Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

EXPECTED VALUE OF INFORMATION METHODS FOR CONTEXTUAL RANKING AND SELECTION: CLINICAL TRIALS AND SIMULATION OPTIMIZATION

Andres Alban
Stephen E. Chick

Technology and Operations Management
INSEAD
Boulevard de Constance
Fontainebleau, 77300, FRANCE

Spyros I. Zoumpoulis

Decision Sciences
INSEAD
Boulevard de Constance
Fontainebleau, 77300, FRANCE

ABSTRACT

We consider the contextual ranking and selection problem that aims to learn the best treatment as a function of covariates when expected outcomes are unknown but can be learned from noisy observations. We develop a sequential allocation policy based on Bayesian expected value of information methods, called *f*EVI, to learn the best treatment for a finite set of covariates. We observe good performance of the *f*EVI allocation policy in simulation experiments and find that prior distributions which accurately reflect correlations across treatments and patient types can improve sampling effectiveness with limited sample sizes. We compare the performance between the case when covariates are random arrivals from a population, and the case when the allocation policy chooses covariates. In experiments, the benefit of *f*EVI over allocation policies that sample randomly is much larger than the benefit from being able to choose covariates, or from using a prior that accurately reflects correlations.

1 INTRODUCTION

Consider a heterogeneous population that requires treatment for a condition. We have a set of available treatments and would like to learn how to personalize the treatments to observable covariates of the subjects in the population. The treatment effects are unknown, but noisy observations can be obtained by assigning subjects to treatment and observing the outcome of interest. We call this the *contextual Ranking and Selection* (contextual R&S) problem. We tackle the problem of allocating sequentially arriving subjects to treatments in order to learn, as accurately as possible, the strategy as a function of subject covariates that maximizes the expected value of the outcomes of the population.

This problem is found in simulation optimization (Xiong 2020) and clinical trial design (e.g., Zhou et al. 2008). Compared to simulation optimization problems, in clinical trial settings there is generally less control over the conditions in which the noisy observations are obtained. Clinical trials are restricted to the covariates of patients that randomly arrive at recruitment sites, they generally face longer delays in observing patient outcomes (noisy observations), and they require randomization to prevent biases from the attending physicians. This paper focuses on the case of random arrival of covariates and discusses the implications compared to simulation optimization problems.

The Ranking and Selection (R&S) literature has extensively studied the problem of selecting the best among a set of alternative systems (Kim and Nelson 2006; Chen et al. 2015). Until recently, the R&S literature has focused on systems without external covariates to which the best alternative can be tailored. Recently, the R&S problem of choosing the best system as a function of covariates (contextual R&S) has started to receive attention. Shen et al. (2021) study the problem from the frequentist indifference zone perspective and develop a two-stage approach. Gao et al. (2019) develop a sequential procedure

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

based on the optimal computing budget allocation that achieves the asymptotically optimal rate function in the decay of the probability of incorrect selection. They show that the rate of decay is the same when the probability of incorrect selection is defined as the average, maximum, or worst-case over the set of covariates. Xiong (2020) proposes two algorithms based on the expected improvement and confidence bound criteria. Li et al. (2020) provide a Bayesian dynamic scheme to maximize the minimum probability of correct selection among a finite set of covariates. Pearce and Branke (2017) develop algorithms to estimate the *Expected Improvement*, a sequential sampling method often used in Bayesian optimization. Finally, Zhang et al. (2019) provide an algorithm using the Expected Value of Information (EVI) method and prove its consistency when the patient outcomes follow a Gaussian process in the covariate space. While all these papers on contextual R&S have focused on the simulation optimization setting, we are, to the best of our knowledge, the first to apply R&S methods to the clinical trial design setting in which the random arrival of patients constrains covariates.

The contextual bandit literature (e.g., Goldenshluger and Zeevi 2013, Bastani and Bayati 2020, Villar and Rosenberger 2018) tackles a similar problem, with the most notable applications being in online advertising. In contextual bandits, an arm among many is pulled under a random context to obtain a reward. The objective is to maximize the cumulative rewards by balancing the exploration-exploitation trade-off. Contextual bandits are online learning problems, as they maximize online rewards. The contextual R&S is an offline learning problem because it focuses on the best selection at the end of sampling, disregarding in its objective the rewards accrued online. In this paper, we focus on the offline learning problem.

In this paper, we take the EVI approach. The EVI is a Bayesian approach that has been extensively applied in the R&S problem (e.g., Chick and Inoue 2001, Branke et al. 2007, Frazier et al. 2009). Unlike Zhang et al. (2019), we allow the model to have random covariates and correlated mean outcomes across treatments as is found in clinical trials. Moreover, while they use a model with continuous covariates, we assume a finite set of covariate values (patient types) which simplifies the model and allows for a prior covariance matrix that does not only depend on the difference of covariates. The benefits of using discrete covariates are twofold. First, we can formulate an algorithm that is significantly easier to compute. Second, we develop a more practical understanding of the problem, particularly concerning the impact of the prior distribution and the correlation structure that can be imposed. Xie et al. (2016) and Chick et al. (2021) consider correlated means, but do not allow for selecting the best treatment as a function of covariates.

Because of the focus of contextual R&S on the simulation context where both the treatment and covariates can be selected, we show how to adapt our algorithm to additionally select covariates. This allows us to compare the settings with random versus with selected covariates in simulation experiments (see Section 6.4) and observe that the ability to choose the covariates translates into a significant increase in performance in selecting the best treatment strategy.

Section 2 introduces the model, and Section 3 presents the Bayesian inference procedure. Section 4 presents our *f*EVI algorithms for the contextual R&S problem with a finite set of covariate values. Section 5 discusses the correlation structure of the prior distribution and its consequences on the design of effective allocation policies. Section 6 presents simulation experiments that investigate the empirical performance of the *f*EVI algorithm, and Section 7 concludes with a summary and directions for future work.

2 PROBLEM FORMULATION

We formulate a model of a sequential trial with a budget to enroll a fixed number of patients T. For each patient $t = 1,\ldots,T$, we observe $X_t \in \mathscr{X}$, where $\mathscr{X} = \{1,\ldots,m\}$ is a finite set of *patient types* that reflect the subgroups in the population that have relevant interactions with treatment and patient outcomes. The patients arrive in a random fashion such that the type of each patient is drawn i.i.d., and the probability of observing a patient of type x is assumed to be $p(x)$. After observing the type of a patient, we allocate the patient to treatment $W_t \in \mathscr{W}$, where $\mathscr{W} = \{1,\ldots,n\}$ is a finite set of potential treatments. After treating a patient, we observe the outcome $Y_t \in \mathscr{Y} \subseteq \mathbb{R}$, where we assume that larger values represent better health outcomes for the patient.

At the conclusion of the trial, we select a *treatment strategy* $f \in \mathbf{f}$, where $\mathbf{f}$ is the set of functions that map patient types to treatments, i.e., $\mathbf{f} = \{f : \mathscr{X} \to \mathscr{W}\}$. The size of $\mathbf{f}$ is given by all possible allocations of treatments to patient types so that $|\mathbf{f}| = n^m$. Our objective is to find the treatment strategy that maximizes the expected health outcomes of each patient type in the population.

To formally define the objective, let $\mu(w,x)$ be the unknown expected outcome for a patient of type x who receives treatment w, i.e., $\mathbb{E}[Y_t \mid W_t, X_t, \mu(W_t, X_t)] = \mu(W_t, X_t,)$. We let $\boldsymbol{\mu}$ be the nm-dimensional vector of expected outcomes arranged in lexicographic order:

$$\boldsymbol{\mu} = (\mu(1,1), \ldots, \mu(1,m), \ldots, \mu(n,1), \ldots, \mu(n,m))^\top \tag{1}$$

We say that $(\mu(w,1), \mu(w,2), \ldots, \mu(w,m)) \in \mathbb{R}^m$ are the elements of $\boldsymbol{\mu}$ *associated* with treatment w and that $(\mu(1,x), \mu(2,x), \ldots, \mu(n,x)) \in \mathbb{R}^n$ are the elements of $\boldsymbol{\mu}$ *associated* with type x. We use the lexicographic ordering and this terminology in defining all vectors.

We use a Bayesian model with normal conjugate priors by assuming the following statistical model:

$$Y_t \mid X_t, W_t, \boldsymbol{\mu} \overset{i.i.d.}{\sim} \mathscr{N}(\mu(W_t, X_t), \sigma^2(W_t, X_t)) \ \forall t \in \{1, \ldots, T\} \tag{2}$$

$$\boldsymbol{\mu} \sim \mathscr{N}(\boldsymbol{\theta}_0, \Sigma_0), \tag{3}$$

where $\sigma^2(w,x)$ are presumed to be known sampling variances for any $w \in \mathscr{W}$ and any $x \in \mathscr{X}$, $\boldsymbol{\theta}_0$ is a known prior mean vector, and Σ_0 is a known prior covariance matrix. In practice, $\sigma^2(w,x)$, $\boldsymbol{\theta}_0$, and Σ_0 are unknown but might be fit using pilot data and adaptations of empirical Bayes estimation techniques (e.g., Xie et al. 2016; Chick et al. 2021) if (w,x) is embeddable in Euclidean space.

In Section 3, we show how these assumptions lead to a conjugate model where all the information about $\boldsymbol{\mu}$ obtained after observing the type, treatment, and outcome of the first t patients can be summarized in posterior updates of $\boldsymbol{\theta}_t$ and Σ_t such that

$$\boldsymbol{\mu} \mid W_1, X_1, Y_1, \ldots, W_t, X_t, Y_t \sim \mathscr{N}(\boldsymbol{\theta}_t, \Sigma_t).$$

The treatment strategy $f_{\boldsymbol{\theta}_t}$ that maximizes the posterior means for each type is:

$$f_{\boldsymbol{\theta}_t}(x) = \underset{w \in \mathscr{W}}{\arg\max}\, \theta_t(w,x), \tag{4}$$

where $\theta_t(w,x)$ is the posterior mean outcome of a patient of type x treated with w. At the conclusion of the trial, the treatment strategy $f_{\boldsymbol{\theta}_T}$ is implemented in the population.

An allocation policy $\pi = (\pi_t)_{t=1,\ldots,T}$ is composed of a set of T functions that map the available information prior to allocating the treatment of the patient — including the type of patient t — to the treatment to be allocated W_t:

$$W_t = \pi_t(\boldsymbol{\theta}_{t-1}, \Sigma_{t-1}, X_t).$$

Let $\tilde{p}(x) \in [0,1]$ be the fraction of the population that is of type x such that $\sum_{x \in \mathscr{X}} \tilde{p}(x) = 1$, and $\tilde{X}$ be a random patient type from the population. The probabilities $\tilde{p}(x)$ may differ from $p(x)$ if the patients enrolled in the trial are not fully representative of the population of patients. We can interpret $\tilde{p}(x)$ as the relative importance we assign to finding the best treatment for type x.

Our objective is to find an allocation policy π that maximizes the expected health benefit to future patients treated with strategy $f_{\boldsymbol{\theta}_T}$ identified on the basis of the T patients in the trial. We refer to this as the value function:

$$V^\pi = \mathbb{E}^\pi \left[\mu(f_{\boldsymbol{\theta}_T}(\tilde{X}), \tilde{X})\right] \tag{5}$$

where $\mathbb{E}^\pi$ is the expectation induced by allocation policy π. By conditioning on $\boldsymbol{\mu}$ and $\tilde{X}$, and using the definition of $f_{\boldsymbol{\theta}_T}$, we can rewrite (5) as

$$V^\pi = \mathbb{E}^\pi \left[\sum_{x \in \mathscr{X}} \tilde{p}(x) \max_{w \in \mathscr{W}} \theta_T(w,x)\right]. \tag{6}$$

We end this section by summarizing the aspects that our objective and decision variables do not consider compared to other related models. Our objective does not maximize the probability of correct selection, which is presumed to be high when the value function is maximized. In addition, the objective does not include a cost of sampling. Thus, the decision variables of this base model do not consider a stopping time or an option to not enroll a patient based on the type.

3 INFERENCE MODEL

Let $\mathbf{e}(w,x)$ be the nm-dimensional vector composed of all zeros except for a one in the entry associated with treatment w and type x — the associated entry uses the same lexicographic ordering introduced in (1) to define the vector of expected outcomes. We can recursively update the prior mean vector and covariance matrix using the following recursive equations (Powell and Ryzhov 2012, Section 8.2.2):

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \frac{(Y_{t+1} - \theta_t(W_{t+1}, X_{t+1}))}{\sigma^2(W_{t+1}, X_{t+1}) + \mathbf{e}^\top(W_{t+1}, X_{t+1})\Sigma_t \mathbf{e}(W_{t+1}, X_{t+1})} \Sigma_t \mathbf{e}(W_{t+1}, X_{t+1}), \tag{7}$$

$$\Sigma_{t+1} = \Sigma_t - \frac{\Sigma_t \mathbf{e}(W_{t+1}, X_{t+1}) \mathbf{e}^\top(W_{t+1}, X_{t+1}) \Sigma_t}{\sigma^2(W_{t+1}, X_{t+1}) + \mathbf{e}^\top(W_{t+1}, X_{t+1})\Sigma_t \mathbf{e}(W_{t+1}, X_{t+1})}. \tag{8}$$

From the recursive equations, we can also derive the distribution of $\boldsymbol{\theta}_{t+1}$ given that we have allocated a patient of type X_{t+1} to treatment W_{t+1} but have not observed the patient's outcome. We define the so-called posterior predictive standard deviation as

$$\tilde{\boldsymbol{\sigma}}(\Sigma, w, x) = \frac{\Sigma \mathbf{e}(w,x)}{\sqrt{\sigma^2(w,x) + \mathbf{e}^\top(w,x)\Sigma_{t-1}\mathbf{e}(w,x)}}. \tag{9}$$

One can show that $\boldsymbol{\theta}_{t+1} \mid \boldsymbol{\theta}_t, \Sigma_t, X_{t+1}, W_{t+1} \sim \mathcal{N}(\boldsymbol{\theta}_t, \tilde{\boldsymbol{\sigma}}^\top(\Sigma_t, W_{t+1}, X_{t+1})\tilde{\boldsymbol{\sigma}}(\Sigma_t, W_{t+1}, X_{t+1}))$, or equivalently

$$\boldsymbol{\theta}_{t+1} \overset{D}{=} \boldsymbol{\theta}_t + \tilde{\boldsymbol{\sigma}}(\Sigma_t, W_{t+1}, X_{t+1}) Z \tag{10}$$

where $\overset{D}{=}$ represents equality in distribution and $Z \sim \mathcal{N}(0,1)$. That is, $\tilde{\boldsymbol{\sigma}}(\Sigma_t, W_{t+1}, X_{t+1})$ is the standard deviation of the (random) posterior mean to be observed at time $t+1$, given the information at time t, and the patient type and treatment choice for patient $t+1$, but before the outcomes for patient $t+1$ is observed.

4 EXPECTED VALUE OF INFORMATION HEURISTIC

Expected value of information (EVI) allocation policies maximize the value gained each time a decision is made. These allocation policies are not optimal; they are myopic in the sense that they maximize the current value gained even when larger gains in future decisions would compensate for lower current gains. However, computing an optimal allocation policy is generally computationally infeasible, and the EVI heuristics have been found to perform well in practice for related problems (Branke et al. 2007; Frazier et al. 2008). This section defines the *f*EVI allocation policy, particularly tailored to the setting where the aim is to find the best treatment strategy f under random arrival of patients. We show how to efficiently compute the *f*EVI allocation policy, extending ideas from Frazier et al. (2009) as described below. The algorithmic contribution is a modest — yet non-trivial — extension of the algorithm in Frazier et al. (2009).

The *f*EVI allocation policy relies on computing an index for each treatment in $\mathscr{W}$ (*f*EVI-indices) and sampling the treatment with the largest index. The *f*EVI-indices represent the expected value gained from sampling one more patient and then selecting a treatment strategy, over selecting a treatment strategy without taking any additional samples. When the next patient to treat is of type $X_{t+1} = x$, the indices are

defined as

$$\nu_t(w) = \underbrace{\mathbb{E}_{\boldsymbol{\mu},Y_{t+1},\tilde{X}}[\mu(f_{\boldsymbol{\theta}_{t+1}}(\tilde{X}),\tilde{X})) \mid \boldsymbol{\theta}_t,\Sigma_t,W_{t+1}=w,X_{t+1}=x]}_{\text{Expected value of implementing after one additional observation}} - \underbrace{\mathbb{E}_{\tilde{X}}[\mu(f_{\boldsymbol{\theta}_t}(\tilde{X}),\tilde{X}) \mid \boldsymbol{\theta}_t,\Sigma_t]}_{\text{Expected value of implementing now}}$$

$$= \mathbb{E}_{Y_{t+1}}\left[\sum_{\tilde{x}\in\mathscr{X}} \tilde{p}(\tilde{x}) \max_{\tilde{w}\in\mathscr{W}} \theta_{t+1}(\tilde{w},\tilde{x}) \,\middle|\, \boldsymbol{\theta}_t,\Sigma_t,W_{t+1}=w,X_{t+1}=x\right] - \sum_{\tilde{x}\in\mathscr{X}} \tilde{p}(\tilde{x}) \max_{\tilde{w}\in\mathscr{W}} \theta_t(\tilde{w},\tilde{x}), \tag{11}$$

and the *f*EVI allocation policy samples

$$W_{t+1} = \arg\max_{w\in\mathscr{W}} \nu_t(w).$$

We compute the following auxiliary indices that represent the value gained within the subpopulation of patients of type $\tilde{x}$:

$$\tilde{\nu}_t(w,\tilde{x}) = \mathbb{E}_{Y_{t+1}}\left[\max_{\tilde{w}\in\mathscr{W}} \theta_{t+1}(\tilde{w},\tilde{x}) \,\middle|\, \boldsymbol{\theta}_t,\Sigma_t,W_{t+1}=w,X_{t+1}=x\right] - \max_{\tilde{w}\in\mathscr{W}} \theta_t(\tilde{w},\tilde{x}). \tag{12}$$

The *f*EVI-index is then calculated as a weighted sum of the auxiliary indices: $\nu_t(w) = \sum_{\tilde{x}\in\mathscr{X}} \tilde{p}(\tilde{x})\tilde{\nu}_t(w,\tilde{x})$. While the auxiliary indices look very similar to the computation of the correlated Knowledge Gradient (cKG) in Frazier et al. (2009), we cannot immediately execute their algorithm because the set of actions is not the same as the set of alternatives to select for implementation — the patient to be allocated to treatment is of type x while we are evaluating the value gained for implementation in the subpopulation of type $\tilde{x}$, which may differ. However, we use the procedure in Frazier et al. (2009) to compute the function

$$h(\mathbf{a},\mathbf{b}) = \mathbb{E}\left[\max_i a_i + b_i Z\right] - \max_i a_i,$$

where $Z \sim \mathscr{N}(0,1)$, and a_i and b_i are the ith entries of the vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. Frazier et al. (2009) provide an efficient procedure to compute $h(\mathbf{a},\mathbf{b})$ with complexity $O(M\log(M))$, where M is the length of the vectors $\mathbf{a}$ and $\mathbf{b}$, which is detailed in Steps 3-11 of their Algorithm 2.

By using (10), one finds that

$$\tilde{\nu}_t(w,\tilde{x}) = \mathbb{E}\left[\max_{\tilde{w}\in\mathscr{W}} [\theta_t(\tilde{w},\tilde{x}) + \tilde{\sigma}(\Sigma_t,w,x)(\tilde{w},\tilde{x})Z] \,\middle|\, \boldsymbol{\theta}_t,\Sigma_t,W_{t+1}=w,X_{t+1}=x\right] - \max_{\tilde{w}\in\mathscr{W}} \theta_t(\tilde{w},\tilde{x}) = h(\tilde{\mathbf{a}},\tilde{\mathbf{b}}),$$

where $\tilde{\mathbf{a}} \in \mathbb{R}^n$ and $\tilde{\mathbf{b}} \in \mathbb{R}^n$ are the entries of $\boldsymbol{\theta}_t$ and $\tilde{\boldsymbol{\sigma}}(\Sigma_t,w,x)$, respectively, that are associated with type $\tilde{x}$. Algorithm 1 summarizes the procedure we just described to compute $\nu_t(w)$ and choosing the treatment according to the *f*EVI allocation policy. The algorithm has complexity $O(mn^2\log(n))$. In actual implementation, we compute the logarithms of $\tilde{\nu}_t(w,\tilde{x})$ and $\nu_t(w)$ for numerical stability.

The differences between Algorithm 1 to compute the *f*EVI and the algorithm to compute the cKG (Algorithm 2 in Frazier et al. 2009) are subtle but important. In our setup, the cKG algorithm with inputs $\boldsymbol{\theta}_t$ and Σ_t would aim to find the treatment-type combination that obtains the largest health outcome by selecting a treatment-type combination at each time step. The *f*EVI differs in two aspects. First, its goal is to find the best treatment for each patient type. Therefore, the loop in Steps 4-8 of Algorithm 1 computes an index associated with each patient type, which are then aggregated through a weighted sum in Step 9. Second, the action space is constrained by $X_{t+1} = x$, the type of patient that randomly arrived, in the sense that we cannot allocate a patient of a different type. Thus, the loop starting in Step 2 is over the set of treatments, instead of the set of treatment-type combinations, and requires the additional input x.

We now comment on several features of the *f*EVI allocation policy. Interestingly, it does not depend on the patient type probabilities of arrival $p(x)$. This is because the allocation policy is myopic and does

Algorithm 1 *f*EVI allocation policy.

1: **function** *f*EVI($\boldsymbol{\theta}_t, \Sigma_t, x$)
2: **for all** $w \in \mathscr{W}$ **do**
3: $\mathbf{b} \leftarrow \tilde{\boldsymbol{\sigma}}(\Sigma_t, w, x)$
4: **for all** $\tilde{x} \in \mathscr{X}$ **do**
5: $\tilde{\mathbf{a}} \leftarrow \boldsymbol{\theta}_t(\mathscr{W}, \tilde{x})$ ▷ subset posterior means to keep only the entries associated with type $\tilde{x}$
6: $\tilde{\mathbf{b}} \leftarrow \mathbf{b}(\mathscr{W}, \tilde{x})$ ▷ subset predictive standard deviations to keep only the entries associated with type $\tilde{x}$
7: $\tilde{\nu}_t(w, \tilde{x}) \leftarrow h(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ ▷ h is the procedure in Frazier et al. (2009)
8: **end for**
9: $\nu_t(w) \leftarrow \sum_{\tilde{x} \in \mathscr{X}} \tilde{p}(\tilde{x}) \tilde{\nu}_t(w, \tilde{x})$
10: **end for**
11: **return** $W_{t+1} = \arg\max_{w \in \mathscr{W}} \nu_t(w)$
12: **end function**

not consider the learning that will be gained from future patients. By design, the fEVI allocation policy is optimal when a single patient is left to be assigned to treatment. Moreover, several allocation policies based on EVI ideas are asymptotically optimal as the sample size T goes to infinity (e.g., Frazier et al. 2009, Zhang et al. 2019). Conditions under which the *f*EVI allocation policy is asymptotically optimal is a potential area for future work.

A scenario where we can select the type and treatment to allocate next may not be appropriate in clinical trials because trials cannot select who arrives at the recruitment centers next. However, if the supply of patients willing to join the trial is large enough to select the patient type that we would like to enroll next, then such a scenario is relevant. Moreover, the literature has studied this problem as discussed in Section 1. We can adapt the *f*EVI algorithm to this setting with little additional work and present it here for completeness. Zhang et al. (2019) present the *Integrated Knowledge Gradient* (IKG) allocation policy, based on EVI concepts, to find the best treatment strategy in a population with continuous patient types using a Gaussian process model. The allocation policy we present here is a special case of the IKG with finite patient types. However, we provide an algorithm that efficiently computes the IKG allocation policy when the patient types are finite. We refer to this algorithm as the *f*EVI choosing covariates and present it in Algorithm 2. The *f*EVI choosing covariates allocation policy computes the same indices as defined in (11) but does so for all possible X_{t+1}:

$$\nu_t(w,x) = \mathbb{E}\left[\sum_{\tilde{x} \in \mathscr{X}} \tilde{p}(\tilde{x}) \max_{\tilde{w} \in \mathscr{W}} \theta_{t+1}(\tilde{w}, \tilde{x}) \,\middle|\, \boldsymbol{\theta}_t, \Sigma_t, W_{t+1} = w, X_{t+1} = x\right] - \sum_{\tilde{x} \in \mathscr{X}} \tilde{p}(\tilde{x}) \max_{\tilde{w} \in \mathscr{W}} \theta_t(\tilde{w}, \tilde{x}).$$

The only differences between Algorithms 1 and 2 are that Algorithms 2 does not take x as an input but instead loops over all treatment-type combinations in Step 2, and selects the largest index over the set of treatment-type combinations in Step 11. Because we loop over all treatment-type combinations, the complexity of Algorithm 2 is $O(m^2 n^2 \log(n))$.

5 MODELING CORRELATION ACROSS TREATMENTS AND PATIENT TYPES

In this section, we discuss how different correlation structures in the prior covariance matrix impact our model's optimal solution and the *f*EVI algorithm. We first discuss a special correlation structure (correlations only within patient types) in which the optimal allocation policy can be obtained by breaking up the problem into independent subproblems that can be solved with classical R&S methods. The *f*EVI algorithm resembles EVI algorithms designed to solve the R&S problem (not contextual) in this special case. We then discuss the correlation structure used in a well-known clinical trial and how it fits our model's scope. Finally, we describe a correlation structure that illustrates correlations across treatments and patient types.

Algorithm 2 *f*EVI choosing covariates.

function *f*EVI choosing covariates($\boldsymbol{\theta}_t, \Sigma_t$)
  for all $(w,x) \in \mathscr{W} \times \mathscr{X}$ **do**
    $\mathbf{b} \leftarrow \tilde{\boldsymbol{\sigma}}(\Sigma_t, w, x)$
    for all $\tilde{x} \in \mathscr{X}$ **do**
      $\tilde{\mathbf{a}} \leftarrow \boldsymbol{\theta}_t(\mathscr{W}, \tilde{x})$ ▷ subset posterior means to keep only the entries associated with type $\tilde{x}$
      $\tilde{\mathbf{b}} \leftarrow \mathbf{b}(\mathscr{W}, \tilde{x})$ ▷ subset predictive standard deviations to keep only the entries associated with type $\tilde{x}$
      $\tilde{\nu}_t(w, \tilde{x}) \leftarrow h(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ ▷ h is the procedure in Frazier et al. (2009)
    end for
    $\nu_t(w,x) \leftarrow \sum_{\tilde{x} \in \mathscr{X}} \tilde{p}(\tilde{x}) \tilde{\nu}_t(w, \tilde{x})$
  end for
  return $(W_{t+1}, X_{t+1}) = \arg\max_{(w,x) \in \mathscr{W} \times \mathscr{X}} \nu_t(w,x)$
end function

5.1 Correlation Only Within Patient Types

Consider a trial in which patient types have highly different characteristics, but the treatments have similar targets. For instance, consider types as sepsis patients with different organ dysfunctions and treatments as inhibitors of different molecules that drive the same immune response. A prior distribution on $\boldsymbol{\mu}$ that would appropriately model such a trial would impose elements associated with the same type to be correlated and elements associated with different types to be uncorrelated. That is, the non-zero correlations in the prior distribution are only for elements that are associated with the same patient type: if $x_1 \neq x_2$, then $\mathrm{Cov}(\mu(w_1,x_1), \mu(w_2,x_2)) = 0$ for any $w_1, w_2 \in \mathscr{W}$. Here, the problem can be simplified to m independent R&S subproblems with random sample sizes for each patient type because, by learning about one patient type, we do not learn about other patient types. Thus, for each patient type, we search for an allocation policy π_x, which is only concerned with the allocation of patients of type x, that solves the following R&S problem:

$$\max_{\pi_x} \mathbb{E}^{\pi_x} \left[\max_{w \in \mathscr{W}} \theta_T(w,x) \right].$$

For such a prior distribution, the *f*EVI allocation policy is equivalent to an allocation policy that implements the *Knowledge Gradient with correlated beliefs* (cKG) of Frazier et al. (2009) for each patient type independently. If $\boldsymbol{\mu}$ has fully uncorrelated elements, i.e., Σ_0 is a diagonal matrix, the *f*EVI allocation policy further simplifies to the *Knowledge Gradient with uncorrelated beliefs* (iKG) of Frazier et al. (2008).

When prior correlation is only within patient types, it is appropriate to have independent allocation policies for each type. However, if we can choose the patient types to be observed, as in simulation optimization problems, it is necessary to perform a joint optimization for all patient types because the allocation policy has to allocate the budget of observations for each type.

5.2 Correlation Only Within Treatments

Consider now the setting of the BATTLE trial (Zhou et al. 2008) for lung cancer. The patient types were all lung cancer patients with differences in a small number of genes in tumor biopsies. The treatments were previously approved drugs for other types of cancer. Thus, patient types were similar, but the treatments had substantial differences. A prior distribution on $\boldsymbol{\mu}$ that would appropriately model such a trial would impose elements associated with the same treatment to be correlated and elements associated with different treatments to be uncorrelated. That is, the non-zero correlations in the prior distribution are only for elements that are associated with the same treatment: if $w_1 \neq w_2$, then $\mathrm{Cov}(\mu(w_1,x_1), \mu(w_2,x_2)) = 0$ for any $x_1, x_2 \in \mathscr{X}$. For this prior distribution, we cannot reduce the problem into smaller subproblems as in

Section 5.1 because learning about one patient type may lead to learning about other patient types, and a good allocation policy needs to take into account the gains for all patient types.

The BATTLE trial implemented a hierarchical model to pool observations with the same treatment:

$$\mu(w,x) \mid \phi(w) \overset{i.i.d.}{\sim} \mathcal{N}(\phi(w), \sigma_\mu^2) \ \forall x \in \mathscr{X}$$
$$\phi(w) \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\phi^2) \ \forall w \in \mathscr{W},$$

where $\phi(w)$ are the hyperparameters associated with the expected outcomes of patients treated with w, and σ_μ^2 and σ_ϕ^2 are given variance parameters. This hierarchical model can be formulated as a special case of our model in (3), in the sense that the prior and posterior marginal distributions of $\boldsymbol{\mu}$ are the same, if we specify the prior parameters as follows (Gelman et al. 2013, Section 15.1, *Intraclass correlation*). Let Σ_0 be blockdiagonal with each block having size m, i.e., there is a correlation within treatment and no correlation across different treatments. Let each diagonal element be equal to $\sigma_\mu^2 + \sigma_\phi^2$ and each off-diagonal element within the block diagonal be equal to σ_ϕ^2. Finally, let $\boldsymbol{\theta}_0$ be the zero vector. Thus, the prior distribution used in the BATTLE trial is a special case of our prior in (3) with only positive correlations within treatments.

5.3 Correlation Both Within Treatments and Patient Types

The more general case has correlations both within treatments and patient types, i.e., there are similarities among the patient types and the treatments. Here is when we expect to see the most benefits from the *f*EVI allocation policy. For example, if patient types have similar or overlapping characteristics and treatments are similar, we could see a prior distribution that assigns a positive correlation for the parameters associated with the same treatment or the same patient type. In Section 6, we use this prior to illustrate the benefits of the *f*EVI allocation policy.

Instead of assigning correlation in the prior, it would be desirable to learn correlations from the observations. However, learning correlations would require more structure than our model's assumptions. For instance, we could use a linear model with multiple patient covariates as in Shen et al. (2021) or model treatment correlations as in Chick et al. (2021). We leave this as an opportunity for future research.

6 NUMERICAL RESULTS

In this section, we specify the model parameters to simulate the trial value and draw further insight into the following questions:

1. How does the *f*EVI allocation policy perform compared to other baseline policies? How does the correlation structure affect performance?
2. What is the performance gap between being able to choose covariates and observing covariates at random?

In Section 6.1, we describe the experimental setup and metrics used to assess the performance of the *f*EVI allocation policy. Section 6.2 describes the competing allocation policies. In Section 6.3, we simulate the performance of the policies with random instances drawn from the prior distribution to assess the performance of the *f*EVI allocation policy. Section 6.4 compares the performance between when random patient types are observed and when the allocation policy chooses the patient types. Finally, Section 6.5 assesses the performance of the allocation policies when specific instances of the true mean are fixed.

6.1 Experimental Setup

We use the following parameters as a proof of concept motivated from different clinical trial settings. We assume four patient types ($m = 4$), consistent with medical applications, such as sepsis (Scicluna et al. 2017), and close to the BATTLE trial that identified five types. We assume eight treatments ($n = 8$) representing

the 2^3 combinations of three aspects of treatment where each aspect has two options. We assume a zero vector for the prior mean, i.e., there is no belief that any treatment of patient type has better outcomes than the rest. The prior covariance matrix is assumed to have the following structure representing a setting in which the *f*EVI is particularly beneficial, as discussed in Section 5.3. The variance of each parameter is fixed to one. The correlation between two parameters is equal to ρ if the parameters are associated with the same treatment or patient type. As an illustration, for two treatment and two patient types, the prior covariance matrix would be as follows:

$$\Sigma_0 = \begin{pmatrix} 1 & \rho & \rho & 0 \\ \rho & 1 & 0 & \rho \\ \rho & 0 & 1 & \rho \\ 0 & \rho & \rho & 1 \end{pmatrix}.$$

We assume $\rho = 0.3$, a moderate correlation, and assume the sampling variance to be $\sigma^2 = 1$. We assume that the probability of observing any patient type in the population and in the trial is the same: $p(x) = \tilde{p}(x) = 1/m$.

We report as the main performance metric the Expected Opportunity Cost (EOC), defined as the difference between the value function with a perfect treatment strategy and the value function of the allocation policy (consistent with extant literature with random problem instances):

$$\text{EOC}^{\pi} = \mathbb{E}^{\pi}\left[\max_{w \in \mathscr{W}} \mu(w, \tilde{X})\right] - V^{\pi}.$$

The EOC only shifts and mirrors our value function in (6), so minimizing the EOC is equivalent to maximizing the value function. We report the EOC because it is guaranteed to be positive and approaches zero with good policies. Thus, it is easier to visualize and compare among competing allocation policies.

We also report the instance-dependent $\text{EOC}(\boldsymbol{\mu})$ in Section 6.5:

$$\text{EOC}^{\pi}(\boldsymbol{\mu}) = \mathbb{E}\left[\max_{w \in \mathscr{W}} \mu(w, \tilde{X}) \;\middle|\; \boldsymbol{\mu}\right] - \mathbb{E}^{\pi}\left[\mu(f_{\boldsymbol{\theta}_T}(\tilde{X}), \tilde{X}) \;\middle|\; \boldsymbol{\mu}\right].$$

consistent with extant literature for specific problem instances. Here, $\boldsymbol{\mu}$ is fixed and the expectation is over the patient types $\tilde{X}$ and posterior mean $\boldsymbol{\theta}_T$. The instance-independent EOC is the expectation over $\text{EOC}(\boldsymbol{\mu})$ given the prior distribution of $\boldsymbol{\mu}$: $\text{EOC} = \mathbb{E}[\text{EOC}(\boldsymbol{\mu}) \mid \boldsymbol{\theta}_0, \Sigma_0]$. We consider specific instances discussed in the literature to assess the performance of the allocation policies that we describe in Section 6.5.

6.2 Competing Allocation Policies

The *f*EVI allocates patients to treatment as described in Section 4 using the prior distribution defined in Section 6.1. The iKG makes allocations using the *f*EVI with the same prior mean but with the identity matrix as the prior covariance, i.e., assuming there is no correlation among parameters. The iKG makes the same allocations as the iKG in Frazier et al. (2008), assuming that each patient type is a separate R&S problem. The allocation policy *random* samples a treatment uniformly at random and *round robin* samples in a round robin fashion, regardless of the patient type.

In Section 6.4, we also report the "*f*EVI choosing covariates" that is only valid if the allocation policy chooses the patient types. It makes allocations using Algorithm 2.

6.3 Expected Opportunity Cost Comparison

Figure 1 (left panel) shows the EOC in a log-scale as a function of the sample size. The *f*EVI allocation policy performs best among all competing allocation policies. Notice that iKG has equivalent performance to *f*EVI for larger sample sizes but a significant gap for smaller sample sizes. While the correlation structure speeds up learning in earlier stages, the correlation (off-diagonal entries) of the posterior covariance matrix

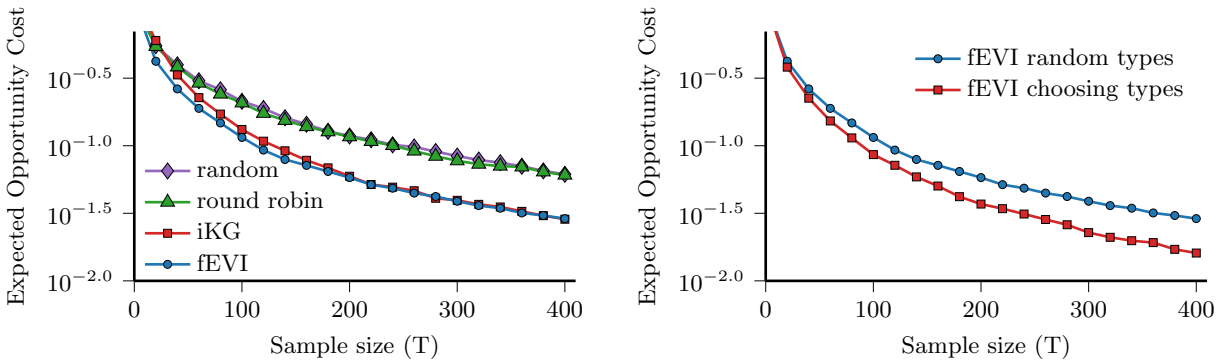

Figure 1: Expected opportunity cost with random instances drawn from the prior distribution. Standard errors were too small to be plotted. The left panel compares allocation policies when patient types are random. The right panel compares the performance of the *f*EVI allocation policy when being able to choose the patient types, versus when observing patient types at random.

diminishes with the sample size and loses relevance for larger sample sizes. We assessed the EOC when the number of treatments and number of types were in the ranges between two and ten (results not shown due to space constraints). We observed that the same insights hold and that the *f*EVI benefits become larger with more available treatments.

The distinction between *f*EVI and iKG is the prior distribution. The results we have presented in this section draw random instances from the prior used by the *f*EVI. We also obtained results by drawing random samples from the prior used by the iKG (i.e., independent mean outcomes). We found that the iKG performs better than the *f*EVI by about the same margin by which *f*EVI outperforms iKG in Figure 1 (left panel). Thus, using the prior distribution that generates the random instances leads to improvements in EOC, particularly in small samples.

6.4 Comparison Between Choosing and Random Patient Types

Figure 1 (right panel) shows a comparison between the *f*EVI allocation policy when covariates are random (our new Algorithm 1) and when covariates are chosen (Algorithm 2, a special case of Zhang et al. 2019). Far fewer samples are required to achieve a given EOC threshold when choosing covariates. When patients arrive through a random process, the allocation policy can only allocate the patient type that would yield the greatest value of information with probability $1/m$ in this example.

To better compare the performance between the allocation policies presented in Figure 1, we now report the sample size required to reach an EOC lower than 10^{-1}: *f*EVI choosing covariates(91), *f*EVI(116), iKG(130), and *random*(253). At this level of EOC, being able to choose covariates reduces the required sample size (from 116 to 91) by more than using the correct prior distribution (from 130 to 116). However, the benefits of using the EVI approach compared to *random* is much larger (from 253 to 116). Furthermore, the benefits of using the correct prior diminish for larger sample sizes (lower EOC). We do not observe any statistically significant difference between *f*EVI and iKG for sample sizes larger than 148 (EOC < 0.076).

While choosing covariates may be infeasible in most clinical trial contexts, this result suggests that innovative approaches that allow to select covariates — at least partially, e.g. by selectively enrolling patients — could allow to capture significant additional benefits.

6.5 Instance-Dependent Expected Opportunity Cost

To further understand the instances where the *f*EVI has benefits over the other allocation policies, we consider the following instances of the true mean to evaluate the instance-dependent EOC.

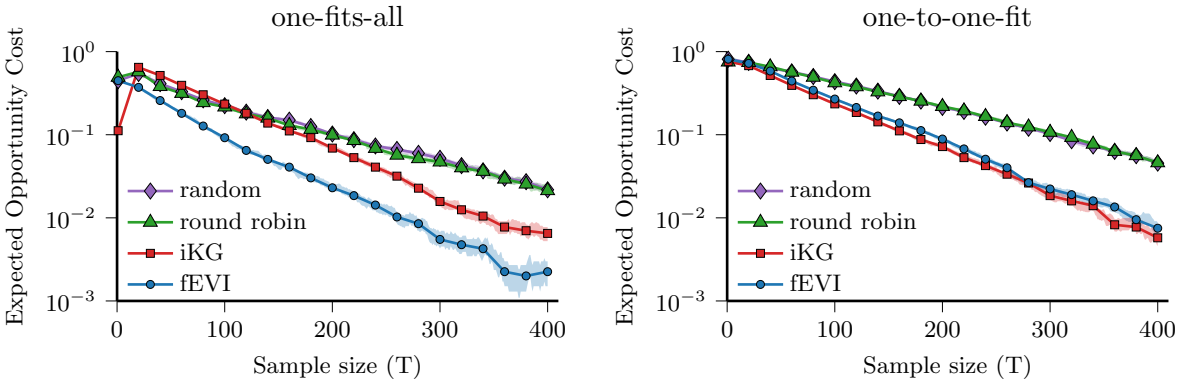

Figure 2: Instance-dependent EOC for the one-fits-all (left panel) and one-to-one-fit (right panel) instances. The shaded area represents standard errors. *random* and *round robin* are mostly overlapping.

One treatment works for all patients. Figure 2 (left panel) reports the EOC under random patient types for the *generalized slippage configuration* (GSC) of Shen et al. (2021). In this instance, Treatment 1 is the best treatment for any type by a constant margin. We assume that the expected outcome for any patient receiving Treatment 1 is $\delta = 1$, and zero for all other treatments: $\mu(w,x) = \mathbf{1}_{w=1}\delta$, where $\mathbf{1}$ is the indicator function. We refer to this instance as the *one-fits-all*. The *f*EVI outperforms the iKG because it starts with a prior that suggests that if Treatment 1 works for any type, it works for other types as well. Thus, *f*EVI learns that Treatment 1 works for all types with fewer observations.

A different treatment works for one and only one patient type. Figure 2 (right panel) reports an instance in which each patient type has one treatment with high expected outcomes while the remaining treatments have zero mean: $\mu(w,x) = \mathbf{1}_{w=x}\delta$, $\delta = 1$. This resembles some instances reported by Zhou et al. (2008) and Lai et al. (2013). We refer to this instance as the *one-to-one-fit*. In this instance, iKG outperforms *f*EVI by a small margin. The correlation structure of the prior used by *f*EVI may be misleading because each type has a different treatment that obtains the largest expected outcomes. However, because it is just a moderate correlation, the *f*EVI still learns the optimal treatment strategy with a performance comparable to iKG.

We observe the same qualitative results when $\delta = 0.5$ or $\delta = 2$ (data not shown).

7 CONCLUSION

The contextual R&S literature has focused on the simulation optimization problem. Clinical trials pose additional practical challenges that require specific modeling. In this paper, we have focused on modeling the random arrival of patients instead of the typical assumption in simulation optimization that covariates are under the control of the experimenter. We have developed the *f*EVI allocation policy and characterized scenarios (different kinds of prior distributions) in which contextual R&S is and is not beneficial. We have shown through simulations that choosing covariates may lead to significant gains in selecting treatment strategies. While it may not be feasible to choose every patient's type, we may obtain better performance by denying enrollment to certain patient types, in order to save budget for other patients that are more informative. Modeling such decisions requires investigating different trade-offs because denying enrollment increases the length of the trial and entails costs of screening for patients.

Future work could address other practical challenges. One very relevant extension for clinical trial applications is to allow for a delay in observing the outcomes of the patients. In addition, the high cost of clinical trials suggests that policies should also be after an optimal stopping time that saves costs without sacrificing a good selection of treatment strategy.

REFERENCES

Bastani, H., and M. Bayati. 2020. "Online Decision Making with High-Dimensional Covariates Online Decision Making with High-Dimensional Covariates". *Operations Research* 68(1):276–294.

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure". *Management Science* 53(12):1916–1932.

Chen, C.-H., S. E. Chick, L. H. Lee, and N. A. Pujowidianto. 2015. "Ranking and Selection: Efficient Simulation Budget Allocation". In *Handbook of Simulation Optimization*, edited by M. C. Fu, Volume 216 of *International Series in Operations Research & Management Science*, Chapter 3, 45–80. Springer.

Chick, S. E., N. Gans, and O. Yapar. 2021. "Bayesian Sequential Learning for Clinical Trials of Multiple Correlated Medical Interventions". *Management Science*:to appear.

Chick, S. E., and K. Inoue. 2001. "New Two-Stage and Sequential Procedures for Selecting the Best Simulated System". *Operations Research* 49(5):732–743.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2008. "A Knowledge-Gradient Policy for Sequential Information Collection". *SIAM Journal on Control and Optimization* 47(5):2410–2439.

Frazier, P. I., W. B. Powell, and S. Dayanik. 2009. "The Knowledge-Gradient Policy for Correlated Normal Beliefs". *INFORMS Journal on Computing* 21(4):599–613.

Gao, S., J. Du, and C.-H. Chen. 2019. "Selecting the Optimal System Design under Covariates". In *IEEE 15th International Conference on Automation Science and Engineering*, 547–552. Institute of Electrical and Electronics Engineers, Inc.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. CRC press.

Goldenshluger, A., and A. Zeevi. 2013. "A Linear Response Bandit Problem". *Stochastic Systems* 3(1):230–261.

Kim, S.-H., and B. L. Nelson. 2006. "Selecting the Best System". In *Handbook in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson. Elsevier.

Lai, T. L., O. Y.-w. Liao, and D. W. Kim. 2013. "Group Sequential Designs for Developing and Testing Biomarker-guided Personalized Therapies in Comparative Effectiveness Research". *Contemporary Clinical Trials* 36:651–663.

Li, H., H. Lam, Z. Liang, and Y. Peng. 2020. "Context-Dependent Ranking and Selection Under a Bayesian Framework". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 2060–2070. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pearce, M., and J. Branke. 2017. "Efficient Expected Improvement Estimation for Continuous Multiple Ranking and Selection". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 2161–2172. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Powell, W. B., and I. O. Ryzhov. 2012. *Optimal Learning*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Scicluna, B. P. et al. 2017. "Classification of Patients with Sepsis According to Blood Genomic Endotype: A Prospective Cohort Study". *The Lancet Respiratory Medicine* 5(10):816–826.

Shen, H., L. J. Hong, and X. Zhang. 2021. "Ranking and Selection with Covariates for Personalized Decision Making". *INFORMS Journal on Computing* 33:to appear.

Villar, S. S., and W. F. Rosenberger. 2018. "Covariate-adjusted Response-Adaptive Randomization for Multi-arm Clinical Trials using a Modified Forward Looking Gittins Index Rule". *Biometrics* 74(1):49–57.

Xie, J., P. I. Frazier, and S. E. Chick. 2016. "Bayesian Optimization via Simulation with Pairwise Sampling and Correlated Prior Beliefs". *Operations Research* 64(2):542–559.

Xiong, S. 2020. "Personalized optimization and its implementation in computer experiments". *IISE Transactions* 52(5):528–536.

Zhang, X., H. Shen, L. J. Hong, and L. Ding. 2019. "Knowledge Gradient for Selection with Covariates: Consistency and Computation". https://arxiv.org/pdf/1906.05098v1.pdf.

Zhou, X., S. Liu, E. S. Kim, R. S. Herbst, and J. J. J. Lee. 2008. "Bayesian Adaptive Design for Targeted Therapy Development in Lung Cancer - A Step Toward Personalized Medicine". *Clinical Trials* 5(3):181–193.

AUTHOR BIOGRAPHIES

ANDRES ALBAN is a PhD candidate in Technology and Operations Management at INSEAD. He has a BS degree from New Jersey Institute of Technology in Applied Mathematics and Physics. His email address is andres.alban@insead.edu.

STEPHEN E. CHICK is a Professor of Technology and Operations Management and the Novartis Chair of Healthcare Management at INSEAD. He works in the areas of simulation analysis, sequential optimization, health care management, and Bayesian inference. His email address is stephen.chick@insead.edu.

SPYROS I. ZOUMPOULIS is an Assistant Professor of Decision Sciences at INSEAD. He works on optimal personalized decision and learning problems; and on applications in marketing and healthcare. His email address is spyros.zoumpoulis@insead.edu.