



## Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### A Contextual Ranking and Selection Method for Personalized Medicine

Jianzhong Du, Siyang Gao, Chun-Hung Chen

To cite this article:

Jianzhong Du, Siyang Gao, Chun-Hung Chen (2024) A Contextual Ranking and Selection Method for Personalized Medicine. Manufacturing & Service Operations Management 26(1):167-181. <https://doi.org/10.1287/msom.2022.0232>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# A Contextual Ranking and Selection Method for Personalized Medicine

Jianzhong Du,<sup>a</sup> Siyang Gao,<sup>b,\*</sup> Chun-Hung Chen<sup>c</sup>

<sup>a</sup>School of Management, Fudan University, Shanghai 200433, China; <sup>b</sup>Department of Systems Engineering and School of Data Science, City University of Hong Kong, Hong Kong, China; <sup>c</sup>Department of Systems Engineering and Operations Research, George Mason University, Fairfax, Virginia 22030

\*Corresponding author

Contact: jianzhodu2-c@my.cityu.edu.hk,  <https://orcid.org/0000-0002-5355-5902> (JD); siyangao@cityu.edu.hk,  <https://orcid.org/0000-0002-3574-6393> (SG); cchen9@gmu.edu,  <https://orcid.org/0000-0001-7824-1814> (C-HC)

Received: May 16, 2022

Revised: February 26, 2023; June 24, 2023

Accepted: July 24, 2023

Published Online in Articles in Advance:  
August 22, 2023

<https://doi.org/10.1287/msom.2022.0232>

Copyright: © 2023 INFORMS

**Abstract.** *Problem definition:* Personalized medicine (PM) seeks the best treatment for each patient among a set of available treatment methods. Because a specific treatment does not work well on all patients, traditionally, the best treatment was selected based on the doctor's personal experience and expertise, which is subject to human errors. In the meantime, stochastic models have been well developed in the literature for a lot of major diseases. This gives rise to a simulation-based solution for PM, which uses the simulation tool to evaluate the performance for pairs of treatment and patient biometric characteristics and, based on that, selects the best treatment for each patient characteristic. *Methodology/results:* In this research, we extend the ranking and selection (R&S) model in simulation-based decision making to solving PM. The biometric characteristics of a patient are treated as a context for R&S, and we call it contextual ranking and selection (CR&S). We consider two formulations of CR&S with small and large context spaces, respectively, and develop new techniques for solving them and identifying the rate-optimal budget allocation rules. Based on them, two selection algorithms are proposed, which can be shown to be numerically superior via a set of tests on abstract and real-world examples. *Managerial implications:* This research provides a systematic way of conducting simulation-based decision-making for PM. To improve the overall decision quality for the possible contexts, more simulation efforts should be devoted to contexts in which it is difficult to distinguish between the best treatment and non-best treatments, and our results quantify the optimal trade-off of the simulation efforts between the pairs of contexts and treatments.

**Funding:** J. Du is partially supported by the National Natural Science Foundation of China [Grant 72091211]. C.-H. Chen is partially supported by the National Science Foundation under Awards FAIN212368.

**Supplemental Material:** The e-companion is available at <https://doi.org/10.1287/msom.2022.0232>.

**Keywords:** personalized medicine • contextual ranking and selection • simulation optimization • OCBA • convergence rate

## 1. Introduction

Personalized medicine (PM) (also known as precision medicine or P4 medicine) is an emerging healthcare problem. Benefiting from the advance of medical knowledge and technology, patients usually have access to a set of competing and sometimes complementary medical treatment methods for their diseases. However, the treatment used for a patient should be carefully chosen, because the effectiveness of the treatment might depend heavily on the patient's biometric characteristics. For instance, the highly active antiretroviral therapy (a standard treatment for AIDS) has substantially different profiles in efficacy and toxicity across subgroups of patients, influenced by the virus level at the time of receiving treatments and the gender and behavior pattern of the

patient (Cai et al. 2011). In this research, we call such biometric characteristics contexts. PM aims to determine the best treatments for contexts that might appear in practice and thus provides tailored treatment for each patient. This is substantially different from the traditional evaluation of treatment performance (World Health Organization 2003, chapter 1).

PM involves evaluating the effectiveness of medical treatments under different contexts. Typically, there are two ways to do it, by trial-based and model-based approaches. The trial-based approach applies statistical analysis to a series of well-designed clinical trials and is capable of supporting personalized medicine with a large set of contexts (Schork 2015). However, this approach suffers from major ethical issues. Statistical

analysis and inference rely on comparing the results of the treatment group and control group. It is unethical if the patients in the control group have worsening progression and are not allowed to take experimental drugs (Mok 2011). In addition, this approach is further complicated by the prohibitively large amount of resources needed for following up the test results and making decisions. These drawbacks of the trial-based approach can limit its implementation in the real world (Hamburg and Collins 2010).

The model-based approach employs mathematical models to depict the progression of the disease and, based on it, assesses the effectiveness of the treatment (Garnett et al. 2011). It does not involve making experiments on humans and thus can avoid the above-mentioned ethical and resource-related issues in trial-based approaches. The evaluation model is generally stochastic because of the uncertainty in the model structure and transition and the estimation of the model parameters. From the personalized perspective, the effectiveness of a treatment is also random among individuals, even under the same patient context (Brennan et al. 2006). There has been a rich body of literature on the application of stochastic models to healthcare problems, for example, the epidemiological transmission dynamics (Chick et al. 2001, 2008; Alonso et al. 2007) and HIV preventions for susceptible populations (Tan 2012).

In this research, we will focus on the model-based approach for PM and study the problem of efficiently identifying the best treatment under all the possible patient contexts within a finite computing time. In the view that stochastic models for practical problems can be large-scaled, complex, and not analytical, we will use the generic tool of simulation to evaluate the performance of them.

In simulation experiments, designs (a terminology in systems engineering, analogous to treatments in medical decision problems) are simulated for multiple replications. Their performance estimators (typically sample means) are compared, and the estimated best design is selected. This practice imposes two challenges for the purpose of PM. First, for a given patient context, the probability that we correctly select the true best treatment is always less than one with a finite simulation budget. The randomness in the model will cause a non-best treatment to occasionally outperform the best, leading us to a suboptimal decision. Second, the difficulties in correctly selecting the best treatment vary among contexts. The best treatment under some contexts is easy to identify, whereas for some other contexts, it can be highly difficult.

To address these two challenges, in this research, we propose to utilize the simulation budget to maximize the chance of identifying the best treatment under each possible context. It is achieved by smartly controlling the number of simulation replications

allocated to each pair of contexts and treatments so as to concentrate the computing efforts on contexts where the best treatment is more difficult to identify. By doing so, the best treatment under each context can be correctly selected with a higher confidence in a limited time. We call this problem contextual ranking and selection (CR&S).

CR&S is closely related to two streams of literature. The first is ranking and selection (R&S). R&S is a well-established model in the field of simulation optimization. It aims to allocate the simulation budget to a set of competing system designs in order to efficiently select the best one. Representative R&S methods include the optimal computing budget allocation (OCBA) (Chen et al. 2000; Gao et al. 2017a,b), value of information procedures (VIP) (Frazier et al. 2008, Chick et al. 2010), and indifference-zone (IZ) methods (Kim and Nelson 2001, Nelson et al. 2001). However, these procedures do not consider contexts and thus cannot be applied to CR&S (Goodwin et al. 2022).

Recently, Shen et al. (2021) considered the R&S problem in the presence of continuous contexts and used the linear models to predict the design performance. Li et al. (2018) generalized the method of Shen et al. (2021) to handle high-dimensional context spaces. However, these two studies pursued a different goal from this research, which was to provide performance guarantees for the designs (treatments) selected instead of optimizing the design performance. Compared with them, the OCBA-type method is less conservative in the sense that it can achieve better design performance with a less-simulation budget (Branke et al. 2007) at the cost of losing the performance guarantee on the designs selected. Therefore, our model and method are more appropriate when efficiency is important (e.g., when the simulation budget available is relatively small).

In addition, Hu and Ludkovski (2017) and Pearce and Branke (2017) considered the large-scale problem of CR&S and used the method of Bayesian optimization to solve it. They employed the stochastic kriging model for predictions of the design performance and focused on how to search the design and context spaces rather than developing the budget allocation rules. The performances of their algorithms were studied only empirically. Ding et al. (2022) extended the algorithm in Pearce and Branke (2017) and showed that the new algorithm was consistent. Compared with these works, this research aims at the budget allocations of the small-scale and large-scale problems and shows that the proposed budget allocation rules and selection algorithms are asymptotically optimal, which is a stronger result than consistency.

The second stream of literature related to CR&S is the best arm identification (BAI). BAI is studied more in the fields of statistics and machine learning, but it is a very similar model to R&S, aiming to identify the best

arm from a finite set by adaptively pulling the arms and learning their rewards without consideration of contexts (Audibert et al. 2010, Kaufmann et al. 2016, Russo 2020). Recently, BAI has also been extended to the context environment, known as contextual bandits (Li et al. 2022). In bandit problems, the sampling object is typically the real system, whereas in R&S, the sampling object is the simulation model. The different sampling objects do not cause any differences when solving BAI and R&S problems but make the structures of contextual bandits and CR&S problems fundamentally different. In contextual bandits, contexts are associated with the real system and are thus out of the experimenter's control; that is, the experimenter can decide only which arm to sample given the context that appears but cannot decide which context to appear or sample (Tewari and Murphy 2017). In contrast, CR&S considers an entirely simulated environment, in which contexts are also input variables to the simulation model and are controlled by the experimenter. As a result, the experimenter needs to decide both the context and design (the context-design pair) to sample. It leads to a different and more complex decision problem.

Our contributions in this research are fourfold. First, we study three measures for evaluating the evidence of correct selection over the context space. These measures are extensions of the probability of correct selection (PCS) used in R&S to the contextual setting and are capable of depicting the quality of the estimated best treatment under all the possible contexts. We show that the three measures are asymptotically equivalent in the sense that they have the same rate function.

Second, we propose two formulations for the PM problem. Both formulations optimize the rate function of the three measures under a simulation budget constraint. One formulation samples all treatment-context pairs and is suitable for a small context space. In the other formulation, treatment performance and context are assumed to have linear relationship. This is suitable for a large context space.

Third, for both formulations, we develop the rate-optimal selection rules and devise easily implementable selection algorithms, called CR&S Algorithms 1 and 2. We show that the two algorithms can recover the rate-optimal selection rules.

Last, we conduct extensive numerical experiments to assess the performances of the two algorithms. We first test them on a set of benchmark functions and demonstrate their superiority in solving different types of problems. Next, we apply the algorithms to two real-world PM problems and obtain the medical decision maps for them.

The rest of the paper is organized as follows. Section 2 introduces the basic notation and assumptions. Section 3 studies three objective measures of CR&S and their

rate functions. Sections 4 and 5 consider the PM problem with small and large context spaces, respectively. They formulate and solve the selection problems, develop selection algorithms for implementation, and theoretically study their performances. Numerical examples and computational results are provided in Section 6, followed by conclusions and discussion in Section 7. The allocation rule and selection algorithm for the small-scale problem have been presented without proof in Gao et al. (2019).

## 2. Preliminaries

Suppose there are  $k$  different treatments. The performance of each treatment depends on  $\mathbf{X} = (X_1, \dots, X_d)^\top$ , a vector of random contexts with support  $\mathcal{X} \subseteq \mathbb{R}^d$ . For each treatment  $i = 1, 2, \dots, k$ , let  $Y_{il}(\mathbf{x})$  be the  $l$ th simulation sample from treatment  $i$  and context  $\mathbf{x}$  and  $y_i(\mathbf{x})$  be the mean performance of this treatment. We have  $Y_{il}(\mathbf{x}) = y_i(\mathbf{x}) + \epsilon_{il}(\mathbf{x})$ , where  $\epsilon_{il}(\mathbf{x})$  is the random noise incurred in the simulation. Denote  $n_i(\mathbf{x})$  as the number of simulation replications for treatment  $i$  and context  $\mathbf{x}$ . The sample mean  $\bar{Y}_i(\mathbf{x}) = \frac{1}{n_i(\mathbf{x})} \sum_{l=1}^{n_i(\mathbf{x})} Y_{il}(\mathbf{x})$ . Without loss of generality, we let the best treatment  $i^*(\mathbf{x})$  under context  $\mathbf{x}$  be the treatment with the smallest mean performance.

Throughout the paper, we assume that  $\mathcal{X}$  has a finite number of  $m$  possible contexts  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ . This setting aligns with context spaces that are finite in nature. For infinite context spaces (continuous or discrete and unbounded), we usually do not need to find the best treatment for each context; instead, a common practice is to classify the values of context variables into a number of categories/levels. For example, when treating diabetic patients, a key context variable is the body mass index (BMI) of the patients, and it takes real values. Two possible ways to process BMI is to classify it into categories <18.5 underweight, 18.5–24.9 normal weight, 25.0–29.9 overweight, and  $\geq 30.0$  obesity (World Health Organization 2010) or, more accurately, into levels <18, 18, 19,  $\dots$ , 29, 30 and  $>30$ . To this end, the finite setting provides great flexibility in the level of contextual discrepancy we want to distinguish when formulating the problem.

Specifically, we consider two cases for the context space. The first case is when the context space is small, and we have time to simulate all the treatment-context pairs. The second case is when the context space is large, and we only have time to simulate treatments under a fraction of contexts. In this case, we further assume that all contexts lie on a grid, and the relationship between the treatment performance and contexts can be described by linear models, so the performances of treatments under unsimulated contexts can be interpolated. In this research, we call them *small-scale problem* and *large-scale problem*, respectively.

Suppose  $n$  is our total simulation budget (number of simulation replications), and  $n_{i,j}$  is the number of simulation replications that we allocate to treatment  $i$  under context  $\mathbf{x}_j$ . Let  $\alpha_{i,j} = n_{i,j}/n$  and  $\boldsymbol{\alpha} = (\alpha_{1,1}, \alpha_{2,1}, \dots, \alpha_{k,1}, \alpha_{1,2}, \alpha_{2,2}, \dots, \alpha_{k,2}, \dots, \alpha_{1,m}, \alpha_{2,m}, \dots, \alpha_{k,m})$  be the vector of  $\alpha_{i,j}$ . We make the following technical assumptions in our analysis.

**Assumption 1.** The best treatment  $i^*(\mathbf{x})$  is unique for all  $\mathbf{x} \in \mathcal{X}$ .

**Assumption 2.**  $Y_{il}(\mathbf{x})$ 's are independent across different  $i$ ,  $l$ , and  $\mathbf{x}$ .

**Assumption 3.**  $Y_{il}(\mathbf{x})$ 's are normally distributed with mean  $y_i(\mathbf{x})$  and variance  $\sigma^2(\mathbf{x})$ .

Assumption 1 assumes that the best treatment under each of the  $m$  contexts is unique, because two treatments with the same mean performance cannot be distinguished. The assumptions of independence and normality of samples in Assumptions 2 and 3 are standard in the simulation optimization literature (Law and Kelton 2000). The independence between simulation samples can be achieved by using independent sequences of random numbers in different simulation runs. The normality assumption is typically satisfied in simulation because the output is obtained from an average performance or batch means. According to the Central Limit Theorem, it is approximately normal.

### 3. Objective Measures

In this section, we discuss three objective measures for PM. Next, we analyze the rate functions of the three measures and establish their equivalence.

Suppose performance  $y_i(\mathbf{x})$  of treatment  $i$  under context  $\mathbf{x}$  is estimated by  $\hat{y}_i(\mathbf{x})$ . For context  $\mathbf{x}$ , a correct selection happens when the estimated best treatment  $\hat{i}^*(\mathbf{x})$  is identical to the real best treatment  $i^*(\mathbf{x})$ . However, the correct selection can never be guaranteed in practice with a finite simulation budget. Under a fixed context  $\mathbf{x}$ , traditional R&S typically assesses the quality of the selection for the best treatment by the probability of correct selection (PCS)

$$\begin{aligned} \text{PCS}(\mathbf{x}) &= \mathbb{P}(\hat{i}^*(\mathbf{x}) = i^*(\mathbf{x})) \\ &= \mathbb{P}\left(\bigcap_{i=1, i \neq i^*(\mathbf{x})}^k (\hat{y}_{i^*(\mathbf{x})}(\mathbf{x}) < \hat{y}_i(\mathbf{x}))\right), \end{aligned}$$

and seeks to either maximize this probability or guarantee a prespecified level for it. The probability here is taken with respect to the random noises in the simulation samples.

In CR&S, each context  $\mathbf{x}$  is associated with an R&S problem. We want to provide the best treatments for all the  $m$  contexts, and therefore, we need measures for evaluating the quality of the selection over the entire

context space  $\mathcal{X}$ . To fulfill this need, we consider the following three measures based on PCS:

$$\text{PCS}_E = \mathbb{E}[\text{PCS}(\mathbf{X})] = \sum_{j=1}^m p_j \text{PCS}(\mathbf{x}_j),$$

$$\text{PCS}_M = \min_{\mathbf{x} \in \mathcal{X}} \text{PCS}(\mathbf{x}),$$

$$\text{PCS}_A = \mathbb{P}\left(\bigcap_{j=1, i \neq i^*(\mathbf{x}_j)}^m (\hat{y}_{i^*(\mathbf{x}_j)}(\mathbf{x}_j) < \hat{y}_i(\mathbf{x}_j))\right).$$

In  $\text{PCS}_E$ ,  $p_j$  is the probability of  $\mathbf{X} = \mathbf{x}_j$ ,  $j = 1, 2, \dots, m$ .  $\text{PCS}_E$  describes the expected probability of correct selection over  $\mathcal{X}$ , where the expectation is taken with respect to the randomness of  $\mathbf{X}$ .  $\text{PCS}_M$  shows the worst-case performance of  $\text{PCS}(\mathbf{x})$  over  $\mathcal{X}$ . This measure is, in some sense, similar to the worst-case performance in robust optimization (Bertsimas et al. 2011) and R&S with input uncertainty (Fan et al. 2020).

$\text{PCS}_A$  is defined in a different way from the two measures above. It is not based on  $\text{PCS}(\mathbf{x})$ ; instead, it requires correctness for all of the comparisons of interest, that is, comparisons between the estimated best treatment and the alternatives under all of the possible contexts.  $\text{PCS}_A$  sets the highest standard for the quality of the selection among the three and is appropriate to be used by conservative decision makers. It is obvious that  $\text{PCS}_A \leq \text{PCS}_M \leq \text{PCS}_E$ . Intuitively,  $\text{PCS}_E$  and  $\text{PCS}_M$  are the average and the minimum probabilities of the best treatment being identified among all the patient contexts;  $\text{PCS}_A$  is the probability of the best treatment being identified for all the contexts. Note that  $\text{PCS}_A$  is newly proposed for CR&S, whereas  $\text{PCS}_E$  and  $\text{PCS}_M$  have been used and discussed in Shen et al. (2021) as measures for R&S with covariates.

Because of the lack of analytical expressions of  $\text{PCS}_E$ ,  $\text{PCS}_M$ , and  $\text{PCS}_A$ , it is challenging to find the exact optimizers of them. As a result, it is common to instead pursue their asymptotic optimizers (optimizers as  $n \rightarrow \infty$ ) in the R&S literature (Chen et al. 2000, Frazier et al. 2008, Ryzhov 2016). Asymptotic optimizers become close to the real optimizers when the simulation budget  $n$  is large and often demonstrate very good empirical performance when  $n$  is small (Branke et al. 2007). To find asymptotic optimizers of  $\text{PCS}_E$ ,  $\text{PCS}_M$ , and  $\text{PCS}_A$ , we can look for solutions that maximize the asymptotic performance of the three measures, that is, solutions that maximize the rates at which they converge to 1. The following theorem characterizes these rates of the three measures.

**Theorem 1.** Define probabilities of false selection  $\text{PFS}_E = 1 - \text{PCS}_E$ ,  $\text{PFS}_M = 1 - \text{PCS}_M$ , and  $\text{PFS}_A = 1 - \text{PCS}_A$ . Under Assumptions 1–3, the three measures  $\text{PFS}_E$ ,  $\text{PFS}_M$ , and  $\text{PFS}_A$  converge exponentially and have the same rate function  $\mathcal{R}(\boldsymbol{\alpha})$ .

That is,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \text{PFS}_E &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \text{PFS}_M \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \text{PFS}_A = -\mathcal{R}(\alpha). \end{aligned}$$

Moreover, it can be shown that  $\mathcal{R}(\alpha) = \min_{j \in \{1, \dots, m\}} \min_{i \in \{1, \dots, k\}, i \neq i^*(x_j)} - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\hat{y}_{i^*(x_j)}(x_j) \geq \hat{y}_i(x_j))$ .

To interpret Theorem 1, we pick  $i_o$  and  $j_o$  such that

$$(i_o, j_o) \in \arg \min_{i \in \{1, \dots, k\}, i \neq i^*(x_j), j \in \{1, \dots, m\}} - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\hat{y}_{i^*(x_j)}(x_j) \geq \hat{y}_i(x_j)).$$

The theorem shows that the three measures, although defined from different perspectives, converge at the same exponential rate  $\mathcal{R}(\alpha)$ , where  $\alpha$  is the sampling rate of each treatment-context pair. The rate function  $\mathcal{R}(\alpha)$  is characterized by the most difficult comparison among comparisons between the best treatment and non-best treatments under each context, that is, the comparison of sample means between treatments  $i^*(x_{j_o})$  and  $i_o$  under context  $j_o$ . The reason for this effect is that, the most difficult comparison has the slowest convergence rate, which dominates the convergence rates of the other comparisons and thus represents the rate at which these measures converge. Theorem 1 lays the foundation of this paper; instead of considering the three measures separately, we can solve them once and for all by directly optimizing the rate function  $\mathcal{R}(\alpha)$ .

## 4. Small-Scale Problem

In this section, we consider the small-scale problem, where our simulation budget is sufficient for simulating all the treatment-context pairs. The estimate  $\hat{y}_i(x_j)$  for the performance  $y_i(x_j)$  of treatment  $i$  and context  $x_j$  is the sample mean  $\bar{Y}_i(x_j)$ ,  $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots, m$ .

### 4.1. Rate-Optimal Budget Allocation Rule

For the small-scale problem, optimization of the rate function  $\mathcal{R}(\alpha)$  is given by

$$\begin{aligned} \min -\mathcal{R}(\alpha) \quad \text{s.t.} \quad & \sum_{i=1}^k \sum_{j=1}^m \alpha_{i,j} = 1, \alpha_{i,j} \geq 0, \\ & i = 1, 2, \dots, k, j = 1, 2, \dots, m. \end{aligned} \quad (1)$$

The simulation budget constraint  $\sum_{i=1}^k \sum_{j=1}^m \alpha_{i,j} = 1$  is equivalent to  $\sum_{i=1}^k \sum_{j=1}^m n_{i,j} = n$ . This is an OCBA-like formulation (Chen et al. 2000), which finds a simulation budget allocation strategy to optimize the measure of interest, that is, the rate function  $\mathcal{R}(\alpha)$  in our problem.

Before we solve (1), we carry out more analysis on the rate function  $\mathcal{R}(\alpha)$ . According to Theorem 1,  $\mathcal{R}(\alpha) = \min_{j \in \{1, 2, \dots, m\}} \min_{i \in \{1, \dots, k\}, i \neq i^*(x_j)} - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\hat{y}_{i^*(x_j)}(x_j) \geq \hat{y}_i(x_j))$ . We denote

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{Y}_{i^*(x_j)}(x_j) \geq \bar{Y}_i(x_j)) \doteq \mathcal{G}_{i^*(x_j), i, j}(\alpha_{i^*(x_j), j}, \alpha_{i, j}).$$

From the Gärtner-Ellis Theorem (Dembo and Zeitouni 1998), for i.i.d. normal samples  $Y_{il}(x)$ ,

$$\mathcal{G}_{i^*(x_j), i, j}(\alpha_{i^*(x_j), j}, \alpha_{i, j}) = \frac{(y_{i^*(x_j)}(x_j) - y_i(x_j))^2}{2(\sigma_{i^*(x_j)}^2(x_j)/\alpha_{i^*(x_j), j} + \sigma_i^2(x_j)/\alpha_{i, j})}.$$

Then, an equivalent formulation of problem (1) is given by

$$\begin{aligned} \max \quad & z \\ \text{s.t.} \quad & \mathcal{G}_{i^*(x_j), i, j}(\alpha_{i^*(x_j), j}, \alpha_{i, j}) \geq z, \quad i = 1, 2, \dots, k \text{ and} \\ & i \neq i^*(x_j), j = 1, 2, \dots, m, \\ & \sum_{i=1}^k \sum_{j=1}^m \alpha_{i,j} = 1, \alpha_{i,j} \geq 0, \\ & i = 1, 2, \dots, k, j = 1, 2, \dots, m. \end{aligned} \quad (2)$$

Note that  $\mathcal{G}_{i^*(x_j), i, j}(\alpha_{i^*(x_j), j}, \alpha_{i, j})$  is a concave function, so  $\mathcal{G}_{i^*(x_j), i, j}(\alpha_{i^*(x_j), j}, \alpha_{i, j}) \geq z$  forms a convex set, and problem (2) is a convex optimization model. We can investigate the KKT conditions (Boyd and Vandenberghe 2004) of this model to solve it.

**Theorem 2.** The optimal solution to problem (2) is given by

$$\begin{aligned} \frac{\alpha_{i^*(x_j), j}^2}{\sigma_{i^*(x_j)}^2(x_j)} &= \sum_{i=1, i \neq i^*(x_j)}^k \frac{\alpha_{i, j}^2}{\sigma_i^2(x_j)}, \quad j = 1, 2, \dots, m, \\ \frac{(y_i(x_j) - y_{i^*(x_j)}(x_j))^2}{\sigma_{i^*(x_j)}^2(x_j)/\alpha_{i^*(x_j), j} + \sigma_i^2(x_j)/\alpha_{i, j}} &= \frac{(y_{i'}(x_{j'}) - y_{i^*(x_{j'})}(x_{j'}))^2}{\sigma_{i^*(x_{j'})}^2(x_{j'})/\alpha_{i^*(x_{j'})}, j' + \sigma_{i'}^2(x_{j'})/\alpha_{i', j'}}, \\ & \quad j, j' = 1, \dots, m, i, i' = 1, \dots, k, i \neq i^*(x_j) \\ & \quad \text{and } i' \neq i^*(x_{j'}). \end{aligned} \quad (3)$$

Theorem 2 indicates that the solution satisfying conditions (3) and (4) corresponds to the budget allocation rule that maximizes the convergence rate of  $\text{PFS}_E$ ,  $\text{PFS}_M$ , and  $\text{PFS}_A$ . Condition (3) establishes for each context  $x_j$  a certain balance between the proportions of simulation replications allocated to the best treatment  $\alpha_{i^*(x_j), j}$  and

those allocated to non-best treatments  $\alpha_{i,j}$  for  $i \neq i^*(\mathbf{x}_j)$  in the sense that  $\frac{\alpha_{i^*(\mathbf{x}_j),j}^2}{\sigma_{i^*(\mathbf{x}_j)}^2}$  (representing the simulation replications allocated to the best treatment) should be equal to  $\sum_{i=1, i \neq i^*(\mathbf{x}_j)}^k \frac{\alpha_{i,j}^2}{\sigma_i^2}$  (representing the simulation replications allocated to the non-best treatments). Condition (4) further adjusts the ratios of the simulation replications allocated to any two non-best treatments under the same context and across different contexts. This condition suggests that the difficulty of correctly identifying a non-best treatment  $i$  under context  $\mathbf{x}_j$  as non-best can be reflected by the index  $\frac{(y_i(\mathbf{x}_j) - y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j))^2}{\sigma_{i^*(\mathbf{x}_j)}^2(\mathbf{x}_j)/\alpha_{i^*(\mathbf{x}_j),j} + \sigma_i^2(\mathbf{x}_j)/\alpha_{i,j}}$ , which represents a comparison between the non-best treatment  $i$  and the optimal treatment  $i^*(\mathbf{x}_j)$  under context  $\mathbf{x}_j$ . To optimize the rate function, we should allocate the simulation budget to the treatment-context pairs such that this index remains equal for all the treatment-context pairs.

#### 4.2. Selection Algorithm

In this section, we develop a selection algorithm based on optimality conditions (3) and (4) for implementation and analyze its asymptotic performance.

For simplicity of presentation, define

$$\mathcal{U}_j^b = \frac{\alpha_{i^*(\mathbf{x}_j),j}^2}{\sigma_{i^*(\mathbf{x}_j)}^2}, \quad \mathcal{U}_j^{\text{non}} = \sum_{i=1, i \neq i^*(\mathbf{x}_j)}^k \frac{\alpha_{i,j}^2}{\sigma_i^2}, \quad j = 1, 2, \dots, m,$$

$$\mathcal{V}_{i,j} = \frac{(y_i(\mathbf{x}_j) - y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j))^2}{\sigma_{i^*(\mathbf{x}_j)}^2(\mathbf{x}_j)/\alpha_{i^*(\mathbf{x}_j),j} + \sigma_i^2(\mathbf{x}_j)/\alpha_{i,j}}, \quad j = 1, 2, \dots, m,$$

$$i = 1, 2, \dots, k \text{ and } i \neq i^*(\mathbf{x}_j).$$

Note that  $\mathcal{U}_j^b$  represents the simulation replications allocated to the best treatment  $i^*(\mathbf{x}_j)$  under context  $\mathbf{x}_j$ ,  $\mathcal{U}_j^{\text{non}}$  represents the simulation replications allocated to the non-best treatments  $i$  under the same context, and  $\mathcal{V}_{i,j}$  represents the difficulty of correctly identifying the non-best treatment  $i$  under context  $\mathbf{x}_j$  as non-best. Then, conditions (3) and (4) can be rewritten as

$$\mathcal{U}_j^b = \mathcal{U}_j^{\text{non}}, \quad j = 1, 2, \dots, m, \quad (5)$$

$$\mathcal{V}_{i,j} = \mathcal{V}_{i',j'}, \quad j, j' = 1, \dots, m, \quad i, i' = 1, \dots, k, \quad i \neq i^*(\mathbf{x}_j) \text{ and } i' \neq i^*(\mathbf{x}_{j'}), \quad (6)$$

Because Equations (5) and (6) do not have an analytical solution, we will design the algorithm in a simple and cost-effective manner that gradually reduces the error terms  $|\mathcal{U}_j^b - \mathcal{U}_j^{\text{non}}|$  and  $|\mathcal{V}_{i,j} - \mathcal{V}_{i',j'}|$  in (5) and (6).

Let  $(i_*, j_*) \in \arg \min_{j \in \{1, 2, \dots, m\}, i \in \{1, \dots, k\} \setminus \{i^*(\mathbf{x}_j)\}} \mathcal{V}_{i,j}$ . Note that

$$\frac{d\mathcal{U}_j^b}{d\alpha_{i^*(\mathbf{x}_j),j}} = \frac{2\alpha_{i^*(\mathbf{x}_j),j}}{\sigma_{i^*(\mathbf{x}_j)}^2} > 0, \quad j = 1, 2, \dots, m;$$

$$\frac{\partial \mathcal{U}_j^{\text{non}}}{\partial \alpha_{i,j}} = \frac{2\alpha_{i,j}}{\sigma_i^2} > 0, \quad j = 1, 2, \dots, m, \quad i = 1, 2, \dots, k \text{ and } i \neq i^*(\mathbf{x}_j);$$

$$\frac{\partial \mathcal{V}_{i,j}}{\partial \alpha_{i^*(\mathbf{x}_j),j}} = \frac{(y_i(\mathbf{x}_j) - y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j))^2}{(\sigma_{i^*(\mathbf{x}_j)}^2(\mathbf{x}_j)/\alpha_{i^*(\mathbf{x}_j),j} + \sigma_i^2(\mathbf{x}_j)/\alpha_{i,j})^2} \frac{\sigma_{i^*(\mathbf{x}_j)}^2(\mathbf{x}_j)}{\alpha_{i^*(\mathbf{x}_j),j}^2} > 0,$$

$$\frac{\partial \mathcal{V}_{i,j}}{\partial \alpha_{i,j}} = \frac{(y_i(\mathbf{x}_j) - y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j))^2}{(\sigma_{i^*(\mathbf{x}_j)}^2(\mathbf{x}_j)/\alpha_{i^*(\mathbf{x}_j),j} + \sigma_i^2(\mathbf{x}_j)/\alpha_{i,j})^2} \frac{\sigma_i^2(\mathbf{x}_j)}{\alpha_{i,j}^2} > 0,$$

$$j = 1, 2, \dots, m, \quad i = 1, 2, \dots, k \text{ and } i \neq i^*(\mathbf{x}_j).$$

That is, we can choose to increase the values of  $\mathcal{U}_j^b$  and  $\mathcal{U}_j^{\text{non}}$  by allocating more replications to treatment  $i^*(\mathbf{x}_j)$  and treatment  $i$  for any  $i \neq i^*(\mathbf{x}_j)$  under context  $\mathbf{x}_j$ . We can also choose to increase the value of  $\mathcal{V}_{i,j}$  by allocating additional replications to either treatment  $i^*(\mathbf{x}_j)$  or treatment  $i$  under context  $\mathbf{x}_j$ ,  $i \in \{1, 2, \dots, k\}$  and  $i \neq i^*(\mathbf{x}_j)$ .

To design a selection algorithm based on (5) and (6), suppose for a budget allocation, (6) cannot be fulfilled. To fix it, we will provide a small incremental budget to improve  $\mathcal{V}_{i_*,j_*}$  so that the gap between  $\min_{j \in \{1, 2, \dots, m\}} \min_{i \in \{1, \dots, k\} \setminus \{i^*(\mathbf{x}_j)\}} \mathcal{V}_{i,j}$  and  $\max_{j \in \{1, 2, \dots, m\}} \max_{i \in \{1, \dots, k\} \setminus \{i^*(\mathbf{x}_j)\}} \mathcal{V}_{i,j}$  can be reduced. As discussed above, allocating more replications to treatment  $i^*(\mathbf{x}_{j_*})$  or  $i_*$  under context  $\mathbf{x}_{j_*}$  achieves this goal. To further decide which of the treatments  $i^*(\mathbf{x}_{j_*})$  or  $i_*$  receives the incremental budget, we check condition (5). If  $\mathcal{U}_{j_*}^b < \mathcal{U}_{j_*}^{\text{non}}$ , the additional replications should be allocated to the best treatment  $i^*(\mathbf{x}_{j_*})$  in order to balance the equation; otherwise, the additional replications should be allocated to the non-best treatment  $i_*$ . This idea is summarized in CR&S Algorithm 1 below.

#### Algorithm 1 (CR&S Algorithm 1)

**Input:** Specify the number of contexts  $m$ , the number of treatments  $k$ , the total simulation budget  $n$ , and the initial number of simulation replications  $n_0$ . Iteration counter  $r \leftarrow 0$ .

**Initial Sampling:** Perform  $n_0$  replications for treatment  $i$  under context  $\mathbf{x}_j$ ,  $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots, m$ , and calculate sample means  $\bar{Y}_i(\mathbf{x}_j)$  and sample variances  $\hat{\sigma}_i^2(\mathbf{x}_j)$ . Set  $\hat{n}_{i,j} = n_0$ ,  $n^{(r)} = \sum_{j=1}^m \sum_{i=1}^k \hat{n}_{i,j}$  and  $\hat{\alpha}_{i,j} = \hat{n}_{i,j}/n^{(r)}$ .

repeat

**Step 1:** Obtain  $\hat{U}_j^b, \hat{U}_j^{non}, \hat{V}_{i,j}$  and  $\hat{i}^*(x_j)$  for  $i = 1, 2, \dots, k, i \neq \hat{i}^*(x_j)$  and  $j = 1, 2, \dots, m$ . Let  $(\hat{i}_*, j^*) \in \arg \min_{j \in \{1, 2, \dots, m\}, i \in \{1, \dots, k\} \setminus \{\hat{i}^*(x_j)\}} \hat{V}_{i,j}$ .

**Step 2:** If  $\hat{U}_{j^*}^b < \hat{U}_{j^*}^{non}$ ,  $i^r = \hat{i}^*(x_{j^*})$ ; otherwise,  $i^r = \hat{i}_*$ . Provide one more replication to treatment  $i^r$  under context  $x_{j^*}$ . Update  $\bar{Y}_{i^r}(x_{j^*})$  and  $\hat{\sigma}_{i^r}^2(x_{j^*})$ .

**Step 3:** Update  $\hat{n}_{i,j}, n^{(r+1)}$  and  $\hat{\alpha}_{i,j}$ .  $r \leftarrow r + 1$ .

until  $n^{(r)} = n$ .

At the beginning of the algorithm, we simulate each treatment-context pair for the same number of replications and acquire initial estimates for their means and variances. In each of the subsequent iterations, we sample more on a certain treatment-context pair determined by  $\hat{U}_j^b, \hat{U}_j^{non}$  and  $\hat{V}_{i,j}$  and update its sample mean and sample variance. Although we have set the incremental budget  $\Delta n = 1$  in this generic algorithm, in practice,  $\Delta n$  can be larger than 1 to reduce the number of iterations. The algorithm terminates when the total simulation budget is exhausted.

This idea for designing CR&S Algorithm 1 does not involve solving the set of nonlinear Equations (3) and (4) and is thus cost-effective; more importantly, this algorithm can recover the optimality conditions (3) and (4). It can be established in the following theorem.

**Theorem 3.** Suppose Assumptions 1–3 hold. For  $\hat{\alpha}_{i,j}$  generated by CR&S Algorithm 1,  $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots, m$ , we have almost surely that

$$\lim_{r \rightarrow \infty} \left| \frac{\hat{\alpha}_{i^r(x_j),j}^2}{\hat{\sigma}_{i^r(x_j)}^2(x_j)} - \sum_{i=1, i \neq i^*(x_j)}^k \frac{\hat{\alpha}_{i,j}^2}{\hat{\sigma}_i^2(x_j)} \right| = 0, j = 1, 2, \dots, m,$$

$$\lim_{r \rightarrow \infty} \left| \frac{(y_i(x_j) - y_{i^r(x_j)}(x_j))^2}{\hat{\sigma}_{i^r(x_j)}^2(x_j) / \hat{\alpha}_{i^r(x_j),j} + \hat{\sigma}_i^2(x_j) / \hat{\alpha}_{i,j}} - \frac{(y_{i'}(x_j) - y_{i^r(x_j)}(x_j))^2}{\hat{\sigma}_{i^r(x_j)}^2(x_j) / \hat{\alpha}_{i^r(x_j),j} + \hat{\sigma}_{i'}^2(x_j) / \hat{\alpha}_{i',j}} \right| = 0, j, j' = 1, \dots, m,$$

$$i, i' = 1, \dots, k, i \neq i^*(x_j) \text{ and } i' \neq i^*(x_{j^*}).$$

In other words, when  $\alpha_{i,j}$ 's are replaced by the sample allocation  $\hat{\alpha}_{i,j}$ 's generated by CR&S Algorithm 1, conditions (3) and (4) still hold almost surely as the iteration  $r \rightarrow \infty$ . A byproduct of this theorem is that the number of simulation replications  $\hat{n}_{i,j} = \hat{\alpha}_{i,j}n$  allocated to treatment  $i$  under context  $j$  by the algorithm will go to infinity as the total budget  $n$  goes to infinity. It ensures that all the estimators in this algorithm, such as  $\bar{Y}_i(x_j), \hat{\sigma}_i^2(x_j), \hat{i}^*(x_j)$ , etc., will converge to their true values. Particularly, the estimated best treatment  $\hat{i}^*(x_j)$  will converge to the true best  $i^*(x_j)$  in the long term for all patient contexts  $x_j, j = 1, \dots, m$ .

## 5. Large-Scale Problem

In this section, we consider the large-scale problem. Suppose the contexts lie on a grid, and the relationship between treatment performance  $y_i(x)$  and context  $x$  can be described by the linear model

$$y_i(x) = f(x)^\top \beta_i, \quad i = 1, 2, \dots, k,$$

where  $\beta_i = (\beta_{i1}, \dots, \beta_{iq})^\top$  is a vector of unknown parameters that need to be estimated and  $f(x) = (f_1(x), \dots, f_q(x))^\top$  is a vector of known basis functions. A common choice of  $f_i(x)$ 's is  $f_i(x) = x, i = 1, 2, \dots, q$ . They can also be set as other functional forms to achieve a potential better fit. Although the linear models are simple and require the knowledge of adequate basis functions, they are robust to model misspecification and often have good performance in prediction (Thompson 1982, James et al. 2013).

The large-scale problem looks similar to but is in essence different from a class of BAI problems known as linear bandits (Soare et al. 2014, Jedra and Proutiere 2020). In linear bandits, it is assumed that treatment  $i$  can be represented by a characteristic vector  $v_i$ , and the mean performance of treatment  $i$  is given by  $v_i^\top \beta_L$ , where  $\beta_L$  is a vector of unknown parameters. In other words, linear bandits are concerned with only one linear model, with independent variables of the model representing information of the treatments. No contexts are involved in linear bandits. Our large-scale problem is concerned with  $k$  linear models, with independent variables of the models representing information of the contexts.

For the large-scale problem, we need only to simulate the treatments under a small fraction of contexts. Suppose the contexts we simulate are  $x_h^\circ, h = 1, \dots, p$  and  $p \ll m$ . As before,  $n_{i,h}$  denotes the number of simulation replications that we allocate to the treatment-context pair  $(i, x_h^\circ)$ . Let  $\alpha_{i,h} = n_{i,h}/n$ ,  $\alpha$  be the vector of  $\alpha_{i,h}$ 's and  $F = (f(x_1^\circ), \dots, f(x_p^\circ))^\top$  be the  $p \times q$  design matrix. For treatment  $i$ , let  $\bar{Y}_i = (\bar{Y}_i(x_1^\circ), \dots, \bar{Y}_i(x_p^\circ))^\top$  be the sample means of the treatments under the  $p$  simulated contexts, and let  $\bar{\epsilon}_i = (\bar{\epsilon}_i(x_1^\circ), \dots, \bar{\epsilon}_i(x_p^\circ))^\top$  be the averaged observation errors, where  $\bar{\epsilon}_i(x_h^\circ) = \frac{1}{n_{i,h}} \sum_{l=1}^{n_{i,h}} \epsilon_{il}(x_h^\circ)$ .

We use the method of least squares to estimate  $\beta_i$ , that is,  $\hat{\beta}_i = (F^\top F)^{-1} F^\top \bar{Y}_i$ . Then, the estimate  $\hat{y}_i(x)$  for the mean performance  $y_i(x)$  of treatment  $i$  under context  $x$  is  $\bar{Y}_i^L(x) = f(x)^\top \hat{\beta}_i$ .

### 5.1. Rate-Optimal Budget Allocation Rule

For the large-scale problem, optimization of the rate function  $\mathcal{R}(\alpha)$  is given by

$$\min -\mathcal{R}(\alpha) \quad \text{s.t.} \quad \sum_{i=1}^k \sum_{h=1}^p \alpha_{i,h} = 1, \quad \alpha_{i,h} \geq 0, \\ i = 1, 2, \dots, k, \quad h = 1, 2, \dots, p. \quad (7)$$

Model (7) has the same structure as (1). They both optimize the rate function  $\mathcal{R}(\alpha)$  subject to the simulation budget constraint. The difference is that in (7), mean performance  $y_i(\mathbf{x})$  is predicted by  $\bar{Y}_i^L(\mathbf{x})$  from the linear models.

By Theorem 1,  $\mathcal{R}(\alpha) = \min_{j \in \{1, \dots, m\}} \min_{i \in \{1, \dots, k\}, i \neq i^*(\mathbf{x}_j)} - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{Y}_{i^*(\mathbf{x}_j)}^L(\mathbf{x}_j) \geq \bar{Y}_i^L(\mathbf{x}_j))$ . Here, we denote

$$- \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\bar{Y}_{i^*(\mathbf{x}_j)}^L(\mathbf{x}_j) \geq \bar{Y}_i^L(\mathbf{x}_j)) \doteq \mathcal{G}_{i^*(\mathbf{x}_j), i, j}^L(\alpha).$$

Obviously,  $\mathcal{G}_{i^*(\mathbf{x}_j), i, j}^L(\alpha)$  is different from the rate function  $\mathcal{G}_{i^*(\mathbf{x}_j), i, j}(\alpha_{i^*(\mathbf{x}_j), j}, \alpha_{i, j})$  in Section 4. We next derive  $\mathcal{G}_{i^*(\mathbf{x}_j), i, j}^L(\alpha)$  in the following lemma.

**Lemma 1.** Suppose Assumptions 1–3 hold. With the linear models, the rate function of the three measures PFS<sub>E</sub>, PFS<sub>M</sub>, and PFS<sub>A</sub> is  $\mathcal{R}(\alpha) = \min_{j \in \{1, \dots, m\}} \min_{i \in \{1, \dots, k\}, i \neq i^*(\mathbf{x}_j)} \mathcal{G}_{i^*(\mathbf{x}_j), i, j}^L(\alpha)$ , where

$$\begin{aligned} \mathcal{G}_{i^*(\mathbf{x}_j), i, j}^L(\alpha) &= \frac{[\mathbf{f}(\mathbf{x}_j)^\top (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i^*(\mathbf{x}_j)})]^2}{2\mathbf{f}(\mathbf{x}_j)^\top (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top (\boldsymbol{\Sigma}_{\epsilon, i}^{(p)} + \boldsymbol{\Sigma}_{\epsilon, i^*(\mathbf{x}_j)}^{(p)}) \mathbf{F} (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{f}(\mathbf{x}_j)} \end{aligned}$$

and  $\boldsymbol{\Sigma}_{\epsilon, i_1}^{(p)}$  is the diagonal matrix with  $(\boldsymbol{\Sigma}_{\epsilon, i_1}^{(p)})_{hh} = \frac{\sigma_{i_1}^2(\mathbf{x}_h^\circ)}{\alpha_{i_1, h}}$ ,  $i_1 = i, i^*(\mathbf{x}_j)$ .

A model equivalent to (7) is given by

$$\begin{aligned} \max \quad & z \\ \text{s.t.} \quad & \mathcal{G}_{i^*(\mathbf{x}_j), i, j}^L(\alpha) \geq z, \quad i = 1, 2, \dots, k \text{ and} \\ & i \neq i^*(\mathbf{x}_j), \quad j = 1, 2, \dots, m, \\ & \sum_{i=1}^k \sum_{h=1}^p \alpha_{i, h} = 1, \quad \alpha_{i, h} \geq 0, \quad i = 1, 2, \dots, k, \\ & h = 1, 2, \dots, p. \end{aligned} \quad (8)$$

Although (8) is a convex optimization model, its KKT conditions cannot be easily analyzed as for its counterpart (2) in the small-scale problem. To solve (8), we will consider the dual problem of it. For simplicity of notation, let  $\bar{\mathbf{Y}} = (\bar{\mathbf{Y}}_1^\top, \dots, \bar{\mathbf{Y}}_k^\top)^\top$ ,  $\hat{\boldsymbol{\sigma}}_i^2 = (\sigma_{i1}^2(\mathbf{x}_1^\circ), \dots, \sigma_{ip}^2(\mathbf{x}_p^\circ))^\top$ ,  $\hat{\boldsymbol{\sigma}}^2 = (\hat{\boldsymbol{\sigma}}_1^{2\top}, \dots, \hat{\boldsymbol{\sigma}}_k^{2\top})^\top$ , and  $\boldsymbol{\lambda}$  be the vector of  $\lambda_{i, j}$ 's for  $i = 1, \dots, k$  and  $i \neq i^*(\mathbf{x}_j)$  and  $j = 1, \dots, m$ . Let the mean of  $(\bar{\mathbf{Y}}, \hat{\boldsymbol{\sigma}}^2)$  be  $(\mathbf{y}, \boldsymbol{\sigma}^2)$ .

**Theorem 4.** The optimal solution to (8) is

$$\begin{aligned} \alpha_{i, h} &= \frac{\sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)}}{\sum_{i=1}^k \sum_{h=1}^p \sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)}}, \\ i &= 1, \dots, k, \quad h = 1, \dots, p. \end{aligned} \quad (9)$$

In (9),  $\boldsymbol{\lambda}$  is the optimal solution to

$$\begin{aligned} \min_{\boldsymbol{\lambda}} \quad & a(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2) = - \sum_{i=1}^k \sum_{h=1}^p \sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)} \\ \text{s.t.} \quad & \sum_{j=1}^m \sum_{i=1, i \neq i^*(\mathbf{x}_j)}^k \lambda_{i, j} = 1, \quad \lambda_{i, j} \geq 0, \quad i = 1, \dots, k \text{ and} \\ & i \neq i^*(\mathbf{x}_j), \quad j = 1, \dots, m, \end{aligned} \quad (10)$$

where  $\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)$  is defined as

$$\begin{aligned} \chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2) &= \sigma_i^2(\mathbf{x}_h^\circ) \left( \sum_{j \in \mathcal{C}_i} \sum_{i'=1, i' \neq i}^k \frac{2\lambda_{i', j}(\mathbf{f}(\mathbf{x}_j)^\top (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{f}(\mathbf{x}_h^\circ))^2}{[\mathbf{f}(\mathbf{x}_j)^\top (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i'})]^2} \right. \\ &\quad \left. + \sum_{j \notin \mathcal{C}_i} \frac{2\lambda_{i, j}(\mathbf{f}(\mathbf{x}_j)^\top (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{f}(\mathbf{x}_h^\circ))^2}{[\mathbf{f}(\mathbf{x}_j)^\top (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i^*(\mathbf{x}_j)})]^2} \right) \end{aligned} \quad (11)$$

for  $i = 1, \dots, k$  and  $h = 1, \dots, p$ , and  $\mathcal{C}_i = \{j : i^*(\mathbf{x}_j) = i\}$ .

Intuitively, allocating more replications to treatment  $i$  under context  $\mathbf{x}_h^\circ$  can increase the accuracy of estimate  $\hat{\boldsymbol{\beta}}_{i'}$ , and the more accurate  $\hat{\boldsymbol{\beta}}_i$  and  $\hat{\boldsymbol{\beta}}_{i'}$  are, the more likely  $\bar{Y}_i^L(\mathbf{x}_j) = \mathbf{f}(\mathbf{x}_j)^\top \hat{\boldsymbol{\beta}}_i < \mathbf{f}(\mathbf{x}_j)^\top \hat{\boldsymbol{\beta}}_{i'} = \bar{Y}_{i'}^L(\mathbf{x}_j)$  given  $y_i(\mathbf{x}_j) < y_{i'}(\mathbf{x}_j)$  for all  $j = 1, \dots, m$ . Each term in the summation of (11) can be seen as the contribution of allocating replications to treatment  $i$  under context  $\mathbf{x}_h^\circ$  to the correct comparison between the best and non-best treatments under different contexts. Therefore,  $\sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)}$  can be seen as the total contribution of allocating replications to treatment  $i$  under context  $\mathbf{x}_h^\circ$  to maximizing the rate function  $\mathcal{R}(\alpha)$ , and Theorem 4 indicates that the number of replications allocated to treatment  $i$  under context  $\mathbf{x}_h^\circ$  should be proportional to  $\sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)}$ .

## 5.2. Selection Algorithm

In this section, we devise a selection algorithm for the large-scale problem based on Theorem 4.

The parameters  $\mathbf{y}$  and  $\boldsymbol{\sigma}^2$  in Theorem 4 are unknown in practice and can be estimated by  $\bar{\mathbf{Y}}$  and  $\hat{\boldsymbol{\sigma}}^2$ . Given  $\bar{\mathbf{Y}}$  and  $\hat{\boldsymbol{\sigma}}^2$ , (10) is a convex optimization problem, and we develop a gradient descent algorithm to find its optimal solution  $\hat{\boldsymbol{\lambda}}$ . In each iteration, we compute a descent direction  $\tilde{\mathbf{d}}$  and a descent step-size  $\tilde{s}$  and update  $\hat{\boldsymbol{\lambda}}$  by letting it move along the direction  $\tilde{\mathbf{d}}$  with the step-size  $\tilde{s}$ . Different from most gradient descent algorithms that conduct this movement for multiple times, our algorithm conducts the movement only once and then plugs the updated  $\hat{\boldsymbol{\lambda}}$ ,  $\bar{\mathbf{Y}}$ ,  $\hat{\boldsymbol{\sigma}}^2$  into  $\alpha_{i, h} = \frac{\sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)}}{\sum_{i=1}^k \sum_{h=1}^p \sqrt{\chi_{i, h}(\boldsymbol{\lambda}, \mathbf{y}, \boldsymbol{\sigma}^2)}}$  to

compute the estimated optimal allocation  $\hat{\alpha}_{i,h}$ . This algorithm design considerably reduces the computation associated with gradient descent while still ensuring that  $\bar{Y}$ ,  $\hat{\sigma}^2$  and  $\hat{\alpha}_{i,h}$  converge to the correct values. Next, we provide a small incremental budget and allocate it to the treatment-context pairs based on  $\hat{\alpha}_{i,h}$  and update  $\bar{Y}$  and  $\hat{\sigma}^2$  of the treatment-context pairs that receive additional replications. Then, the algorithm proceeds to the next iteration. This process is repeated until the simulation budget is consumed.

#### Algorithm 2 (CR&S Algorithm 2)

**Input:** Specify the number of contexts  $m$ , the number of treatments  $k$ , the total simulation budget  $n$ , and the initial number of simulation replications  $n_0$ . Calculate  $\mathbf{f}(\mathbf{x}_j)$  for each context  $\mathbf{x}_j$  and  $(\mathbf{F}^\top \mathbf{F})^{-1}$ . Choose a small constant  $\kappa_0$  and  $\eta < \frac{1}{(k-1)m}$ . Iteration counter  $r \leftarrow 0$ .

**Initial Sampling:** Perform  $n_0$  replications on each pair of treatment  $i$  and context  $\mathbf{x}_h^\circ$ , calculate sample means and sample variances  $\bar{Y}_i(\mathbf{x}_h^\circ)$  and  $\hat{\sigma}_i^2(\mathbf{x}_h^\circ)$ , and estimate  $\hat{\beta}_i$  by  $\hat{\beta}_i = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \bar{Y}_i$ . Let  $\hat{n}_{i,h} = n_0$ ,  $n^{(r)} = \sum_{i=1}^k \sum_{h=1}^p \hat{n}_{i,h}$  and  $\hat{\alpha}_{i,h} = \hat{n}_{i,h} / n^{(r)}$ . Find the best treatment  $\hat{i}^*(\mathbf{x}_j) = \arg \min_i \mathbf{f}(\mathbf{x}_j)^\top \hat{\beta}_i$  for each  $\mathbf{x}_j$ . Set  $\hat{\lambda}_{i,j} = \frac{1}{(k-1)m}$ ,  $i = 1, \dots, k$ ,  $i \neq \hat{i}^*(\mathbf{x}_j)$ ,  $j = 1, \dots, m$ .

#### Repeat

**Step 1:**  $r \leftarrow r + 1$ . Obtain  $\chi_{i,h}(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)$  by plugging  $\hat{\lambda}$ ,  $\bar{Y}$ , and  $\hat{\sigma}^2$  into  $\chi_{i,h}(\lambda, \bar{y}, \sigma^2)$ .

**Step 2:** Randomly choose a  $(i^*, j^*)$  from  $\{(i, j) : \hat{\lambda}_{i,j} \text{ exists and } \hat{\lambda}_{i,j} \geq \eta\}$ .

**Step 3:** Compute the descent direction  $\tilde{\mathbf{d}}^{(r)} = \arg \min_{\tilde{\mathbf{d}} \in \mathcal{D}^{(i^*, j^*)}(\hat{\lambda})} \tilde{s}^{\max}(\tilde{\mathbf{d}}, \hat{\lambda}) \nabla a(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)^\top \tilde{\mathbf{d}}$ , where  $\mathcal{D}^{(i^*, j^*)}(\hat{\lambda}) = \{e_{i,j} - e_{i',j'} : i \neq i' \text{ or } j \neq j'\} \cup \{e_{i',j'} - e_{i,j} : i \neq i' \text{ or } j \neq j', \hat{\lambda}_{i,j} > 0\}$ ,  $e_{i,j}$  is obtained by letting  $\hat{\lambda}_{i,j}$  equal to one and all of the other elements of  $\hat{\lambda}$  equal to zero, and  $\tilde{s}^{\max}(\tilde{\mathbf{d}}, \hat{\lambda}) = \hat{\lambda}_{i_2, j_2}$  for  $\tilde{\mathbf{d}} = e_{i_1, j_1} - e_{i_2, j_2} \in \mathcal{D}^{(i^*, j^*)}(\hat{\lambda})$ . Let  $W^{(r)} = \nabla a(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)^\top \tilde{\mathbf{d}}^{(r)}$ .

**Step 4:** If  $W^{(r)}$  satisfies  $W^{(r)} < \max\left\{-\kappa_0, -\left(\frac{\log r}{r}\right)^{1/4}\right\}$  and  $\tilde{s}^{\max}(\tilde{\mathbf{d}}^{(r)}, \hat{\lambda}) W^{(r)} < \max\left\{-\kappa_0, -\left(\frac{\log r}{r}\right)^{1/2}\right\}$ , choose  $\tilde{s}^{(r)} = \text{LineSearch}(\tilde{\mathbf{d}}^{(r)}, \tilde{s}^{\max}(\tilde{\mathbf{d}}^{(r)}, \hat{\lambda}), \hat{\lambda}, \bar{Y}, \hat{\sigma}^2)$ , and let  $\hat{\lambda} = \hat{\lambda} + \tilde{s}^{(r)} \tilde{\mathbf{d}}^{(r)}$ . Otherwise,  $\hat{\lambda}$  remains unchanged.

**Step 5:** Update  $\chi_{i,h}(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)$ . Compute  $\hat{\alpha}_{i,h}^*$  using (9) with  $\chi_{i,h}(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)$ .

**Step 6:** Choose  $(i^r, h^r) = \arg \max_{(i,h)} \frac{\hat{\alpha}_{i,h}^*}{\hat{n}_{i,h}}$ . Provide one more replication to treatment  $i^r$  under context  $\mathbf{x}_{h^r}^\circ$ . Update  $\bar{Y}_{i^r}(\mathbf{x}_{h^r}^\circ)$ ,  $\hat{\sigma}_{i^r}^2(\mathbf{x}_{h^r}^\circ)$ ,  $\hat{\beta}_{i^r}$ , and  $\hat{i}^*(\mathbf{x}_j)$ . If  $\hat{i}^*(\mathbf{x}_j)$  is

changed for any  $\mathbf{x}_j$ , set  $\hat{\lambda}_{i,j} = \frac{1}{(k-1)m}$ ,  $i = 1, \dots, k$ ,  $i \neq \hat{i}^*(\mathbf{x}_j)$ ,  $j = 1, \dots, m$ .

**Step 7:** Update  $\hat{n}_{i,h}$ ,  $n^{(r)}$  and  $\hat{\alpha}_{i,h}$ .

**until**  $n^{(r)} = n$ .

This idea is summarized in CR&S Algorithm 2. Note that this way of algorithm design has appeared in the literature, for example, in Zhou et al. (2023). The calculation of the step-size  $\tilde{s}$  in Step 4 of CR&S Algorithm 2 calls for a line search, which is provided in Algorithm 3 below. For the input parameters  $\tilde{s}_1$  and  $\tilde{s}_2$  in Algorithm 3, the recommended values are  $10^{-4}$  and  $10^{-1}$  (Nocedal and Wright 2006, chapter 3).

#### Algorithm 3 (LineSearch( $\tilde{\mathbf{d}}, \tilde{s}^{\max}, \hat{\lambda}, \bar{Y}, \hat{\sigma}^2$ ))

**Initialization:** Specify the descent direction  $\tilde{\mathbf{d}}$ , maximum feasible step-size  $\tilde{s}^{\max}$ , dual solution  $\hat{\lambda}$ , estimate of coefficients  $\hat{\beta}$ , and parameters for line search  $\tilde{s}_1$ ,  $\tilde{s}_2$ , and  $\tau \in (0, 1)$ . Let  $\tilde{s} = \tilde{s}^{\max}$ .

**where** any of the conditions

$$\min_i \min_j \chi_{i,j}(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2) > 0, \quad (12)$$

$$a(\hat{\lambda} + \tilde{s} \cdot \tilde{\mathbf{d}}, \bar{Y}, \hat{\sigma}^2) \leq a(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2) + \tilde{s}_1 \tilde{s} \nabla a(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)^\top \tilde{\mathbf{d}}, \quad (13)$$

$$\nabla a(\hat{\lambda} + \tilde{s} \cdot \tilde{\mathbf{d}}, \bar{Y}, \hat{\sigma}^2)^\top \tilde{\mathbf{d}} \leq \tilde{s}_2 |\nabla a(\hat{\lambda}, \bar{Y}, \hat{\sigma}^2)^\top \tilde{\mathbf{d}}|, \quad (14)$$

are not satisfied, **do**  $\tilde{s} \leftarrow \tau \tilde{s}$ .

**Output:** Step-size  $\tilde{s}$ .

Similar to CR&S Algorithm 1, CR&S Algorithm 2 can recover the optimal solution to problem (8). This is established in the following theorem.

**Theorem 5.** Suppose Assumptions 1–3 hold. For  $\hat{\alpha}_{i,h}$  generated by CR&S Algorithm 2, we have that  $\hat{\alpha}_{i,h}$  converges to the optimal solution to problem (8) almost surely,  $i = 1, \dots, k$  and  $h = 1, \dots, p$ .

## 6. Numerical Experiments

In this section, we conduct two sets of numerical experiments. The first set tests the performances of CR&S Algorithms 1 and 2 on a series of benchmark functions, and the second set applies them to two real-world PM problems.

### 6.1. Performance Comparison on the Benchmark Functions

In this test, we numerically assess the performances of the CR&S Algorithms 1 and 2 on some benchmark functions. We use the following algorithms for comparison:

- *Equal allocation.* The number of simulation replications allocated to any treatment-context pair is equal. This

is a naive method and can serve as a baseline against which improvement from other methods might be measured.

- *Successive rejection with equal allocation among contexts (equal SR)*. The original SR was designed for a single context and has been shown to be highly efficient for BAI problems with bounded sampling distributions (Carpentier and Locatelli 2016, Gabillon et al. 2012). In this test, we apply SR to treatments under the same context while equally distributing the simulation budget among different contexts. Under each context, the simulation budget available  $n/m$  is divided into  $k-1$  phases. Every treatment that has not been rejected receives  $n_{(i)} - n_{(i-1)}$  more replications, and the estimated worst treatment is rejected in phase  $i, i = 1, \dots, k-1$ .

- *Optimal computing budget allocation with equal allocation among contexts (equal OCBA)*. Similarly to SR, the original OCBA was designed for a single context. In this test, we apply OCBA to treatments under the same context while equally distributing the simulation budget among different contexts:

$$\frac{n_{i_1,j}}{n_{i_2,j}} = \frac{\sigma_{i_1}^2(\mathbf{x}_j) (y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j) - y_{i_2}(\mathbf{x}_j))^2}{\sigma_{i_2}^2(\mathbf{x}_j) (y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j) - y_{i_1}(\mathbf{x}_j))^2},$$

$$i_1, i_2 \in \{1, 2, \dots, k\} \setminus \{i^*(\mathbf{x}_j)\}, j = 1, 2, \dots, m,$$

$$n_{i^*(\mathbf{x}_j),j} = \sigma_{i^*(\mathbf{x}_j)}(\mathbf{x}_j) \left( \sum_{i=1, i \neq i^*(\mathbf{x}_j)}^k \left( \frac{n_{i,j}}{\sigma_i(\mathbf{x}_j)} \right)^2 \right)^{\frac{1}{2}}, j = 1, 2, \dots, m,$$

$$\sum_{i=1}^k n_{i,j_1} = \sum_{i=1}^k n_{i,j_2}, j_1, j_2 = 1, 2, \dots, m.$$

- *The two-stage procedure (TS; Shen et al. 2021)*. TS considers R&S in the presence of contexts and also assumes a linear relationship between treatment performance and contexts as CR&S Algorithm 2. It allocates a small fraction of the simulation budget to some selected treatment-context pairs in the first stage and, based on the sample estimates, decides the number of replications that these treatment-context pairs should receive in the second stage. TS is based on the IZ method. The ultimate goal of it is to make a guarantee of the quality of the selected design over the context space instead of maximizing the quality. When stopped, the total simulation budget consumed by TS is random. To add TS into comparison, we use the allocation  $\alpha_{TS}$  obtained from the first stage of TS as a reference to allocate the remaining fixed simulation budget.

- *Optimal allocation matching (OAM, Hao et al. 2020)*. OAM is an algorithm for contextual bandit problems. Suppose  $y_i(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}_i, i = 1, \dots, k$ . OAM shows that the optimal budget allocation of contextual bandits

satisfies

$$\begin{aligned} & \inf_{\alpha_{i,j} \in [0, \infty]} \sum_{j=1}^m \sum_{i=1}^k \tilde{\alpha}_{i,j} (y_i(\mathbf{x}_j) - y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j)), \\ \text{s.t. } & \mathbf{f}(\mathbf{x}_j)^\top \left( \sum_{j'=1}^m \alpha_{i,j'} \mathbf{f}(\mathbf{x}_{j'}) \mathbf{f}(\mathbf{x}_{j'})^\top \right)^{-1} \\ & \mathbf{f}(\mathbf{x}_j) \leq \frac{(y_i(\mathbf{x}_j) - y_{i^*(\mathbf{x}_j)}(\mathbf{x}_j))^2}{2}, \forall j \neq i^*(\mathbf{x}_j), i = 1, \dots, k. \end{aligned}$$

Intuitively, the left-hand side of the constraint represents the width of the confidence interval to compare  $\mathbf{f}(\mathbf{x}_j)^\top \hat{\boldsymbol{\beta}}_i$  and  $\mathbf{f}(\mathbf{x}_j)^\top \hat{\boldsymbol{\beta}}_{i^*(\mathbf{x}_j)}$ . In each iteration of OAM, with the given context, it decides which treatment to sample based on an approximated optimal allocation. As discussed in Section 1, contextual bandit algorithms (including OAM) do not decide which context to sample. In this test, we set each context to be sampled with the same probability of  $1/m$  for OAM.

The test will be conducted on the benchmark functions below, where  $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$  is the context,  $\mathbf{z} = (z_1, \dots, z_d)^\top \in \mathbb{R}^d$  is the solution for the benchmark function (treatments in PM), and  $\epsilon$  is a normally distributed noise that is independent across different solutions, contexts, and simulation replications.

1. Sphere function:  $Y(\mathbf{z}, \mathbf{x}) = f(\mathbf{z}, \mathbf{x}) + \epsilon = \sum_{l=1}^d (z_l - x_l)^2 + \epsilon$ . The global minimum of  $f(\mathbf{z}, \mathbf{x})$  is 0 obtained at  $\mathbf{z} = \mathbf{x}$ . We consider the one-dimensional case ( $d=1$ ) of this problem with four contexts  $\mathbf{x} \in \{-0.45, -0.15, 0.15, 0.45\}$  and 11 solutions  $\mathbf{z} \in \{-1.25, -1.00, -0.75, \dots, 1.25\}$ . The noise  $\epsilon$  follows the normal distribution  $N(0, 0.05)$ .

2. Rosenbrock function:  $Y(\mathbf{z}, \mathbf{x}) = f(\mathbf{z}, \mathbf{x}) + \epsilon = \sum_{l=1}^{d-1} [100((z_{l+1} - x_{l+1}) - (z_l - x_l))^2 + (1 - (z_l - x_l))^2] + \epsilon$ . The global minimum of  $f(\mathbf{z}, \mathbf{x})$  is 0 obtained at  $z_l = x_l + 1, l = 1, 2, \dots, d, d > 1$ . We consider the two-dimensional case ( $d=2$ ) of this problem with 25 contexts  $\mathbf{x} \in \{-0.30, -0.15, 0, 0.15, 0.30\} \times \{-0.30, -0.15, 0, 0.15, 0.30\}$  and nine solutions  $\mathbf{z} \in \{0, 0.75, 1.5\} \times \{0, 0.75, 1.5\}$ . The noise  $\epsilon$  follows the normal distribution  $N(0, 2.25)$ .

3. Randomly generated linear functions:  $Y(\mathbf{z}, \mathbf{x}) = f(\mathbf{z}, \mathbf{x}) + \epsilon = \boldsymbol{\beta}(\mathbf{z})^\top (1, \mathbf{x}^\top)^\top + \epsilon$ , where components of  $\boldsymbol{\beta}(\mathbf{z})$  are randomly generated from  $\text{Unif}[0, 5]$ , and  $\epsilon$  follows the normal distribution  $N(0, 1)$ . We consider context space dimensions  $d=1$  and  $d=3$ ,  $4^d$  contexts  $\mathbf{x} \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}^d$ , and five solutions  $\mathbf{z} \in \{1, 2, \dots, 5\}$ .

In the first two examples, we modified the original benchmark functions  $f(\mathbf{z})$  to  $f(\mathbf{z} - \mathbf{x})$  to incorporate context  $\mathbf{x}$ . These two examples align with the structure of the small-scale problem and will be used to compare equal allocation, equal SR, equal OCBA, OAM, and CR&S Algorithm 1. The third example is not a typical benchmark function. It is built with a linear structure

that aligns with the large-scale problem. It will be used to compare equal allocation, equal allocation, OAM, TS, and CR&S Algorithms 1 and 2.

We assess the average performances of the compared algorithms based on  $10^4$  macro-replications for the sphere and Rosenbrock functions and 2, 500 macro-replications for the randomly generated linear functions. Figure 1 shows the comparison result under different simulation budgets. The four rows in Figure 1 correspond to the sphere function, Rosenbrock function, and one-dimensional and three-dimensional randomly generated linear functions. The three columns correspond to the three measures under study. Because the linear functions in each macro-replication are randomly generated instead of being fixed, the average performances of the algorithms in the third and fourth rows are with respect to randomness from the function instances and simulation noises.

The proposed CR&S Algorithm 1 performs the best under the sphere and Rosenbrock functions and the three measures, followed by equal OCBA, OAM and equal SR. The advantage of CR&S Algorithm 1 is particularly big under the Rosenbrock function. Although equal OCBA and equal SR have been shown to be efficient for R&S problems, they do not have any good mechanisms to balance the budgets allocated among contexts, causing the performances of them to be inferior to CR&S Algorithm 1. The equal allocation performs the worst.

Under the two linear functions, CR&S Algorithm 2 performs the best. When the context space is one-dimensional, CR&S Algorithm 1 outperforms TS and OAM. The goal of TS is to guarantee the quality of the estimated best treatment under each context. The budget allocation of it is not so effective in optimizing the quality of the estimated best treatments as CR&S Algorithm 1. OAM lacks efficient mechanisms to balance the simulation budgets allocated among contexts. When the context space is three-dimensional, TS and OAM outperform CR&S Algorithm 1. Although the budget allocations of TS and OAM are not optimal for the large-scale problem, they have a major advantage over CR&S Algorithm 1 in that they utilize prediction models. This advantage becomes more obvious when the total number of contexts is larger, as with the three-dimensional context space. The equal allocation again performs the worst.

## 6.2. Case Studies

In this test, we apply our proposed algorithms to two real-world PM problems, namely the prevention of cervical cancer (Levin et al. 2015) and treatment of chronic obstructive pulmonary disease (Hoogendoorn et al. 2019, Corro Ramos et al. 2020). Because of the space limitation, here we will provide only the numerical results for the cervical cancer example and leave

the test of the chronic obstructive pulmonary disease in Section S.1 of the Online Companion.

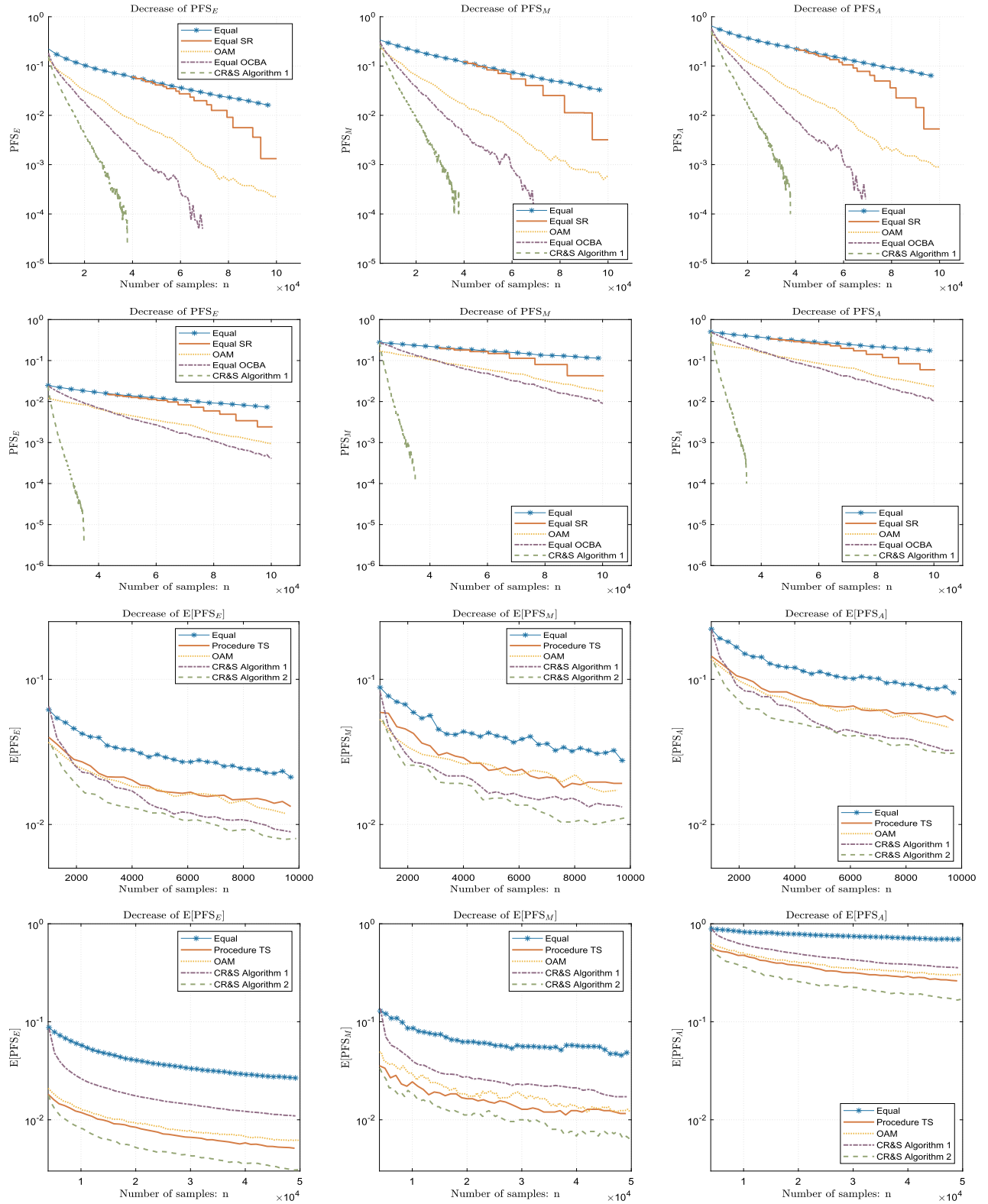
Cervical cancer is the fifth-leading cause of cancer globally. Unlike most cancers, cervical cancer has only one direct cause, the human papillomavirus (HPV), and is thus preventable. Although the widespread screening practice has led to a dramatic decrease in the cervical cancer mortality in developed countries, the cost of it is prohibitive, especially for women with low income in developing countries (Levin et al. 2015).

The incidence rate of cervical cancer evolves with age and reaches its peak when the patients are around 45 years old (GLOBOCAN 2019). There are two ways to prevent the progression of it: the traditional screening and the newly invented HPV vaccine. Traditional screening programs (distinguished by the frequency) conduct the examination at regular time points (McLay et al. 2010), including no screening, low-frequency screening (two times per lifetime at age 35 and 45), and high-frequency screening (one screening every three years from 30 to 60).

The HPV vaccine came to the market in recent years and is usually expensive. Despite the high price, the vaccine could effectively prevent the infection of the most risky types of HPV (e.g., HPV 16/18), and the immunization period is life-long. The perfect time for HPV vaccination is before the start of any sexual behaviors (usually at age 12) (Westra et al. 2011). The decision on HPV vaccination is a trade-off between the current economic loss and future risk. Vaccination or not, combined with the screening policy, forms six possible treatment methods: HPV vaccination alone, HPV vaccination with a low-frequency screening, HPV vaccination with a high-frequency screening, low-frequency screening alone, high-frequency screening alone, and no prevention.

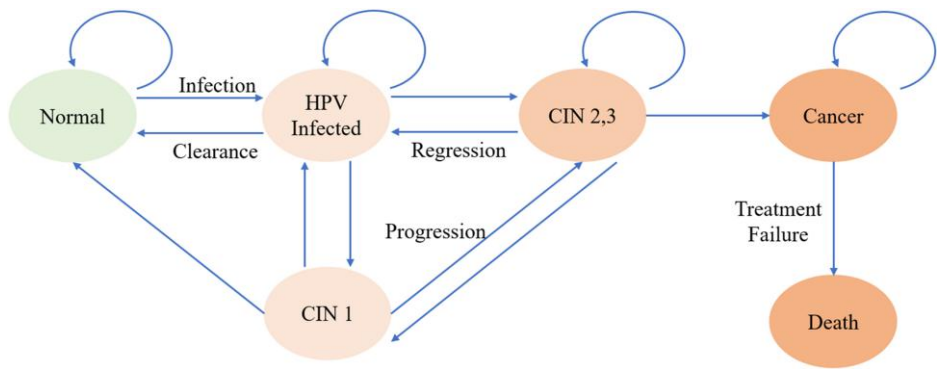
The simulation model of this problem is developed based on the Markov chain in Figure 2. Cervical intraepithelial neoplasia at grade 1 (CIN 1) and its deteriorating grades (CIN 2,3) result from the human papillomavirus (HPV) infection, and they may regress to the normal state. However, once the lesions are at CIN 2,3 and are not detected, the illness would develop into cancer at substantial risk.

The context variables that we consider include income, age, and HPV progression risk of the patients and the price of HPV vaccine. The vaccine is assumed to have four possible prices: \$5, \$20, \$35, and \$50. Income is classified into five levels, representing the five income quintiles in a population. Age is classified into four five-year groups (11 – 15, 16 – 20, 21 – 25, 26 – 30). The HPV progression risk has low, medium, and high levels, corresponding to different multipliers on the baseline progression rate. As a result, we have 240 possible contexts. The cancer treatment cost and state transition probabilities for each context are determined based on Levin et al.

**Figure 1.** (Color online) Comparison on the Benchmark Functions

*Note.* From top to bottom, benchmark functions being tested are the sphere function, Rosenbrock function, and 1-dimensional and 3-dimensional randomly generated linear functions.

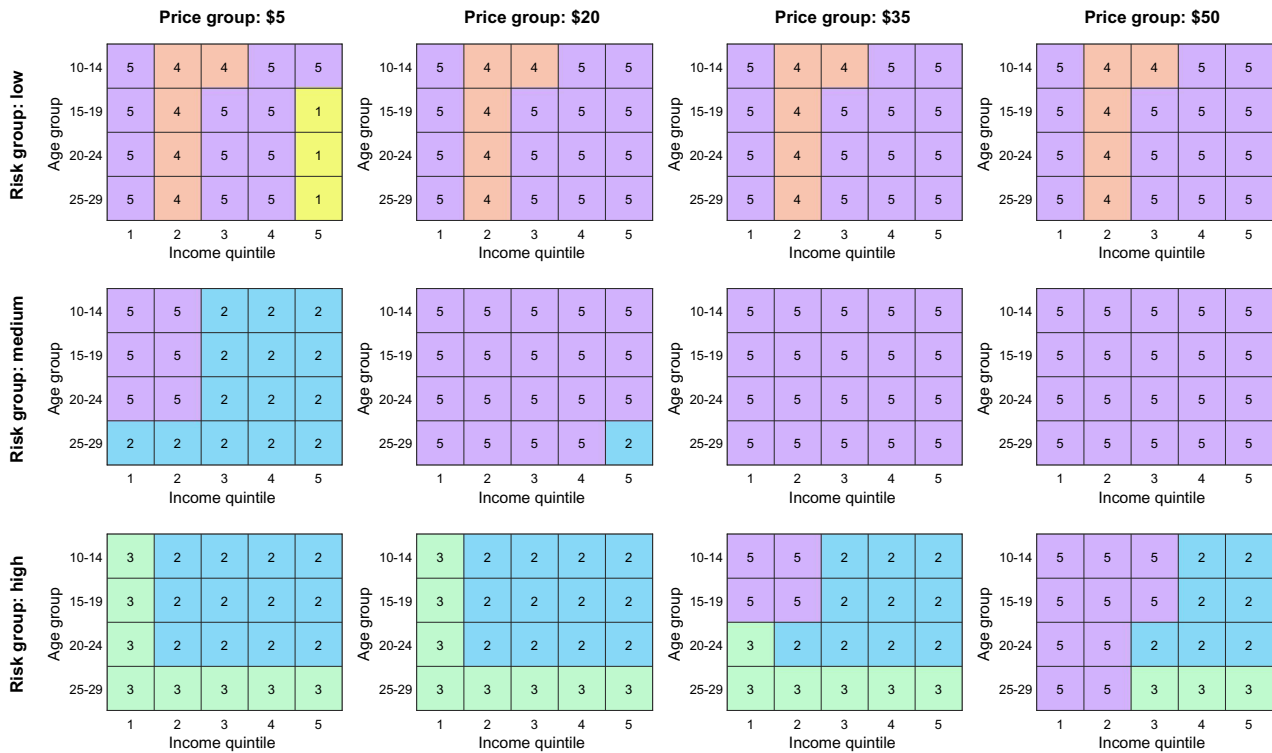
Figure 2. (Color online) Simulation Model for the Cervical Cancer



Notes. This figure is adapted from Levin et al. (2015). Each state may suffer from age-related all-cause mortality. “Normal” state means the individual is not infected; “CIN 1” means the individual has cervical intraepithelial neoplasia at grade 1; “CIN 2,3” means the individual has cervical intraepithelial neoplasia at grade 2 or 3.

(2015). The treatment performance is measured by the expected cost per quality-adjusted life years (QALY). We apply CR&S Algorithm 1 to this problem. The algorithm determines the number of simulation replications for each treatment-context pair and, based on it, estimates the best treatment under each context. The result is reported in Figure 3. It can be observed that the HPV vaccine with a price higher than \$20 is the best treatment for high-risk women only. The cost per QALY of it is too high for medium- and low-risk women. The best treatment for individuals with high income is mostly vaccination based. In terms of age, the best treatment for individuals at ages 11 – 20 does not involve vaccination, whereas for individuals at ages 21 – 30 the best treatment becomes vaccination based. This is because a female individual is most likely to get cervical cancer between age 30 and 60. When they are at ages 11 – 20, they are not exposed to the high risk of it, and

Figure 3. (Color online) Medical Decision Map for the Cervical Cancer Prevention Problem



Notes. In each subfigure, the horizontal and vertical axes represent the income quintiles and age groups, respectively. Numbers in cells show the best treatment method under different contexts. Specifically, numbers 1 – 6 mean HPV vaccination alone, HPV vaccination with a low-frequency screening, HPV vaccination with a high-frequency screening, low-frequency screening alone, high-frequency screening alone, and no-prevention.

there is no need for HPV vaccination. When they are at ages 21 – 30, although there is a certain probability of failure in getting immunization from the vaccines, the cancer-prone period that is coming soon makes the vaccination-based treatment methods the best choices for them.

## 7. Conclusions and Discussion

In this study, we consider the problem of personalized medicine. We adopt the tool of simulation for assessing the performances of the treatment methods and aim to efficiently utilize the computing time to select the best treatment for each patient context that might appear. To do so, we start by introducing three measures for evaluating the evidence of correct selection over the context space and showing that these measures have the same convergence rate function. Next, we propose two simulation budget allocation models that are appropriate for small and large context spaces. For the two models, we identify the rate-optimal budget allocation rules that optimize the rate function, develop convenient selection algorithms for implementation, and show the consistency of the algorithms. A series of numerical experiments on benchmark functions and real-world problems demonstrates the superior empirical performances of the proposed algorithms.

In this research, we have focused on the one-time treatment, where only one treatment decision is expected to be made for the patients. In practice, there is a class of PM problems that require multiple decisions during the progression of the disease, and the goal is to find the optimal treatment policy that maximizes the cumulative rewards over the decision periods (Negoescu et al. 2018, Lee et al. 2019). These problems are based on more complex context and decision structures, and our proposed CR&S algorithms cannot be applied in general. This is a good future research direction. In terms of methodology, we have solved the PM problem based on the OCBA method. Recently, Russo (2020) proposed three simple context-free Bayesian algorithms under a top-two framework for BAI, which have been shown to have nice theoretical properties and empirical performances. We believe it is also a promising research avenue to extend the top-two framework and algorithms to the PM problems.

## Acknowledgments

The authors thank the department editor, associate editor, and two anonymous reviewers for their valuable comments that have greatly improved the paper.

## References

Alonso D, McKane AJ, Pascual M (2007) Stochastic amplification in epidemics. *J. R. Soc. Interface* 4(14):575–582.

- Audibert JY, Bubeck S, Munos R (2010) Best arm identification in multi-armed bandits. *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, 41–53.
- Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Rev.* 53:464–501.
- Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, New York).
- Branke J, Chick SE, Schmidt C (2007) Selecting a selection procedure. *Management Sci.* 53(11):1916–1932.
- Brennan A, Chick SE, Davies R (2006) A taxonomy of model structures for economic evaluation of health technologies. *Health Econom.* 15(12):1295–1310.
- Cai T, Tian L, Wong PH, Wei LJ (2011) Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics* 12(2):270–282.
- Carpentier A, Locatelli A (2016) Tight (lower) bounds for the fixed budget best arm identification bandit problem. *Conference on Learning Theory*, 590–604 (PMLR, New York).
- Chen CH, Lin J, Yücesan E, Chick SE (2000) Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dyn. Syst.* 10:251–270.
- Chick SE, Branke J, Schmidt C (2010) Sequential sampling to myopically maximize the expected value of information. *INFORMS J. Comput.* 22(1):71–80.
- Chick SE, Mamani H, Simchi-Levi D (2008) Supply chain coordination and influenza vaccination. *Oper. Res.* 56(6):1493–1506.
- Chick SE, Koopman JS, Soorapanth S, Brown ME (2001) Infection transmission system models for microbial risk assessment. *Sci. Total Environ.* 274(1):197–207.
- Corro Ramos I, Hoogendoorn M, Rutten-van Mölken MP (2020) How to address uncertainty in health economic discrete-event simulation models: An illustration for chronic obstructive pulmonary disease. *Medical Decis. Making* 40(5):619–632.
- Dembo A, Zeitouni O (1998) *Large Deviations Techniques and Applications*, 2nd ed. (Springer, New York).
- Ding L, Hong LJ, Shen H, Zhang X (2022) Knowledge gradient for selection with covariates: Consistency and computation. *Naval Res. Logist.* 69(3):496–507.
- Fan W, Hong LJ, Zhang X (2020) Distributionally robust selection of the best. *Management Sci.* 66:190–208.
- Frazier PI, Powell WB, Dayanik S (2008) A knowledge-gradient policy for sequential information collection. *SIAM J. Control Optim.* 47(5):2410–2439.
- Gabillon V, Ghavamzadeh M, Lazaric A (2012) Best arm identification: A unified approach to fixed budget and fixed confidence. *Adv. Neural Inf. Process. Syst.* 25:1–9.
- Gao S, Chen W, Shi L (2017a) A new budget allocation framework for the expected opportunity cost. *Oper. Res.* 65:787–803.
- Gao S, Du J, Chen CH (2019) Selecting the optimal system design under covariates. *IEEE 15th International Conference on Automation Science and Engineering (CASE)* (IEEE, Piscataway, NJ), 547–552.
- Gao S, Xiao H, Zhou E, Chen W (2017b) Robust ranking and selection with optimal computing budget allocation. *Automatica J. IFAC.* 81:30–36.
- Garnett GP, Cousens S, Hallett TB, Steketee R, Walker N (2011) Mathematical models in the evaluation of health programmes. *Lancet* 378(9790):515–525.
- GLOBOCAN (2019) *Cancer over Time*. Accessed June 2019, <https://gco.iarc.fr/>.
- Goodwin T, Xu J, Celik N, Chen CH (2022) Real-time digital twin-based optimization with predictive simulation learning. *J. Simul.* 16(6):1–18.
- Hamburg MA, Collins FS (2010) The path to personalized medicine. *N. Engl. J. Med.* 363(4):301–304.
- Hao B, Lattimore T, Szepesvari C (2020) Adaptive exploration in linear contextual bandit. *Proc. 23rd Internat. Conf. Artificial Intelligence Statist.*, vol. 108 (PMLR, New York), 3536–3545.

- Hoogendoorn M, Ramos IC, Baldwin M, Guix NGR, Rutten-van Mölken MP (2019) Broadening the perspective of cost-effectiveness modeling in chronic obstructive pulmonary disease: A new patient-level simulation model suitable to evaluate stratified medicine. *Value Health* 22(3):313–321.
- Hu R, Ludkovski M (2017) Sequential design for ranking response surfaces. *SIAM/ASA J. Uncertain. Quantif.* 5(1):212–239.
- James G, Witten D, Hastie T, Tibshirani R (2013) *An Introduction to Statistical Learning*, vol. 112 (Springer, New York).
- Jedra Y, Proutiere A (2020) Optimal best-arm identification in linear bandits. *Adv. Neural Inform. Process. Syst.* 33:10007–10017.
- Kaufmann E, Cappé O, Garivier A (2016) On the complexity of best-arm identification in multi-armed bandit models. *J. Machine Learning Res.* 17(1):1–42.
- Kim SH, Nelson BL (2001) A fully sequential procedure for indifference-zone selection in simulation. *ACM Trans. Model. Comput. Simul.* 11(3):251–273.
- Law AM, Kelton WD (2000) *Simulation Modeling and Analysis*, 3rd ed. (McGraw-Hill, New York).
- Lee E, Lavieri M, Volk M (2019) Optimal screening for hepatocellular carcinoma: A restless bandit model. *Manufacturing. Service Oper. Management* 21(1):198–212.
- Levin CE, Sharma M, Olson Z, Verguet S, Shi JF, Wang SM, Qiao YL, Jamison DT, Kim JJ (2015) An extended cost-effectiveness analysis of publicly financed hpv vaccination to prevent cervical cancer in China. *Vaccine* 33(24):2830–2841.
- Li X, Zhang X, Zheng Z (2018) Data-driven ranking and selection: High-dimensional covariates and general dependence. *2018 Winter Simulation Conference (WSC)*, 1933–1944 (IEEE, Piscataway, NJ).
- Li Z, Ratliff L, Nassif H, Jamieson K, Jain L (2022) Instance-optimal PAC algorithms for contextual bandits. *Adv. Neural Inf. Processing Systems* 37590–37603.
- McLay LA, Foufoulides C, Merrick JRW (2010) Using simulation-optimization to construct screening strategies for cervical cancer. *Health Care Management Sci.* 13(4):294–318.
- Mok TSK (2011) Personalized medicine in lung cancer: What we need to know. *Nat. Rev. Clin. Oncol.* 8:661–668.
- Negoescu D, Bimpikis K, Brandeau M, Iancu D (2018) Dynamic learning of patient response types: An application to treating chronic diseases. *Management Sci.* 64(8):3469–3488.
- Nelson BL, Swann J, Goldsman D, Song W (2001) Simple procedures for selecting the best simulated system when the number of alternatives is large. *Oper. Res.* 49(6):950–963.
- Nocedal J, Wright S (2006) *Numerical Optimization* (Springer Science & Business Media, New York).
- Pearce M, Branke J (2017) Efficient expected improvement estimation for continuous multiple ranking and selection. *2017 Winter Simulation Conference (WSC)* (IEEE, Piscataway, NJ), 2161–2172.
- Russo D (2020) Simple Bayesian algorithms for best-arm identification. *Oper. Res.* 68(6):1625–1647.
- Ryzhov IO (2016) On the convergence rates of expected improvement methods. *Oper. Res.* 64(6):1515–1528.
- Schork NJ (2015) Personalized medicine: Time for one-person trials. *Nature* 520(7549):609–611.
- Shen H, Hong LJ, Zhang X (2021) Ranking and selection with covariates for personalized decision making. *INFORMS J. Comput.* 33(4):1500–1519.
- Soare M, Lazaric A, Munos R (2014) Best-arm identification in linear bandits. *Adv. Neural Inform. Process. Syst.* 27:828–836.
- Tan WY (2012) *Stochastic Models with Applications to Genetics, Cancers, AIDS and Other Biomedical Systems*, volume 19 of Series on Concrete and Applicable Mathematics (World Scientific, Singapore).
- Tewari A, Murphy SA (2017) From ads to interventions: Contextual bandits in mobile health. Rehag JM, Murphy SA, Kumar S, eds. *Mobile Health* (Springer, Cham, Switzerland), 495–517.
- Thompson M (1982) Regression methods in the comparison of accuracy. *Analyst (Lond.)*. 107(1279):1169–1180.
- Westra TA, Rozenbaum MH, Rogoza RM, Nijman HW, Daemen T, Postma MJ, Wilschut JC (2011) Until which age should women be vaccinated against HPV infection? Recommendation based on cost-effectiveness analyses. *J. Infect. Dis.* 204(3):377–384.
- World Health Organization (2003) *Making Choices in Health: WHO Guide to Cost-Effectiveness Analysis* (World Health Organization, Geneva).
- World Health Organization (2010) A healthy lifestyle – WHO recommendations. <https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle—who-recommendations>.
- Zhou Y, Fu MC, Ryzhov IO (2023) Sequential learning with a similarity selection index. *Oper. Res.*, ePub ahead of print May 17, <https://doi.org/10.1287/opre.2023.2478>.