Contextual Constrained Learning for Dose-Finding Clinical Trials

Hyun-Suk Lee
University of Cambridge
Yonsei University

Cong Shen
University of Virginia

James Jordon
University of Oxford

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

Abstract

Clinical trials in the medical domain are constrained by budgets. The number of patients that can be recruited is therefore limited. When a patient population is heterogeneous, this creates difficulties in learning subgroup specific responses to a particular drug and especially for a variety of dosages. In addition, patient recruitment can be difficult by the fact that clinical trials do not aim to provide a benefit to any given patient in the trial. In this paper, we propose C3T-Budget, a contextual constrained clinical trial algorithm for dose-finding under both budget and safety constraints. The algorithm aims to maximize drug efficacy *within* the clinical trial while *also* learning about the drug being tested. C3T-Budget recruits patients with consideration of the remaining budget, the remaining time, and the characteristics of each group, such as the population distribution, estimated expected efficacy, and estimation credibility. In addition, the algorithm aims to avoid unsafe dosages. These characteristics are further illustrated in a simulated clinical trial study, which corroborates the theoretical analysis and demonstrates an efficient budget usage as well as a balanced learning-treatment trade-off.

1 INTRODUCTION

Traditionally, dose-finding studies are early stage clinical trials used to evaluate the safety of a new treatment, and are often referred to as Phase I clinical trials. Thus, traditional dose-finding methods aim to find the most appropriate dose for use in further phases of clinical trials by considering only toxicity of the treatment (Storer, 1989; O'Quigley et al., 1990). Recently, however, both toxicity and efficacy have been considered in dose-finding methods to accelerate the development process of new treatments and to reduce costs (Zang et al., 2014). As the efficacy is considered in dose-finding, the dose-finding methods should be designed for both *clinical research* (i.e., to find the appropriate dose) and *clinical practice* (i.e., to benefit the patients participating) (Berry et al., 2004). This gives rise to an inevitable trade-off between information gathering and treatment effectiveness, which makes the dose-finding problem more complex.

Heterogeneous groups of patients also make the dose-finding problem more complex (Wages et al., 2015). In many cases, the efficacy and toxicity of a dose will vary between groups as shown in Table 1. Thus, with the heterogeneous groups, the goal of dose-finding clinical trials is to find the most appropriate dose for each group. On the other hand, the presence of heterogeneous responses can also be a blessing – a drug need only be effective for a single subgroup for it to be approved. Furthermore, in general, different groups will be of different sizes (i.e. a different number of people belong to each group on a population level), which may result in inaccurate dose-finding for under-represented groups (Hussain-Gambles et al., 2004). To avoid this, an algorithm for dose-finding should account for differing group sizes (which manifest themselves in terms of patient enrolment/arrival rates).

In clinical trials, there are several ethical constraints involved with treating human subjects (O'Quigley et al., 1990) such as a limited number of patients and a careful consideration for the safety of the patients. These constraints highly motivate a clinical trial that effectively utilizes the limited budget for successfully balancing the aforementioned trade-off while avoiding unsafe doses. It is worth noting that such a clinical trial might be more ethical than the traditional methods since it more effectively uses the budget and exposes fewer participants to unsafe doses (Park et al., 2018; Pallmann

Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

Table 1: Examples of Studies with Heterogeneous Groups in the Literature

Study	Treatment	Groups	Results
(Polyzos, 2008)	Sunitinib	Type of cancer	37% RR, 48% SD for renal cell carcinoma, 8% RR, 70% SD for gastrointestinal stromal tumors, 9% RR, 19% SD for lung cancer (RR: response rate, SD: stable disease)
(Kim et al., 2009)	Irinotecan	3 Groups according to genetic characteristics	350 mg/m^2 for group 0 and 1, 200 mg/m^2 for group 2
(Dasari et al., 2013)	Sorafenib 400 mg bid with vorinostat 300 mg daily days 1–14 of a 21-day cycle	Type of cancer	Treatment is not tolerable in renal cell carcinoma and non-small cell lung cancer patients while tolerable in other cancer types
(Moss et al., 2015)	Ivacaftor	Adult and children	Treatment significantly improves lung function only in adult patients

et al., 2018). In particular, with a joint consideration of heterogeneous groups and limited budget, the relationship between clinical research and clinical practice should be carefully studied. From the point of view of *research*, a successful clinical trial with a given limited budget should concentrate on the groups for which the estimate of the optimal dose is poor. However, a new treatment may not be effective for all groups as shown in Table 1. Thus, from the point of view of *practice*, precious budget is being wasted if the trial concentrates on improving estimates for groups for which the new treatment is ineffective. Moreover, the differing populations of each group should also be considered while balancing the trade-off, since what can be achieved, in terms of both research and practice, depends on the population (Hussain-Gambles et al., 2004).

In this paper, we study a contextual constrained clinical trial (C3T) problem for dose-finding with both budget and safety constraints. The goal is to effectively balance the aforementioned trade-off while avoiding unsafe doses. In the C3T problem, patients arrive sequentially (with arrival rates for different groups assumed to be different), and an agent that administrates the clinical trial determines how a new patient will be treated. In other words, in the C3T problem, the agent's first decision is whether or not the patient is treated (i.e. accepted into the trial), and if so, the agent should then determine the dose that will be allocated to the patient.

To answer these two questions, we propose a novel contextual multi-armed bandit (MAB) algorithm for C3Ts with budget constraint called *C3T-Budget*. In the algorithm, different patient groups are modeled as the context. In each round, to maximize the effectiveness of *clinical practice*, the dose for each group is chosen to maximize the expected efficacy while satisfying the safety constraint based on information learned from previously treated patients. Then, given the chosen dose, the algorithm determines whether the patient is treated or not with the aim of maximizing the information from *clinical research*. Specifically, we provide a linear programming (LP) approximation to the C3T problem based on the optimal dose, remaining budget, remaining time, and population of the groups. With the LP approximation, we can determine whether to treat a patient or not by maximizing the accuracy improvement of the efficacy estimation, for which a Bayesian approach is adopted. C3T-Budget can, therefore, systematically balance the trade-off between treatment and learning by recruiting patients from appropriate subgroups. However, when the budget is very limited, it is hard to balance the trade-off for all subgroups. In this case, clinical trials should focus on subgroups with high efficacies since an attempt to focus on all subgroups may lead to a completely failed trial. To address this, we propose a modified version of C3T-Budget called *C3T-Budget-E*, which concentrates on the subgroups with estimated high efficacies. Through numerical results, we show that our algorithms outperform state of the art in terms of efficacy, safety, and recommendation errors.

1.1 Related Works

MAB models are particularly appropriate for clinical trials since the trade-off between *clinical research* and *clinical practice* can be seen as the well-known tradeoff between *exploration* and *exploitation* in the context of MAB, which has been very well-studied. Under this pretext, the application of novel MAB models tailored to clinical trials has been widely studied. In Villar et al. (2015), a patient allocation strategy for clinical trials is proposed based on the Gittins index. Villar and Rosenberger (2018) extend the strategy as the modified forward looking Gittins index to utilize the covariate information. In Aboutalebi et al. (2019), a regret model that considers the safety of dosages is proposed. Based on the regret model, a dose-finding algorithm is developed, and the regret model helps the algorithm avoid unsafe dosages when finding the optimal dosage. In Aziz et al. (2019), a Thompson sampling-based Phase I clinical trial algorithm is proposed. In the algorithm, both efficacy and toxicity are considered when finding the optimal dosage. However, none of these algorithms consider heterogeneous groups of patients.

As the interest in heterogeneous groups of patients in clinical trials has rapidly grown, clinical trial designs that consider heterogeneous groups have begun

to see some attention. In Wages et al. (2015), an adaptive clinical trial design for heterogeneous groups was proposed by extending a well-known continual reassessment method (CRM). However, this work does not consider a limited budget. Another adaptive clinical trial algorithm with heterogeneous groups is developed in Atan et al. (2019), which is based on the knowledge gradient policy. In this work, the patient recruitment is determined by considering the budget, but all groups are assumed to have the same arrival distribution. Moreover, the primary goal is to label the effectiveness of treatments, not dose-finding. To the best of our knowledge, Varatharajah et al. (2018) is the only work using contextual MAB for clinical trials with heterogeneous groups, and is is probably the closest to our work. In the algorithm, a simple contextual MAB model is applied to clinical trials by considering groups as contexts. However, the algorithm does not address several of the challenges associated with heterogeneous groups such as the differing populations of the groups and the limited budget.

2 CONTEXTUAL CONSTRAINED CLINICAL TRIAL (C3T) MODEL

We consider a contextual constrained clinical trial (C3T) for dose-finding with S subgroups (contexts) over a time-horizon, T, and limited budget, B, indicating the maximum number of patients that can be admitted into the trial. The set of subgroups is defined as $\mathcal{S} = \{1, 2, ..., S\}$. We consider K candidate doses, with the set of candidate doses given by $\mathcal{K} = \{1, 2, ..., K\}$. The efficacy, $X_{s,k}$, and toxicity, $Y_{s,k}$, of dose k for subgroup s[1] are modelled as Bernoulli random variables with unknown parameters $q_{s,k}$ and $p_{s,k}$ respectively. $X_{s,k} = 1$ indicates that dose k is effective for subgroup s and $Y_{s,k} = 1$ indicates that dose k is unsafe for subgroup s. We define a dose to be unsafe for subgroup s if the expected toxicity ($p_{s,k}$) for subgroup s exceeds a pre-defined toxicity threshold ζ, which is referred to as the MTD (maximum tolerated dose) threshold in the clinical trial literature. We also define a minimum efficacy threshold, θ, below which we do not wish to administrate treatment because the efficacy is too low.

For $p_{s,k}$, we introduce the logistic dose-toxicity model $p_{s,k}(a) = \left(\frac{\tanh u_k + 1}{2}\right)^a$, as in O'Quigley et al. (1990), where u_k is an actual dose level of dose k and a is a parameter used for all doses. Then, due to the monotonicity of the dose-toxicity model, we can partition doses into safe/unsafe doses as $\{p_{s,1}, ..., p_{s,U_s}\}$ and $\{p_{s,U_s+1}, ..., p_{s,K}\}$, where $U_s = \max\{k \in \mathcal{K} : p_{s,k} < \zeta\}$.

[1] For ease of exposition, we use "subgroup s" and "patient in subgroup s" interchangeably in the rest of the paper.

At the beginning of each round, t, of the clinical trial, a patient arrives. We denote the subgroup of the patient in round t by $H(t) \in \mathcal{S}$. The arrival rate of the patients in each subgroup s is given by ξ_s. Then, the probability of the patient in round t being from subgroup s is given by $\pi_s = \mathbb{P}(H(t) = s) = \frac{\xi_s}{\sum_{s \in \mathcal{S}} \xi_s}$. For the observed subgroup $H(t)$, an agent of the clinical trial chooses a dose $k \in \{0\} \cup \mathcal{K}$, where "0" represents a "no-dose" action corresponding to the agent skipping the new patient. We denote the dose allocated in round t by $I(t)$. Let X_t, Y_t, and Z_t be the efficacy, toxicity, and cost in round t, respectively. When $I(t) \geq 1$, we have $X_t = X_{H(t),I(t)}$, $Y_t = Y_{H(t),I(t)}$, and $Z_t = 1$ (we will generalize this to heterogeneous costs across subgroups in Section 4.1). When $I(t) = 0$, we have no efficacy or toxicity and the cost $Z_t = 0$. The clinical trial ends when the budget is exhausted or at the end of time-horizon T.

3 C3T WITH LIMITED BUDGET

3.1 Problem Formulation

Let Π be a bandit algorithm that maps historical observations, $\{(X_\tau, Y_\tau, H(\tau), I(\tau)\}_{\tau=1}^{t-1}$, and the current subgroup, $H(t)$, to a dose $I(t) \in \{0\} \cup \mathcal{K}$. We define the average toxicity of subgroup s to be $S_{\Pi,s}(T, B) = \mathbb{E}\left[\frac{\sum_{t=1}^{T} \mathbb{I}\{H(t)=s\} Y_t}{\sum_{t=1}^{T} \mathbb{I}\{H(t)=s\} \mathbb{I}\{I(t)\neq 0\}}\right]$. We define the total expected cumulative efficacy as $E_\Pi(T, B) = \mathbb{E}\left[\sum_{t=1}^{T} X_t\right]$. To make a recommendation, the MTD threshold and minimum efficacy threshold should be considered. We define the set of candidate doses for subgroup s by $\mathcal{K}_s = \{k \in \mathcal{K} : q_{s,k} \geq \theta, p_{s,k} \leq \zeta\}$. We define the optimal dose-to-recommend for subgroup s by

$$k_s^* = \begin{cases} \text{argmax}_{k \in \mathcal{K}_s} q_{s,k}, & \text{if } \mathcal{K}_s \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

We define the recommended dose to subgroup s at the end of the clinical trial as $\hat{k}_s^*(T, B)$. Then, the dose recommendation error can be written as $D_\Pi(T, B) = \sum_{s \in \mathcal{S}} \mathbb{E}\left[\mathbb{I}[\hat{k}_s^*(T, B) \neq k_s^*]\right]$. Our C3T problem of minimizing the dose recommendation error while satisfying the budget and safety constraints is formally presented as

$$\begin{aligned} \text{minimize} \quad & D_\Pi(T, B) \\ \text{subject to} \quad & \mathbb{P}\left[S_{\Pi,s}(T, B) \leq \zeta\right] \geq 1 - \delta_s, \ \forall s \in \mathcal{S} \\ & \textstyle\sum_{t=1}^{T} Z_t \leq B. \end{aligned} \quad (2)$$

where δ_s is the maximum probability with which the toxicity for subgroup s can exceed the MTD threshold.

3.2 Addressing the Budget Constraint

In clinical trials with a single subgroup, maximizing the cumulative efficacy with the well-known principle of optimism in the face of uncertainty is also effective

to recommend dose (Aziz et al., 2019). Such objective makes the agent more frequently choose the doses that are more likely to be the optimal dose. The dose recommendation accuracy is improved since the accuracy for any dose depends primarily on the number of patients that were given that dose. Since patients are more likely to receive the optimal dose, the accuracy of the optimal dose is better. However, in the case of multiple subgroups and limited budget, maximizing the cumulative efficacy is no longer effective at recommending doses across *all* subgroups due to the varying efficacies and arrival rates. For example, if one subgroup has much higher efficacy than others, the optimal strategy for the agent would involve skipping all patients that arrive from subgroups other than the one with the highest efficacy. Little will therefore be learned about the response of other subgroups to the drug, resulting in a poor learning performance. Hence, in each round, the agent should decide to skip the new patient or not based *also* on the estimation accuracy for the patient's subgroup, remaining budget, and remaining time.

We consider an oracle problem with budget B and time-horizon T to highlight how to address the budget constraint. We assume S subgroups with different arrival rates π_s, $s = 1,...,S$. Each subgroup s has an arbitrary expected reward d_s^*. For convenience, we assume $d_1^* > d_2^* > ... > d_S^*$. In general, the expected rewards d_s^* are unknown to the agent. However, we begin by considering an oracle agent that knows d_s^* and has only two actions: skip or accept the patient. In each round t, the agent will determine whether to skip the new patient or not based on the remaining rounds $\tau = T - t + 1$ and the remaining budget b_τ.

Although the oracle problem seems simple, it is known to be computationally intractable (Wu et al., 2015). We thus approximate this problem as a linear program (LP). We first relax the problem by substituting the hard budget constraint with an average budget constraint using $\rho = \frac{B}{T}$. We define the probability that the agent does not skip a patient in subgroup s by ψ_s, and write $\boldsymbol{\psi} = \{\psi_1, \psi_2, ...\psi_S\}$. We now formulate the following LP problem:

$$\underset{\boldsymbol{\psi}}{\text{maximize}} \sum_{s\in\mathcal{S}} \psi_s \pi_s d_s^* \quad \text{subject to} \sum_{s\in\mathcal{S}} \psi_s \pi_s \le \rho. \tag{3}$$

The optimal solution of this problem can be easily derived:

$$\psi_s(\rho) = \begin{cases} 1, & \text{if } 1 \le s \le \tilde{s}(\rho), \\ \frac{\rho - \eta_{\tilde{s}(\rho)}}{\pi_{\tilde{s}(\rho)+1}}, & \text{if } s = \tilde{s}(\rho)+1, \\ 0, & \text{if } s > \tilde{s}(\rho)+1, \end{cases} \tag{4}$$

where $\eta_s = \sum_{s'=1}^{s} \pi_{s'}$ (with the convention that $\eta_0 = 0$) and $\tilde{s}(\rho) = \max\{s \in \mathcal{S} \cup \{0\} : \eta_s \le \rho\}$.

We can adapt this policy to our more general problem by deciding to skip a patient or not by replacing d_s^* and ρ in (3) with the estimated recommendation accuracy or estimated efficacy and the remaining average budget $\rho_\tau = \frac{b_\tau}{\tau}$, respectively.

3.3 C3T-Budget

In Algorithm 1, we propose a bandit solution for the budget-limited C3T problem in (2) which we call C3T-Budget. As noted above, when there is only 1 subgroup, maximizing the cumulative efficacy using the optimism principle is effective for recommending dose. Thus, in the algorithm, the dose for each subgroup is chosen according to the upper confidence bound (UCB) principle using the estimated efficacy and toxicity of the subgroup. On the other hand, the decision to skip a patient is based on how convincing the estimated efficacy of the chosen dose for each subgroup is.

Algorithm 1 C3T-Budget

Input: Time-horizon T, budget B, subgroup arrival distributions π_s's, ϕ for credible intervals
Initialize: $\tau = T$, $b = B$, $t = 1$, $\alpha_{s,k}^{\text{Beta}}(0) = 1$, $\beta_{s,k}^{\text{Beta}}(0) = 1, \forall s \in \mathcal{S}, \forall k \in \mathcal{K}$.
while $t \le T$ **do**
  $\hat{a}_s(t) \leftarrow \frac{\sum_{k=1}^{K} \hat{a}_{s,k}(t-1) N_{s,k}(t-1)}{N_s(t-1)}, \forall s \in \mathcal{S}$
  $\mathcal{K}_s(t) = \{k \in \mathcal{K}: \hat{q}_{s,k}(t) \ge \theta, p_{s,k}(\hat{a}_s(t) + \alpha_s(t)) \le \zeta\}, \forall s \in \mathcal{S}$
  if $b > 0$ **then**
    if $N_{H(t)}(t) \le K$ **then**
      Sample each dose once $I(t) = N_{H(t)}(t)$
    else
      $k_s^*(t) \leftarrow \text{argmax}_{k \in \mathcal{K}_s(t)} \hat{q}_{s,k}(t), \forall s \in \mathcal{S}$
      Calculate $B_s^*(t)$ as in (5)
      Obtain $\hat{\boldsymbol{\psi}}(b/\tau)$'s by solving the LP problem in (3) with ordered $B_s^*(t)$'s
      Allocate dose
      $$I(t) = \begin{cases} k_{H(t)}^*(t), & \text{with probability } \hat{\psi}_{H(t)}(b/\tau), \\ 0, & \text{otherwise.} \end{cases}$$
    end if
  end if
  Observe the efficacy X_t and toxicity Y_t
  Update τ, b, $N_s(t)$, $N_{s,k}(t)$, $\bar{q}_{s,k}(t)$, $\hat{q}_{s,k}(t)$, $\bar{p}_{s,k}(t)$, $\alpha_{s,k}^{\text{Beta}}(t)$, $\beta_{s,k}^{\text{Beta}}(t)$
  $\hat{a}_{s,k}(t) \leftarrow \text{argmin}_a |p_{s,I(t)}(a) - \bar{p}_{s,I(t)}(t)|, \forall s \in \mathcal{S}, \forall k \in \mathcal{K}$
  $t = t + 1$
end while
Output: Recommended dose
$$\hat{k}_s^* = \begin{cases} \text{argmax}_{k \in \mathcal{K}_s'(T)} \bar{q}_{s,k}, & \text{if } \mathcal{K}_s'(T) \ne \emptyset, \\ 0, & \text{otherwise} \end{cases}, \forall s \in \mathcal{S}$$

We define the empirical efficacy estimation of dose k for subgroup s at round t by $\bar{q}_{s,k}(t) = \frac{\sum_{\tau=1}^{t} \mathbb{I}\{H(\tau)=s, I(\tau)=k\} X_\tau}{N_{s,k}(t)}$, where $N_{s,k}(t)$ is the number of times that dose k has been allocated to subgroup s up to round t (i.e., $N_{s,k}(t) = \sum_{\tau=1}^{t} \mathbb{I}\{H(\tau) = s, I(\tau) = k\}$). Similarly, we define the empirical toxicity estimation of dose k for subgroup s at round t by $\bar{p}_{s,k}(t) = \frac{\sum_{\tau=1}^{t} \mathbb{I}\{H(\tau)=s, I(\tau)=k\} Y_\tau}{N_{s,k}(t)}$.

Then, the UCB of $q_{s,k}$ at round t is defined as $\hat{q}_{s,k}(t) = \bar{q}_{s,k} + \sqrt{\frac{c \log N_s(t)}{N_{s,k}(t-1)}}$, where $N_s(t)$ is the number of times that subgroup s has arrived up to round t (i.e., $N_s(t) = \sum_{\tau=1}^{t} \mathbb{I}\{H(\tau) = s\}$). The confidence interval a_s for the dose-toxicity model of subgroup s is given by $\alpha_s(t) = CK\left(\frac{\log \frac{2K}{\delta_s}}{2N_s(t)}\right)^{\frac{\gamma}{2}}$, where C and γ are hyper-parameters of our model. Details of hyper-parameter selection are given in the supplementary material.

In each round t, the algorithm constructs a set of candidate doses for each subgroup s, $\mathcal{K}_s$, by considering the estimated expected efficacy and toxicity of each dose for the subgroup and using the UCB principle. Among the candidate doses, the algorithm selects the estimated optimal dose in round t, $k_s^*(t)$, which has the largest UCB of the expected efficacy for subgroup s, as $k_s^*(t) = \text{argmax}_{k \in \mathcal{K}_s(t)} \hat{q}_{s,k}(t)$. Then, the agent determines whether the patient in round t is to be skipped or not by considering how convincing the estimation of the efficacy of k_s^* is. To do this, we use the credible interval of the estimation of $\bar{q}_{s,k}$.

We adopt a Bayesian approach to estimate $\bar{q}_{s,k}$. We first consider a uniform distribution (i.e., Beta(1,1)) as the prior for $q_{s,k}$. Then, the maximum posterior estimation of $q_{s,k}$ becomes the empirical efficacy estimation $\bar{q}_{s,k}$. This allows us to combine the Bayesian approach and the UCB principle without conflict. We define the parameters of the prior distribution of $q_{s,k}$ in round t as $\alpha_{s,k}^{\text{Beta}}(t)$ and $\beta_{s,k}^{\text{Beta}}(t)$. Then, at the end of round t, we update the posterior distribution of $q_{H(t),I(t)}$ as $\alpha_{H(t),I(t)}^{\text{Beta}}(t) = \alpha_{H(t),I(t)}^{\text{Beta}}(t-1) + X(t)$ and $\beta_{H(t),I(t)}^{\text{Beta}}(t) = \beta_{H(t),I(t)}^{\text{Beta}}(t-1) + (1 - X(t))$. By using the posterior distribution of $q_{s,k}$, we can obtain a credible interval for a given probability ϕ. Let $f(\phi, \alpha^{\text{Beta}}, \beta^{\text{Beta}})$ be a function that calculates the interval length to achieve the probability ϕ. We then define the expected improvement of the credible interval for an additional patient in subgroup s with dose k as

$$B_{s,k} = \bar{q}_{s,k}(f(\phi, \alpha_{s,k}^{\text{Beta}}, \beta_{s,k}^{\text{Beta}}) - f(\phi, \alpha_{s,k}^{\text{Beta}} + 1, \beta_{s,k}^{\text{Beta}})) + (1 - \bar{q}_{s,k})(f(\phi, \alpha_{s,k}^{\text{Beta}}, \beta_{s,k}^{\text{Beta}}) - f(\phi, \alpha_{s,k}^{\text{Beta}}, \beta_{s,k}^{\text{Beta}} + 1)). \quad (5)$$

This value quantifies the improvement in estimation accuracy provided by including an additional patient from subgroup s into the trial. Thus, we can use it to determine which subgroups should be skipped to maximize the dose-recommendation accuracy. Specifically, in round t, we rearrange the set $\{B_{s,k_s^*(t)}(t) : s = 1, ..., S\}$ in descending order. The algorithm then solves the LP problem in (3), substituting d_s^* and ρ with $B_{s,k_s^*(t)}(t)$ and ρ_τ, respectively, and obtains a vector of probabilities $\hat{\psi}(\rho_\tau)$. Finally, at the end of the clinical trial, the dose recommended to subgroup s is given by

$$\hat{k}_s^*(T, B) = \begin{cases} \text{argmax}_{k \in \mathcal{K}_s(T)} \bar{q}_{s,k}, & \text{if } \mathcal{K}_s'(T) \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

3.4 Extension: C3T-Budget-E

In a clinical trial where the budget is particularly limited, identifying even just one subgroup for which the drug is effective can be more important than understanding its effect on the whole population of interest. Trying to learn on the whole population might lead to the responses *all* subgroups being poorly estimated and the drug failing to progress at all. Thus, a different problem formulation is used when the budget is small, allowing the trial to focus on subgroups with high efficacies:

$$\begin{aligned} &\text{maximize} \quad E_\Pi(T, B) \\ &\text{subject to} \quad \mathbb{P}\left[S_{\Pi,s}(T, B) \leq \zeta\right] \geq 1 - \delta_s, \ \forall s \in \mathcal{S} \\ &\qquad\qquad \textstyle\sum_{t=1}^{T} Z_t \leq B. \end{aligned} \quad (6)$$

where we have simply substituted the objective function $D_\Pi(T, B)$ in (2) with $E_\Pi(T, B)$. With this formulation, the agent tries to accurately recommend doses for subgroups with high efficacies while also achieving high total cumulative efficacy.

We now propose a bandit algorithm for the budget-limited C3T problem in (6) to maximize the total cumulative efficacy called C3T-Budget-E. This algorithm focuses more on the subgroups that have high efficacies, which might result in inaccurate dose recommendation for subgroups with low efficacies.

C3T-Budget-E is a modified version of C3T-Budget. The key difference is that the decision to skip a patient is based on the expected efficacy of each subgroup rather than the expected improvement of the credible interval. Hence, in C3T-Budget-E, the parameters for the Bayesian posterior distribution are not used. Specifically, the algorithm first finds the largest UCB of the estimated expected efficacy for subgroup s in round t as $\hat{q}_s^*(t) = \max_{k \in \mathcal{K}_s(t)} \hat{q}_{s,k}(t)$. Then, it re-orders $\{\hat{q}_s^*(t) : s = 1, ..., S\}$ and obtains the vector of probabilities $\hat{\psi}(\rho(t))$ by solving the LP problem in (3) with the ordered $\hat{q}_s^*(t)$'s and ρ_τ. A detailed description of C3T-Budget-E can be found in the supplementary material.

3.5 Theoretical Analysis

We now analyze the theoretical performance of the proposed algorithms, which we refer to collectively as C3T-Budgets (i.e., C3T-Budget and C3T-Budget-E). All proofs can be found in the supplementary material. We begin by bounding the safety constraint violation.

Theorem 1. *For any given T, the average toxicity of subgroup s observed from C3T-Budgets satisfies*

$$\mathbb{P}\left[\frac{\sum_{t=1}^{N_s(T)} p_{s,I(N_s^{-1}(t))}}{N_s(T)} - \zeta \leq C\epsilon^\gamma\right] \geq 1 - \delta_s,$$

for an arbitrary $\epsilon > 0$, where C and γ are hyper-parameters (described in the supplementary material).

C3T-Budget and C3T-Budget-E therefore satisfy the safety constraints given in (2) and (6), respectively.

The following theorem bounds the recommended dose error of C3T-Budgets.

Theorem 2. *The probability that C3T-Budgets recommends an incorrect dose for subgroup s is bounded according to*

$$\mathbb{P}\left[\hat{k}_s^* \neq k_s^*\right] \leq M_{R1} e^{-M_{R2} N_s(T)}$$

where M_{R1} and M_{R2} are non-negative constants (provided in the supplementary material).

Theorem 2 indicates that the accuracy of the recommended dose depends strongly on the number of samples accepted in each subgroup. The worst-case regret bound for the total efficacy of C3T-Budget-E is provided in the supplementary material. We also corroborate these theoretical behaviors of C3T-Budgets with empirical results in the experiments section.

4 DISCUSSION

4.1 Heterogeneous Costs for Subgroups

In practice, the cost of recruiting a patient into a clinical trial may be different across subgroups. To address this, an extension of C3T-Budgets is needed in which heterogeneous costs for different subgroups can be accounted for. Let c_s be a non-negative cost associated with subgroup s that occurs when any dose is allocated to a patient in subgroup s. We can then reformulate the LP problem in (3) as

$$\underset{\psi}{\text{maximize}} \sum_{s \in \mathcal{S}} \psi_s \pi'_s \frac{d_s^*}{c_s} \quad \text{subject to} \sum_{s \in \mathcal{S}} \psi_s \pi'_s \leq \rho,$$

where $\pi'_s = \pi_s c_s$. Then we apply C3T-Budget as in Algorithm 1 by substituting π'_s and $B_s^*(t)/c_s$ ($q_s^*(t)/c_s$ for C3T-Budget-E) into line 12, whenever appropriate.

4.2 Sequential Patient Recruitment

In clinical trials, a "complete" list of candidate patients is often given in advance, and then patients are sequentially selected from among the candidate patients. For this sequential patient recruitment, C3T-Budgets can be applied. Let $\tilde{\pi}_s$ be proportion of candidate patients that belong to subgroup s. Based on $\tilde{\pi}_s$, a patient arrival sequence can be virtually generated and the problem reduces to the same problem as originally discussed, so that C3T-Budgets can be applied to the virtually generated arrival sequence.

5 EXPERIMENTS

In this section, we demonstrate the performance of our algorithm through a series of experimental results.[2] We consider a dose-finding clinical trial with 3 subgroups (i.e. $S = 3$) and 6 doses (i.e. $K = 6$). In the clinical trial, 400 patients can be recruited in total (i.e., $B = 400$) and a maximum of 1200 rounds can be performed (i.e. $T = 1200$). The arrival distributions for each of the 3 subgroups is given by $\pi_1 = \frac{5}{12}$, $\pi_2 = \frac{4}{12}$, and $\pi_3 = \frac{3}{12}$. The MTD threshold is set to be 0.35 (i.e. $\zeta = 0.35$) and the minimum efficacy threshold is set to be 0.2 (i.e. $\theta = 0.2$). The expected efficacy, $q_{s,k}$, and expected toxicity, $p_{s,k}$, of each dose for each subgroup are provided in Table 2. We highlight the optimal dose for each subgroup in bold. Note that there is no optimal dose for subgroup 1 as in (1) since their expected efficacy is below the minimum efficacy threshold (and so the optimal action is to not treat them at all). All experimental settings are similar to Riviere et al. (2018); Aziz et al. (2019) and we use them unless otherwise mentioned. All experiments are repeated 500 times and the results are averaged.

Table 2: Settings of Synthetic Model

S	Type	Characteristics					
1	$q_{1,k}$'s	0.01	0.02	0.05	0.1	0.1	0.1
	$p_{1,k}$'s	0.01	0.01	0.05	0.15	0.2	0.45
2	$q_{2,k}$'s	0.1	0.2	0.3	**0.5**	0.6	0.65
	$p_{2,k}$'s	0.01	0.05	0.15	**0.2**	0.45	0.6
3	$q_{3,k}$'s	0.2	0.5	0.6	**0.8**	0.84	0.85
	$p_{3,k}$'s	0.01	0.05	0.15	**0.2**	0.45	0.6

For benchmarks, we modify the following baseline algorithms to include contexts as in Zhou (2015) and implement them for contextual clinical trials: 3+3 design (Storer, 1989), Contextual UCB (Auer et al., 2002; Varatharajah et al., 2018), Contextual KL-UCB (Garivier and Cappé, 2011; Varatharajah et al., 2018), and Contextual independent Thompson sampling (TS) (Aziz et al., 2019). The details of the algorithms are provided in the supplementary material.

5.1 Recommended Dose Error Rates

In Table 3, we report the recommended dose error rate of each algorithm both as an average across all subgroups (Total) and individually for each subgroup (SG1, SG2, SG3). From the results, we can see that C3T-Budget achieves the lowest total error rate, with a lower error within each subgroup than all benchmarks except: C-UCB and C-Indep-TS whose error is lower in SG1 due to the fact that SG1 has the highest arrival

[2]The implementation of C3T-Budgets is available at: `https://bitbucket.org/mvdschaar/mlforhealthlabpub/src/master/alg/c3t_budgets`

rate and C-UCB and C-Indep-TS recruits all patients, C3T-Budget instead makes better use of the budget to *balance* the learning across all subgroups; and C3T-Budget-E, which notably only achieves lower error on SG3, the subgroup with the highest efficacy. This is exactly as we would expect, since the goal of C3T-Budget-E is to focus on the highest efficacy subgroups, thus resulting in better estimations of the optimal dose for those subgroups. The superior performance of C3T-Budget is due to its capability to effectively recruit patients from subgroups that will best improve recommended dose errors while remaining within the budget constraint. This will be investigated further in Section 5.4.

Table 3: Recommended Dose Error Rates

Algorithm	Errors			
	SG 1	SG 2	SG 3	Total
C3T-Budget	0.050	**0.056**	0.036	**0.047**
C3T-Budget-E	0.204	0.138	**0.020**	0.121
C-UCB	**0.012**	0.414	0.316	0.247
C-KL-UCB	0.204	0.440	0.368	0.337
C-Indep-TS	**0.012**	0.354	0.372	0.246
C-3+3	0.874	0.746	0.780	0.800

5.2 Safe Dose Estimation Error Rates

We define the type-I and type-II errors for dose safety as follows: Type-I error occurs when a safe dose is estimated to be unsafe and type-II error occurs when an unsafe dose is estimated to be safe. We report these errors for C3T-Budgets and each of the benchrmark algorithms in Table 4. We see that C3T-Budget achieves the lowest total error rates. In C3T-Budget, subgroups are recruited to maximize the confidence of our estimations, which in turn leads to confident estimates of safe doses.

Table 4: Safe Dose Estimation Error Rates

Algorithm	Errors		
	Type-I	Type-II	Total
C3T-Budget	0.0226	**0.0198**	**0.0212**
C3T-Budget-E	**0.0200**	0.0304	0.0252
C-UCB	0.0248	0.0289	0.0268
C-KL-UCB	0.0371	0.0437	0.0404
C-Indep-TS	0.1006	0.0366	0.0686
C-3+3	0.0400	0.1081	0.0741

5.3 Efficacy and Toxicity Per Patient

We now report the efficacy and toxicity per patient for each algorithm in Table 5. From the results, we can see that both C3T-Budgets, but particularly C3T-Budget-E, achieve a much higher efficacy than other algorithms. In addition, C3T-Budgets achieve a lower toxicity than other algorithms except for the 3+3 design. Note, however, that the 3+3 design is designed specifically to avoid safety violations - it ends when a certain number of toxicities have occurred. Hence, it naturally has a low toxicity compared to the bandit algorithms, but it achieves a much lower efficacy as well as shown in Table 5 and inaccurately recommends doses as shown in Table 3.

Table 5: Efficacy and Toxicity Per Patient

Algorithm	Efficacy	Toxicity
C3T-Budget	0.4975	0.1881
C3T-Budget-E	**0.5791**	0.1911
C-UCB	0.4154	0.3235
C-KL-UCB	0.3823	0.2621
C-Indep-TS	0.4101	0.3157
C-3+3	0.3081	**0.1537**

5.4 Recruited Patients for Each Subgroup and Efficacy Estimation Errors

To understand the recruitment behaviors of C3T-Budgets and how they relate to efficacy estimation error, Fig. 1 shows the number of patients recruited from each subgroup and the mean squared error (MSE) of the efficacy estimate of the optimal dose in each subgroup (for subgroup 1, we provide the averaged MSE of all safe doses since there is no optimal dose) over time. Note that the recruitment numbers of the other bandit algorithms are equal to those with π_s since they do not address the budget constraint.

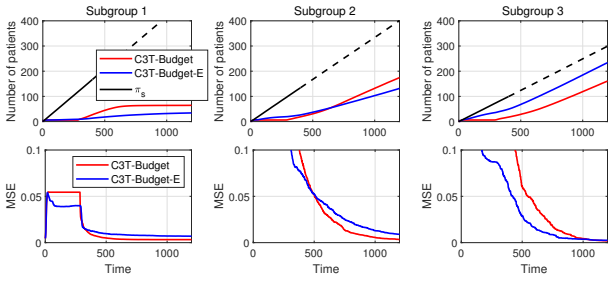


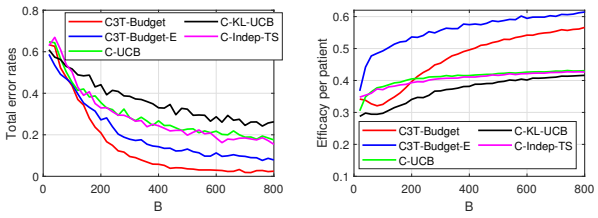
Figure 1: Number of recruited patients and the MSE of the efficacy estimation of each subgroup.

From the figure, we can see that both C3T-Budgets recruit patients independently of the arrival rate of each subgroup. In C3T-Budget, the patients are recruited to achieve low enough estimation errors for all subgroups, demonstrated by the fact that recruitment for subgroup 1 is stopped after the estimation errors of the doses for subgroup 1 become low enough. C3T-Budget then starts to recruit more patients from subgroups 2 and 3. Subgroup 2 ends up having the most patients recruited for it, due to the fact that the expected efficacy (0.5) of its optimal dose is more difficult to estimate than those of subgroups 1 and 3 (0.1 and 0.8, respectively). On the other hand, C3T-Budget-E recruits more patients from subgroup 3 due to it having the highest expected efficacy. C3T-Budget-E recruits only enough patients from subgroup 1 to allow for reasonable belief that it

has the lowest expected efficacy. More patients are recruited from subgroup 2 than from subgroup 1, due to the fact that it requires more samples to be sure that the efficacy is lower than that of subgroup 3, but once the algorithm has confidence that the efficacy is lower, recruitment slows, in favour of recruiting from subgroup 3, thus leading to higher estimation error for subgroup 2. We see from these figures that the recruitment behavior of C3T-Budget and C3T-Budget-E are as intended.

5.5 Impact of Budget

In this experiment, we report the recommended dose error rates and efficacy per patient of each algorithms while we vary the budget, B, in Fig. 2. Since the ratio B/T affects C3T-Budgets, we set the time-horizon T such that the ratio B/T is constant and 1/3. The results of the 3+3 design are not provided here since its performance does not depend on the total number of patients to dose.



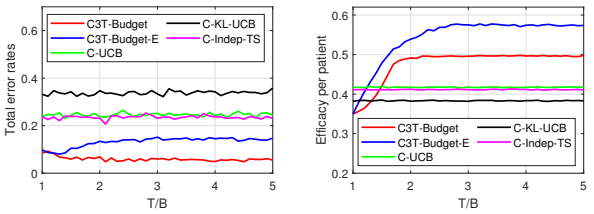
(a) Total recommended dose error rate. (b) Efficacy per patient.

Figure 2: Total recommended dose error rate and efficacy per patient varying the budget.

From Figs. 2a and 2b, we can see that the recommended dose error rates of all algorithms decrease and the efficacy per patient of all algorithms increases, as B increases. However, we see that for all $B > 80$, C3T-Budget achieves the lowest error, while for $B < 80$ (i.e. when the budget is particularly limited), we see that C3T-Budget-E achieves the lowest error. For all B, we see that C3T-Budget-E achieves the best efficacy. Notably, as B increases, the error of C3T-Budget drastically decreases and achieves near-zero error when $B \geq 500$. These results clearly demonstrate that C3T-Budgets do indeed effectively utilize the budget constraints unlike the other algorithms.

5.6 Impact of Time Horizon

We show the impact of varying the time-horizon T on the performance of our algorithm. If the time-horizon T is set to be the budget B, then C3T-Budgets do not skip any patients. Thus, C3T-Budget and C3T-Budget-E become the same algorithm. Note that all the other algorithms do not depend on the time-horizon T.



(a) Total recommended dose error rate. (b) Efficacy per patient.

Figure 3: Total recommended dose error rate and efficacy per patient varying the time-horizon.

In Fig. 3, we show the recommended dose error rates and efficacy per patient of the algorithms when varying T/B from 1 to 5. From Fig. 3a, we can see that the recommended dose error rate of C3T-Budget decreases while that of C3T-Budget-E increases as T/B increases. As T/B increases, C3T-Budgets can recruit subgroups more flexibly, since they have more time in which to "wait" for patients from a more desirable subgroup (they can afford to skip more patients). Thus, the error rate of C3T-Budget decreases since it recruits patients that will best reduce the error. On the other hand, the increased recruitment policy allows C3T-Budget-E to focus its recruitment more on subgroup 3 (so that it achieves high cumulative efficacy), but this in turn results in a higher error rate for subgroups 1 and 2, thus resulting in a higher total error. From Fig. 3b, we can see that as T/B increases, the efficacies of both C3T-Budgets increase thanks to the increased recruitment flexibility. From the results, we can see that both error rate and efficacy are saturated. If T/B is large enough, even very small arrival rates for a given group can be accommodated. The limiting factor then becomes the budget, B, and increasing T/B will have no (or very little) further effect.

6 CONCLUSION

In this paper, we have studied dose-finding clinical trials with heterogeneous groups and a limited budget. We proposed C3T-Budget in which, for each group, the dose with the highest estimated efficacy is chosen, but the group may be skipped if the algorithm is sufficiently confident in its current estimation of the efficacy. In addition, we proposed an extension, C3T-Budget-E, which is particularly applicable in very low-budget settings, where accurately estimating efficacy for the most promising subgroups is more important than accurate estimation for all groups. These designs can effectively utilize the limited budget to balance the trade-off between learning and treatment. We demonstrated that our proposed algorithms outperform state-of-the-art algorithms through a series of experiments.

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 *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1891–1900.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256.

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

Berry, D. A. et al. (2004). Bayesian statistics and the efficiency and ethics of clinical trials. *Statistical Science*, 19(1):175–187.

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.

Garivier, A. and Cappé, O. (2011). The KL-UCB algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual conference on learning theory*, pages 359–376.

Hussain-Gambles, M., Atkin, K., and Leese, B. (2004). Why ethnic minority groups are under-represented in clinical trials: a review of the literature. *Health & social care in the community*, 12(5):382–388.

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.

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.

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.

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.

Polyzos, A. (2008). Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor and platelet-derived growth factor receptor, in patients with metastatic renal cell carcinoma and various other solid tumors. *The Journal of steroid biochemistry and molecular biology*, 108(3-5):261–266.

Riviere, M.-K., Yuan, Y., Jourdan, J.-H., Dubois, F., and Zohar, S. (2018). Phase I/II dose-finding design for molecularly targeted agent: plateau determination using adaptive randomization. *Statistical methods in medical research*, 27(2):466–479.

Storer, B. E. (1989). Design and analysis of phase I clinical trials. *Biometrics*, 45(3):925–937.

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, H., Srikant, R., Liu, X., and Jiang, C. (2015). Algorithms with logarithmic or sublinear regret for constrained contextual bandits. In *Advances in Neural Information Processing Systems*, pages 433–441.

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

Zhou, L. (2015). A survey on contextual multi-armed bandits. *arXiv preprint arXiv:1508.03326*.