Contents lists available at ScienceDirect

Operations Research for Health Care

journal homepage: www.elsevier.com/locate/orhc

Assigning treatment rooms at the Emergency Department

Maartje van de Vrugt [a,b,c,*], Richard J. Boucherie [a,b]

[a] *Centre for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Enschede, The Netherlands*
[b] *Stochastic Operations Research, Department of Applied Mathematics, University of Twente, Enschede, The Netherlands*
[c] *Department of Quality and Safety, Jeroen Bosch Hospital, 's Hertogenbosch, The Netherlands*

ARTICLE INFO

Article history:
Received 2 November 2014
Accepted 29 June 2015
Available online xxxx

Keywords:
Queueing theory
Emergency department
Discrete event simulation

ABSTRACT

Increasing efficiency at the Emergency Department (ED) reduces overcrowding. At the ED in typical Dutch Hospitals treatment rooms are mostly shared by two residents of different specialties: a Surgeon and an Internist. Each resident uses multiple rooms in parallel; while one patient awaits test results in a treatment room, the resident visits other patients. The assignment of rooms among the residents is often unbalanced, which affects the blocking probability and waiting and sojourn times of patients. Invoking a queueing model in a random environment, we analytically investigate expected sojourn times of (semi-urgent) patients for both types of residents for different room assignment policies and working routines of the residents. We determine the Pareto efficient policies and working routines for all performance measures. We conduct a Discrete Event Simulation to validate our model and present numerical results for a large Dutch teaching hospital and other illustrative cases.

1. Introduction

Emergency department (ED) overcrowding has many negative effects, for both patients and ED staff. ED overcrowding leads for example to an increase in the number of medical errors, long waiting times and high levels of stress for ED staff (see e.g. [1–3]). Therefore, governments have set legal norms on the waiting and sojourn times of patients visiting the ED. Meeting these norms is often difficult due to an increasing demand for acute care and the closure of many EDs and hospital beds [4].

The waiting and sojourn times of ED patients are influenced by many factors, among others the patient's urgency class and type. In the Netherlands, patients are both triaged and assigned to a specialty upon arrival at the ED. The triage category of a patient determines the order in which patients are treated. Treatment can start when there is a treatment room and the patient's physician is available. The specialties with the highest number of patients are represented at the ED at all times, either by a specialist or a resident (together with physicians collectively called 'doctors' in this paper). Other specialties have to be called for assistance, implying that patients at least have to wait for the travel time of the doctor.

Typically at an ED, one doctor occupies multiple rooms in parallel; when a patient requires diagnostic tests, the doctor visits other patients. Each specialty/doctor has a different working routine: Emergency Physicians are for example specially trained for only providing urgent care, while other doctors tend to provide care more like they are used to outside the ED [4]. As a consequence, sojourn and waiting times of different patient types differ significantly, and the room utilization can be unbalanced over the doctor types. The unbalanced room utilization possibly also affects the sojourn and waiting times of the patients.

Besides the sojourn and waiting times, the probability that the ED is 'full' (*blocking probability*) is an important performance measure. When the number of patients at the ED exceeds a certain level, the ED staff will call in an extra doctor and/or defer new patients. The first will result in additional costs, while the latter negatively affects the image of the hospital.

In this paper we investigate several room assignment policies and doctor working routines at an ED and determine the Pareto-efficient combinations with respect to the performance measures mentioned above. To this end, we invoke a continuous time queueing model in which patients may require diagnostic tests, implying that doctors visit their patients a random number of times and interact in sharing the available treatment rooms. We investigate a case study of a large Dutch teaching hospital (the Jeroen Bosch Hospital, JBH) and several additional illustrative cases. By means of a Discrete Event Simulation, we investigate how the obtained policies perform in a more realistic setting.

* Correspondence to: Centre for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Postbox 217, 7500 AE, Enschede, The Netherlands.
E-mail address: n.m.vandevrugt@utwente.nl (M. van de Vrugt).

http://dx.doi.org/10.1016/j.orhc.2015.06.004
2211-6923/

We also study the sojourn time of an accepted patient conditioned on the state found in the ED upon arrival. This conditional waiting and sojourn times information may be provided to the patients to inform them about their expected length of stay at the ED, which adds to patient satisfaction [5].

In the next section we will provide an overview of related literature. The model is described in Section 3. Section 4 provides the analytical results, Section 5 the numerical results, and Section 6 the conclusion and discussion.

2. Related literature

The literature that is related to this paper consists of papers on EDs and of queueing models that are similar to the one considered in this paper. Both types of literature will be summarized below.

EDs are modeled most often by means of queueing theory and computer simulation. Typical focus areas of these works are capacity dimensioning and patient routing, with the objectives to minimize patient waiting time, maximize patient throughput or increase staff utilization [6]. In patient routing the effects of for example parallelization of tasks, a 'fast-track' system, and letting patients wait outside the treatment room are studied.

There are many papers that apply queueing models to an ED, but only a few papers are closely related to this paper as they consider patients who are visited more than once by the doctor, cf. [7–9]. These studies consider a multi-server queueing network in which customers can re-enter a queue, and the authors aim to determine appropriate staffing levels to achieve certain performance targets. Differences between the models in these references (and the references therein) and the model considered here, are that patients do not return to their own doctor and that it is assumed that treatment can start once the doctor is available. This implies that either there always is a sufficient number of treatment rooms, or patients await their test results outside the room.

In a recent review, Wiler et al. [3] state that most queueing approaches to modeling an ED assumed that: (1) arrival rates are constant, (2) patients do not deteriorate, and (3) patients are not seen by same doctor again. Here we also assume (1) and (2), but allow for patients to be seen by the same doctor more than once. For more literature on modeling EDs, the reader is referred to recent literature reviews [2,3,6] and the references therein.

A different but related topic is the optimal assignment of servers to two types of calls in a callcenter, studied by e.g. Bhulai and Koole [10]. Here the question is when to assign servers to outbound calls in such a way that inbound calls still achieve high service levels. Differences with our work are that in [10] each call visits the server only once, and each server has its own line (opposed to doctors sharing multiple treatment rooms).

Our contribution to both fields of literature is that we incorporate the ordering of diagnostic tests and doctors revisiting their patients in a queueing model for an ED. We consider a model with two resource types (doctors and rooms) in which a customer can only get service when both resource types are available at the same time. To the best of our knowledge, this type of model has not been analyzed before. This model is applied to an ED, but could easily be adapted to, for example, an outpatient clinic in which multiple physicians share the available treatment rooms.

The method used in this paper has been applied before to telecommunication systems, see for the most related papers [11–14]. These systems typically have two call types: inelastic (speech) calls, and elastic (data) calls. The inelastic calls require one unit of capacity (bandwidth) throughout their entire service time. The elastic calls share the remaining capacity in a processor sharing fashion. Important performance measures are the (average) throughput, sojourn time and blocking probability. At an ED, multiple types of patients also have to share the available capacity (treatment rooms). Although the processor sharing service discipline is not applicable for an ED, it will be shown that the methods applied to telecommunication systems can still be applied.

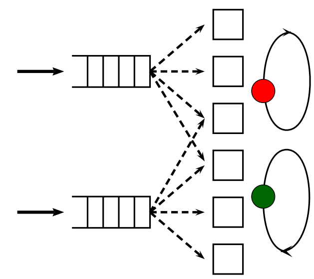

Fig. 1. Schematic representation of the queueing network.

3. The model

In this paper, we focus on a system with two doctors, since in many Dutch hospitals such as the JBH most of the non-urgent patients require a Surgeon or an Internist and both specialties are represented at the ED at all times. Patients that require other specialties are either treated by an Emergency Physician (working at the ED 7:00–23:00 h) or by a specialist doctor who is called for assistance.

Upon arrival at the ED, patients are triaged, assigned to a doctor, and join a waiting area. We only consider patients that are triaged as non-urgent, and are therefore treated on a first come, first serve (FCFS) basis. The triage is typically not the bottleneck at an ED, and therefore we only focus on the process after triage.

To model this ED in which two doctors share multiple treatment rooms to treat two types of patients with equal priority and each type of patient has its own queue and doctor, we consider a queueing system as depicted in Fig. 1. There are more treatment rooms than doctors, depicted by squares and circles respectively. The number of patients in the system is finite for both patient types and patients are assumed to arrive at their assigned queue according to a random process with constant arrival rate (we numerically investigate this assumption in Section 5). When the required doctor is inactive upon arrival of a patient and a treatment room is available, the treatment commences immediately. Otherwise, the patient is either blocked (if the maximum number of patients of this type is reached) or placed in a waiting area.

The treatment of a patient consists of at least one visit by the doctor; between two visits of the doctor, patients take diagnostic tests. At the instant the treatment of a patient begins, he is assigned a treatment room. This room will be assigned to this patient during the entire treatment time, even if the patient is not always physically in the room, which is the common policy at Dutch hospitals. We call each time a doctor visits a patient a 'phase' and the 'consultation time' of a patient denotes the total time a doctor has to spend with this patient in a room (so the sum of all phases). The treatment time thus includes the consultation time and possibly time between two doctor visits and time for taking diagnostic tests. To enhance tractability, we assume that both types of doctors visit their patients in random order. The sojourn time equals the waiting time before a room is assigned plus the treatment time.

During each visit the doctor can decide that the patient requires (further) diagnostic tests. Therefore, after each phase completion a patient stays in the system with a certain probability. When a patient's phase is completed but this patient does not leave the ED, the doctor can consult either a patient from the waiting area ('new patient') or in another treatment room ('existing patient'). The first option requires an extra room. The decision which of the two options to choose, depends on the working routine of the doctor and will be varied. If at phase completion the patient cannot leave the system and there is no other patient in the system, the doctor will become 'inactive', e.g. perform administrative tasks, for a random time. If a new patient arrives during an inactive