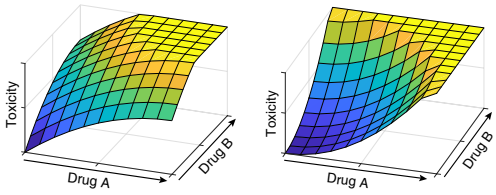
Supplementary Material for SDF-Bayes: Cautious Optimism in Safe Dose-Finding Clinical Trials with Drug Combinations and Heterogeneous Patient Groups

A Joint Dose-Toxicity Models for Drug Combinations

To reduce the search space in dose-finding for drug combinations, various joint dose-toxicity models have been proposed. They help us to efficiently investigate the toxicities of combinations of drugs. We denote the vector of the parameters of joint-toxicity model as $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$, where $\boldsymbol{\alpha}$ is a vector of the parameters that represents the relation between the toxicity effect and the dosage of drug A, $\boldsymbol{\beta}$ is a vector of the parameters that represents the relation between the toxicity effect and the dosage of drug B, and $\boldsymbol{\gamma}$ is a vector of the parameters that represents the relation between the toxicity effect and both dosages (i.e., drug-drug interaction). The type of functions suitable for a joint dose-toxicity model is defined as the following admissibility conditions [1]:

1. $\pi(j, k, \boldsymbol{\theta})$ is increasing separately in both j and k
2. There are functions $\pi_1(j, \boldsymbol{\alpha})$ and $\pi_2(k, \boldsymbol{\beta})$, which are called marginal dose-toxicity model such that $\pi(j, 0, \boldsymbol{\theta}) = \pi_1(j, \boldsymbol{\alpha})$ and $\pi(0, k, \boldsymbol{\theta}) = \pi_2(k, \boldsymbol{\beta})$
3. $\pi_1(0, \boldsymbol{\alpha}) = \pi_2(0, \boldsymbol{\beta}) = 0$

In joint dose-toxicity models, the drug-drug interaction such as synergy and antagonism shown in Fig. 1 should be captured. To this end, various joint dose-toxicity models have been proposed in the literature as in Table 1. In some models, the joint dose-toxicity model is defined by using the marginal model (i.e., $\pi_1(j, \boldsymbol{\alpha})$ and $\pi_2(k, \boldsymbol{\beta})$). In this case, any dose-toxicity model for a single drug such as two-parameter logistic model can be used. In the logistic model, the standardized dosage levels are used (i.e., d_j and u_k do not indicate actual dosage levels). Thus, there is a parameter ζ that is not related to any drugs to ensure the appropriate toxicity modeling [2]. All the models except for the no interaction model can capture the drug-drug interaction and the parameter $\boldsymbol{\gamma}$ implies the characteristics of the interaction. For the detailed description and discussion on drug-drug interactions, we refer the readers to [1].



(a) Synergistic. (b) Antagonistic.

Figure 1: Examples of drug-drug interactions.

Table 1: Summary of joint dose-toxicity models

Model	$\pi(j, k, \boldsymbol{\theta})$
No interaction	$1 - \{1 - \pi_1(j, \boldsymbol{\alpha})\}\{1 - \pi_2(k, \boldsymbol{\beta})\}$
Constant log-odds difference	$\frac{1}{1+\gamma^{-1}[\{\pi_1(j,\boldsymbol{\alpha})+\pi_2(k,\boldsymbol{\beta})-\pi_1(j,\boldsymbol{\alpha})\pi_2(k,\boldsymbol{\beta})\}^{-1}-1]}$
Copula-based [3]	$1 - C\{1 - \pi_1(j, \boldsymbol{\alpha}), 1 - \pi_2(k, \boldsymbol{\beta}), \gamma\}$
Thall [4]	$\frac{\alpha_1 d_j^{\alpha_2} + \beta_1 u_k^{\beta_2} + \gamma\alpha_1 d_j^{\alpha_2}\beta_1 u_k^{\beta_2}}{1+\alpha_1 d_j^{\alpha_2} + \beta_1 u_k^{\beta_2} + \gamma\alpha_1 d_j^{\alpha_2}\beta_1 u_k^{\beta_2}}$
Exponential	$1 - \exp\{-(\alpha d_j + \beta u_k + \alpha\beta\gamma d_j u_k)\}$
Logistic [2]	$\frac{1}{1-\exp(-\zeta-\alpha d_j-\beta u_k-\gamma d_j u_k)}$

B Proof of Proposition 1

From the definition of $F_a^{\mathcal{O}}(v)$, in round t, we have

$$\mathbb{P}\left[p_a(\boldsymbol{\theta}) \leq F_a^{\mathcal{O}(t)}(v) \middle| \mathcal{O}(t)\right] = v.$$

Note that the event $\bigcap_{\tau=1}^{t}\left\{p_{a(\tau)}(\boldsymbol{\theta}) \leq F_{a(\tau)}(\boldsymbol{\theta})^{\mathcal{O}(t)}(v) \middle| \mathcal{O}(t)\right\}$ is a subset of the event $\left\{\sum_{\tau=1}^{t} p_{a(\tau)}(\boldsymbol{\theta}) \leq \sum_{\tau=1}^{t} F_a^{\mathcal{O}(t)}(v) \middle| \mathcal{O}(t)\right\}$ clearly. Then, with $v = (1-\delta)^{1/t}$, we have

$$\mathbb{P}\left[\sum_{\tau=1}^{t} p_{a(\tau)}(\boldsymbol{\theta}) \leq \sum_{\tau=1}^{t} F_{a(\tau)}^{\mathcal{O}(t)}(v) \middle| \mathcal{O}(t)\right] \geq \prod_{\tau=1}^{t} \mathbb{P}\left[p_{a(\tau)}(\boldsymbol{\theta}) \leq F_{a(\tau)}^{\mathcal{O}(t)}(v) \middle| \mathcal{O}(t)\right] = v^t = 1-\delta.$$

From the proposition, the residual is non-negative, which implies that $(\xi+\epsilon_s)t \geq \sum_{\tau=1}^{t} F_a^{\mathcal{O}(t)}(v)$. We then have

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

C StructMAB: Safe DC-Finding Bandits Based on Dose-Toxicity Structure

C.1 DC-Finding Bandits Based on Structured Bandits

In our manuscript, we propose a DC-finding algorithm which allows the agent to effectively find the best recommendation for the MTD. To this end, our algorithm exploits a dose-toxicity structure for the drugs from the joint dose-toxicity model. Moreover, it is able to exploit arbitrary joint dose-toxicity models to ensure its practical use. Since the conventional MAB-based dose-finding clinical trial algorithm in [5] does not consider the dose-toxicity structure for drug combinations, we develop an advanced MAB-based algorithm by using structured bandits [6].

We first define a confidence set for the parameters as

$$\hat{\Theta}(t) \coloneqq \left\{\boldsymbol{\theta} : \forall a \in \mathcal{C},\ |\bar{p}_a(t) - p_a(\boldsymbol{\theta})| < \sqrt{\frac{\alpha \log t}{2n_a(t)}}\right\}, \tag{1}$$

where $\bar{p}_a(t)$ is the empirical toxicity of DC a in round t and $n_a(t)$ is the number of observed samples for DC a until round t. Since there is no assumption on the parameters $\boldsymbol{\theta}$, this confidence set can be obtained for arbitrary joint dose-toxicity models. Then, by using the confidence set, we define a set of candidate MTDs as

$$\mathcal{A}(t) \coloneqq \left\{a \in \mathcal{C} : \operatorname*{argmin}_{a' \in \mathcal{C}} |p_{a'}(\boldsymbol{\theta}) - \xi| \text{ for some } \boldsymbol{\theta} \in \hat{\Theta}(t)\right\}. \tag{2}$$

For the candidate MTDs, we choose the DC by using Thompson sampling. To this end, we assume that the expected toxicity of the DCs, p_a's, are random variables. The posterior probability of the expected toxicity of DC a in round t is updated as $\mathrm{Beta}(s_a(t)+1, n_a(t)-s_a(t)+1)$. Then, Thompson sampling is conducted on the set of candidate MTDs as

$$\tilde{p}_a(t) \sim \mathrm{Beta}(s_a(t)+1, n_a(t)-s_a(t)+1),\ \forall a \in \mathcal{A}(t). \tag{3}$$

Based on the samples, the DC whose the sample is closest to the threshold is chosen as

$$a(t) = \operatorname*{argmin}_{a \in \mathcal{A}(t)} |\tilde{p}_a(t) - \xi|. \tag{4}$$

At the end of the algorithm, the agent chooses the dose whose empirical toxicity is closest to ξ as the MTD

$$\hat{a}^*(T) = \operatorname*{argmin}_{a \in \mathcal{C}} |\bar{p}_a(T) - \xi| \tag{5}$$

or pick $\hat{a}^*$ uniformly at random among the allocated doses.

Algorithm 1 StructMAB

while $t \leq T$ **do**
  Obtain a confidence set $\hat{\Theta}(t)$ as in (1)
  Obtain a set of candidate MTDs $\mathcal{A}(t)$ as in (2)
  Sample the expected toxicities for all candidate MTDs as in (3)
  $\bar{a} \leftarrow \mathrm{argmin}_{a \in \mathcal{A}(t)} |\tilde{p}_a(t) - \xi|$
  if $r(t) - \max_{\boldsymbol{\theta} \in \hat{\Theta}(t)} p_{\bar{a}} \geq 0$ **then**
    $a(t) \leftarrow \bar{a}$
  else
    Obtain a set of conservative DCs $\mathcal{A}_c(t)$ as in (6)
    Sample the expected toxicities for all conservative DCs
    $a(t) \leftarrow \mathrm{argmin}_{a \in \mathcal{A}_c(t)} |\tilde{p}_a(t) - \xi|$
  end if
  Observe the DLT Y_t
  Update $m_{a(t)}(t+1)$ and $n_{a(t)}(t+1)$
  $t \leftarrow t + 1$
end while

C.2 Safe DC-Finding Bandits

In clinical trials, it is important to avoid testing the unsafe DCs due to the ethical issues. This can be achieved by considering the safety constraint in (1). Here, we propose a safe DC-finding approach that satisfies the safety constraint as in the cautiousness of SDF-Bayes.

For the safe DC-finding approach, we first define a set of conservative DCs as

$$\mathcal{A}_c(t) \coloneqq \left\{ a \in \mathcal{C} : \max_{\boldsymbol{\theta} \in \hat{\Theta}(t)} p_a(\boldsymbol{\theta}) \leq \xi \right\}. \tag{6}$$

The DCs in the set of conservative DCs are safe (i.e., toxicity probability does not exceed the MTD threshold) with a high probability. Thus, by choosing the DCs in the set only, we can avoid the trials with unsafe doses with a high probability. However, this is too conservative and we cannot expect a good MTD recommendation.

To resolve this issue, we consider a residual for the safety constraint in each round. The residual for the safety constraint in round t is defined as

$$r(t) = (\xi + \epsilon_s)t - \sum_{\tau=1}^{t-1} \max_{\boldsymbol{\theta} \in \hat{\Theta}(t)} p_{a(\tau)}(\boldsymbol{\theta}).$$

Note that the confidence set for the parameters in round t, $\hat{\Theta}(t)$, is used for the calculation of the sum of the expected toxicities until round t. By using the residual, we can infer whether the safety constraint will be violated or not based on the current confidence set for the parameters. If we still have the non-negative residual with the chosen DC in (4) (i.e., $r(t) - \max_{\boldsymbol{\theta} \in \hat{\Theta}(t)} p_{a(t)}(\boldsymbol{\theta}) \geq 0$), then we can expect the safety constraint not to be violated with the chosen DC. Thus, in that case, the safe DC-finding approach accepts the chosen DC, and otherwise, it rejects the chosen DC and chooses the DC in the conservative set $\mathcal{A}_c(t)$ instead to ensure the positive residual. We summarize the safe DC-finding algorithm (StructMAB) in Algorithm 1.

C.3 Sampling-Based Implementation of StructMAB

To use StructMAB in practice, we need to identify the confidence set $\hat{\Theta}$ and the set of candidate MTDs $\mathcal{A}$ based on the confidence set, which requires a high computational complexity in general. Besides, we need different identification methods for different joint dose-toxicity model used in the algorithm. Thus, to resolve these issues, we propose a simple sampling method for StructMAB that can efficiently approximate the confidence set and set of candidate MTDs regardless of which joint dose-toxicity model is used.

Here, we sample $\boldsymbol{\theta}$ from the posterior distribution of $\boldsymbol{\theta}$, $f(\boldsymbol{\theta}|\mathcal{O})$, by using Gibbs sampling which is one of most representative Bayesian sampling algorithms for multidimensional sampling as in SDF-Bayes. The posterior

distribution of $\boldsymbol{\theta}$ can be found in Section 3.1. of our manuscript. Details of the Gibbs sampling procedure can be found in the following section. We denote the number of samples from the posterior distribution $p(\boldsymbol{\theta}|\mathcal{O}(t))$ by L and the samples by $\tilde{\Theta}(t) = \{\boldsymbol{\theta}^{(l)}\}_{l\in[L]}$. Then, from the samples, we can define an approximated confidence set for the parameters as

$$\hat{\Theta}'(t) \coloneqq \left\{ \boldsymbol{\theta}^{(l)} \in \tilde{\Theta}(t) : \forall a \in \mathcal{C},\ |\bar{p}_a(t) - p_a(\boldsymbol{\theta}^{(l)})| < \sqrt{\frac{\alpha \log t}{2n_a(t)}} \right\}$$

and an approximated set of candidate MTDs as

$$\mathcal{A}'(t) \coloneqq \left\{ a \in \mathcal{C} : \operatorname*{argmin}_{a'\in\mathcal{C}} |p_{a'}(\boldsymbol{\theta}^{(l)}) - \xi| \text{ for some } \boldsymbol{\theta}^{(l)} \in \hat{\Theta}'(t) \right\}. \tag{7}$$

Similarly, we can define an approximated conservative DCs as

$$\mathcal{A}'_c(t) \coloneqq \left\{ a \in \mathcal{C} : p_{a'}(\boldsymbol{\theta}^{(l)}) \leq \xi,\ \forall \boldsymbol{\theta}^{(l)} \in \hat{\Theta}'(t) \right\}.$$

We can use these approximated sets for StructMAB. Note that these approximations converges to the true sets as the number of samples goes to infinity. Moreover, for the residual, we use $\max_{\boldsymbol{\theta}^{(l)}\in\tilde{\Theta}(t)} p_{a(t)}(\boldsymbol{\theta}^{(l)})$.

D Description of Sampling Procedure

We describe the sampling procedure used in SDF-Bayes.

- **Gibbs sampling**: Gibbs sampling is one of the most representative Markov chain Monte Carlo (MCMC) sampling methods to generate a sequence of samples for multiple variables from a multivariate joint probability distribution. In SDF-Bayes, it is used to generate a sequence of samples of the parameter vector $\boldsymbol{\theta} = \{\theta_0, \theta_1, ..., \theta_{D_{\boldsymbol{\theta}}-1}\}$ from $f(\boldsymbol{\theta}|\mathcal{O})$, where $D_{\boldsymbol{\theta}}$ is the dimension of the parameter vector. Note that the dimension of the parameter vector depends on which joint toxicity model is used for SDF-Bayes. In Gibbs sampling, we start from an arbitrary initial sample $\boldsymbol{\theta}$ in the distribution. We then samples each component of the parameter vector in turn based on each of the full conditional distribution with updated parameter samples. In Gibbs sampling, we discard L_b samples at the beginning which is so-called burn-in period, and then, retain the following L samples. We formally summarize Gibbs sampling as following Algorithm 2.

Algorithm 2 GIBBS SAMPLING

1: Initialize $\boldsymbol{\theta}^{(0)}$
2: **for** $l = 1$ **to** $L_b + L$ **do**
3: **for** $d = 0$ **to** $D_{\boldsymbol{\theta}} - 1$ **do**
4: Sample $\theta_d^{(l)}$ from its full conditional distribution $f\left(\theta_d \middle| \theta_0^{(l)}, ..., \theta_{d-1}^{(l)}, \theta_{d+1}^{(l-1)}, ..., \theta_{D_{\boldsymbol{\theta}}-1}^{(l-1)}, \mathcal{O}\right)$
5: **end for**
6: **end for**

- **Adaptive rejection metropolis sampling**: Gibbs sampling relies on the complete full conditional distributions of all components. However, in general, we cannot access to closed forms of such distributions. Thus, in SDF-Bayes, we use adaptive rejection metropolis sampling (ARMS) within Gibbs sampling [7]. ARMS is a MCMC sampling method as well, and it used to sample from a univariate target distribution specified by (unnormalized) log density.

  Thus, we utilize ARMS to sample each component of the parameter vector from its univariate full conditional distribution (i.e., for line 4 of Algorithm 2). We assume that the prior distributions of the parameters are independent. We denote the parameter vector except for d-th component by $\boldsymbol{\theta}_{-d}$. Then, we can obtain the unnormalized density of the full conditional distribution of d-th parameter, $f(\theta_d|\boldsymbol{\theta}'_{-d}, \mathcal{O})$, from the likelihood of $\boldsymbol{\theta}$ and the prior distribution as

$$f(\theta_d|\boldsymbol{\theta}'_{-d}, \mathcal{O}) \propto L(\theta_d, \boldsymbol{\theta}'_{-d}|\mathcal{O}) f(\theta_d) \prod_{d'\in\{1,...,d-1,d+1,...D_{\boldsymbol{\theta}}\}} f(\theta'_{d'}).$$

  By using this density, we can sample each component of the parameters within Gibbs sampling. For the detail procedure of ARMS, we refer to [7].

E Description of SDF-Bayes-AR

Algorithm 3 SDF-Bayes-AR

1: **while** $t \leq T$ **do**
2: Obtain $a_m(t)$'s by SDF-Bayes (lines 2–13)
3: $g(t) \leftarrow \text{argmax}_{m \in \mathcal{M}} \tilde{H}^{\mathcal{O}(t)}_{m,a_m(t)}(u)$
4: $a(t) \leftarrow a_{g(t)}(t)$
5: Observe DLT Y_t
6: Update $s^{g(t)}_{a(t)}(t+1)$ and $n^{g(t)}_{a(t)}(t+1)$
7: $t \leftarrow t+1$
8: **end while**
9: **Output:** $\hat{a}^*_m = \text{argmax}_{a \in \mathcal{A}} \tilde{G}^{\mathcal{O}(T)}_{m,a}(u),\ \forall m \in \mathcal{M}$

F Description of Detailed Experiment Settings

F.1 Real-World Dataset

We describe the real-world dataset used in the experiments. We consider a Phase I drug combination clinical trial dataset and its corresponding dose-toxicity model provided in [8]. In the dataset, the drug combination of nilotinib and imatinib is considered and the DLT observations from 50 patients are provided to assess the toxicity of different dose combinations. Furthermore, in [8], the dose-toxicity model of the combination is constructed in a Bayesian way based on a logistic regression model by using the DLT observations and the prior information about the drugs. In specific, the model is given by

$$\text{logit}(d, X_1, X_2, X_3) = \log(\alpha) + \beta \log(d/d^*) + \xi_1 X_1 + \xi_2 X_2 + \xi_3 X_3,$$

where α, β, ξ_n's are dose-toxicity model parameters, d represents the doses of nilotinib and d^* is the reference dose of nilotinib (400mg) for scaling. Also, (X_1, X_2, X_3) represents the dose of imatinib by taking the form $(0,0,0)$, $(1,0,0)$, $(1,1,0)$, and $(1,1,1)$ for imatinib doses 0, 400, 600, and 800mg, respectively. The prior distributions of the parameters are tuned by using the historical data such as previous clinical data of each drug without combinations. We rounded up the toxicities at the third decimal place. This trial is originally designed to find the DCs whose toxicities belong to the target interval $(0.20, 0.35)$. This implies that the potential DCs of the trial are more conservative in terms of safety (toxicity) compared with typical synthetic datasets designed to find the DC whose toxicity is close to the target toxicity 0.3 because it should find the DCs whose toxicities is lower than 0.3 in average.

F.2 Description of Algorithms in Experiments

Here, we describe the algorithms used in the experiments. We set $u = 0.1$ for all the Bayesian algorithms.

- **SDF-Bayes**: We implement SDF-Bayes based on Algorithm 1 in our manuscript. Here, we provide the settings of SDF-Bayes in the experiments. We set $v = 0.9$ unless mentioned. For warm-start of SDF-Bayes, the residual at the early stage of trials can be given by $r(t) = \max\left((\xi + \epsilon_s)t - \sum_{\tau=1}^{t-1} \tilde{F}^{\mathcal{O}(t)}_{a(\tau)}(v), R\right)$, where R is a constant. Typically, ξT works well for the cases with a single group. For the joint dose-toxicity model, we use the following logistic dose-toxicity model proposed in [2]:

  $$\texttt{logit}(\pi_{jk}) = \theta_0 + \theta_1 u_j + \theta_2 v_k + \theta_3 u_j v_k,$$

  where θ_0, θ_1, θ_2, and θ_3 are parameters and u_j and v_k are the standardized dose of drugs. In the literature [2], u_j's and v_k's are defined as $u_j = \log\left(\frac{p_j}{1-p_j}\right)$ and $v_k = \log\left(\frac{q_k}{1-q_k}\right)$, where p_j and q_k are the prior estimates of the toxicity probabilities of the j-th dosage of drug A and k-th dosage of drug B. However, such prior information is not always available, and thus, in the algorithm, we simply use $(-2,-1,0)$ for u_j's and $(-3,-2,-1,0)$ for v_k's assuming without any prior information. In this model, the parameters are defined as $\theta_1 > 0$ and $\theta_2 > 0$ and $-\infty < \theta_0 < \infty$. In addition, θ_3 should satisfy $\theta_1 + \theta_3 v_k > 0$ and $\theta_2 + \theta_3 u_j$ for all

$k \in \mathcal{K}$ and $j \in \mathcal{J}$. This ensures that the increasing toxicity probability with the increasing dose of a single drug. For the prior of the parameters, we use a normal distribution $\mathcal{N}(0, 10)$ for θ_0 and θ_3. This prior is vague with relatively high variance. For θ_1 and θ_2, we use an exponential distribution Exp(1) since they are positive. These default prior settings are same with used in the literature [2] and used for the other algorithms with the joint dose-toxicity model. For dataset RW, we use $v = 0.85$ because the potential DCs of the dataset are designed to be more conservative in terms of safety compared with other datasets as described in the previous section.

- **SOTA Bayes** [2]: We implement a Bayesian dose-finding algorithm in [2]. We denote the probability threshold for dose escalation by c_e and the probability threshold for dose de-escalation by c_d. To avoid that the dose is determined to be escalated and de-escalated at the same time, we need to ensure $c_e + c_d > 1$. We use the joint dose-toxicity model with the same setting from SDF-Bayes for fair comparison. The dose-finding algorithm proposed in [2] is described in following:
  - *Dose escalation*: Let the current DC be (j, k). If $\mathbb{P}[p_{jk} < \xi|\mathcal{O}] > c_e$, then the current DC is escalated to an adjacent DC among $\{(j+1, k), (j, k+1), (j+1, k-1), (j-1, k+1)\}$ that has a toxicity probability higher than p_{jk} and closest to ξ. If the current DC is the highest one, the same DC is allocated. With the samples from the posterior distribution $f(\boldsymbol{\theta}|\mathcal{O})$, we approximate the condition as $\frac{1}{L}\sum_{l=1}^{L} \mathbb{I}[p_{jk}(\boldsymbol{\theta}^{(l)}) < \xi] > c_e$.
  - *Dose de-escalation*: Let the current DC be (j, k). If $\mathbb{P}[p_{jk} > \xi|\mathcal{O}] > c_d$, then the current DC is de-escalated to an adjacent DC among $\{(j-1, k), (j, k-1), (j+1, k-1), (j-1, k+1)\}$ that has a toxicity probability lower than p_{jk} and closest to ξ. If the current DC is the lowest one, the same DC is allocated. With the samples from the posterior distribution $f(\boldsymbol{\theta}|\mathcal{O})$, we approximate the condition as $\frac{1}{L}\sum_{l=1}^{L} \mathbb{I}[p_{jk}(\boldsymbol{\theta}^{(l)}) > \xi] > c_d$.
  - *Dose retainment*: Let the current DC be (j, k). If $\mathbb{P}[p_{jk} < \xi|\mathcal{O}] \leq c_e$ and $\mathbb{P}[p_{jk} > \xi|\mathcal{O}] \leq c_d$, the current DC is allocated. With the samples from the posterior distribution $f(\boldsymbol{\theta}|\mathcal{O})$, we approximate the condition as $\frac{1}{L}\sum_{l=1}^{L} \mathbb{I}[p_{jk}(\boldsymbol{\theta}^{(l)}) < \xi] \leq c_e$ and $\frac{1}{L}\sum_{l=1}^{L} \mathbb{I}[p_{jk}(\boldsymbol{\theta}^{(l)}) > \xi] \leq c_d$.

  At the end of the trial, the DC recommendation is determined as same with SDF-Bayes.

- **StructMAB**: We implement StructMAB based on Algorithm 1. For the algorithm, we use the same joint dose-toxicity model and settings for the model used in SDF-Bayes. We set the exploration parameter α in the confidence set for the parameters in (1) to be 1. Moreover, when constructing the set of candidate MTDs in (7), its condition to include a DC into the set of candidate MTDs is too sensitive in practice since a DC will be included into the set even with only one ground sample that asserts that the DC is optimal. Thus, to mitigate such sensitivity of the condition, we add the following condition: from the set, remove the DCs whose number of ground samples is below the 20% percentile of the number of ground samples of each DC. This helps to remove only the outliers from the set of candidate MTDs. For example, when all the DCs in the set have only one ground sample, then the condition does not remove any DC from the set. Similarly, to avoid StructMAB being too conservative, we use the 80% percentile of the toxicity of the allocated DC among the samples when calculating the residual for the safety constraint.

- **IndepTS** [5]: We extend an independent Thompson sampling algorithm in [5] for drug combinations. To this end, we simply expand an arm space of IndepTS into two dimensional space for drug combinations. In IndepTS, Thompson sampling is used to estimate the toxicity of DC a as follows:

$$\tilde{p}_a(t) \sim \text{Beta}(\alpha_a(t), \beta_a(t)),$$

  where $\alpha_a(t) = s_a(t) + 1$ and $\beta_a(t) = n_a(t) - s_a(t) + 1$. In each round t, the expected toxicities of all DCs are sampled based on the above equation. Then, the DC $a(t)$ that has the maximum toxicity sample $\tilde{p}_a(t)$ is chosen (i.e., $a(t) = \text{argmin}_{a \in \mathcal{C}} |\tilde{p}_a(t) - \xi|$). At the end of the trial, it recommends a DC as: $\hat{a}^*(T) = \text{argmin}_{a \in \mathcal{C}} |\hat{p}_{s,k}(T) - \xi|$. Note that the implicit safety consideration in IndepTS is not exploited in IndepTS-DC since it is for a single drug.

- **SDF-Bayes-AR**: We implement SDF-Bayes-AR based on Algorithm 3 in this supplementary material. In a practical implementation of SDF-Bayes-AR, we adopt uniform patient recruitment in an early phase of clinical trial with the rounds $t \leq T/4$. This enables a warm-start of adaptive patient recruitment and avoids biased recruitment. In addition, we adopt an early stopping strategy to adaptive patient recruitment that

has been widely considered in clinical literatures [2]. If the posterior probability of the most likely DC to be the MTD for group m exceeds a threshold (i.e., $\max_{a\in\mathcal{C}} \tilde{G}^{\mathcal{O}(t)}_{m,a}(u) > p_{es}$, where p_es is the threshold), then SDF-Bayes-AR stops recruiting the patients from group m. This prevents unnecessary patient recruitment for the groups whose MTD is confidently discovered.

- **SOTA Bayes-AR**: We implement SOTA Bayes-AR by adopting our proposed adaptive patient recruitment (AR) to SOTA Bayes. For this, we can simply change line 2 in Algorithm 3 as "Obtain $a_m(t)$'s by SOTA Bayes".

F.3 Experiments with Heterogeneous Groups

In the experiments with heterogeneous groups, we apply the Bayesian algorithms to the entire population by treating it as a single group called EP. The toxicity probabilities for EP are the averages of the toxicity probabilities for group A and B, which are provided in Table 2. In the results, the safety violations of each group and the entire trial are considered. For each simulated trial, the safety violation of each group occurs if $S_m(T) > \xi + \epsilon$ as described in Section 4.1. On the other hand, the safety violation of the entire trial occurs if $\frac{1}{T}\sum_{t=1}^{T} Y_t > \xi + \epsilon$. For prior information, we run a simulated trial for group B with T_p patients using SDF-Bayes. Then, we use the observations from the simulated trial as the prior information.

Table 2: True toxicity of EP

	Synthetic EP			
	1	2	3	4
3	0.12	0.21	**0.30**	0.40
2	0.075	0.125	0.215	**0.30**
1	0.035	0.09	0.125	0.205

G More Experiment Results with Homogeneous Group

G.1 DLT Observation Rates with Datasets and Distribution of DLT Observation Rates

Table 3: DLT observation rates with datasets A,B,C,D,RW.

	Synthetic A	Synthetic B	Synthetic C	Synthetic D	Real-World
SDF-Bayes	0.296±.001	0.245±.001	0.286±.001	0.305±.001	0.274±.001
DF-Bayes	0.342±.002	0.260±.001	0.343±.001	0.351±.001	0.303±.001
SOTA Bayes	0.268±.001	0.204±.001	0.268±.001	0.275±.001	0.238±.001
StructMAB	0.282±.001	0.184±.001	0.279±.001	0.301±.001	0.238±.001
IndepTS	0.239±.001	0.118±.001	0.233±.001	0.278±.001	0.176±.001

From Table 3, SDF-Bayes satisfies the safety constraint but it does not achieve the lowest DLT observation rate because it is more often testing DC's that are believed close to being unsafe; this allows it to more effectively find the true MTD. This exploration-toxicity trade-off can be seen even more clearly by looking at the distribution of DLT observation rates in the following.

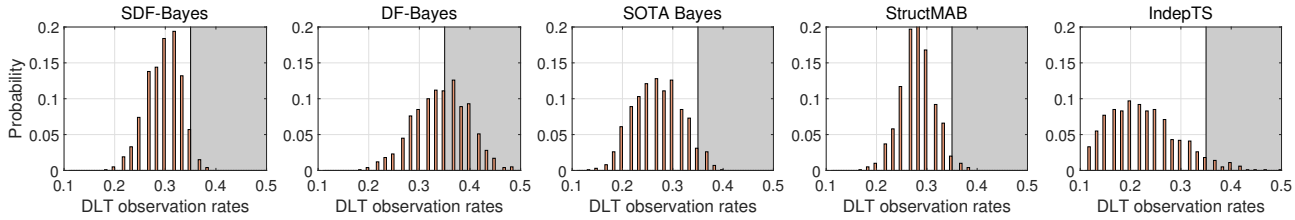


Figure 2: The distribution of DLT observation rates. The region shaded in gray is the region of the DLT observation rates that violate the safety constraint.

In Fig. 2, we provide the histogram of the DLT observation rates of the trials with synthetic dataset A. From the figure, we can see that the DLT observation rates with SDF-Bayes are distributed close to the safety threshold, but only a few trials violate the safety constraint. This clearly shows that SDF-Bayes effectively balances the trade-off between the exploration of the toxicities of the DCs and the DLT observation as intended in its cautious optimism. On the other hand, DF-Bayes frequently violates the safety constraint. This shows the effectiveness of the risk management by the cautiousness in SDF-Bayes. The safety constraint is rarely violated with SOTA Bayes as much as SDF-Bayes. However, their DLT observation rates do not concentrate as in SDF-Bayes since

its dose escalation does not have a capability to balance the trade-off explicitly. StructMAB has the similar distribution of DLT observation rates with SDF-Bayes, since the cautiousness principle of SDF-Bayes is adopted in StructMAB. The DLT observation rates of IndepTS are distributed over a low rate region since it fails to estimate the toxicities of DCs.

G.2 DC Allocation Ratios

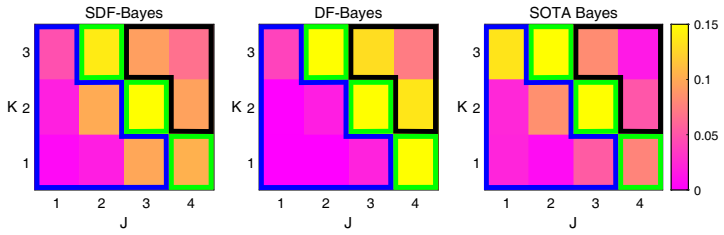


Figure 3: Heatmaps for DC allocation ratios of Bayesian algorithms. The regions with blue, green, and black borders are underdosing, target-dose, and overdosing regions, respectively.

Fig. 3 shows heatmaps for the DC allocation ratios of the Bayesian algorithms. We can see that the optimism of DF-Bayes results in allocating the most DCs to the target-dose and overdosing. SOTA Bayes concentrates DCs near the MTD $(3, 2)$ because of the way it escalates doses. On the other hand, SDF-Bayes allocates DCs in a less concentrated fashion because it tempers its optimism with caution. The result is that SDF-Bayes obtains more information than the other algorithms (and hence makes fewer recommendation errors) while managing risk better.

G.3 Results with More Toxicity Probability Models

Table 4: Synthetic models.

	Synthetic E				Synthetic F				Synthetic G				Synthetic H				Synthetic I			
	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	0.10	0.15	**0.35**	0.50	0.17	**0.35**	0.45	0.50	0.10	0.15	**0.25**	0.40	0.10	**0.25**	0.40	0.60
2	0.09	0.12	0.15	**0.30**	0.07	0.12	0.16	**0.35**	0.10	0.17	**0.35**	0.45	0.07	0.12	0.16	**0.25**	0.07	0.12	**0.25**	0.40
1	0.05	0.08	0.10	0.13	0.03	0.06	0.08	0.10	0.05	0.10	0.17	**0.35**	0.03	0.06	0.08	0.10	0.03	0.08	0.18	**0.25**

Table 5: Safety constraint violation rates and DC recommendation error rates with different datasets.

	Synthetic E		Synthetic F		Synthetic G		Synthetic H		Synthetic I	
Algorithms	Safety vio.	Errors	Safety vio.	Errors	Safety vio.	Errors	Safety vio.	Errors	Safety vio.	Errors
SDF-Bayes	0.010±.002	**0.295±.011**	0.006±.002	0.251±.010	0.035±.004	0.285±.011	0.003±.001	**0.331±.011**	0.033±.004	**0.298±.011**
DF-Bayes	0.239±.010	0.235±.010	0.245±.010	0.195±.009	0.530±.012	0.277±.010	0.112±.007	0.359±.011	0.300±.011	0.392±.011
SOTA Bayes	0.005±.002	0.318±.011	0.005±.002	**0.214±.010**	0.049±.005	**0.283±.010**	0.001±.001	0.338±.011	0.014±.003	0.361±.011
StructMAB	0.001±.001	0.592±.011	0.001±.001	0.547±.012	0.033±.004	0.616±.011	0.000±.000	0.490±.012	0.010±.002	0.445±.012
IndepTS	0.000±.000	0.726±.010	0.000±.000	0.697±.011	0.029±.004	0.632±.011	0.000±.000	0.712±.011	0.004±.001	0.582±.011

Here, we introduce 5 more different synthetic datasets to consider various possible combination toxicities. The toxicity probabilities of each synthetic model are summarized in Table 4 and the MTDs are highlighted in bold. Datasets F and G describe the scenarios in which the MTD's toxicity probability is higher than the target toxicity ($\xi = 0.3$) and datasets H and I describe the scenarios in which the MTD's toxicity probability is lower than the target toxicity. In Table 5, we provide the performance of the algorithms with the additional synthetic model in Table 4. From the results, we can see that in terms of recommendation error rates, SDF-Bayes outperforms SOTA Bayes in datasets E, H, and I. On the other hand, in datasets F and G, SOTA Bayes achieves lower error rates. This is because SDF-Bayes cautiously chooses the DC to be allocated considering the safety violation; it is hard to follow the optimism principle in those datasets because the MTDs have the higher toxicity than the target toxicity. Hence, SDF-Bayes has an advantage in terms of safety as shown in the results of dataset G; SOTA Bayes is very close to the boundary of a failure to satisfy the safety constraint while SDF-Bayes is not. In the results of MAB-based algorithms (i.e., StructMAB and IndepTS), StructMAB outperforms IndepTS in terms of error rates and achieves similar or lower safety violation rates. However, their error rates are too high compared with those of the other Bayesian algorithms.

G.4 Impact of Budgets

Fig. 4 reports the error rate as a function of the total patient budget. We see that in the practical regime of budgets (less than 100), the error rates of the MAB-based algorithms are significantly higher than the Bayesian algorithms. We note that this is even by using *StructMAB*, which already exploits the dose-toxicity structure to speed up learning. This clearly shows that the Bayesian designs are more suitable than the MAB-based designs in practice in terms of error rates.

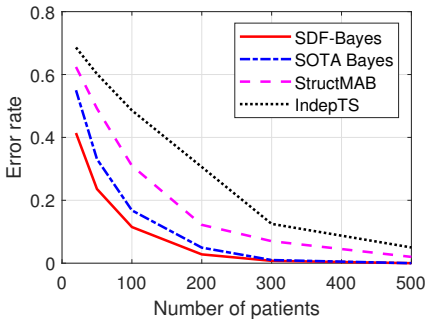


Figure 4: Error rates varying budgets.

G.5 Impact of Hyperparameter v

To show the impact of hyperparameter v on the performance of SDF-Bayes, we provide the performance of SDF-Bayes varying v from 0.80 to 0.95 in Table 6. Note that the hyperparameter v controls how conservative the algorithm is. From the results, we can see that the recommendation error rate increases as v increases. On the other hand, the safety constraint violation rate decreases as v increases. These results clearly show that the hyperparameter v controls the conservativeness of SDF-Bayes well. When v is increasing, SDF-Bayes chooses the DCs more conservatively. Thus, the risk of safety constraint violation decreases. On the other hand, it results in a less exploration. Then, the recommendation error rate increases.

Table 6: Performances of SDF-Bayes varying v

v	Safety vio.	Rec. errors	DLT observ.
0.80	0.135	0.191	0.313
0.85	0.056	**0.188**	0.313
0.90	0.022	0.203	0.296
0.95	**0.005**	0.222	**0.279**

G.6 Sensitivity of Prior Distribution

We now provide the performances of SDF-Bayes varying the prior distribution of the parameters of the dose-toxicity model in Table 7. To show the sensitivity of the algorithms, we consider the non-informative prior distribution (NonInfo) (i.e., a uniform distribution) and the distributions with higher variances (HiVar) compared with the prior distributions in our default setting. Specifically, for the non-informative prior, we use a uniform distribution that is truncated according to the domain of each parameter. For the high variance prior, we use a normal distribution $\mathcal{N}(0, 50)$ for θ_0 and θ_3 and use a gamma distribution with mean 1 and variance 10 (i.e., $\Gamma(0.1, 0.1)$) for θ_1 and θ_2.

Table 7: Performances of SDF-Bayes varying the prior distribution

Prior	Safety vio.	Rec. errors	DLT observ.
Default	**0.023**	**0.203**	0.296
HiVar	0.008	0.212	0.288
NonInfo	0.008	0.208	**0.261**

From the results, we can see that the performance of SDF-Bayes is similar regardless of the prior distribution of the parameters. This implies that SDF-Bayes is robust to the prior distribution since the posterior distribution of the parameters is effectively constructed thanks to the cautious optimism. Beside, it is worth noting that the prior distribution in the default setting is not a precise one as well [2]. Thus, the case with the non-informative prior distribution is an extreme case in clinical trials for drug combinations since the prior information on the parameters can be usually acquired from historical information, such as laboratory tests and clinical trials for a single drug, and characteristics of dose-combination models. In conclusion, we can use SDF-Bayes in practice even in case with a lack of the prior information of drugs.

G.7 Impact of Target Safety

In clinical trials, the target toxicity threshold is a standard of safety, and thus, the target safety is usually defined by the target toxicity threshold ξ ($\xi + \epsilon_s$ in our manuscript). In SOTA Bayes, dose (de-)escalation is determined based on the target toxicity threshold. Thus, its safety naturally focuses on the target toxicity threshold, and the target safety threshold cannot be arbitrarily determined. On the other hand, SDF-Bayes can consider arbitrary target safety ψ_s in its cautious principle by substituting the safety threshold, $\xi + \epsilon_s$, in residual $r(t)$ and the set of conservatice DCs $\mathcal{A}_c(t)$ in our manuscript, respectively, to any other target safety ψ_s.

Table 8: Performances of SDF-Bayes varying ψ_s.

ψ_s	Safety viol. (ψ_s)	Safety viol. ($\xi + \epsilon_s$)	Rec. errors	DLT observ.
0.20	0.048	0.000	0.421	0.169
0.25	0.056	0.000	0.328	0.221
0.30	0.022	0.000	0.257	0.263
0.35	0.019	0.019	0.203	0.297
0.40	0.027	0.193	0.193	0.321

Here, we show the impact of the safety threshold on SDF-Bayes. In Table 8, the recommendation error rates, the safety violation rates with ψ_s, the safety violation rates with $\xi + \epsilon_s$, and the DLT observation rates are provided. From the table, we can see that the safety violation rates with ψ_s are similar regardless of the target safety, ψ_s. This implies that SDF-Bayes effectively manages the risk of safety violation for any given target safety. Consequently, the DLT observation rates increase according to the target safety. This effective risk management in SDF-Bayes makes the target safety ψ_s become a safety budget for the exploration of the toxicities of the DCs. Thus, the larger ψ_s allows SDF-Bayes to choose the DCs more optimistically during clinical trials. Consequently, in the results, the recommendation error rates decreases as ψ_s increases. However, the target safety satisfying $\psi_s > \xi + \epsilon_s$ is hard to use in practice since the safety violation rate with the standard target safety in clinical trials (i.e., $\xi + \epsilon_s$) significantly increases. On the other hand, the target safety satisfying $\psi_s < \xi + \epsilon_s$ can be used to make trials safer while sacrificing the error rates.

G.8 Runtime and Scalability of SDF-Bayes

We evaluated runtime of SDF-Bayes in our simulations with homogeneous group; on average, it took approximately 2.7 milliseconds to complete a one-round update. (The implementation employed MATLAB with Intel Core i7-8700 3.2GHz CPU but without parallel computing.) By contrast, in a clinical trial, it takes days or weeks (sometimes months) to enroll patients for each round. Besides, actual clinical trials have limited numbers of patients and drug combinations. Even if the trial allowed for a large number of patients, testing hundreds of drug combinations would not be realistic because of safety issues. Hence, the runtime of SDF-Bayes is short enough that it can be used in any realistic trial.

H More Experiment Results with Heterogeneous Groups

H.1 Understanding of Behavior of Bayesian Algorithms Applied to Entire Population

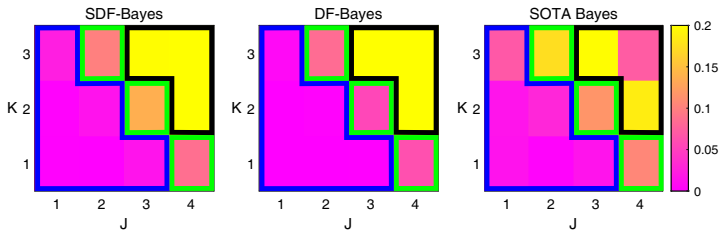


Figure 5: Heatmaps for DC allocation ratios of the patients of group A with the Bayesian algorithms applied to entire population.

Given the toxicities of group EP, the toxicities to group A are under-estimated because the toxicities to group B are very low. As a result, SDF-Bayes and DF-Bayes choose the DCs close to the MTDs of the averaged synthetic model for the entire population (i.e., (2,4) and (3,3) as shown in Table 2). Also, SOTA Bayes escalates dose above the MTDs of group A. These imply that the algorithms applied to group EP frequently select overdosing DCs for the patients of group A; see Fig. 5. Then, as shown in Table 4 in the manuscript, the safety violation rates of group A become excessively high.

H.2 Expected Improvements and Patient Recruitment Ratios

To understand the recruitment behavior of SDF-Bayes-AR, Fig. 6 plots the expected improvement (EI) probability for the most likely DC for each group and the patient recruitment ratio of each group as a function of the amount of prior information (T_p). The EI for group B decreases as the amount of the prior information increases. (Previous observations render later observations less useful.) SDF-Bayes-UR recruits patients uniformly; SDF-Bayes-AR adaptively recruits patients in order to maximize the EI gain from the next patient.

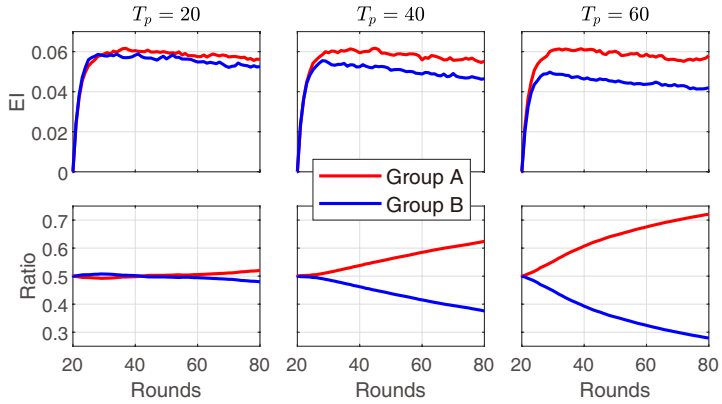


Figure 6: EI of the probability measures and patient recruitments ratio of groups varying the amount of the prior information.

H.3 Results in Section 5.2. with DLT Observation Rates

Here, we provide the DLT observation rates of the results for heterogeneous groups. Tables 9 and 10 provides the DLT observation rates aligned to the results in Tables 3 and 4 in our manuscript, respectively. Table 9 shows that the DLT observation rates of variants of SDF-Bayes are slightly higher than those of SOTA Bayes-UR due to the exploration-toxicity trade-off as in the single group case. (It is shown that they satisfy the safety constraint in Table 3 in our manuscript.) Table 10 shows that SDF-Bayes-AR outperforms SDF-Bayes-UR in terms of total DLT observation rates.

Table 9: DLT observation rates with heterogeneous groups (from Table 3 in the paper).

Algorithms	Total	Group A	Group B
SDF-Bayes-AR	0.263±.001	0.292±.001	0.236±.001
SDF-Bayes-UR	0.263±.001	0.297±.001	0.229±.001
DF-Bayes-UR	0.302±.001	0.352±.002	0.252±.002
SOTA Bayes-UR	**0.233±.001**	**0.270±.001**	**0.196±.001**
SDF-Bayes-EP	0.278±.001	0.391±.002	0.166±.002
DF-Bayes-EP	0.302±.001	0.427±.002	0.178±.002
SOTA Bayes-EP	0.247±.001	0.352±.002	0.142±.001

Table 10: DLT observation rates varying the amount of the prior information (from Table 5 in the paper).

		$T_p = 20$	$T_p = 40$	$T_p = 60$
AR	Group A	0.282±.001	0.261±.001	0.250±.001
	Group B	0.243±.002	0.259±.002	0.269±.003
	Entire	0.262±.001	0.255±.001	0.248±.001
UR	Group A	0.296±.001	0.296±.001	0.296±.001
	Group B	0.238±.002	0.248±.002	0.255±.002
	Entire	0.267±.001	0.272±.001	0.275±.001

References

[1] Mauro Gasparini. 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, 2013.

[2] Marie-Karelle Riviere, Ying Yuan, Frédéric Dubois, and Sarah Zohar. A bayesian dose-finding design for drug combination clinical trials based on the logistic model. *Pharmaceutical statistics*, 13(4):247–257, 2014.

[3] Guosheng Yin and Ying Yuan. 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, 2009.

[4] Peter F Thall, Randall E Millikan, Peter Mueller, and Sang-Joon Lee. Dose-finding with two agents in phase I oncology trials. *Biometrics*, 59(3):487–496, 2003.

[5] Maryam Aziz, Emilie Kaufmann, and Marie-Karelle Riviere. On multi-armed bandit designs for dose-finding clinical trials. *arXiv preprint arXiv:1903.07082*, 2019.

[6] Samarth Gupta, Shreyas Chaudhari, Subhojyoti Mukherjee, Gauri Joshi, and Osman Yağan. A unified approach to translate classical bandit algorithms to the structured bandit setting. *arXiv preprint arXiv:1810.08164*, 2019.

[7] Wally R Gilks, Nicky G Best, and KKC Tan. Adaptive rejection metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(4):455–472, 1995.

[8] Stuart Bailey, Beat Neuenschwander, Glen Laird, and Michael Branson. A bayesian case study in oncology phase I combination dose-finding using logistic regression with covariates. *Journal of biopharmaceutical statistics*, 19(3):469–484, 2009.