Petroni, G. R., Wages, N. A., Paux, G., and Dubois, F. Implementation of adaptive methods in early-phase clinical trials. *Statistics in Medicine*, 36(2):215–224, 2017.

Postel-Vinay, S. et al. Clinical benefit in phase-I trials of novel molecularly targeted agents: does dose matter? *British Journal of Cancer*, 100(9):1373–1378, May 2009.

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 (2):466–479, 2018.

Robbins, H. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58:527–535, 1952.

Roberts, T. G. et al. Trends in the risks and benefits to patients with cancer participating in phase 1 clinical trials. *JAMA*, 292(17):2130–2140, Nov. 2004.

Storer, B. E. Design and analysis of phase I clinical trials. *Biometrics*, 45:925–37, 1989.

Sui, Y., Gotovos, A., Burdick, J. W., and Krause, A. Safe exploration for optimization with Gaussian processes. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 997–1005, 2015.

Thiessen, B. et al. A phase I/II trial of GW572016 (lapatinib) in recurrent glioblastoma multiforme: clinical outcomes, pharmacokinetics and molecular correlation. *Cancer Chemotherapy and Pharmacology*, 65(2):353–361, Jan 2010.

Thompson, W. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, December 1933.

Tighiouart, M., Liu, Y., and Rogatko, A. Escalation with overdose control using time to toxicity for cancer phase I clinical trials. *PLOS ONE*, 9:1–13, 03 2014.

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

Villar, S. S., Bowden, J., and Wason, J. Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statistical Science*, 30(2):199–215, May 2015a.

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

Wang, Z., Zhou, R., and Shen, C. Regional multi-armed bandits. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 510–518, Playa Blanca, Lanzarote, Canary Islands, Apr. 2018.

Whitehead, J. et al. A novel phase I/IIa design for early phase oncology studies and its application in the evaluation of MK-0752 in pancreatic cancer. *Statistics in Medicine*, 31(18):1931–1943, 2012.

Yahyaa, S. and Manderick, B. Thompson sampling for multi-objective multi-armed bandits problem. In *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 47–52, Bruges, Belgium, April 2015.

Yan, F., Thall, P. F., Lu, K. H., Gilbert, M. R., and Yuan, Y. Phase I–II clinical trial design: a state-of-the-art paradigm for dose finding. *Annals of Oncology*, 29(3):694–699, Dec. 2017.

Yap, C. et al. Implementation of adaptive dose-finding designs in two early phase haematological trials: clinical, operational, and methodological challenges. *Trials*, 14 (1):O75, Nov 2013.

Yoshida, K. *Emax Model Analysis with 'Stan'*. Columbia University, New York, USA, 2019. URL `https://cran.r-project.org/web/packages/rstanemax`.

Zang, Y., Lee, J. J., and Yuan, Y. Adaptive designs for identifying optimal biological dose for molecularly targeted agents. *Clinical Trials*, 11(3):319–327, 2014.