

Research Article

Emergency Department Staffing: A Separated Continuous Linear Programming Approach

Xiaoqing Wang

Lingnan (University) College, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

Correspondence should be addressed to Xiaoqing Wang; xqwang1@gmail.com

Received 2 August 2013; Revised 12 September 2013; Accepted 12 September 2013

Academic Editor: Pui-Sze Chow

Copyright © 2013 Xiaoqing Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The emergency department (ED) overcrowding is becoming serious in recent years. The shortage of ED staff is considered to be a common factor to contribute to the ED overcrowding. In this paper, we present a new model to address the ED staffing from a macroscopic perspective. The model is a kind of optimization model called separated continuous linear programming (SCLP). By using the efficient algorithm for SCLP, we can get the ED staffing level which minimizes the total cost of the patients and the ED.

1. Introduction

Emergency medical service is an important part of medical and health care systems. Its quality of service and availability directly affect the quality of health care. In recent years, the emergency department (ED) in many hospitals has experienced overcrowding [1, 2], and patients, health care workers, and hospitals in which the ED exists are adversely affected.

ED overcrowding leads to the delay in the assessment and treatment of patients [3], including the delay in hospitalization. Long residence time in ED adds additional pain, burden on patients and sometimes even leads to the unexpected death.

ED overcrowding makes health care workers overworked [4] and greatly influences their enthusiasm for work and efficiency in work, the chances of medical errors and medical disputes increasing. At the same time, ED overcrowding also affects the normal operation of the entire hospital in which the ED exists, leading to a longer hospitalization time for the emergency patients and lowering the capability of the hospital to meet the need for hospitalization of other nonurgent patients.

Many problems lead to ED overcrowding; one of them is the shortage of physicians to meet the fast-growing demand for ED. There are several reasons for the shortage of physicians; one of them is the high operational cost of ED. Physician salaries represent a large part of the operational cost in

ED and there is not enough motivation for the hospital to improve their service by increasing the amount of physicians in ED in the past. The increasing degree of ED overcrowding in recent years introduces so many troubles to the hospitals and the management authority in hospitals found that they have no other choice but trying to find effective methods to alleviate ED overcrowding. Obviously, improving the quality of ED staffing decision is one approach. The management authority wants to know the least amount of physicians ED needs to provide improved quality of service, so that the degree of ED overcrowding can be alleviated to an acceptable level.

In recent years, research efforts are made to improve the ED staffing. This paper provides a new approach to balance the waiting cost of the patients and the operational cost of the EDs. The objective is to find an appropriate amount of physicians and their scheduling which can meet the ED demand with the least total cost of the patients and the EDs. The advantage of our approach is that the time needed to find the ED staffing level is shorter compared with the existing methodologies. Also, we do not make strong assumptions on the distribution of the arrival rate and the distribution of the service time of the demand to ED.

The paper is organized as follows. In Section 2, we provide a literature review on ED staffing; we also give a brief introduction to separated continuous linear programming (SCLP) on which our approach is based. In Section 3, we

describe the emergency service process that we focus on. In Section 4 we construct an SCLP model to minimize the total cost of patients and the ED given a specific ED staffing level. We then solve the SCLP model with varying staffing level in ED and finally find the appropriate staffing level which results in the least total cost of patients and the ED among the possible staffing levels. In Section 5, we summarize what we get and point out some future research directions.

2. The Literature Review

2.1. Literature on ED Staffing. There are two main research methodologies to tackle the staffing in ED. One is simulation and the other is the stochastic models based on the queueing theory.

The most commonly used research method for ED staffing is simulation [5]. The reason is that the ED of a modern hospital is a highly complex system and it is difficult to use the analytic model to describe the dynamic ED reality accurately. Simulation offers a natural framework to address this issue within. Simulation models often possess high validity because they track true system behavior fairly accurately, but they yield fewer analytic inference than that of analytic models because they take both input parameters and decisions as fixed. A lot of experiments need to be performed before an approximate optimal solution can be found, and the gap between the approximation optimal solution and the true optimal solution is unknown. Also, the approximate optimal solution obtained is just a numerical solution and we have no knowledge on the interactions among ED staffing and other managerial decisions, thus making it difficult to pinpoint the relationship between the staffing level and the quality of service in ED.

On the other hand, although analytic models can never capture all characteristics of an actual ED process, stochastic models, based on queueing theory, are more appropriate for capturing the volatile and inherently nondeterministic ED reality, and they can be invaluable in providing managerial insight that greatly improves performance of the ED process [6]. Although the operation in ED is very complex, many researchers try to focus on the most important parts of the ED process and build simpler stochastic queueing models to study the ED staffing. These models strive to seek a balance between the tractability and validity and get some valuable results [5, 7–9].

There are three important characteristics of the demand of EDs. Firstly, the arrival rate of the patients has significant variation over the time-of-the-day, day-of-the-week, and month-of-the year, that is, time-varying arrival rate. Secondly, the patients are prioritized into emergent, urgent, and nonurgent categories by use of triage upon arrival, that is, multiclass customers. Thirdly, the patients usually visit the physician several times during their sojourn time in the ED before they are discharged or hospitalized, that is, customer feedback.

Zeltyn et al. [5] use simulation-based models to address the ED staffing with time-varying demand. They incorporate the offered-load technique and classical square-root safety-staffing principle which are based on the M/M/s queueing

model. The staffing recommendations they provided are implemented by a large Israeli hospital and get the satisfactory results.

Green et al. [8] use the M/M/s queueing model as part of a Lag stationary independent period by period (SIPP) approach to determine how to vary staffing level to meet changing demand in an urban hospital ED in the United States. Their staffing recommendations have been implemented in that ED and as a result the proportion of patients who left without being seen (LWBS) decreases significantly.

Recently, Yom-Tov and Mandelbaum [9] analyze a queueing model that they call Erlang-R, where the “R” stands for ReEntrant customers, to accommodate the phenomenon that the patients return to physician several times during their sojourn within the ED. They propose time-varying square-root staffing policy which is based on the modified-offered-load and Erlang-R model. They use simulations to verify that their staffing recommendations work well in the ED settings.

One of the problems for the queueing model approach is that some special distributions for patient arrival and service process, such as the Poisson arrival and exponential service time, and some working discipline, such as First-Come-First-Serve (FCFS), need to be assumed in order to get the explicit closed-form expression for some quality measures (e.g., the fraction of patients who LWBS in an ED) on the ED service. But these assumptions do not always agree with reality.

The purpose of the present paper is to introduce models for dynamic ED staffing without explicitly calculating the quantities of interest like patients’ waiting time, the fraction of patients who LWBS, and so forth, and rather developing models directed towards using efficient algorithms to handle these quantities. We do this by looking at the ED system from a macroscopic perspective and use fluid network to approximate the overloaded ED system. From our knowledge, there is no other work tackling the ED staffing by using this approach.

2.2. A Brief Introduction to SCLP. Linear programming (LP) is probably the most successful mathematical model in terms of its extremely wide range of industrial applications and its superb speed and capacity in solving very large size problems. For this reason, LP has been pushed and extended to an ever-broadening frontier. One of such frontier extensions is the so-called *separated continuous linear programming* (SCLP) which was first introduced by Anderson [10] who used it to model the job-shop scheduling problem. The following is the SCLP due to Weiss [11]:

$$\begin{aligned}
 \text{(SCLP)} \quad & \max \int_0^T [(\gamma + (T-t)c)'u(t) + d'x(t)] dt \\
 \text{s.t.} \quad & \int_0^t Gu(s) ds + Fx(t) \leq \alpha + ta, \\
 & Hu(t) \leq b, \\
 & u(t) \geq 0, \quad x(t) \geq 0, \quad t \in [0, T],
 \end{aligned} \tag{1}$$

where $u(t)$, $x(t)$ are decision variables and are assumed to be bounded measurable functions with the measure of the

break-point set being 0. $\gamma, c, d, \alpha, a, b$ are vectors, G, F, H are matrices. $'$ denotes the transpose operation. The word “separated” refers to the fact that there are two kinds of constraints in SCLP: the constraints involving integration and the instantaneous constraints [10].

SCLP has in recent years attracted considerable research attention in the field of stochastic networks, a part of queueing theory. The multiclass stochastic network is a system consists of different classes (types) of jobs which need to be processed and a set of servers which process the jobs. Jobs arrive to the system randomly. Each server can process one or more classes of jobs and the processing time for every job is different for different class of jobs. The jobs in the same class have the same characteristics such as arrival rate and service requirements. After one job is processed in one server, it may leave the network instantaneously or may become another class of job and go to another server for processing. The multiclass stochastic network is a very useful model for many real systems.

For each multiclass stochastic network, there is a corresponding deterministic fluid network, which takes only the first-order data (means and rates) from the stochastic model and assumes that the jobs circulating in the network are continuous flows instead of discrete units. With appropriate scaling, the fluid network is a limit of the stochastic network, in the sense of strong law of large numbers (refer to, e.g., [12]). Furthermore, the fluid model has played a central role in studying the stability of stochastic networks [13]. Because of these developments, the real-time control (dynamic scheduling) of a stochastic network, which is itself a quite intractable problem, can be turned into the control of a corresponding fluid network, and the latter problem takes exactly the form of SCLP. We will show in Section 4 that the problem of finding the dynamic control of the patient flow, subproblem of optimizing the ED staffing, can be formulated as an SCLP (see (20) in Section 4).

The dual of SCLP is the following problem:

$$\begin{aligned}
 (\text{SCLP}^*) \quad & \min \int_0^T [(\alpha + (T-t)a)'p(t) + b'q(t)] dt \\
 \text{s.t.} \quad & \int_0^t G'p(s) ds + H'q(t) - (\gamma + tc) \geq 0, \\
 & F'p(t) - d \geq 0, \\
 & p(t) \geq 0, \quad q(t) \geq 0, \quad t \in [0, T],
 \end{aligned} \tag{2}$$

where the dual variables, $p(t)$ and $q(t)$, are bounded measurable functions.

There exists the rich literature on duality theory and algorithms for solving SCLP. Interested reader may refer to [14] for the detailed review on the research results on SCLP. Specifically, Wang [15] suggested an algorithm which is a polynomial approximation scheme for solving SCLP. Because that algorithm is very efficient in terms of the computing time, we choose that algorithm to solve the model we propose in this paper. In the next Section 2.3, we will explain in detail how that algorithm works.

2.3. An Approximation Algorithm for SCLP. Before explaining the algorithm in [15], we first introduce some notations and conventions which we will use in the remaining part of this paper.

- (i) By default, all vectors are column vectors. One exception is when we denote the solutions to the SCLP as (u, x) , we mean $(u', x')'$.
- (ii) $\pi = \{t_0, \dots, t_m\}$ denotes a partition of $[0, T]$ into m segments:

$$0 = t_0 < t_1 < \dots < t_m = T. \tag{3}$$

We use π_m to denote π when all the interval lengths in π are the same.

- (iii) Given a partition $\pi = \{t_0, \dots, t_m\}$ and a vector $\hat{f} := (\hat{f}(t_0), \hat{f}(t_1), \dots, \hat{f}(t_m))$, where $\hat{f}(\cdot)$ is a right continuous function, the following (continuous) function

$$f(t) = \left(\frac{t_i - t}{t_i - t_{i-1}} \right) \hat{f}(t_{i-1}) + \left(\frac{t - t_{i-1}}{t_i - t_{i-1}} \right) \hat{f}(t_i), \tag{4}$$

$$\text{for } t \in [t_{i-1}, t_i], \quad i = 1, \dots, m$$

is called a piecewise linear extension of \hat{f} , whereas the following (right-continuous) function

$$f(t) = \begin{cases} \hat{f}(t_{i-1}), & t \in [t_{i-1}, t_i), \text{ for } i = 1, \dots, m \\ \hat{f}(t_{m-1}), & t = T, \end{cases} \tag{5}$$

is called a piecewise constant extension of \hat{f} .

- (iv) For problem (P) , we use $v(P)$ to denote the optimal objective value of (P) .

We start with introducing the following discretization of (SCLP) based on a partition of $[0, T]$, $\pi = \{t_0, \dots, t_m\}$:

$$\begin{aligned}
 (\text{LP}_1(\pi)) \quad & \max \sum_{i=1}^m \left(\left(\gamma + \left(T - \frac{t_i + t_{i-1}}{2} \right) c \right)' \hat{u}(t_{i-1}) \right. \\
 & \left. + d' [\hat{x}(t_i) + \hat{x}(t_{i-1})] \frac{t_i - t_{i-1}}{2} \right) \\
 \text{s.t.} \quad & \alpha + t_i a - [G\hat{u}(t_0) + \dots + G\hat{u}(t_{i-1}) \\
 & \quad + F\hat{x}(t_i)] \geq 0, \\
 & i = 1, 2, \dots, m; \\
 & (t_i - t_{i-1})b - H\hat{u}(t_{i-1}) \geq 0, \\
 & i = 1, \dots, m; \\
 & \hat{u}(t_{i-1}) \geq 0, \quad \hat{x}(t_i) \geq 0, \\
 & i = 1, \dots, m.
 \end{aligned} \tag{6}$$

In addition, assume that

$$\alpha - F\hat{x}(t_0) \geq 0, \quad \hat{x}(t_0) \geq 0. \tag{7}$$

The following Lemma and the algorithm presented later in this subsection are adapted from [15].

Lemma 1. Suppose that (\hat{u}, \hat{x}) is a feasible solution to $(LP_1(\pi))$, with

$$\hat{u} = (\hat{u}(t_0), \dots, \hat{u}(t_{m-1})), \quad \hat{x} = (\hat{x}(t_1), \dots, \hat{x}(t_m)). \quad (8)$$

Let $u(t)$ be the piecewise constant extension of $(\hat{u}(t_0)/(t_1 - t_0), \dots, \hat{u}(t_{m-1})/(t_m - t_{m-1}))$, and let $x(t)$ be the piecewise linear extension of $(\hat{x}(t_0), \hat{x})$. Then, $(u(t), x(t))$ is a feasible solution to (SCLP), and the objective value of (\hat{u}, \hat{x}) in (6) is the same as that of $(u(t), x(t))$ in (1).

The same discretization applies to (SCLP*), the dual problem. Specifically, corresponding to the partition $\pi = \{t_0, \dots, t_m\}$, we have

$$\begin{aligned} (LP_2(\pi)) \quad \min \quad & \sum_{i=1}^m \left(\left(\alpha + \left(T - \frac{t_i + t_{i-1}}{2} \right) a \right)' \hat{p}(t_{i-1}) \right. \\ & \left. + b' [\hat{q}(t_i) + \hat{q}(t_{i-1})] \frac{t_i - t_{i-1}}{2} \right) \\ \text{s.t.} \quad & G' \hat{p}(t_0) + G' \hat{p}(t_1) + \dots + G' \hat{p}(t_{i-1}) \\ & + H' \hat{q}(t_i) - (\gamma + t_i c) \geq 0, \\ & i = 1, \dots, m; \\ & F' \hat{p}(t_{i-1}) - (t_i - t_{i-1}) d \geq 0, \\ & i = 1, \dots, m; \\ & \hat{p}(t_{i-1}) \geq 0, \quad \hat{q}(t_i) \geq 0, \\ & i = 1, \dots, m. \end{aligned} \quad (9)$$

In addition, assume that

$$H' \hat{q}(t_0) - \gamma \geq 0, \quad \hat{q}(t_0) \geq 0. \quad (10)$$

We need the following assumption.

Assumption 2. The following two linear programming problems have optimal solutions:

$$\begin{aligned} (LP_1) \quad \max \quad & c'u + d'x \\ \text{s.t.} \quad & \alpha + Ta - Gu - Fx \geq 0, \\ & Tb - Hu \geq 0 \\ & u \geq 0, \quad x \geq 0, \\ (LP_2) \quad \min \quad & a'p + b'q \\ \text{s.t.} \quad & G'p + H'q - (\gamma + Tc) \geq 0, \\ & F'p - Td \geq 0, \\ & p \geq 0, \quad q \geq 0. \end{aligned} \quad (11)$$

Algorithm 3. Let δ be the predefined precision between the objective value of the solution requested and $v(\text{SCLP})$.

Step 1. Check if Assumption 2 is satisfied. If yes, go to Step 2; otherwise, stop.

Step 2. Solving $(T/2m)(v(LP_1) - v(LP_2) + b'\hat{q}(t_0) - d'\hat{x}(t_0)) \leq \delta$ to get m (number of intervals in partition π_m), find the optimal solution (\hat{u}^*, \hat{x}^*) for $(LP_1(\pi_m))$. Use the extension of this solution to construct a feasible solution for (SCLP). Stop.

Note that the method to construct a feasible solution for (SCLP) from the extension of a feasible solution for $(LP_1(\pi_m))$ is mentioned in Lemma 1.

To get m , we need to first solve the two linear programs (LP_1) and (LP_2) . In addition, we also need to determine $\hat{x}(t_0)$ and $\hat{q}(t_0)$. This can be accomplished by solving the following two linear programs:

$$\begin{aligned} \max \quad & d' \hat{x}(t_0) \\ \text{s.t.} \quad & \alpha - F \hat{x}(t_0) \geq 0, \\ & \hat{x}(t_0) \geq 0, \\ \min \quad & b' \hat{q}(t_0) \\ \text{s.t.} \quad & H' \hat{q}(t_0) - \gamma \geq 0, \\ & \hat{q}(t_0) \geq 0. \end{aligned} \quad (12)$$

After getting m , we solve the linear program $(LP_1(\pi_m))$ to get a feasible solution for (SCLP) which satisfies the predefined precision δ .

In summary, the previous Algorithm 3 amounts to solving five linear programming problems: (LP_1) , (LP_2) , (12), and $(LP_1(\pi_m))$. Linear programs are known to be polynomially solvable. Hence, this algorithm is a polynomial-time algorithm.

3. The Emergency Treatment Process

The treatment process in ED constitutes a multiclass stochastic queueing network with feedback. In this network, patients arrive at the ED randomly. The critically ill patients get the medical service immediately after arrival. The other patients are assigned different priority to see the physician for the first time after triage. The service (treatment) time for each patient is random, and there are several possible outcomes after the patient sees the physician, including revisiting the physician after examination and/or testing, going to the emergency observation room (ER) within the ED (under the care of another group of physicians and nurses) or leaving ED (going home with medicine, moving to ward or dead). In the last two cases, we say the patient *leaves the system*. Figure 1 illustrates a simplified version of such an emergency treatment process in a typical ED which we will focus on in this paper.

In Figure 1, Server 1 represents physicians for treatment (hereafter, doctors); Server 2 represents the physicians for examination and testing (hereafter, testing staff). The patients arriving at the ED from outside are classified into n classes after triage based on patients' health condition. The patients in the same class are assigned the same priority which is

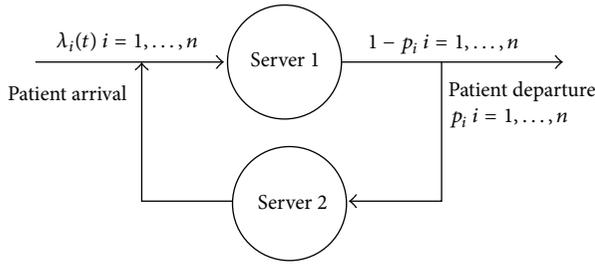


FIGURE 1: The emergency treatment process in a typical ED.

represented by the time deadline to see the doctor for the first time [16]; for example, class 2 patients are told after triage that they will see the doctor within 30 minutes; “30 minutes” here is the time deadline for class 2 patients and is the measure for the quality of service in a specific ED. Patients within each class are served on an FCFS basis. The critically ill patients will not queue in the network, the effect they impose on the network is occupying some critical resources, such as the physicians and the beds, during some time intervals.

The patients queuing before Server 1 and waiting for their first time to see the doctor are said in their *first stage*. After seeing the doctor, the patients either leave the system or enter their *second stage*—queuing before Server 2 and waiting for examination and/or testing. Upon getting their examination and/or testing results, the patients enter their *third stage*—queuing before Server 1 and waiting to see the doctor again. The patients leave the system (discharged or hospitalized) after they see the doctor for the second time.

The physicians and nurses who are in charge of the service within the ER are usually another group of staff. Their staffing is not the subject of this paper and we do not include the service process within the ER in Figure 1.

4. Mathematical Model for ED Staffing

We consider the ED staffing within a time interval $[0, T]$. The criterion for the appropriate ED staffing level is that the total cost for patients and the ED within the time interval $[0, T]$ is minimized. This criterion includes the quality of service requirement in ED by explicitly considering the waiting cost of patients. By balancing the cost of patients and that of the ED, we want to find an ED staffing level to cope with the time-varying demand for the ED service, which can balance the quality of service requirement and the utilization of the ED staff.

We address the ED staffing problem by two steps. In the first step, we construct the queuing model to determine the appropriate scheduling policy for patients to minimize the waiting cost of patients given a specific ED staffing level. The queuing model is an SCLP and we can solve it efficiently by using the algorithm described in Section 2.3. In the second step, by varying the ED staffing level in the SCLP, we can find the least staffing level for the ED which minimize the total cost of patients and the ED for the time interval $[0, T]$.

In building the analytic model of the ED system, one needs to decide how detailed does the description of the system have to be? The more microscopic the model is, the

more accuracy is achieved, but on the other hand, usually the harder it is to handle analytically [17]. In this paper, we choose to view the ED system in a more macroscopic way. We use the deterministic fluid model to describe the performance of the overloaded ED systems. The use of the fluid model in overloaded system is justified in a number of papers, for example, in Whitt [18, 19]. The patients who left without being seen are ignored, and the ED systems try hard to accommodate the patients who wait until seeing the physicians.

In the ED process that we focus on, the arrival rate of the patients and the service time of the patients are random variables; instead of assuming that the arrival rate or the service time obeys some particular distributions (as those in the previous papers [5, 7–9]), we only take the first moments (means of the arrival rate and the service time) to describe these randomness.

Specifically, the corresponding fluid network for the ED treatment process in Figure 1 is as follows: the patients in the ED are regarded as continuous fluid instead of individual person. The arrival rate and the service time of the fluid in each server are deterministic. After visiting Server 1, the fluid proceeds to Server 2 with deterministic percentage (probability). The fluid which do not proceed to Server 2 exits the network. After visiting Server 2, the fluid continues to visit Server 1 again. The fluid exits the network after visiting Server 1 in the second time.

Let $\lambda_i(t)$ denote the arrival rate of class i patients from outside at time t . The mean service time for class i patients in their stage 1 is η_i . After receiving service from Server 1, the class i patients will enter their second stage and queue before Server 2 with probability p_i , or leave the system with probability $1 - p_i$. The mean service time for class i patients in their second stage is γ_i . After receiving service from Server 2, the class i patients enter their third stage and queue before Server 1 again, and the mean service time for them is μ_i .

The service capacity of each doctor or testing staff is 1. Let s_j denote the number of doctors or testing staff in Server j and let $b_j(t)$ denote the service capability of Server j at time t for patients in the queues. If the critically ill patients show during $[0, T]$, they will occupy some capacity of Server j and subsequently $b_j(t)$ becomes $k_j s_j$, $k_j \in [0, 1]$, where the value of k_j depends on the condition of the critically ill patients. If no critically ill patients show during $[0, T]$, $b_j(t)$ is simply s_j .

Let ω_{ij} denote the waiting cost of class i patients in their j th stage (including the cost for the waiting time and the cost of patients’ health condition changes due to the long waiting time, e.g., the additional financial cost). ω_{i1} should be big enough to reflect the cost for different class patients in order to guarantee that each patient in his first stage can get service within the time deadline the ED promises. Usually a senior physician in an ED can tell the approximate waiting cost for a specific class of patients in their specific stage and we can use this piece of information to determine the value of ω_{ij} ’s. Let c_j denote the service cost for each doctor or each testing staff in Server j . The estimation of the ED’s operational cost is relatively straightforward because the salaries paid to the physicians constitute most of the ED operational cost. So we can use the physicians’ salaries to represent the service cost of an ED.

staffing level which minimizes the total cost of patients and the ED.

Our approach accommodates the three characteristics of the demand in EDs, that is, time-varying arrival rate, multi-classes customer, and customer feedback. Comparing with other existing approaches on ED staffing, our approach is more efficient to get the appropriate staffing level.

There are still some interesting problems that need explore. In this paper, we assume that the waiting costs of the patients are the linear functions of the queue lengths before Servers. How to determine the amount of ED physicians if the cost functions are other type of functions, for example, convex and increasing function of the queue lengths, is still unknown.

Also, in this paper, we only consider how to determine the amount of ED physicians who are in charge of the patients arrived from outside. In reality, there is another decision to make; that is, how to determine the amount of physicians who are in charge of the patients in ER. Although this issue is usually translated into determining the appropriate amount of beds in ER in most of the literature, directly handling this issue is still meaningful if there is some cooperation between these physicians and the physicians who are in charge of patients arrived from outside.

References

- [1] R. W. Derlet, J. R. Richards, and R. L. Kravitz, "Frequent overcrowding in U.S. emergency departments," *Academic Emergency Medicine*, vol. 8, no. 2, pp. 151–155, 2001.
- [2] R. W. Derlet and J. R. Richards, "Overcrowding in the nation's emergency departments: complex causes and disturbing effects," *Annals of Emergency Medicine*, vol. 35, no. 1, pp. 63–68, 2000.
- [3] J. Kennedy, K. Rhodes, C. A. Walls, and B. R. Asplin, "Access to emergency care: restricted by long waiting times and cost and coverage concerns," *Annals of Emergency Medicine*, vol. 43, no. 5, pp. 567–573, 2004.
- [4] L. G. Graff, S. Wolf, R. Dinwoodie, D. Buono, and D. Mucci, "Emergency physician workload: a time study," *Annals of Emergency Medicine*, vol. 22, no. 7, pp. 1156–1163, 1993.
- [5] S. Zelty, Y. N. Marmor, A. Mandelbaum et al., "Simulation-based models of emergency departments: operational, tactical, and strategic staffing," *ACM Transactions on Modeling and Computer Simulation*, vol. 21, no. 4, article 24, 2011.
- [6] R. W. Hall, *Queueing Methods for Service and Manufacturing*, Prentice Hall, Upper Saddle River, NJ, USA, 1990.
- [7] L. V. Green, P. J. Kolesar, and J. Soares, "Improving the sipp approach for staffing service systems that have cyclic demands," *Operations Research*, vol. 49, no. 4, pp. 549–564, 2001.
- [8] L. V. Green, J. Soares, J. F. Giglio, and R. A. Green, "Using queueing theory to increase the effectiveness of emergency department provider staffing," *Academic Emergency Medicine*, vol. 13, no. 1, pp. 61–68, 2006.
- [9] G. B. Yom-Tov and A. Mandelbaum, "Erlang-R: a time-varying queue with ReEntrant customers, in support of healthcare staffing," Working Paper, Technion-Israel Institute of Technology, 2011.
- [10] E. J. Anderson, *A Continuous Model for Job-Shop Scheduling [Ph.D. thesis]*, University of Cambridge, Cambridge, UK, 1978.
- [11] G. Weiss, "A simplex based algorithm to solve separated continuous linear programs," *Mathematical Programming*, vol. 115, no. 1, pp. 151–198, 2008.
- [12] H. Chen and D. D. Yao, *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization, Applications of Mathematics*, vol. 46, Springer, New York, NY, USA, 2001.
- [13] J. G. Dai, "On positive harris recurrence of multi-class queueing networks: a unified approach via fluid limit models," *Annals of Applied Probability*, vol. 5, no. 1, pp. 49–77, 1995.
- [14] X. Wang, "Review on the research for separated continuous linear programming: with applications on service operation," *Mathematical Problems in Engineering*, vol. 2013, Article ID 408048, 9 pages, 2013.
- [15] X. Wang, *Theory and algorithms for separated continuous linear programming and its extensions [Ph.D. thesis]*, The Chinese University of Hong Kong, Hong Kong, China, 2005.
- [16] J. Huang, B. Carmeli, and A. Mandelbaum, "Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback," Submitted to *Operations Research*.
- [17] A. Segall, "The modeling of adaptive routing in data-communication networks," *IEEE Transactions on Communications*, vol. 25, no. 1, pp. 85–95, 1977.
- [18] W. Whitt, "Efficiency-driven heavy-traffic approximations for many-server queues with abandonments," *Management Science*, vol. 50, no. 10, pp. 1449–1461, 2004.
- [19] W. Whitt, "Fluid models for multiserver queues with abandonments," *Operations Research*, vol. 54, no. 1, pp. 37–54, 2006.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

