SDF-Bayes: Cautious Optimism in Safe Dose-Finding Clinical Trials with Drug Combinations and Heterogeneous Patient Groups

Hyun-Suk Lee
Sejong University

Cong Shen
University of Virginia

William Zame
UCLA

Jang-Won Lee
Yonsei University

Mihaela van der Schaar
University of Cambridge, UCLA,
The Alan Turing Institute

Abstract

Phase I clinical trials are designed to test the safety (non-toxicity) of drugs and find the maximum tolerated dose (MTD). This task becomes significantly more challenging when multiple-drug dose-combinations (DC) are involved, due to the inherent conflict between the *exponentially* increasing DC candidates and the limited patient budget. This paper proposes a novel Bayesian design, *SDF-Bayes*, for finding the MTD for drug combinations in the presence of safety constraints. Rather than the conventional principle of escalating or de-escalating the current dose of one drug (perhaps alternating between drugs), SDF-Bayes proceeds by *cautious optimism*: it chooses the next DC that, on the basis of current information, is most likely to be the MTD (optimism), subject to the constraint that it only chooses DCs that have a high probability of being safe (caution). We also propose an extension, *SDF-Bayes-AR*, that accounts for patient heterogeneity and enables heterogeneous patient recruitment. Extensive experiments based on both synthetic and real-world datasets demonstrate the advantages of SDF-Bayes over state of the art DC trial designs in terms of accuracy and safety.

1 INTRODUCTION

The use of combinations of drugs is becoming an increasingly valuable–and common–treatment modality (Bamias et al., 2011; Flaherty et al., 2012; Ocana et al., 2019; Kelly and Halabi, 2018), especially in the treatment of cancer, where it has been widely observed that combinations of drugs can be effective when single drugs are not (Paller et al., 2019). (Similar observations have been made for other diseases, including COVID-19.) As a result, there has been an enormous effort to test and validate drug combinations for treatment; indeed combination trials now account for more than 25% of all clinical trials in oncology (Wu et al., 2014).

However, clinical trials of drug combinations face greater challenges than those of single drugs, especially in Phase I trials which are required to find safe doses. The essential problem is that the number of potential dose-combinations (DCs) to be tested in a trial increases *exponentially* with the number of drugs, but the patient budget cannot scale proportionately. Indeed, a typical real-world Phase I trial often recruits fewer than 100 patients; a few examples are shown in Table 1. The limited patient budget constrains the number of DCs that can be throughly tested. Moreover, because drugs interact differently with different body chemistry of different groups of patients (e.g., with the different hormone balances and levels in males and females), it is frequently possible to identify in advance groups of patients who might be expected to exhibit very different tolerances for the same DC (Sun and Braun, 2015; Kim et al., 2009; Dasari et al., 2013; Moss et al., 2015; Wages et al., 2015).

In view of the growing importance of drug combinations in the treatment of disease, it is of enormous importance to design Phase I drug combination trials in a way that is efficient, informative and safe (Hamberg et al., 2010). This design includes the path of DC testing and, in the presence of identified heterogeneous groups, the allocation of patient budgets to groups as well.

This paper develops a new dose-finding Phase I clinical trial method for drug combinations and heterogeneous patient groups and demonstrates that it is superior to existing methods. For a given patient budget, the design objective is to maximize the probability of finding the maximum tolerated dose (MTD), defined to be the DC that is closest to a given target toxicity threshold,

Proceedings of the 24[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

Table 1: Examples of Phase I clinical trial studies with drug combinations

Study	Drugs	No. of DC	No. of Patients	Target disease
(Plummer et al., 2008)	Rucaparib & temozolomide	12	32	Advanced solid tumors
(Bailey et al., 2009)	Nilotinib & imatinib	20	50	Gastrointestinal stromal tumors
(Bagatell et al., 2014)	Temsirolimus & irinotecan & temozolomide	24	71	Solid tumors
(Calvo et al., 2017)	Dacomitinib & Figitumumab	12	74	Advanced solid tumors

subject to constraints on the exposure of patients to unsafe doses throughout the trial. Our safe dose-finding (SDF) method employs an adaptive design: the DC to be tested in the current round is chosen on the basis of past observations. This is possible because Phase I trials are not blind; the trialist knows the DC given to each patient and observes the outcome. The *SDF-Bayes* algorithm builds on a novel learning principle we call *Cautious Optimism*, which manifests by combining two opposing ideas: (1) SDF-Bayes constrains the choice of DC to be tested in the current round to a set of DCs that it estimates to be unlikely to violate the safety constraint; this is the principle of *caution*, and (2) within this constrained set of DCs, SDF-Bayes chooses the DC to be tested in the current round to be the one estimated to be *most likely* to be the MTD; this is the *optimistic belief principle* (Aziz et al., 2019).

To deal with settings in which potentially heterogeneous patient groups can be identified in advance, we also propose a extension, that we call SDF-Bayes-AR, in which both the DC to be tested *and* the patient group to be sampled in the current round are chosen adaptively. To determine the group from which to recruit the next patient, SDF-Bayes-AR uses the criterion of *expected improvement* (EI). Adaptive recruitment is especially useful when there is prior information about one or several groups (Pallmann et al., 2018; Park et al., 2018; Atan et al., 2019), and an appealing feature of our approach is that it can smoothly incorporate prior information, be it from the drug development phase, from dose-toxicity models, from tests *in vitro* and in animals, or from previous trials (Gasparini, 2013; Shen et al., 2019).

We validate the proposed designs via extensive simulated trials using both synthetic and real-world datasets. We show that, using a realistic number of patients, our algorithms provide significantly more accurate recommendations than state-of-the-art designs, while obeying the safety constraints, for both homogeneous and heterogeneous patient populations.

2 DC-FINDING CLINICAL TRIALS

2.1 Dose-Toxicity Model

We consider a dual-agent dose-finding Phase I clinical trial for the combination of agents (drugs) A and B. We assume discrete dose levels $\mathcal{J}=\{1,...,J\}$ for agent A and $\mathcal{K}=\{1,...,K\}$ for agent B. We use (j,k) to denote the combination of dose j of agent A and dose k of agent B so that the set of all DCs is just $\mathcal{A}=\mathcal{J}\times\mathcal{K}$. We model the toxicity event Y_{jk} of DC (j,k) as a Bernoulli random variable with unknown parameter p_{jk}. We set $Y_{jk}=1$ to indicate that a dose-limiting toxicity (DLT) was observed for (j,k), and $Y_{jk}=0$ otherwise.

We assume the toxicities follow a parametric joint dose-toxicity model

$$\pi : \Theta \times \mathcal{J} \times \mathcal{K} \to [0,1] \tag{1}$$

where Θ is some space of parameters, and $\pi(\theta,j,k) = p_{jk}(\theta)$ is the toxicity of the DC (j,k) if the parameter is $\theta \in \Theta$. The *true* vector of parameters θ^* is unknown and must be learned/estimated. The literature has suggested various dose-toxicity models (Gasparini, 2013; Riviere et al., 2014); we present details of some commonly-used models in the Supplementary Material; we focus here on developing a methodology that can be used with *many* parametric joint dose-toxicity models.

2.2 Problem Formulation

We consider an adaptive Phase I clinical trial for drug combinations with a given patient budget T. The nominal trial objective is to find the DC whose estimated toxicity is closest to the threshold ξ; i.e. to find any DC that belongs to

$$\mathcal{A}^* = \operatorname*{argmin}_{jk} \left| p_{jk}(\theta^*) - \xi \right|.$$

For each t, write $\mathcal{O}(t) = \{(Y_\tau, a(\tau))\}_{\tau=1}^{t-1}$ for the history of trial actions and observations *before* t; $a(\tau)$ is the DC that was administered to patient τ and Y_τ is the observed toxicity for DC $a(\tau)$ on patient τ. As in typical adaptive clinical trial designs, all DCs are actionable in the trial. A DC-finding algorithm Π maps the history $\mathcal{O}(t)$ to the DC $a(t) \in \mathcal{A}$ that is to be administered to patient t. At the end of the clinical trial, the algorithm recommends a DC $\hat{a}^*$. The recommendation is *correct* if $\hat{a}^* \in \mathcal{A}^*$; otherwise the recommendation is an *error*. We set the objective of our algorithm to minimize the probability that errors occur. (Because the occurrence of toxicity events is stochastic, there is always some positive probability that errors will occur.)

Because testing unsafe DCs should be avoided, we insist that, with high probability, the average toxicity

should not exceed the toxicity threshold plus a margin of error. We formalize this by defining the *DLT observation rate* to be

$$S(T) = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

and we impose the *safety constraint*

$$\mathbb{P}\big(S(T) \le \xi + \epsilon\big) \ge 1 - \delta, \tag{2}$$

where $\epsilon > 0$ is a prescribed margin of error and $\delta > 0$ is a prescribed acceptable probability of failure.

3 CAUTIOUS OPTIMISM

3.1 Preliminaries

Given the dose-toxicity structure, we begin with a prior distribution $f(\theta)$ on the parameter vector θ and update the posterior distribution, based on the observations. (The assumption of a prior distribution is common in the literature, and is natural, because the toxicities of the individual drugs are usually understood on the basis of prior use, and the various drug combinations will have already been tested in vitro and in animals.) We write $f(\theta|\mathcal{O}(t))$ to denote the posterior distribution of θ following the history $\mathcal{O}(t)$. The likelihood of θ in round t for the observations $\mathcal{O}(t)$ satisfies

$$L(\theta|\mathcal{O}(t)) \propto \prod_{a\in\mathcal{A}} p_a(\theta)^{s_a(t)}(1-p_a(\theta))^{n_a(t)-s_a(t)},$$

where $n_a(t)$ is the number of times DC a has been chosen before round t and $s_a(t) = \sum_{\tau=1}^{t-1} \mathbb{I}[a(\tau) = a]Y_\tau$ is the number of DLTs observed for the DC a before round t. The posterior distribution satisfies

$$f(\theta|\mathcal{O}(t)) \propto L(\theta|\mathcal{O}(t))f(\theta). \tag{3}$$

3.2 Cautious Optimism in Bayesian SDF

We base our algorithm on the principle of "optimism in the face of uncertainty" (Bubeck and Cesa-Bianchi, 2012), which means choosing the DC that is currently estimated to be most likely to be the MTD, but we maintain safety by constraining the set of DCs from which we choose.

Optimism for Efficiency We assess which DC is most likely to be the MTD by using the posterior distribution $f(\theta|\mathcal{O})$ as described in (3). For this, in our algorithm we follow the literature (see (Riviere et al., 2014), for example) and find the DC

$$\operatorname*{argmax}_a \mathbb{P}\big(\{\theta : |p_a(\theta) - \xi| < u\}\big),$$

where u is some prescribed allowable margin of error, and the probability is taken with respect to the posterior distribution on the parameter space Θ.

The probability $G_a^{\mathcal{O}(t)}(u)$ that the toxicity of DC a belongs to the given target toxicity interval $[\xi-u, \xi+u]$ is

$$\begin{aligned} G_a^{\mathcal{O}(t)}(u) &= \mathbb{P}[p_a(\theta) \in [\xi-u, \xi+u]|\mathcal{O}(t)] \\ &= \int_\Theta \mathbb{I}[p_a(\theta) \in [\xi-u, \xi+u]]f(\theta|\mathcal{O}(t))d\theta. \end{aligned} \tag{4}$$

Then, we find a DC $a(t)$ that is deemed most likely to have toxicity in $[\xi-u, \xi+u]$; i.e.

$$a_o(t) = \operatorname*{argmax}_a G_a^{\mathcal{O}(t)}(u).$$

The DC is chosen to be allocated if it is deemed "safe enough" (see the next paragraph) at round t. If the argmax is not a singleton, we choose arbitrarily, subject to maximizing the total dose of both drugs.

Caution for Safety To determine the set of DCs that are "safe enough" in round t we first choose a hyperparameter $v > 0$ that controls conservativeness and define

$$\begin{aligned} F_a^{\mathcal{O}(\tau)}(v) &= \max\Big\{x \in [0,1] : \mathbb{P}\big[p_a(\theta) \le x \,\big|\, \mathcal{O}(\tau)\big] \le v\Big\} \\ \Phi(t,v) &= \sum_{\tau=1}^{t-1} F_{a(\tau)}^{\mathcal{O}(t)}(v) \\ r(t,v) &= (\xi+\epsilon)t - \Phi(t,v). \end{aligned} \tag{5}$$

Roughly speaking, $\Phi(t,v)$ represents the number of DLT observations that would have been expected before round t, given the posterior. For safety, we are "allowed" a DLT observation rate $\xi+\epsilon$ so if $\Phi(t,v) \le (t-1)(\xi+\epsilon)$ the safety constraint holds in expectation after $t-1$ rounds; if we choose $a(t)$ so that $\Phi(t+1,v) \le t(\xi+\epsilon)$ then we will have met the safety constraint in expectation after t rounds. We cannot be assured of choosing such an $a(t)$ because we do not know what the posterior will be after t rounds, but if the posterior after t rounds were the same as the posterior after $t-1$ rounds and we choose $a(t)$ to have expected toxicity less than the *residual* $r(t,v)$ then the safety constraint will be met in expectation after t rounds. So define

$$\mathcal{A}_r(t) \coloneqq \{a \in \mathcal{A} : F_a^{\mathcal{O}(t)}(v) \le r(t,v)\}. \tag{6}$$

Unfortunately it might be that $a_o(t) \notin \mathcal{A}_r(t)$. (That would certainly be the case if $r(t,v) < 0$, which means that the safety constraint had already been violated in expectation.) In that case we can use a *conservative* DC, one whose expected toxicity is no greater than ξ, so that after t rounds the residual $r(t+1,v)$ will be greater than the residual $r(t,v)$. The set of conservative DCs is

$$\mathcal{A}_c(t) \coloneqq \{a \in \mathcal{A} : F_a^{\mathcal{O}(t)}(v) \le \xi\}.$$

We allocate the most likely conservative DC to be the MTD; i.e., $\operatorname{argmax}_{a\in\mathcal{A}_c(t)} G_a^{\mathcal{O}(t)}(u)$. Finally, if $\mathcal{A}_c(t) = \emptyset$ we terminate the procedure with no recommendation.

Theoretical Guarantee The following Proposition shows that, when this cautiously optimistic procedure can be carried out, it maintains the safety constraint by keeping the residual non-negative with a high probability. (The proof is in the Supplementary Materials.)

Proposition 1 *Fix $t \leq T$. For each $\tau \leq t$, set $v_\tau = (1-\delta)^{1/\tau}$. If the residuals $r(\tau, v_\tau)$ are non-negative for all $\tau \leq t$ then the cautious optimism principle in Bayesian SDF satisfies*

$$\mathbb{P}\left[\frac{1}{t}\sum_{\tau=1}^{t} p_{a(\tau)}(\theta) \leq \xi + \epsilon \,\middle|\, \mathcal{O}(t)\right] \geq 1-\delta.$$

3.3 Algorithm Description

Here we provide an implementation of the *cautious optimism* principle described above. Because we are considering an arbitrary joint dose-toxicity model, our implementation relies on Bayesian sampling in order to ensure the universal applicability of the algorithm. We comment that our algorithm can be easily extended for drug combinations with more than two drugs if a corresponding dose-toxicity model is given.

We denote the number of samples from the posterior distribution $f(\theta|\mathcal{O})$ by L, and the samples in round t as $\tilde{\Theta}(t) = \{\theta^{(l)}(t)\}_{l\in[L]}$. We use a Gibbs sampler (Gilks et al., 1995); this is a common multidimensional Bayesian sampling algorithm. Inside the Gibbs sampler, we use the adaptive rejection Metropolis sampling method (Gilks et al., 1995). Details can be found in the Supplementary Material.

In round t, we use the samples to approximate the probability $G_a^{\mathcal{O}(t)}(u)$ by

$$\tilde{G}_a^{\mathcal{O}(t)}(u) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{I}\{t : \xi - u \leq p_a(\theta^{(l)}(t)) \leq \xi + u\}.$$

Using this approximation, the DC most likely to be the MTD in round t is $\tilde{a}_o(t) = \text{argmax}_{a\in\mathcal{A}} \tilde{G}_a^{\mathcal{O}(t)}(u)$. SDF-Bayes then infers whether the safety constraint is violated or not for the chosen DC by evaluating the residual. To calculate the residual in practice, we define $\tilde{F}_a^{\mathcal{O}(t)}(v) \coloneqq \texttt{Prctile}(a, \tilde{\Theta}(t), v)$ that returns the percentile of the toxicities of DC a calculated from the samples $\tilde{\Theta}(t)$ for the percentage $v \in [0,1]$. Then, from (5), we calculate the residual $r(t,v)$ in round t by $r(t,v) = (\xi+\epsilon)t - \sum_{\tau=1}^{t-1} \tilde{F}_{a(\tau)}^{\mathcal{O}(t)}(v)$ and define $\tilde{\mathcal{A}}_r(t)$ as in (6) with $\tilde{F}_a^{\mathcal{O}(t)}(v)$. To keep the residual non-negative, SDF-Bayes accepts the chosen DC $\tilde{a}(t)$ if it does not make the residual negative: $\tilde{a}_o(t) \in \tilde{\mathcal{A}}_r(t)$. Otherwise, it rejects the chosen DC and chooses the most likely DC to be the MTD in the set of conservative DCs; i.e., $\text{argmax}_{a\in\tilde{\mathcal{A}}_c(t)} \tilde{G}_a^{\mathcal{O}(t)}(u)$, where

$$\tilde{\mathcal{A}}_c(t,v) = \{a \in \mathcal{A} : \tilde{F}_a^{\mathcal{O}(t)}(v) \leq \xi\}.$$

Algorithm 1 SDF-Bayes

while $t \leq T$ **do**
  Sample θ from their posterior distribution
  if $\tilde{a}_o(t) \in \tilde{\mathcal{A}}_r(t)$ **then**
    $a(t) \leftarrow \tilde{a}_o(t)$
  else if $\tilde{\mathcal{A}}_c(t,v) \neq \emptyset$ **then**
    $a(t) \leftarrow \text{argmax}_{a\in\tilde{\mathcal{A}}_c(t,v)} \tilde{G}_a^{\mathcal{O}(t)}(u)$
  else
    if $w > \psi$ **then**
      $a(t) \leftarrow \text{argmax}_{a\in\tilde{\mathcal{A}}_c(t,w)} \tilde{G}_a^{\mathcal{O}(t)}(u)$
    else
      Terminate trial without recommendation
    end if
  end if
  Observe the DLT Y_t
  Update $s_{a(t)}(t+1)$ and $n_{a(t)}(t+1)$
  $t \leftarrow t+1$
end while
Output: $\hat{a}^* = \text{argmax}_{a\in\mathcal{A}} \tilde{G}_a^{\mathcal{O}(T)}(u)$

If the set is empty, then the trial is terminated in the cautious optimism principle. However, this may be too conservative in the practical implementation, especially if v is relatively large, because it implies that there is no "safe enough" DC in a conservative view, not all DCs are unacceptably unsafe. Hence, if $\tilde{\mathcal{A}}_c(t,v) = \emptyset$, SDF-Bayes first finds w so that $\tilde{\mathcal{A}}_c(t,w) \neq \emptyset$ and $\tilde{\mathcal{A}}_c(t,v') = \emptyset$ for all $v' > w$. Then, it continues the trial by choosing the most likely DC in $\tilde{\mathcal{A}}_c(t,w)$ if $w > \psi$, where ψ is a predefined parameter for early termination rule. Otherwise, it terminates the trial with no recommendation because it implies all DCs are unsafe. (This early termination rule is widely used in the literature (Riviere et al., 2014; Yin and Yuan, 2009a).) With the chosen DC $a(t)$, the DLT Y_t is then observed, and the observation $\mathcal{O}(t+1)$ is updated. This process repeats until T patients have been administered. At the end of the clinical trial, the DC recommendation is given by $\hat{a}^* = \text{argmax}_{a\in\mathcal{A}} \tilde{G}_a^{\mathcal{O}(T)}(u)$. The pseudocode of the algorithm is summarized in Algorithm 1.

4 HETEROGENEOUS GROUPS

4.1 Problem formulation

We now show how to adapt SDF-Bayes to deal with heterogeneous groups. We continue to assume a total budget of T patients, which now can be distributed across M patient groups $\mathcal{M} = \{1, 2, ..., M\}$. We allow for the possibility that the dose-toxicity model varies across different groups, and model the toxicity Y_{jk}^m of (j,k) for group m as a Bernoulli random variable with unknown parameter p_{jk}^m. For each group m, we assume p_{jk}^m follows a parametric joint dose-toxicity model $p_{jk}^m(\theta) = \pi^m(j,k,\theta)$; we write θ_m^* for the true

parameter and ξ_m for the prescribed toxicity threshold. For patient t, the algorithm first chooses a group $g(t) \in \mathcal{M}$ from which to recruit the next patient, then chooses a DC for the recruited patient based on the history $\mathcal{O}(t) = \{(Y_\tau, g(\tau), a(\tau)\}_{\tau=1}^{t-1}$ prior to round t. The outcome (DLT or not) is then observed and recorded. This continues until the total budget T is exhausted.

At the end of the trial, the algorithm recommends, for each group m, a DC $\hat{a}_m^*$ to be used as the MTD for that group. As before, the (set of) true MTD(s) for group m is

$$\mathcal{A}_m^* = \underset{a\in\mathcal{A}}{\operatorname{argmin}} |p_a^m(\theta_m^*) - \xi_m|,$$

and the recommendation is an *error* if $\hat{a}_m^* \notin \mathcal{A}_m^*$. The safety constraint for group m is $\mathbb{P}[S_m(T) \le \xi_m + \epsilon] \le 1-\delta,\ \forall m \in \mathcal{M}$, where $S_m(T) = \frac{\sum_{t=1}^T Y_t \mathbb{I}[g(t)=m]}{\sum_{t=1}^T \mathbb{I}[g(t)=m]}$.

4.2 SDF-Bayes for Heterogeneous Groups

The goal of a clinical trial with heterogeneous groups is to minimize the DC recommendation error while satisfying the safety constraints of *each* group. Given the flexibility of recruiting patients from different groups, it is intuitive that *uniform* recruitment across groups, which does not utilize the history information prior to each decision time, may be inefficient. For example, if a particular patient group has already accumulated sufficient observations to determine the MTD with high confidence, recruiting more patients for this group is not as beneficial as for other groups. We reinforce this intuition by proposing SDF-Bayes-AR, a modification of SDF-Bayes in which patients are *adaptively* recruited. We will prove that SDF-Bayes-AR utilizes the limited number of patients more efficiently, by adaptively recruiting patients to obtain the best information for each group while satisfying the safety constraints.

Adaptive Patient Recruitment In SDF-Bayes-AR, we use the probability $G_a^{\mathcal{O}(t)}(u)$ in (4) to measure the likelihood of MTD. We denote the probability in round t that the DC a for group m is the MTD by

$$G_{m,a}^{\mathcal{O}(t)}(u) = \int_\Theta \mathbb{I}[p_a^m(\theta) \in [\xi_m - u, \xi_m + u]] f_m(\theta|\mathcal{O}(t)) d\theta,$$

where $f_m(\theta|\mathcal{O}(t))$ is the posterior distribution of θ for group m. The posterior distribution for each group can be calculated as in (3) by using, for each DC a, the number $(n_a^m(t))$ of times a is used for group m before round t and the number $(s_a^m(t))$ of DLTs observed when a is used within group m before round t. The improvement of the probability of the most probable DC for group m in round t for an *additional* patient with DC a' and observed toxicity Y is given by

$$I_{m,a'}^{\mathcal{O},Y}(u) = \left| G_{m,*}^{\{\mathcal{O},(Y,m,a')\}}(u) - G_{m,*}^{\mathcal{O}}(u) \right|,$$

where $G_{m,*}^{\mathcal{O}} = \max_{a\in\mathcal{A}} G_{m,a}^{\mathcal{O}}$. Define the **expected improvement (EI)**

$$H_{m,a'}^{\mathcal{O}(t)}(u) = \int_\Theta \Big\{ p_{a'}^m(\theta) I_{m,a'}^{\mathcal{O}(t),1}(u) + (1 - p_{a'}^m(\theta) I_{m,a'}^{\mathcal{O}(t),0}(u) \Big\} f_m(\theta|\mathcal{O}(t)) d\theta$$

Denote the tentatively allocated DC of group m in round t by $a_m(t)$. Then, in SDF-Bayes-AR, we adaptively recruit a patient of group m^* by using $a_m(t)$'s as $m^*(t) = \operatorname{argmax}_{m\in\mathcal{M}} H_{m,a_m(t)}^{\mathcal{O}(t)}(u)$.

Algorithm Description In round t, we apply SDF-Bayes for each group m to find the DC $a_m(t)$ that should be chosen for that group. Specifically, we use lines 2–13 of Algorithm 1. We then approximate the EI $H_{m,a_m(t)}^{\mathcal{O}(t)}(u)$ for group m. To this end, we sample from three different posterior distributions, $f_m(\theta|\mathcal{O}(t))$, $f_m(\theta|\{\mathcal{O}(t),(1,m,a')\})$, and $f_m(\theta|\{\mathcal{O}(t),(0,m,a')\})$. We denote the sets of the samples by $\tilde{\Theta}_m(t)$, $\tilde{\Theta}_m^{(1,m,a')}(t)$, and $\tilde{\Theta}_m^{(0,m,a')}(t)$, respectively. We approximate the EI $H_{m,a'}^{\mathcal{O}(t)}(u)$ by $\tilde{H}_{m,a'}^{\mathcal{O}(t)}(u) = \tilde{p}_{a'}^m(\mathcal{O}(t))\tilde{I}_{m,a'}^{\mathcal{O}(t),1}(u) + (1 - \tilde{p}_{a'}^m(\mathcal{O}(t)))\tilde{I}_{m,a'}^{\mathcal{O}(t),0}(u)$, where $\tilde{p}_a^m(\mathcal{O}(t)) = \frac{1}{L}\sum_{\theta\in\tilde{\Theta}_m(t)} p_a^m(\theta)$. Finally, the group to be recruited is chosen to be $g(t) = \operatorname{argmax}_{m\in\mathcal{M}} \tilde{H}_{m,a_m(t)}^{\mathcal{O}(t)}(u)$, and the DC to be allocated to that group is $a(t) = a_{g(t)}(t)$. (Ties for $g(t)$ or $a_{g(t)}$ are broken arbitrarily.) We then observe the DLT Y_t from the recruited patient of group $g(t)$ with the allocated DC $a(t)$ and construct the history $\mathcal{O}(t+1)$. At the end of the trial, the DC recommendation for each group is given by $\hat{a}_m^* = \operatorname{argmax}_{a\in\mathcal{A}} \tilde{G}_{m,a}^{\mathcal{O}(T)}(u)$. A more detailed description of SDF-Bayes-AR is provided in the Supplementary Material.

5 EXPERIMENTS

Here we describe a variety of experiments using a real-world dataset and several synthetic data sets. Synthetic datasets are widely used in designing and evaluating novel Phase I clinical trials in order to thoroughly investigate the design before subjecting human subjects to a potentially dangerous regime of drugs. Moreover, because a Phase I trial never identifies toxicity probabilities exactly and never identifies the MTD with certainty, a real-world dataset is not "completely realistic" either.

The real-world dataset RW we use is taken from (Bailey et al., 2009), which reports the observations during a real Phase I clinical trial for the combination of two oncology drugs, nilotinib and imatinib. Bailey et al. (2009) constructs a dose-toxicity model based on DLT observations and on prior information about the drugs. The dose combinations consist of 400, 600,

Table 2: True toxicity probabilities in datasets A,B,C,D,RW

	Synthetic A				Synthetic B				Synthetic C				Synthetic D				Real-World			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	0.15	**0.30**	0.45	0.50	0.09	0.12	0.15	**0.30**	0.08	0.15	0.45	0.60	**0.30**	0.42	0.52	0.62	0.13	0.21	**0.30**	0.43
2	0.10	0.15	**0.30**	0.45	0.05	0.10	0.13	0.15	0.05	0.12	**0.30**	0.55	0.10	0.20	**0.30**	0.40	0.08	0.13	0.20	**0.30**
1	0.05	0.10	0.15	**0.30**	0.02	0.08	0.10	0.11	0.02	0.10	0.15	0.50	0.05	0.12	0.20	**0.30**	0.04	0.07	0.11	0.17

and 800 mg of nilotinib (drug A) and 0, 400, 600, and 800 mg of imatinib (drug B). (Further details are in the Supplementary Materials.) The synthetic datasets A,B,C,D are constructed as variations on RW. (Additional synthetic datasets and results are presented in the Supplementary Materials.) Table 2 reports the true toxicities for the various doses in the datasets A,B,C,D,RW; the true MTD's are shown in boldface.

For comparison purposes, we implemented five dose-finding algorithms: SDF-Bayes; DF-Bayes (SDF-Bayes without caution, chosen to illustrate the effect of optimism without caution); SOTA Bayes (Riviere et al., 2014), a state-of-the-art Bayesian dose-finding algorithm; IndepTS (Aziz et al., 2019), a Thompson sampling-based multi-armed bandit (MAB) clinical trial algorithm; and StructMAB, a structured MAB-based clinical trial algorithm that exploits the joint dose-toxicity model based on the structured bandit method from (Gupta et al., 2019) while taking into account the safety constraint, which is an advanced version of the safe dose allocation method in Shen et al. (2020) to drug combinations. In the algorithms that exploit the joint dose-toxicity model (SDF-Bayes, DF-Bayes, SOTA Bayes and StructMAB), we use the logistic model in Riviere et al. (2014) for fair comparison.

5.1 Homogeneous Groups

For each of these datasets, we conducted 5,000 runs of each algorithm, each with a pool of 60 patients, using $\xi = 0.30, \epsilon = 0.05, \delta = 0.05$ (these values are typical of actual Phase I trials). In any run, we counted a recommendation error if the recommended DC is not the true MTD and a safety violation if the DLT observation rate exceeds the threshold $\xi + \epsilon$. Table 3 reports, for each algorithm and each dataset, the proportions of runs in which there was a safety violation and runs in which the recommended DC was an error, with 95% confidence intervals. Safety violations do not satisfy the safety constraint (i.e., the proportion of runs in which there was a safety violation is more than 0.05) are shown in red; in the absence of those safety violations, the best error performance for each dataset is shown in boldface. As can be seen, for *every* dataset, SDF-Bayes satisfies the safety constraint while making fewer recommendation errors than SOTA Bayes, Indep TS, or StructMAB. Indeed, SOTA Bayes is the only one of these algorithms that is at all competitive with SDF-Bayes; Indep TS and StructMAB are recommending the wrong MTD more than half the time. DF-Bayes makes fewer recommendation errors than SDF-Bayes for dataset B, in which *no DC exceeds the threshold* ξ, but makes wildly unacceptable proportions of safety violations for all other datasets.

SDF-Bayes does best because cautious optimism allows it to more efficiently explore the boundary between safe and unsafe DCs. In the Supplementary Materials, we document that SDF-Bayes is more often testing DC's that are believed to be close to being unsafe, while other algorithms more often test DC's that are believed to be safe. Because the primary objective of a Phase I trial is to find the MTD while maintaining acceptable patient safety, which is determined by the trialist, SDF-Bayes is making the proper trade-off between accurate prediction and safety of patients in the trial.

5.2 Heterogeneous Groups

To evaluate SDF-Bayes for heterogeneous groups, we use the groups A, B whose toxicities are given by Synthetic A and Synthetic B, respectively, provided in Table 2. Because we have two groups, we allow for a total of 80 patients. For each simulated trial, we compute both the safety violations for each group and the overall safety violation for the whole trial. (Other details are in the Supplementary Material.)

To evaluate the effectiveness of adaptive recruitment in SDF-Bayes-AR, we use as baselines the various Bayesian algorithms with uniform recruitment (UR), so that patients from each group are recruited with equal probability, and SOTA Bayes-AR in which the proposed adaptive patient recruitment is adopted to SOTA Bayes. (We have already found that the MAB-based algorithms are not competitive, so we do not use them here.) We also apply the Bayesian algorithms to the entire population, treated as a single group called EP, whose true toxicity probabilities are just the averages of the toxicity probabilities for the two groups A and B. At the end of each trial, the algorithms recommend a *single* DC for EP, which we evaluate as a recommendation for each group separately. As before, we use boldface to indicate the best performance subject to satisfying the safety constraints.

Table 3: Safety Violations and Recommendation Errors for datasets A,B,C,D,RW

	Synthetic A		Synthetic B		Synthetic C		Synthetic D		Real-World	
Algorithms	Safety Viol.	Rec. Errors	Safety Viol.	Rec. Errors	Safety Viol.	RecErrors	Safety Viol.	Rec Errors	Safety Viol.	Rec Errors
SDF-Bayes	0.019±.004	**0.205±.011**	0.001±.001	0.193±.009	0.010±.002	**0.443±.012**	0.040±.005	**0.344±.011**	0.003±.001	**0.368±.011**
DF-Bayes	0.411±.014	0.237±.012	0.020±.003	**0.123±.008**	0.434±.012	0.643±.011	0.460±.012	0.317±.011	0.154±.008	0.341±.011
SOTA Bayes	0.023±.004	0.233±.012	0.000±.000	0.196±.009	0.013±.003	0.570±.012	0.041±.005	0.394±.011	0.002±.001	0.391±.011
StructMAB	0.018±.004	0.516±.014	0.000±.000	0.488±.012	0.011±.002	0.740±.010	0.060±.006	0.564±.012	0.002±.001	0.621±.011
IndepTS	0.013±.003	0.603±.014	0.000±.000	0.847±.008	0.006±.002	0.819±.009	0.072±.006	0.623±.011	0.000±.000	0.749±.010

Table 4: Safety Violations and Recommendation Errors

Algorithm	Safety Violation Rate			Recommendation Error Rate		
	Entire trial	Group A	Group B	Average	Group A	Group B
SDF-Bayes-AR	0.000±.000	0.049±.006	0.002±.001	**0.283±.012**	**0.287±.013**	**0.279±.012**
SDF-Bayes-UR	0.000±.000	0.028±.005	0.001±.001	0.288±.013	**0.287±.013**	0.288±.013
DF-Bayes-UR	0.095±.008	0.452±.014	0.025±.004	0.257±.012	0.325±.013	0.188±.011
SOTA Bayes-AR	0.000±.000	0.064±.006	0.000±.000	0.294±.011	0.301±.011	0.287±.011
SOTA Bayes-UR	0.000±.000	0.036±.005	0.000±.000	0.298±.013	0.314±.013	0.283±.012
SDF-Bayes-EP	0.004±.002	0.734±.012	0.000±.000	0.806±.011	0.706±.008	0.906±.008
DF-Bayes-EP	0.139±.010	0.842±.010	0.007±.002	0.819±.011	0.807±.011	0.830±.010
SOTA Bayes-EP	0.003±.002	0.493±.014	0.000±.000	0.800±.011	0.692±.013	0.908±.008

Safety Violations and Recommendation Errors Table 4 shows that SDF-Bayes-AR achieves the lowest overall error rate among the algorithms that satisfy the safety constraint, but that the improvement of SDF-Bayes-AR over SDF-Bayes-UR is marginal. As we will see below, this is because no prior information is used; as we show, adaptive recruitment is more effective when prior information is available.

When we force the algorithms to treat EP as a homogeneous group, we see that the rate of safety violations in group B is extremely low; as a result, the overall rate of safety violations is also low (for SDF-Bayes and SOTA Bayes, applied to EP) although the rate of safety violations for group A is extremely high. We also see that, in this setting, the recommendation errors are all very high. This is because according to the averages of the toxicity probabilities for the two groups A and B, the true MTDs for EP are (2,4) and (3,3), neither of which is an MTD for either group A or group B. This highlights the danger of treating heterogeneous populations as if they were homogeneous.

Impact of Prior Information As we have noted, useful prior information about one or both groups may be available (Hobbs et al., 2011). To illustrate the impact of prior information on the adaptive recruitment in SDF-Bayes-AR and hence on the results, we assume that the prior information comes from a previous trial for group B; we parametrize the amount/quality of prior information by controlling the number of patients T_p in the previous trial. Table 5 records the fraction of patients recruiting in group A, the safety violation rates, and the recommendation error rates of SDF-Bayes-AR and SDF-Bayes-UR for various sizes of prior trials (and hence various amounts of prior information). We can see that SDF-Bayes-AR adaptively recruits the patients according to the amount of the prior information. In SDF-Bayes-AR, patients from group B are seldom recruited after the most likely MTD DC for group B is determined with high probability. (See the Supplementary Material for more detail.) Because patients from group A are recruited more frequently, more is learned about group A and the error rate for group substantially reduced; the cost is only a marginal increase in the error rate for group B. As a result, the overall error rate is improved as expected. Because adaptive recruitment reduces the number of patients recruited from group B, it also reduces the possibility of balancing the risk of group B and increases the *rate* of safety violations in group B – although *fewer* group B patients are exposed to DCs that are found to be unsafe. Overall, SDF-Bayes-AR outperforms SDF-Bayes-UR in both accuracy and safety: it makes fewer total errors with fewer total safety violations.

6 RELATED WORK

Bayesian methodology for clinical trials Bayesian methodology has been widely used for clinical trials, first to label the effectiveness of treatments (Atan et al., 2019; Berry, 2006). It has also been used in dose-finding clinical trials (Wages et al., 2015; O'Quigley et al., 1990) and for drug combination trials (Riviere et al., 2014; Yin and Yuan, 2009a,b; Yan et al., 2017), which are about online learning for dose-finding clinical trials. In the latter setting it is typically used in conjunction with a dose-toxicity model because it can significantly reduce the size

Table 5: Impact of Prior Information (A: Group A, B: Group B, E: Entire trial)

		Fraction Recruited from Group			Safety Violation Rates			Recommendation Errors		
		$T_p = 20$	$T_p = 40$	$T_p = 60$	$T_p = 20$	$T_p = 40$	$T_p = 60$	$T_p = 20$	$T_p = 40$	$T_p = 60$
AR	A	0.517±.001	0.599±.003	0.719±.004	0.027±.004	0.014±.003	0.014±.003	0.286±.013	0.281±.012	0.258±.012
	B	0.483±.001	0.401±.003	0.281±.004	0.047±.006	0.105±.009	0.207±.011	0.233±.012	0.170±.010	0.131±.009
	E	-	-	-	0.001±.001	0.002±.001	0.001±.001	0.259±.012	0.226±.012	0.195±.011
UR	A	0.500±.000	0.500±.000	0.500±.000	0.031±.005	0.031±.005	0.031±.005	0.303±.013	0.303±.013	0.303±.013
	B	0.500±.000	0.500±.000	0.500±.000	0.021±.004	0.039±.005	0.055±.006	0.234±.012	0.162±.010	0.127±.009
	E	-	-	-	0.006±.002	0.011±.003	0.017±.004	0.268±.012	0.232±.012	0.215±.011

Table 6: Comparison of adaptive clinical trial Phase I methodologies

Reference	Principle	No. of DCs	Toxicity Model	Safety	Heterogeneous Groups
Villar et al. (2015); Garivier et al. (2017); Villar and Rosenberger (2018)	MAB	less than 10	No	No	No
Varatharajah et al. (2018)	MAB	less than 10	No	No	Yes
Aboutalebi et al. (2019)	MAB	less than 10	No	Implicitly considered	No
Aziz et al. (2019)	MAB	less than 10	Fixed	Implicitly considered	No
Shen et al. (2020)	MAB	less than 10	Fixed	δ-safety guaranteed	No
Lee et al. (2020)	MAB	less than 10	Fixed	δ-safety guaranteed	Yes
Yan et al. (2017)	Bayesian	less than 10	No	Dose-escalation	No
Wages et al. (2015)	Bayesian	less than 10	Fixed	Dose-escalation	Yes
Yin and Yuan (2009a); Riviere et al. (2014)	Bayesian	more than 10	Fixed	Dose-escalation	No
This work (SDF-Bayes)	Bayesian	more than 10	Arbitrary	δ-safety guaranteed	Yes

of the search space (Shen et al., 2019). By using the posterior distribution of the parameters of the dose-toxicity model, traditional Bayesian DC-finding algorithms are proposed to find the MTD based on dose escalation and de-escalation from the lowest dose levels for safety (Riviere et al., 2014; Yin and Yuan, 2009a; Yan et al., 2017).

MAB for clinical trials MABs have also been widely proposed for dose-finding clinical trials, especially for trials with a single drug (Aziz et al., 2019; Shen et al., 2020; Varatharajah et al., 2018; Villar and Rosenberger, 2018; Lee et al., 2020). Nevertheless, they do not address the challenging issues of Phase I clinical trials for multiple drugs, which operate in the small sample size regime relative to the number of potential doses. In trials for drug combination with more than ten potential DCs, their *asymptotic* optimality is not meaningful for a practical number (fewer than 100) of patients in Phase I trials. In MAB models, various safety management methods have been studied to address safety issue. However, they deal with the safety issue only implicitly or cannot be applied to the drug combination setting because they rely on a very limited dose-toxicity model with a single drug.

Methods for heterogeneous groups For both Bayesian and MAB methods, there has been only limited work involving heterogeneous groups. In Wages et al. (2015), a well-known Bayesian continual reassessment method (CRM) is extended for an adaptive clinical trial design for heterogeneous groups. However, it does not consider drug combinations and a patient recruitment with a limited number of patients. In Atan et al. (2019), a patient recruitment with a limited number of patients is adaptively determined based on a Bayesian knowledge gradient policy, but its goal is to label the effectiveness of drugs as opposed to dose-finding. (Varatharajah et al., 2018; Lee et al., 2020) adapt a contextual MAB for clinical trials with heterogeneous groups by treating groups as contexts. However, this work does not consider either drug combinations or patient recruitment with a limited number of patients.

Table 6 provides a summary comparison of our work with other methodologies.

7 CONCLUSION

In this paper, we have studied the problem of designing Phase I clinical trials for drug combinations. We have enunciated a principle of cautious optimism and proposed the SDF-Bayes algorithm that applies that principle to effectively balance the trade-off between the exploration of drug combinations and the risk of safety violation. For settings with identified heterogeneous groups, we proposed an extension SDF-Bayes-AR in which both the DC to be allocated and the group from which the next patient is to be recruited are chosen adaptively. On the basis of experiments, we demonstrated that our proposed algorithms outperform previous state-of-the-art algorithms.

Acknowledgments

CS acknowledges the funding support from Kneron, Inc. and the Virginia Commonwealth Cyber Initiative (CCI) Cybersecurity Research Collaboration Grant. We thank all reviewers for their comments and suggestions.

References

Aboutalebi, H., Precup, D., and Schuster, T. (2019). Learning modular safe policies in the bandit setting with application to adaptive clinical trials. *arXiv preprint arXiv:1903.01026*.

Atan, O., Zame, W. R., and Schaar, M. (2019). Sequential patient recruitment and allocation for adaptive clinical trials. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Aziz, M., Kaufmann, E., and Riviere, M.-K. (2019). On multi-armed bandit designs for dose-finding clinical trials. *arXiv preprint arXiv:1903.07082*.

Bagatell, R., Norris, R., Ingle, A. M., Ahern, C., Voss, S., Fox, E., Little, A. R., Weigel, B. J., Adamson, P. C., and Blaney, S. (2014). Phase 1 trial of temsirolimus in combination with irinotecan and temozolomide in children, adolescents and young adults with relapsed or refractory solid tumors: a children's oncology group study. *Pediatric blood & cancer*, 61(5):833–839.

Bailey, S., Neuenschwander, B., Laird, G., and Branson, M. (2009). A bayesian case study in oncology phase I combination dose-finding using logistic regression with covariates. *Journal of biopharmaceutical statistics*, 19(3):469–484.

Bamias, A., Aravantinos, G., Kastriotis, I., Alivizatos, G., Anastasiou, I., Christodoulou, C., Gyftaki, R., Kalofonos, H. P., and Dimopoulos, M. A. (2011). Report of the long-term efficacy of two cycles of adjuvant bleomycin/etoposide/cisplatin in patients with stage I testicular nonseminomatous germ-cell tumors (NSGCT): a risk adapted protocol of the Hellenic Cooperative Oncology Group. *Urologic Oncology: Seminars and Original Investigations*, 29(2):189–193.

Berry, D. A. (2006). Bayesian clinical trials. *Nature reviews Drug discovery*, 5(1):27.

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122.

Calvo, E., Soria, J.-C., Ma, W. W., Wang, T., Bahleda, R., Tolcher, A. W., Gernhardt, D., O'Connell, J., Millham, R., Giri, N., et al. (2017). A phase I clinical trial and independent patient-derived xenograft study of combined targeted treatment with dacomitinib and figitumumab in advanced solid tumors. *Clinical Cancer Research*, 23(5):1177–1185.

Dasari, A., Gore, L., Messersmith, W., Diab, S., Jimeno, A., Weekes, C., Lewis, K., Drabkin, H., Flaig, T., and Camidge, D. (2013). A phase I study of sorafenib and vorinostat in patients with advanced solid tumors with expanded cohorts in renal cell carcinoma and non-small cell lung cancer. *Investigational new drugs*, 31(1):115–125.

Flaherty, K. T., Infante, J. R., Daud, A., Gonzalez, R., Kefford, R. F., Sosman, J., Hamid, O., Schuchter, L., Cebon, J., Ibrahim, N., et al. (2012). Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *New England Journal of Medicine*, 367(18):1694–1703.

Garivier, A., Ménard, P., Rossi, L., and Menard, P. (2017). Thresholding bandit for dose-ranging: The impact of monotonicity. *arXiv preprint arXiv:1711.04454*.

Gasparini, M. (2013). General classes of multiple binary regression models in dose finding problems for combination therapies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1):115–133.

Gilks, W. R., Best, N. G., and Tan, K. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(4):455–472.

Gupta, S., Chaudhari, S., Mukherjee, S., Joshi, G., and Yağan, O. (2019). A unified approach to translate classical bandit algorithms to the structured bandit setting. *arXiv preprint arXiv:1810.08164*.

Hamberg, P., Ratain, M. J., Lesaffre, E., and Verweij, J. (2010). Dose-escalation models for combination phase I trials in oncology. *European Journal of Cancer*, 46(16):2870–2878.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056.

Kelly, W. K. and Halabi, S. (2018). *Oncology clinical trials: successful design, conduct, and analysis*. Springer Publishing Company.

Kim, T., Sym, S., Lee, S., Ryu, M., Lee, J., Chang, H., Kim, H., Shin, J., Kang, Y., and Lee, J. (2009). A UGT1A1 genotype-directed phase I study of irinotecan (CPT-11) combined with fixed dose of capecitabine in patients with metastatic colorectal cancer (mCRC). *Journal of Clinical Oncology*, 27(15_suppl):2554–2554.

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

Moss, R. B., Flume, P. A., Elborn, J. S., Cooke, J., Rowe, S. M., McColley, S. A., Rubenstein, R. C., and Higgins, M. (2015). Efficacy and safety of ivacaftor treatment: randomized trial in subjects with cystic fibrosis who have an R117H-CFTR mutation. *The Lancet. Respiratory medicine*, 3(7):524.

Ocana, A., Gil-Martin, M., Antolín, S., Atienza, M., Montaño, Á., Ribelles, N., Urruticoechea, A., Falcón, A., Pernas, S., Orlando, J., et al. (2019). Efficacy and safety of dasatinib with trastuzumab and paclitaxel in first line HER2-positive metastatic breast cancer: results from the phase II GEICAM/2010-04 study. *Breast cancer research and treatment*, 174(3):693–701.

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

Paller, C. J., Huang, E., Luechtefeld, T., Masset, H., Williams, C., Zhao, J., Gravell, A., Tamashiro, T., Reeves, S., Rosner, G., et al. (2019). Factors affecting combination trial success (FACTS): Investigator survey results on early-phase combination trials. *Frontiers in Medicine*, 6:122.

Pallmann, P., Bedding, A. W., Choodari-Oskooei, B., Dimairo, M., Flight, L., Hampson, L. V., Holmes, J., Mander, A. P., Sydes, M. R., Villar, S. S., et al. (2018). Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC medicine*, 16(1):29.

Park, J. J., Thorlund, K., and Mills, E. J. (2018). Critical concepts in adaptive clinical trials. *Clinical epidemiology*, 10:343.

Plummer, R., Jones, C., Middleton, M., Wilson, R., Evans, J., Olsen, A., Curtin, N., Boddy, A., McHugh, P., Newell, D., et al. (2008). Phase I study of the poly (ADP-ribose) polymerase inhibitor, AG014699, in combination with temozolomide in patients with advanced solid tumors. *Clinical cancer research*, 14(23):7917–7923.

Riviere, M.-K., Yuan, Y., Dubois, F., and Zohar, S. (2014). A bayesian dose-finding design for drug combination clinical trials based on the logistic model. *Pharmaceutical statistics*, 13(4):247–257.

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

Shen, Y., Liu, T., Chen, J., Li, X., Liu, L., Shen, J., Wang, J., Zhang, R., Sun, M., Wang, Z., et al. (2019). Harnessing artificial intelligence to optimize long-term maintenance dosing for antiretroviral-naive adults with HIV-1 infection. *Advanced Therapeutics*, page 1900114.

Sun, Z. and Braun, T. M. (2015). A two-dimensional biased coin design for dual-agent dose-finding trials. *Clinical Trials*, 12(6):596–607.

Varatharajah, Y., Berry, B., Koyejo, S., and Iyer, R. (2018). A contextual-bandit-based approach for informed decision-making in clinical trials. *arXiv preprint arXiv:1809.00258*.

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

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

Wages, N. A., Read, P. W., and Petroni, G. R. (2015). A phase I/II adaptive design for heterogeneous groups with application to a stereotactic body radiation therapy trial. *Pharmaceutical statistics*, 14(4):302–310.

Wu, M., Sirota, M., Butte, A. J., and Chen, B. (2014). Characteristics of drug combination therapy in oncology by analyzing clinical trial data on ClinicalTrials. gov. In *Pacific Symposium on Biocomputing Co-Chairs*, pages 68–79. World Scientific.

Yan, F., Mandrekar, S. J., and Yuan, Y. (2017). Keyboard: a novel bayesian toxicity probability interval design for phase I clinical trials. *Clinical Cancer Research*, 23(15):3994–4003.

Yin, G. and Yuan, Y. (2009a). Bayesian dose finding in oncology for drug combinations by copula regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(2):211–224.

Yin, G. and Yuan, Y. (2009b). A latent contingency table approach to dose finding for combinations of two agents. *Biometrics*, 65(3):866–875.