Wait time announcements at hospital emergency departments

Marco Bijvank, Zhankun Sun

Wait time announcements at hospital emergency departments

Overcrowding in the emergency department (ED) is a worldwide problem impairing the ability of hospitals to offer emergency care within a reasonable time frame. Upon arrival at an ED, patients are triaged using a standardized scoring system to decide the severity of a patient. The Canadian Triage and Acuity Scale (CTAS) classifies patients into five levels. For patients who are assessed as requiring immediate or emergent care, there are generally no or very short waits for care. However, for patients with issues in line with CTAS levels 3, 4 and 5, there will typically be a wait to see a physician. More than 40% of 161 Canadian hospitals fell below the suggested three-hour wait time to get assessed by an ED physician (CBC News, 2014). A number of Canadian hospitals have started publishing live emergency department wait times online in an effort to provide patients with expectations on how long they will have to wait to be seen for non-urgent care after initial assessment by a triage nurse. Calgary was the first in Canada to launch such an online system in July 2011, which shows the expected wait times at the major hospitals in the city. The tool is now also available as application for mobile devices. Yip et al. (2012) conducted a survey among 1,211 patients to determine the proportion who accessed wait time information on the website. They concluded that only 10% of patients accessed the website, whereas 45% said that they would like to use such a website. The goal of this paper is to accurately predict the real-time ED wait times and to study the impact of different prediction techniques on patient flow and patient care.

The study of real-time wait times in the literature is mostly found in a call center setting. Ibrahim and Whitt (2009, 2011) explore the performance of different real-time delay estimators based on recent delay experience by customers. Most researchers study the performance impact of making delay announcements to arriving customers in a multi-server queue setting with customer abandonment (Guo and Zipkin, 2007; Armony et al., 2009; Jouini et al., 2009; Huang et al., 2014).

Jouini et al. (2011) is the first paper to consider delay announcements in a multiclass setting with priorities. The customer behavior to delay announcements in call centers is empirically studied by Yu et al. (2014), whereas Batt and Terwiesch (2015) study the abandonment in an ED based on visual queue information.

There are a number of models developed in the literature that are closely related to our work. Yom-Tov and Mandelbaum (2014) study staffing for an emergency ward where patients are treated by a physician, then they may have to receive treatment by a nurse or undergo tests/scans after which the physician needs to reassess the patient again. They called such a model with reentrant patients the Erlang-R model. Zayas-Caban et al. (2014) model the triage and treatment operation at an ED as a multi-server two-stage tandem queueing system for which the authors study the service discipline which station to serve next when patients can abandon between the two stages. The physician's decision which patient to treat or task to perform next is also studied by Dobson et al. (2013) when there are interruptions that are indirect patient care tasks.

Even though these models allow for a more realistic modeling of the treatment process in the ED, there are two important characteristics not included. First, there is no priority of patients based on severity. Yom-Tov and Mandelbaum (2014) use a first-come-first-service (FCFS) disciple for all patients, whereas Dobson et al. (2013) and Zayas-Caban et al. (2014) study different service policies for the patients that require service in the second stage but not for the first stage. We consider patients of two priority classes to arrive to the system, namely patients that request emergent care (CTAS level 1 and 2) or non-emergent care (CTAS level 3 to 5). Class-1 patients have service priority over class-2 patients. Second, patients usually receive treatment from the same physician that gave them their first treatment. In the abovementioned papers, patients can be treated by any of the physicians. We consider a system closer related to Campello et al. (2013), where patients are seen as cases that require multiple rounds of service by a physician (or case manager as they call it) with a random delay in between.

The system we study is illustrated in Figure 1. Rather than analyzing the stationary behavior of the system, we focus on wait time predictions for low-priority patients to be seen by a physician for the first time. We develop a procedure to predict the state-dependent wait time based on an analysis of the busy period for class-*j* patients, where a busy period of class-*j* (*j=1,2*) is defined as the time interval during which the physicians continue to treat patients of class-*j*.

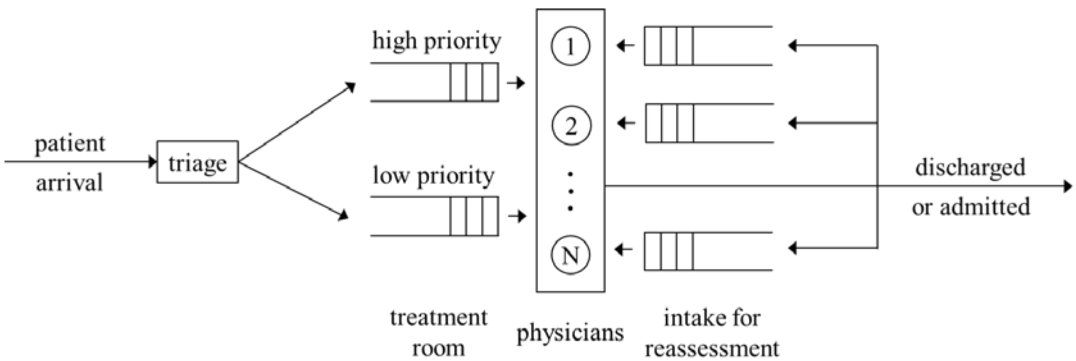

Figure 1: The ED queueing system.

We illustrate the performance of our wait time predictions with a case study at the four major hospitals in the Calgary area. We compare this with other predictors observed in call center settings (e.g., the wait time of the last patient who is seen for the first time) as well as the current prediction method (which divides the number of patients waiting for their first treatment by the speed of the physicians). More importantly, we use a simulation model to illustrate the impact of the online wait times on the patient flow within the Calgary area. It seems intuitive that real-time information would improve the flow, since patients and ambulances might avoid busier EDs if they knew there were prolonged wait times at that site. In practice the wait times can vary between 1 to 3 hours from hospital to hospital at the same time. We illustrate the robustness and impact of the predictors for different proportions of patients that consult the online tool and for different update frequencies of the tool (currently the wait times are updated every 2 minutes).