

CONSTRAINED RANKING AND SELECTION FOR OPERATIONS OF AN EMERGENCY DEPARTMENT

Lee, M. L.*; Park, I.**; Park, D. U.** & Park, C.**#

* College of Business Administration, Hongik University, Seoul, 04066, Republic of Korea

** Department of Industrial Engineering, Hanyang University, Seoul, 04763, Republic of Korea

E-Mail: parkej@hanyang.ac.kr (# Corresponding author)

Abstract

We consider a problem of finding the optimal number of medical staffs for an emergency department through stochastic simulation. Specifically, the objective is to maximize the expected net profit per period of the emergency department while confining the maximum duration of expected waiting times of critical and noncritical patients. This is formulated as a simulation optimisation problem with two stochastic constraints on waiting times. To find the optimal solution of the problem under statistical guarantees, we introduce a two-step fully sequential ranking and selection framework: we first determine a set of strictly feasible solutions and then select the best solution among them. We implement the proposed framework via Simio and apply it to an emergency department of a university hospital.

(Received in December 2016, accepted in September 2017. This paper was with the authors 6 months for 1 revision.)

Key Words: Healthcare Management, Emergency Department, Simulation, Ranking and Selection, Simio

1. INTRODUCTION

Analysis for healthcare service systems has gathered much attention recently. Due to the systems' high complexity and variability, problems for healthcare system management include significant challenges, and many scholars and practitioners have adapted operations research techniques to solve the problems. Especially, for finding optimal policies with respect to healthcare resource planning and scheduling, various models such as simulation, optimisation, and mathematical programming are used and studied. Côté [1] presents a simulation model to analyse patient flows and resource utilization in a medical clinic. Swisher et al. [2] design a discrete-event simulation model to analyse and improve an existing physician network. Angelis et al. [3] suggest a framework combining simulation, neural network, and optimisation models to determine the optimal number of staffs in a healthcare system. Beaulieu et al. [4] formulate an integer programming for a scheduling problem in an emergency room. Baesler et al. [5] use a simulation optimisation algorithm to handle the uncertainties in schedules of an operating room. Cardoen et al. [6] provide a thorough review for operating room planning and scheduling.

As a part of the healthcare service systems in hospitals, emergency departments have been studied as well. Since the decisions made in an emergency department (ED) should be able to manage high uncertainties related to critical situations, like deaths, many of the studies select stochastic simulations as their solution approaches. Blasak et al. [7] present an example of a simulation model for an ED by using ARENA. Shim and Kumar [8] construct a simulation model to study the impact of changes in emergency care processes. Mielczarek [9] constructs a discrete event simulation model in order to evaluate costs of emergency services in a hospital. Sinreich and Marmor [10] show how a simulation tool can be applied to plan and re-design an existing emergency department. Carmen et al. [11] propose a decision support system for EDs by combining simulation modelling with data envelop analysis.

Although emergency medical services are indispensable to the public, many hospitals are reluctant to expand or invest in their EDs. It is usual because the high uncertainties in the EDs are difficult to deal with, and the profits from the EDs are usually low when compared to other departments. While it is not easy to increase the prices of the services, the costs of an ED, mainly caused by its medical staffs, keep increasing year by year. Therefore, in order to help the ED's efficient operation while maintaining the service quality, this paper addresses a problem of finding the optimal number of staffing levels that maximize the expected net profit per period in a steady state while guaranteeing designated levels of expected waiting times of critical and noncritical patients. For handling the high uncertainties in EDs, we construct a discrete-event simulation model via Simio and apply a constrained Ranking and Selection (R&S) framework to the model.

Constrained R&S frameworks have been developed to find the best feasible solution among finite (usually under 1000) number of simulated solutions with a statistical guarantee. The frameworks mainly belong to two different types: optimal computing budget allocation (OCBA) frameworks and fully sequential indifference-zone (IZ) frameworks. The OCBA frameworks are designed to allocate a given simulation budget to maximize the probability of returning the true best feasible solution correctly under stochastic constraint(s) (see Lee et al. [12], Hunter and Pasupathy [13], and Pasupathy et al. [14] for more details). On the other hand, the fully sequential IZ frameworks are designed to guarantee the pre-specified probability of selecting the true best feasible (or nearly feasible) solution correctly under stochastic constraint(s) (see Andradóttir and Kim [15], Healey et al. [16-18] for more details). The existing fully sequential IZ frameworks for the constrained R&S employ the feasibility check procedure of Batur and Kim [19] that returns a set of feasible or nearly feasible solutions with a pre-specified probability. Meanwhile, Lee et al. [20] present an adaptive feasibility check procedure recently that return a set of strictly feasible solutions with a pre-specified probability. Since our objective is to find the best strictly feasible solution of a constrained R&S problem with a given pre-specified probability, we focus on fully sequential IZ frameworks with the adaptive feasibility check procedure rather than the OCBA frameworks.

The most closely related work to ours is from Ahmed and Alkhamis [21]. The paper considers a problem to decide the optimal number of medical staffs that maximize patient throughput under various constraints on budgets and waiting times of patients of an ED. They also consider simulation tools to solve the problem. Nevertheless, due to the limitation of the feasibility check procedure from Batur and Kim [19] they used in their problem, their solution approach does not guarantee to return strictly feasible solutions. Therefore, the averaged waiting times of critical patients can be longer than the designated threshold values under their solution, and it may result in critical situations related to the patient's survival. In order to prevent such a risk, we introduce a constrained R&S procedure, namely $AF+KN$, in this paper. Specifically, the $AF+KN$ procedure consists of two phases: the first phase AF adopts the adaptive feasibility check procedure from Lee et al. [20], and the second phase KN adopts the procedure of selection of the best in Kim and Nelson [22]. As a result, the procedure provides a statistical guarantee to select the best among the strictly feasible solutions. We implement the procedure $AF+KN$ as an add-in program of Simio, and use the add-in for finding the optimal number of staffing levels in an ED.

Possible challenges in this problem are as follows: 1) the work process in an ED are very complicated to construct a simulation model, 2) performance measures in the objective and constraints should be estimated via stochastic simulations under various scenarios and the task requires huge computational budget, and 3) pursuing the maximum profit of the ED may lead to increase patients' waiting times up to their threshold values.

This paper is organized as follows: Section 2 provides a brief description of the simulation model regarding an emergency department we targeted. In Section 3, we formulate our

problem as a stochastically constrained R&S problem and provide a procedure to solve the problem. Section 4 presents experimental results, and concluding remark follows in Section 5.

2. SIMULATION MODELLING

In this section, we describe key features of our simulation model. The targeted emergency department (ED) for modelling is a part of a university hospital in an Asian city: Seoul, Republic of Korea. The ED is operating 24 hours every day with three shifts. Five nurses, two interns, three residents, and three medical specialists per shift are the current staffing level of the ED.

2.1 Overall process

In order to describe process flow of the targeted ED, we first observe real incidences of patient flow during a week and take interviews with doctors and nurses in the hospital. Based on the observation and interviews, we depict the process flow shown in Fig. 1. When any patient arrives at the ED, the process starts. After the arrival, patients are categorized based on their level of severity. Note that the lower the level is, the higher the severity is. Patients with severity level 1 or 2 are directly transferred to a resuscitation room and immediately treated. If the patient survives in the resuscitation room, proper medical image services such as X-ray, CT, or MRI, are provided, and the patient is admitted to the main hospital. Patients with other severity levels need to see a receptionist for the registration and a nurse for the triage. While patients with severity level 5 take a single-step treatment in a casualty room, patients with severity level 3 or 4 take a two-step treatment with a basic examination. Patients with severity level 3 or 4 are first classified to pediatric patients, gynecologic patients, and others. Pediatric patients and gynecologic patients use special beds due to their physical characteristics while other patients use normal beds. After the basic examination and the first treatment, additional steps for medical image services and observations are suggested to the patients if needed, and then the second treatment is provided. The process in the ED ends with a patient release due to death, hospitalization, or discharge.

2.2 Patients

For modelling the patients, we study anonymous patient data visiting the ED for one year. Since the seasonal and week effects on patient arrival are not significant, we focus on hourly arrival rate in a day. We partition 24 hours into 1-hour intervals and count the number of patient arrivals for each interval separately. By taking the averages of the numbers over a year, we obtain the average hourly arrival rate as shown in Fig. 2. Based on the rates, the patient arrival process is modelled as a non-homogeneous Poisson process.

After generating the patients through the non-homogeneous Poisson process, the simulation model assigns characteristics of each patient as states of entities. Two main characteristics considered are the severity level (1-2, 3-4, 5) and the physical class (normal, pediatric, gynecologic). Table I provides the description of each severity level for the triage and shows proportions of the patients with severity level 1-2, 3-4, and 5 respectively. Among the patients with severity level 3 or 4, 11.4 % of the patients are pediatric patients and 3.5 % of the patients are gynecologic patients. The simulation model assigns the severity level and the physical class of a patient randomly according to the proportions.

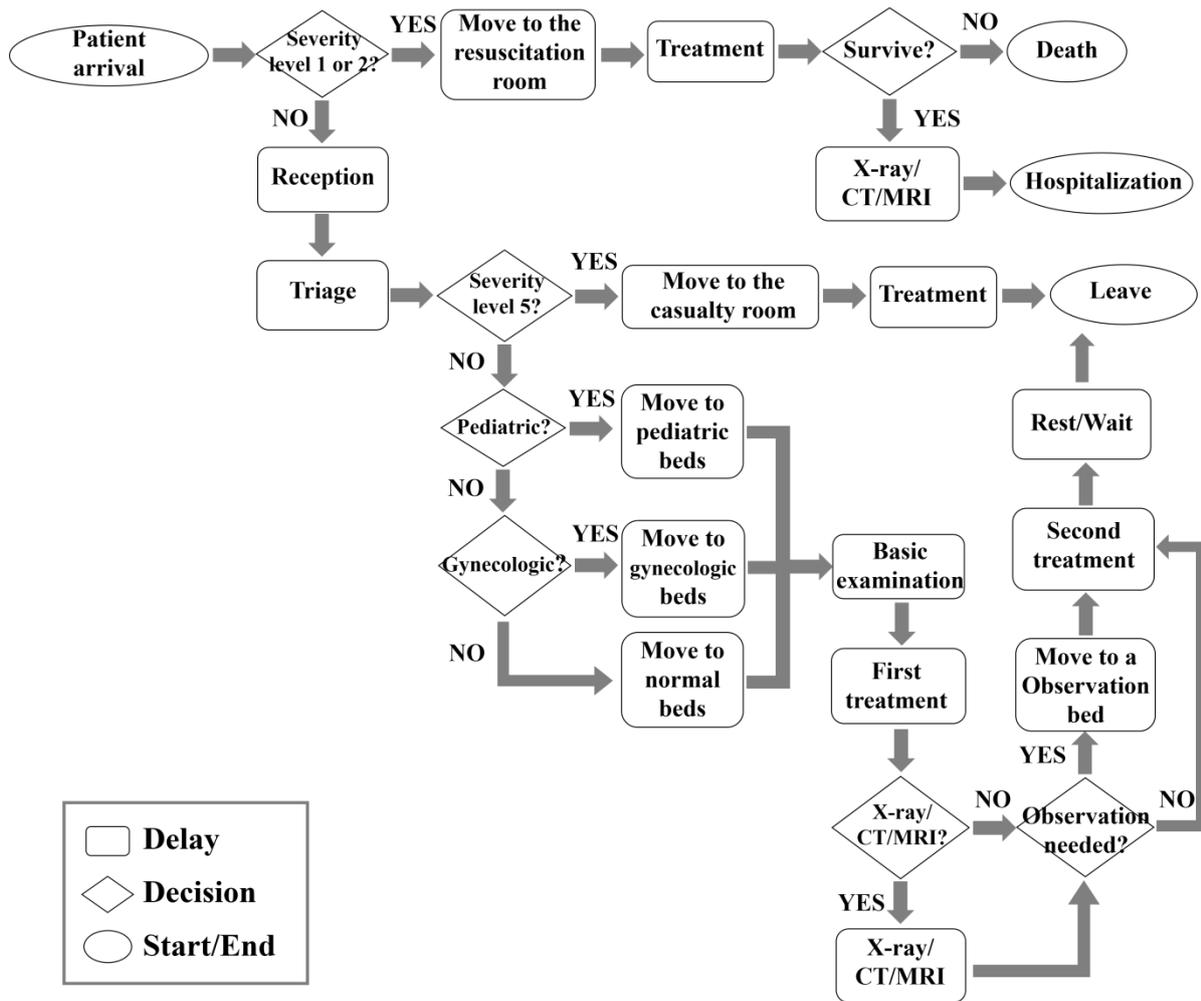


Figure 1: The medical treatment process of the emergency department.

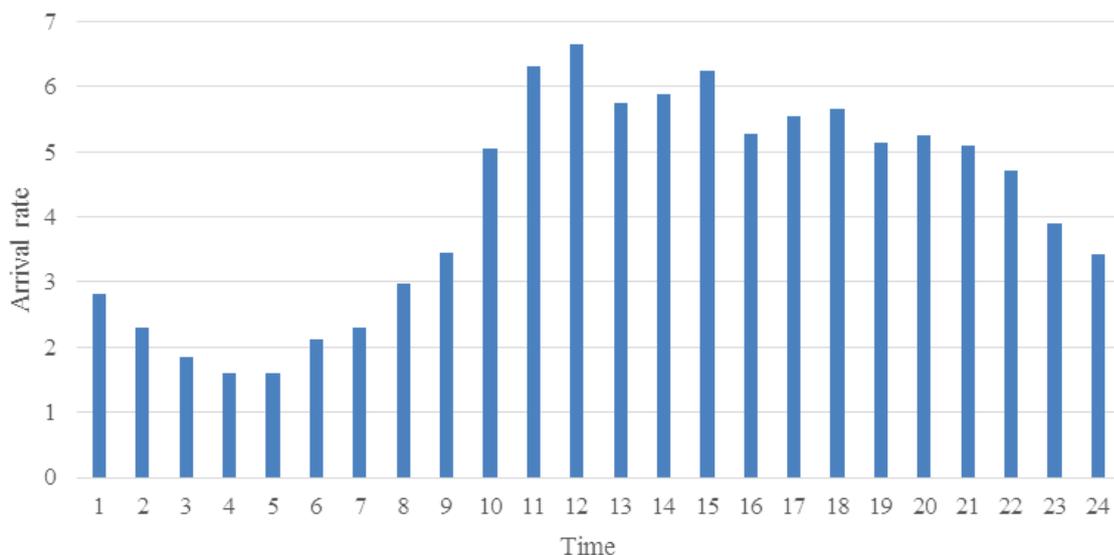


Figure 2: The estimated hourly patient arrival rate for a day.

2.3 Resources

The main non-human resources of the ED and their capacities are shown in Table II. Beds for normal patients can be used by any type of the patients with severity level 3 or 4, while beds

for pediatric and gynecologic patients are reserved for their own purposes. The resuscitation room can be occupied by the most urgent patients (i.e., patients with severity level 1 or 2) only and the casualty room can be occupied by patients with severity level 5 only. Observation rooms operate for patients with severity level 3 or 4 so that these patients may receive concentrated cares such as treatments for geriatric diseases or possible side-effects.

The human resources in the ED are categorized into four different types: 1) nurse, 2) medical intern, 3) resident, and 4) medical specialist. Each type of staffs covers a different range of jobs as in Table III.

Table I: Severity levels of patients.

Severity Level	Description	Proportion (%)
1	Dead or nearly dead	3.9
2	Alive but unconscious	
3	Conscious mostly with musculoskeletal disorders	84.5
4	Slightly wounded with cuts and bruises	
5	Similar to outpatients	11.6

Table II: The main resources of the ED and their capacities.

Resource	Capacity
Bed for normal patients	20
Bed for pediatric patients	3
Bed for gynecologic patients	1
Casualty room	1
Resuscitation room	1
Observation room	4
X-ray, CT, MRI	2 for each

Table III: Range of a job for each type of medical staffs.

Job	Nurse	Medical Intern	Resident	Medical Specialist
Triage	X			
Treatment (for severity level 1-2)	X		X	X
Basic examination (for severity level 3-4)	X	X		
First treatment (for severity level 3-4)	X		X	
Second treatment (for severity level 3-4)	X		X	X
Treatment (for severity level 5)		X		

Through checking blood pressure, pulse, body temperature, and respiration, a nurse determines the severity level of patients. A patient with severity level 1-2 should be treated by a nurse, a resident, and a medical specialist. A patient with severity level 3-4 needs a nurse and a medical intern for the basic examination, a nurse and a resident for the first treatment, and a nurse, a resident, and a medical specialist for the second treatment. A patient with

severity level 5 is treated only by a medical intern. Any medical service in the process is provided only when all required staffs and resources become available. After analysing the service time data for a year of the ED, we model the distributions of the service times as shown in Table IV.

Table IV: Distributions of service times in minutes.

Activity	Distribution
Reception	Exponential with mean 1
Triage	Exponential with mean 1
Treatment at the resuscitation room	Exponential with mean 30
Treatment at the casualty room	Exponential with mean 30
Basic examination	Exponential with mean 10
First treatment for patients with severity level 3 or 4	Exponential with mean 30
Second treatment for patients with severity level 3 or 4	Exponential with mean 15
Medical Imaging (X-ray/CT/MRI)	Uniform (5,10)

3. OPTIMISATION MODEL

The purpose of the problem considered in this paper is finding the optimal number of staffs that maximize the ED’s expected net profit per period under constraints on the expected waiting times of patients. In this section, we first provide the problem formulation and then explain a solution approach based on a constrained R&S framework.

3.1 Problem formulation

Let x_1, x_2, x_3 and x_4 denote the number of nurses, medical interns, residents, and medical specialists, respectively. The ED is operated by 3 shifts a day, and a shift requires 8 hours of work approximately. Let \mathbf{x} define a solution vector (i.e., $\mathbf{x} = (x_1, x_2, x_3, x_4)$) while Θ denotes the set of whole solution candidates. We consider three performance measures $P(\mathbf{x})$, $W_1(\mathbf{x})$, and $W_2(\mathbf{x})$ for a solution \mathbf{x} . $P(\mathbf{x})$ represents the net profit of the ED per period. $W_1(\mathbf{x})$ denotes the waiting time per critical patient (i.e., a patient with severity level 1 or 2) and $W_2(\mathbf{x})$ denotes the waiting time per noncritical patient (i.e., a patient with severity level 3, 4, or 5). The waiting time is defined by the time duration between the arrival of a patient and the start of the first treatment of the patient. Therefore, for example, the waiting time for a noncritical patient even include the service times for the reception, triage, and basic examination. Since the waiting time of a critical patient can cause serious medical situations (e.g., deaths), the ED treats $W_1(\mathbf{x})$ more importantly than $W_2(\mathbf{x})$ and thus manages $W_1(\mathbf{x})$ separately from $W_2(\mathbf{x})$.

In order to calculate the net profit per period $P(\mathbf{x})$, we formulate the revenue per period $R(\mathbf{x})$, and the cost per period $C(\mathbf{x})$. The revenue of the ED mainly consists of the total payments from the patients and the fixed operational expense allocated by the main hospital. The payment amount from a patient does not depend on the severity level but depends on when the patient visit the ED. If a patient visits the ED in day time (i.e., from 8 am to 5 pm), the patient pays a general fee and a treatment fee for day time. If a patient visits the ED in night time (i.e., from 5 pm to 8 am), the patient pays a general fee and a treatment fee for night time. Let $L_d(\mathbf{x}, t)$ and $L_n(\mathbf{x}, t)$ define the number of patients who visit the ED in day time and night at time t , respectively. Then the revenue per period of the ED is formulated:

$$R(\mathbf{x}) = \int_{t=0}^T \{p_1(L_d(\mathbf{x}, t) + L_n(\mathbf{x}, t)) + p_2L_d(\mathbf{x}, t) + p_3L_n(\mathbf{x}, t)\}dt + p_4, \quad (1)$$

where T denotes the duration of the period, p_1 denotes the general fee, p_2 denotes the treatment fee in day time, p_3 denotes the treatment fee in night time, and p_4 denotes the total amount of operational expenses from the hospital per period.

Since the labour cost takes the most part of total cost of the ED, the cost per period $C(\mathbf{x})$ is formulated based on the labour cost as follows:

$$C(\mathbf{x}) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4, \quad (2)$$

where $c_1, c_2, c_3,$ and c_4 denote the amount of salary per period for a nurse, an intern, a resident, and a medical specialist, respectively. Therefore, the net profit per period is as follows:

$$P(\mathbf{x}) = R(\mathbf{x}) - C(\mathbf{x}), \quad (3)$$

and our optimisation problem is defined as follows:

$$\operatorname{argmax}_{\mathbf{x} \in \Theta} E[P(\mathbf{x})], \quad (4)$$

$$\text{subject to } E[W_1(\mathbf{x})] \leq q_1, \quad (5)$$

$$E[W_2(\mathbf{x})] \leq q_2, \quad (6)$$

where q_1 and q_2 are the maximum limits the ED considers for the expected value of $W_1(\mathbf{x})$ (i.e., $E[W_1(\mathbf{x})]$), and the expected value of $W_2(\mathbf{x})$ (i.e., $E[W_2(\mathbf{x})]$), respectively. Note that the within-replication averages of simulation observations $P(\mathbf{x}), W_1(\mathbf{x}),$ and $W_2(\mathbf{x})$ follow the normal distribution approximately.

3.2 Solution approach

In order to find the optimal solution of the problem in a constrained R&S framework, we benchmark a fully sequential IZ procedure of Andradóttir and Kim [15]. Our approach *AF+KN* consists of two phases largely. The first phase *AF* finds a set of strictly feasible solutions that satisfy the two waiting time constraints. Among the feasible solutions, the second phase *KN* selects the best or near best solutions with a statistical guarantee.

For the *AF* phase, we employ the adaptive feasibility check procedure of Lee et al. [20]. Fig. 3 describes the procedure named $AF(\mathbf{x}, \ell)$ to check the feasibility of solution \mathbf{x} regarding a constraint ℓ . In Fig. 3, $R(r; v, w, z) = \max\left\{0, \frac{wz}{2v} - \frac{v}{2}r\right\}$ for any $v, w, z \in \mathbb{R}, v \neq 0$. For $\ell = 1$ or 2 , both $W_{\ell j}^1(\mathbf{x})$ and $W_{\ell j}^2(\mathbf{x})$ for $j = 1, 2, \dots,$ are within-averages of $W_{\ell}(\mathbf{x})$ that are independent and identically distributed (i.i.d.). For the *KN* phase, we use the procedure of selection of the best in Kim and Nelson [22].

Fig. 4 introduces detailed steps of the entire solution procedure *AF+KN*. In Fig. 4, k represents the total number of solutions in Θ , and $P_j(\mathbf{x})$ represents the j^{th} within-averaged observation regarding $P(\mathbf{x})$. Note that $P_j(\mathbf{x})$ for $j = 1, 2, \dots,$ are i.i.d. and let $\bar{P}_r(\mathbf{x})$ be the sample average of $P_j(\mathbf{x}), j = 1, 2, \dots, r$ where r is the number of across replications. When two solutions \mathbf{x} and \mathbf{x}' are given, $S_{xx'}^2$ denotes the sample variance of $P_1(\mathbf{x}) - P_1(\mathbf{x}'), \dots, P_{n_0}(\mathbf{x}) - P_{n_0}(\mathbf{x}')$.

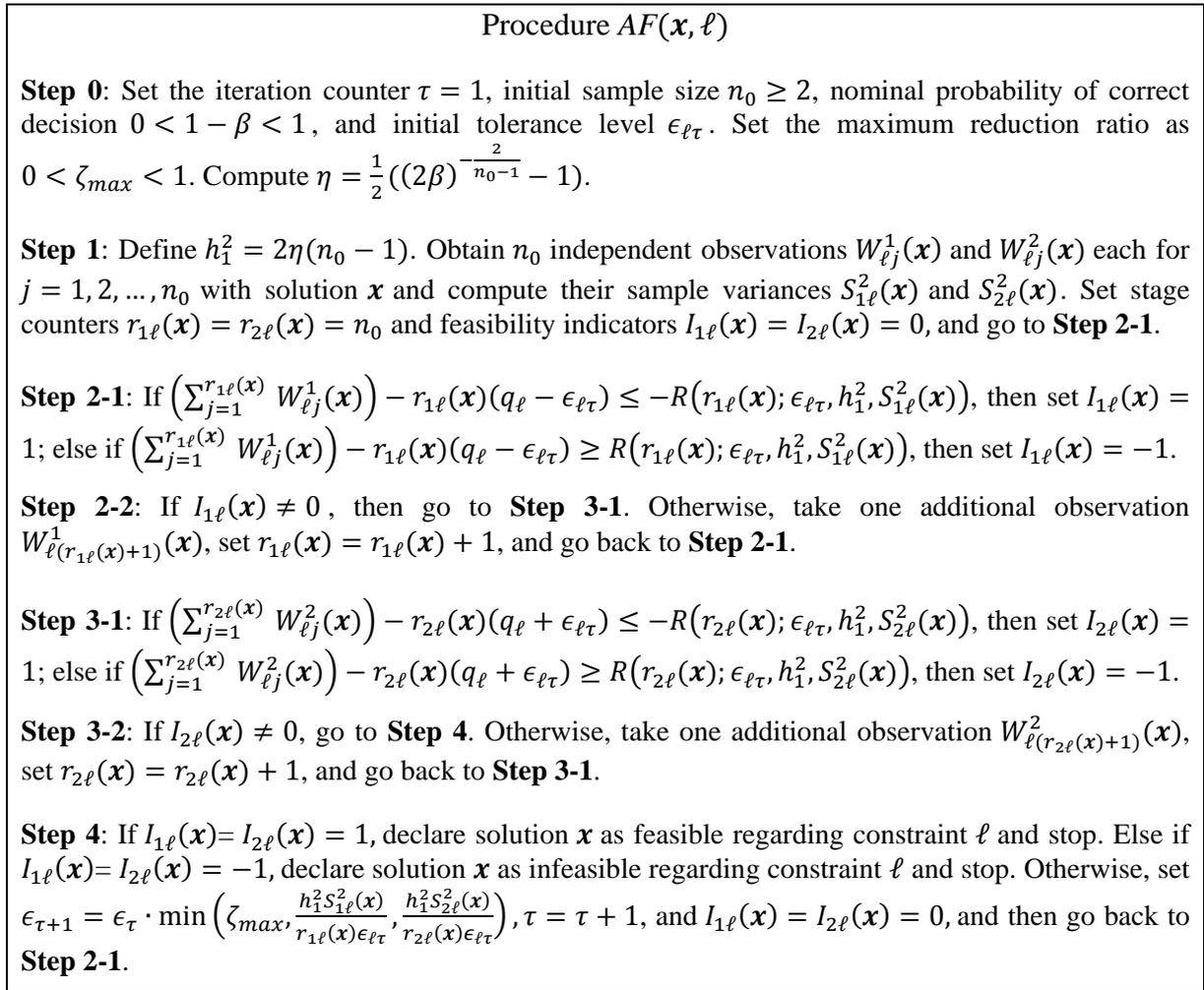


Figure 3: Description of the AF procedure for solution \mathbf{x} regarding constraint ℓ .

Remark. The $AF+KN$ procedure shown in Fig. 4 runs with the common random number (CRN) as in Andradóttir and Kim [15]. If users consider independent random streams over different solutions rather than using CRN, they need to set $\beta = \left[1 - (1 - \alpha_1)^{\frac{1}{k}}\right]/2$ in Phase I and $\beta = \left[1 - (1 - \alpha_2)^{\frac{1}{|I|-1}}\right]$ in Phase II instead of the β values in Fig. 4.

Let \mathbf{x}^* be the best feasible solution selected by $AF+KN$ and \mathbf{x}^b be the true best feasible solution. Then, from a simple extension of Lemma 1 in Andradóttir and Kim [15] with theorems in Lee et al. [20], $AF+KN$ can provide the statistical guarantee as follows:

$$\Pr\{E[P(\mathbf{x}^*)] \geq E[P(\mathbf{x}^b)] - \delta \text{ and } \mathbf{x}^* \text{ is strictly feasible}\} \geq 1 - \alpha. \quad (7)$$

Since Simio provides an add-in program for the KN phase already, we create a customized add-in program for the AF phase to implement $AF+KN$. In practice, we first apply the new add-in for the AF phase to all solutions under consideration. After terminating the job, we apply the existing add-in for the KN phase to the solutions which are declared as feasible by the AF phase.

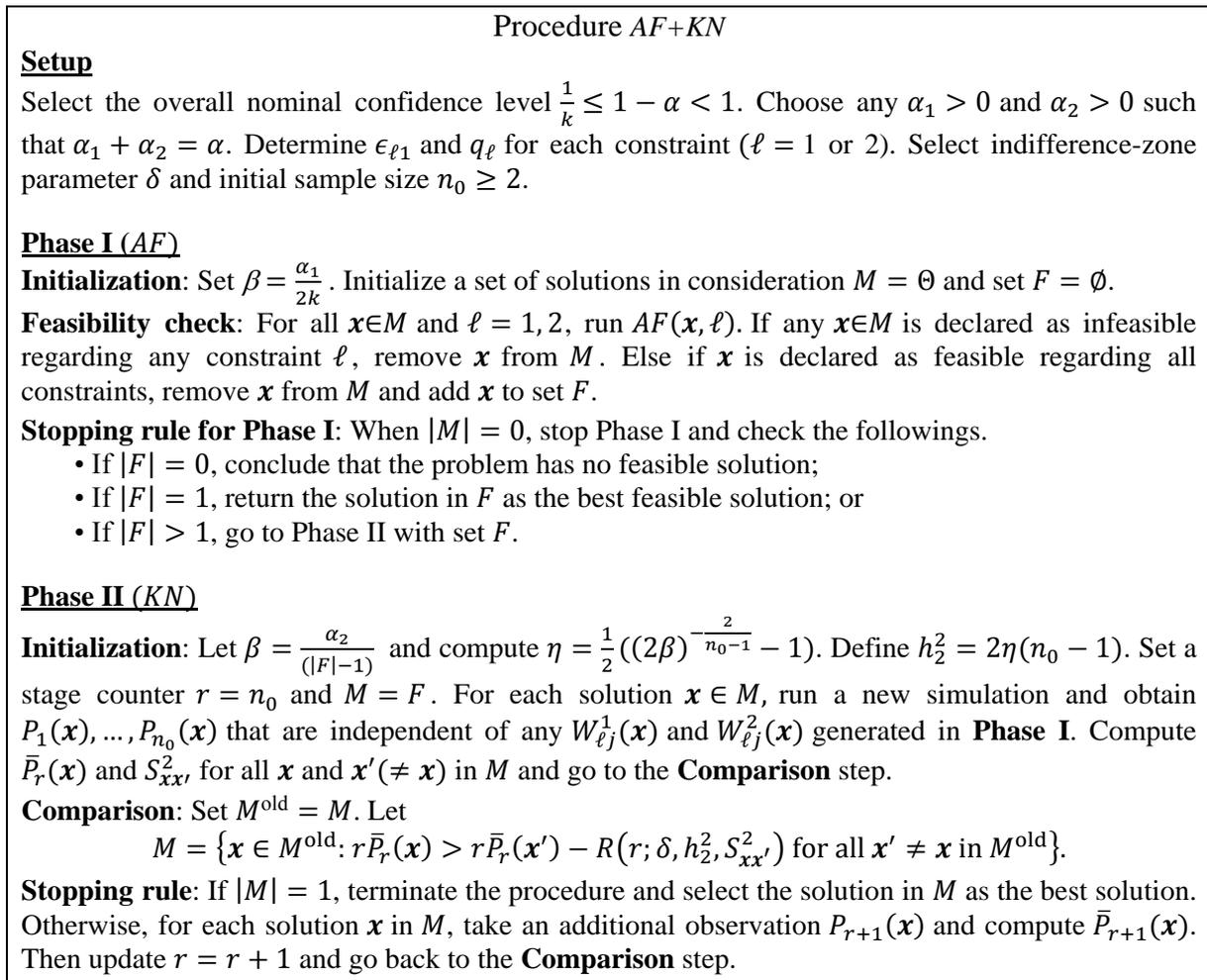


Figure 4: Description of the procedure $AF+KN$.

4. EXPERIMENTAL RESULTS

With the solution approach and the add-ins described in Section 3, we run our simulation model to find the optimal number of medical staffs in the ED. We regard a day as a unit period and run a simulation for 30 days after the warm-up period of 5 days in simulation clock. In order to keep information security of the ED, we use a special budget unit (BU) to represent all money value. For $R(\mathbf{x})$, $p_1 = 49.28$ BU per patient, $p_2 = 17.91$ BU per patient, $p_3 = 21.17$ BU per patient, and $p_4 = 76.71$ BU per day are used. For $C(\mathbf{x})$, $c_1 = 100$ BU per nurse per day, $c_2 = 133$ BU per intern per day, $c_3 = 166$ BU per resident per day, and $c_4 = 200$ BU per medical specialist per day are used. When we run the simulation once, a set of the observations of $P(\mathbf{x})$, $W_1(\mathbf{x})$, and $W_2(\mathbf{x})$ is obtained. By taking the averages of the observations from across replications, we obtain the estimates of $E[P(\mathbf{x})]$, $E[W_1(\mathbf{x})]$, and $E[W_2(\mathbf{x})]$. We consider $\Theta = \{(x_1, x_2, x_3, x_4) \in Z^4 | 4 \leq x_1, x_2, x_3, x_4 \leq 7\}$ and thus, the set Θ includes 256 candidate solutions. Note that R&S frameworks may not properly perform for the problem with more than 1000 candidate solutions.

Firstly, we evaluate the current staffing level of the ED with five nurses, two interns, three residents, and three medical specialists. Based on the results from 10,000 replications, the expected net profit per period (i.e., $E[P(\mathbf{x})]$) is estimated to 4530.50 BU per day, the expected waiting time per critical patient (i.e., $E[W_1(\mathbf{x})]$) is estimated to 0.4315 hour, and the expected waiting time per noncritical patient (i.e., $E[W_2(\mathbf{x})]$) is estimated to 1.8814 hour at the current staffing level. The results are presented in the first row of Table V. For the

validation check of the simulation model, it is confirmed that the averaged profit, waiting times, and throughput estimated through the simulation are similar to the true values of the current ED.

In order to compare the quality of solutions from $AF+KN$ to those from an existing solution approach, we employ the procedure in Andradóttir and Kim [15] which uses a similar fully sequential R&S framework to ours. As a counterpart of $AF+KN$, let us call the procedure as $F+KN$ in this paper. After some interviews with the medical staffs in the hospital, the target threshold values of waiting times are set to $q_1 = 0.15$ hour and $q_2 = 0.5$ hour. For the settings of both $AF+KN$ and $F+KN$, we choose the overall confidence level $\alpha = 0.05$, the initial sample size $n_0 = 10$, and the indifference zone parameter $\delta = 33.33$ BU. The initial tolerance levels of $AF+KN$ are set to $\epsilon_{11} = 0.05$ and $\epsilon_{21} = 0.5/3$. Since the performance of $F+KN$ depends on its tolerance levels ϵ_1 and ϵ_2 designated by users, we first test $\epsilon_1 = 0.05$ and $\epsilon_2 = 0.5/3$ and then test $\epsilon_1 = 0.01$ and $\epsilon_2 = 0.5/3$ as well to show the impact of the tolerance levels.

Table V summarizes the experimental results with the current staffing level and solutions given by $AF+KN$ and $F+KN$. For the procedures $AF+KN$ and $F+KN$, we first run ten preliminary replications to find promising solutions and report them in the 2nd column of Table V. The % of time that each solution is selected as the best by the procedures is provided in the 3rd column of Table V. Then, we obtain the estimates of the performance measures by running 10,000 replications for each solution (see the 4th to 6th columns of Table V). The average number of total observations for each solution is provided in the last column of Table V.

Table V: The current staffing level vs. the solutions selected by $AF+KN$ and $F+KN$.

Procedure	Suggested solution (x_1, x_2, x_3, x_4)	% of time	Estimated $E[P(\mathbf{x})]$	Estimated $E[W_1(\mathbf{x})]$	Estimated $E[W_2(\mathbf{x})]$	Average number of total observations
Current level	(5, 2, 3, 3)	-	4530.50	0.4315	1.8814	-
$AF+KN$	(6, 3, 5, 3)	100 %	3582.50	0.1107	0.2086	37212
$F+KN$ ($\epsilon_1 = 0.05$)	(5, 4, 5, 3)	40 %	3849.63	0.1564	0.3869	3099
	(5, 3, 5, 3)	20 %	3875.60	0.1592	0.3978	3150
	(5, 5, 5, 3)	10 %	3824.37	0.1567	0.3865	3171
	(6, 3, 4, 4)	10 %	3710.60	0.1753	0.3602	3287
	(5, 4, 5, 4)	10 %	3654.07	0.1541	0.3820	3359
$F+KN$ ($\epsilon_1 = 0.01$)	(5, 5, 5, 4)	10 %	3626.40	0.1537	0.3805	3574
	(6, 3, 5, 3)	60 %	3582.50	0.1107	0.2086	11970
	(5, 4, 5, 4)	30 %	3654.07	0.1541	0.3820	11644
	(5, 3, 5, 4)	10 %	3679.70	0.1560	0.3930	12005

When considering the constraint $E[W_1(\mathbf{x})] \leq 0.15$, one can notice that (6, 3, 5, 3) is the only strictly feasible solution among the suggested in Table V. With the solution, we expect to reduce the waiting times of patients significantly. Since $AF+KN$ adjusts its tolerance levels automatically, $AF+KN$ requires relatively large number of observations than $F+KN$. Nevertheless, the solution resulted from $AF+KN$ is strictly feasible for 100 % of time. On the other hand, $F+KN$ uses fixed tolerance levels specified by users and seems to require relatively small number of observations than $AF+KN$. However, we note that all solutions suggested by $F+KN$ are infeasible with $\epsilon_1 = 0.05$, and 40 % of solutions suggested by $F+KN$ are still remained as infeasible with $\epsilon_1 = 0.01$. In order to increase the probability that $F+KN$ returns the strictly feasible solution, one may need to decrease the tolerance level of $F+KN$.

but as the tolerance level becomes smaller, the required computational budget becomes larger. Moreover, in practice, it is usually unknown how small the proper tolerance level should be to guarantee the strict feasibility of the solution returned by $F+KN$. For more computational comparisons between AF and F procedures, see Lee et al. [20].

Let $\overline{W}_1(\mathbf{x})$ denote a within-averaged observation regarding W_1 for solution \mathbf{x} . As shown in Fig. 5, the Simio Measure of Risk and Error (SMORE) plot in Kelton et al. [23] is checked to analyse the variability in $\overline{W}_1(\mathbf{x})$. In Fig. 5, the centre, upper, and lower horizontal lines of the boxes represent the median, upper 5th quantile, and lower 5th quantile, respectively. For example, based on the leftmost plot in Fig. 5, we can conclude that $\Pr(\overline{W}_1(\mathbf{x}) \geq 0.33) \geq 0.95$ approximately and in other words, 95 % of the critical patients should wait longer than 20 minutes on average with the current staffing level.

Fig. 5 also presents the SMORE plots for $W_1(\mathbf{x})$ observations based on the solutions from $AF+KN$ and $F+KN$. In Fig. 5, the second plot from the left corresponds to the solution (6, 3, 5, 3) suggested by $AF+KN$ while the rest of plots correspond to other solutions suggested by $F+KN$ in Table V accordingly. When compared to the current staffing level, one can notice that the solutions suggested by $AF+KN$ and $F+KN$ reduce the variability in $W_1(\mathbf{x})$ significantly. Especially, the solution selected by $AF+KN$ guarantees that $\Pr(\overline{W}_1(\mathbf{x}) \leq 0.17) \geq 0.95$ while other solutions suggested by $F+KN$ guarantee that $\Pr(\overline{W}_1(\mathbf{x}) \leq 0.26) \geq 0.95$. In terms of the risk measures for $W_1(\mathbf{x})$, $AF+KN$ seems to provide the best solution consistently.

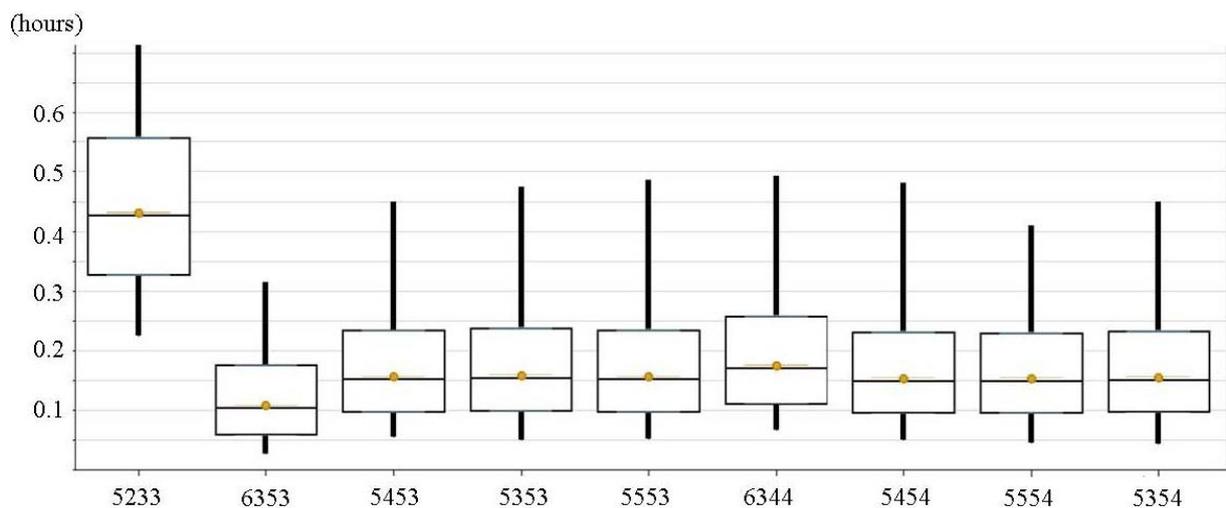


Figure 5: The SMORE plots for $\overline{W}_1(\mathbf{x})$ under the current staffing level and $\overline{W}_1(\mathbf{x})$ based on the solutions suggested by $AF+KN$ and $F+KN$.

5. CONCLUSION

In this paper, we consider a problem for operating an ED. The specific problem is finding the optimal number of medical staffs that maximize the expected net profit while limiting the maximum duration of waiting times of critical and noncritical patients. A simulation model is constructed via Simio to mimic the patients, treatment processes, and resources of a real ED and used to analyse various solution scenarios.

In order to solve the problem with the simulation model, we present a fully sequential Ranking and Selection procedure, namely $AF+KN$, that first identifies the feasible solutions and then selects the best solutions among them. As an advantage, $AF+KN$ guarantees strict feasibility of the returned solution with a pre-specified probability at least, while an existing competitive procedure cannot guarantee it.

By implementing $AF+KN$ to our problem, we find a strictly feasible solution that reduces averaged waiting times of patients by 75 %. The solution improves not only the average of the waiting times but also the distribution of the waiting times. When checking the SMORE plots, the solution returned by $AF+KN$ shows significantly better performance than the current staffing level of the ED in terms of the risk measures of the waiting time for critical patients.

The Simio model and the solution approach suggested in this paper can be applied to resource planning and scheduling of EDs directly. If the patient and resource data are given properly, EDs can use the model as well to estimate patients' waiting times in real time and share the information through the web with patients, hospitals, or other government offices needed. We expect that the EDs improve their service quality while reducing their burden in risk management and profit realization.

Since $AF+KN$ procedure is designed under R&S framework, the number of solutions we consider in Θ is limited. In addition, although the AF phase guarantees the strict feasibility of solutions selected, the KN phase still allows an indifference zone to the objective of the problem. Therefore, future researches can be conducted to relax the limit on the number of solutions via simulation optimisation or provide the optimal solution without the indifference zone as Fan et al. [24] do.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIP) (No.2015R1C1A2A01054115 and No.2016R1C1B2011462). The authors thank Dr. Hyung Goo Kang, the editor, and the anonymous referee for his or her helpful comments and suggestions.

REFERENCES

- [1] Côté, M. J. (1999). Patient flow and resource utilization in an outpatient clinic, *Socio-Economic Planning Sciences*, Vol. 33, No. 3, 231-245, doi:[10.1016/S0038-0121\(99\)00007-5](https://doi.org/10.1016/S0038-0121(99)00007-5)
- [2] Swisher, J. R.; Jacobson, S. H.; Jun, J. B.; Balci, O. (2001). Modeling and analysing a physician clinic environment using discrete-event (visual) simulation, *Computers & Operations Research*, Vol. 28, No. 2, 105-125, doi:[10.1016/S0305-0548\(99\)00093-3](https://doi.org/10.1016/S0305-0548(99)00093-3)
- [3] De Angelis, V.; Felici, G.; Impelluso, P. (2003). Integrating simulation and optimisation in health care centre management, *European Journal of Operational Research*, Vol. 150, No. 1, 101-114, doi:[10.1016/S0377-2217\(02\)00791-9](https://doi.org/10.1016/S0377-2217(02)00791-9)
- [4] Beaulieu, H.; Ferland, J. A.; Gendron, B.; Michelon, P. (2000). A mathematical programming approach for scheduling physicians in the emergency room, *Health Care Management Science*, Vol. 3, No. 3, 193-200, doi:[10.1023/A:1019009928005](https://doi.org/10.1023/A:1019009928005)
- [5] Baesler, F.; Gatica, J.; Correa, R. (2015). Simulation optimisation for operating room scheduling, *International Journal of Simulation Modelling*, Vol. 14, No. 2, 215-226, doi:[10.2507/IJSIMM14\(2\)3.287](https://doi.org/10.2507/IJSIMM14(2)3.287)
- [6] Cardoen, B.; Demeulemeester, E.; Belien, J. (2010). Operating room planning and scheduling: A literature review, *European Journal of Operational Research*, Vol. 201, No. 3, 921-932, doi:[10.1016/j.ejor.2009.04.011](https://doi.org/10.1016/j.ejor.2009.04.011)
- [7] Blasak, R. E.; Starks, D. W.; Armel, W. S.; Hayduk, M. C. (2003). Healthcare process analysis: the use of simulation to evaluate hospital operations between the emergency department and a medical telemetry unit, *Proceedings of the 35th Winter Simulation Conference*, 1887-1893
- [8] Shim, S. J.; Kumar, A. (2010). Simulation for emergency care process reengineering in hospitals, *Business Process Management Journal*, Vol. 16, No. 5, 795-805, doi:[10.1108/14637151011076476](https://doi.org/10.1108/14637151011076476)
- [9] Mielczarek, B. (2014). Simulation modelling for contracting hospital emergency services at the regional level, *European Journal of Operational Research*, Vol. 235, No. 1, 287-299, doi:[10.1016/j.ejor.2013.10.061](https://doi.org/10.1016/j.ejor.2013.10.061)

- [10] Sinreich, D.; Marmor, Y. (2005). Emergency department operations: the basis for developing a simulation tool, *IIE Transactions*, Vol. 37, No. 3, 233-245, doi:[10.1080/07408170590899625](https://doi.org/10.1080/07408170590899625)
- [11] Carmen, R.; Defraeye, M.; Van Nieuwenhuysse, I. (2015). A decision support system for capacity planning in emergency departments, *International Journal of Simulation Modelling*, Vol. 14, No. 2, 299-312, doi:[10.2507/IJSIMM14\(2\)10.308](https://doi.org/10.2507/IJSIMM14(2)10.308)
- [12] Lee, L. H.; Pujowidianto, N. A.; Li, L.-W.; Chen, C.-H.; Yap, C. M. (2012). Approximation simulation budget allocation for selecting the best design in the presence of stochastic constraints, *IEEE Transactions on Automatic Control*, Vol. 57, No. 11, 2940-2945, doi:[10.1109/TAC.2012.2195931](https://doi.org/10.1109/TAC.2012.2195931)
- [13] Hunter, S. R.; Pasupathy, R. (2013). Optimal sampling laws for stochastically constrained simulation optimization on finite sets, *INFORMS Journal on Computing*, Vol. 25, No. 3, 527-542, doi:[10.1287/ijoc.1120.0519](https://doi.org/10.1287/ijoc.1120.0519)
- [14] Pasupathy, R.; Hunter, S. R.; Pujowidianto, N. A.; Lee, L. H.; Chen, C.-H. (2015). Stochastically constrained ranking and selection via SCORE, *ACM Transactions on Modeling and Computer Simulation*, Vol. 25, No. 1, Article No. 1, 26 pages, doi:[10.1145/2630066](https://doi.org/10.1145/2630066)
- [15] Andradóttir, S.; Kim, S.-H. (2010). Fully sequential procedures for comparing constrained systems via simulation, *Naval Research Logistics*, Vol. 57, No. 5, 403-421, doi:[10.1002/nav.20408](https://doi.org/10.1002/nav.20408)
- [16] Healey, M. C.; Andradóttir, S.; Kim, S.-H. (2013). Efficient comparison of constrained systems using dormancy, *European Journal of Operational Research*, Vol. 224, No. 2, 340-352, doi:[10.1016/j.ejor.2012.08.012](https://doi.org/10.1016/j.ejor.2012.08.012)
- [17] Healey, M. C.; Andradóttir, S.; Kim, S.-H. (2014). Selection procedures for simulations with multiple constraints under independent and correlated sampling, *ACM Transactions on Modeling and Computer Simulation*, Vol. 24, No. 3, Article No. 14, doi:[10.1145/2567921](https://doi.org/10.1145/2567921)
- [18] Healey, M. C.; Andradóttir, S.; Kim, S.-H. (2015). A minimal switching procedure for constrained ranking and selection under independent or common random numbers, *IIE Transactions*, Vol. 47, No. 11, 1170-1184, doi:[10.1080/0740817X.2015.1009198](https://doi.org/10.1080/0740817X.2015.1009198)
- [19] Batur, D.; Kim, S.-H. (2010). Finding feasible systems in the presence of constraints on multiple performance measures, *ACM Transactions on Modeling and Computer Simulation*, Vol. 20, No. 3, Article No. 13, doi:[10.1145/1842713.1842716](https://doi.org/10.1145/1842713.1842716)
- [20] Lee, M. L.; Park, C.; Park, D. U. (2016). *Self-adjusting the tolerance level in a fully sequential feasibility check procedure*, Technical Report, Department of Industrial Engineering, Hanyang University, Seoul
- [21] Ahmed, M. A.; Alkhamis, T. M. (2009). Simulation optimization for an emergency department healthcare unit in Kuwait, *European Journal of Operational Research*, Vol. 198, No. 3, 936-942, doi:[10.1016/j.ejor.2008.10.025](https://doi.org/10.1016/j.ejor.2008.10.025)
- [22] Kim, S.-H.; Nelson, B. L. (2001). A fully sequential procedure for indifference-zone selection in simulation, *ACM Transactions on Modeling and Computer Simulation*, Vol. 11, No. 3, 251-273, doi:[10.1145/502109.502111](https://doi.org/10.1145/502109.502111)
- [23] Kelton, W. D.; Smith, J. S.; Sturrock, D. T.; Verbraeck, A. (2011). *Simio and Simulation: Modeling, Analysis, Applications*, 2nd edition, McGraw-Hill, New York
- [24] Fan, W.; Hong, L. J.; Nelson, B. L. (2016). Indifference-zone-free selection of the best, *Operations Research*, Vol. 64, No. 6, 1499-1514, doi:[10.1287/opre.2016.1530](https://doi.org/10.1287/opre.2016.1530)