Journal of Operations Management 58–59 (2018) 15–26

Contents lists available at ScienceDirect

Journal of Operations Management

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

Managing responsiveness in the emergency department: Comparing dynamic priority queue with fast track

Yann B. Ferrand[a,*], Michael J. Magazine[b], Uday S. Rao[b], Todd F. Glass[c,1]

[a] Department of Management, Clemson University, 374E Sirrine Hall, Clemson, SC 29634, United States
[b] Department of Operations, Business Analytics, and Information Systems, University of Cincinnati, 501 Carl H. Lindner Hall, 2925 Campus Green Drive, Cincinnati, OH 45221, United States
[c] Cincinnati Children's Hospital Medical Center, Division of Emergency Medicine, 3333 Burnet Avenue, ML 16022, Cincinnati, OH 45229, United States

ARTICLE INFO

Accepted by: Tyson Browning

Keywords:
Service system design
Focused and flexible resources
Customer prioritization

ABSTRACT

Emergency Departments (EDs) commonly face capacity imbalances and long wait times in a service system handling patients with different priorities. These problems are particularly important for low-priority patients who often remain in the queue for extended periods. We investigate two distinct approaches to address these challenges: fast track (FT) and dynamic priority queue (DPQ). Traditionally, EDs have prioritized patients using an Emergency Severity Index (ESI), in conjunction with FT, to strictly or partially dedicate resources to different ESI patient classes. With our proposed DPQ, patients are prioritized using ESI and additional real-time operational information about the patient, specifically the amount of accumulated wait time and flow time. Using an empirical simulation, we compare the impact of different resource allocation and prioritization approaches on patient length of stay (LOS), including the existing system at the ED, FT with strict and partial dedication and the possibility of shorter and less variable service times, and versions of the proposed DPQ using simple dynamic prioritization. Our main results are that: (i) the DPQ approach dominates the other approaches tested; (ii) for various ED sizes, FT with strict and partial dedication do not reduce average LOS of low-priority patients without significantly increasing average LOS of high-priority patients, unless service time mean and variance are reduced; (iii) DPQ using accumulated wait time or accumulated flow time improves performance. The results are robust to changes in the proportion of patients in each priority level. Overall, expanding decision making about patient prioritization from only considering the patient's clinical condition to also including operational data can improve performance dramatically, even without improved service times.

1. Introduction

Emergency departments (EDs) are service systems handling customers with different priorities. Historically, there has been an increase in the annual patient visit rate to EDs (Tang et al., 2010) and a decrease in the number of hospital EDs (Nawar et al., 2007), resulting in overcrowding, and higher pressure to cost-effectively manage resources. In EDs, after safety, timeliness and efficiency are the two most critical criteria (Graff et al., 2002; Lindsay et al., 2002), yet EDs have shown increasing difficulty in providing timely care (Horwitz and Bradley, 2009). This paper studies approaches to providing timely care, particularly for lower priority (non-urgent and simple) patients that represent the preponderance of patients seen by typical EDs.

A common approach is to prioritize patients as they arrive using a triage protocol and the Emergency Severity Index (ESI, Gilboy et al., 2012). The ESI accounts for patient acuity, which is a combination of urgency and severity, and the anticipated amount of resources needed, which is a proxi for complexity of the patient's case. ESI level is determined at triage, and should not change once assigned. In general, patients are prioritized from most severe (ESI 1) to least severe (ESI 5). However, in this approach the low-priority ESI 4 and 5 patients can wait a long time. To alleviate this wait time, a commonly-used approach, referred to as Fast Track (FT), proactively allocates these low-priority patients some dedicated capacity (Gilboy et al., 2012).

The medical director of the Liberty Township Emergency Department (Liberty ED), a branch ED of Cincinnati Children's Hospital Medical Center (CCHMC), sought to investigate ways of improving patient length of stay (LOS) in this pediatric ED, without adding resources. A key goal of hospital administration was to *significantly reduce LOS for lower priority patients without increasing LOS for higher priority*

* Corresponding author.
E-mail addresses: yferran@clemson.edu (Y.B. Ferrand), mike.magazine@uc.edu (M.J. Magazine), uday.rao@uc.edu (U.S. Rao), todd.glass@nemours.org (T.F. Glass).
[1] Permanent address: Nemours Children's Hospital, Division of Emergency Medicine, 13535 Nemours Parkway, Orlando, FL 32827.

https://doi.org/10.1016/j.jom.2018.03.001
Received 24 July 2016; Received in revised form 28 March 2018; Accepted 29 March 2018
0272-6963/

patients. This is a pressing goal, since low priority patients wait the most, they are the most sensitive to resource availability, and they represent a large proportion of total patient visits. The initial motivation was to investigate modalities of setting up FT at the Liberty ED. This was later expanded to incorporate (1) improved service times (lower mean and variance), which could result from the process focus associated with the dedication of resources to FT (e.g., see Song et al., 2015); and (2) virtual streaming (with logically-segregated resources shared across prioritized patient streams), which could improve system performance when capacity is partitioned (Saghafian et al., 2012). However, the FT approach does not consider dynamic operational information such as the time spent waiting for service.

This ED context provides us with a realistic setting, which we use to contrast the a priori resource allocation FT approach with "Dynamic Priority Queue" (DPQ), a relatively new approach in healthcare Operations Management (OM) research that dynamically updates the sequence of patients waiting, based on the status of the system and specific information about individual patients. Several reasons motivated this approach. In preliminary testing, FT did not achieve the hospital administration's goal of reducing LOS for lower priority patients without increasing LOS for higher priority patients. Achieving this goal required taking some corrective action, one such action considered in this paper includes increasing a patient's priority when the patient had waited for a long time. The latter action is a DPQ approach in line with previous research in Queuing (Kleinrock, 1964; Stanford et al., 2014), in medical applications, (Halsted and Froehle, 2008), and in recent practice/OM research (e.g., Park et al., 2017, found that physicians apply a time-dependent prioritization when deciding which patient to see next). However, further research is needed to study potential DPQ approaches and determine if simple prioritization rules can improve patient LOS and how the wait time of different patient classes should be weighted.

We develop a simulation model to closely represent the Liberty ED and evaluate different configurations such as the existing configuration and FT. We study the effects of the FT resource allocation with and without improved service times and virtual streaming, and we contrast FT with specific DPQ approaches. To our knowledge, this is the first comparison that incorporates features of a real ED setting (including data, work shifts, and various multi-step care paths) and compares several different resource allocation and patient prioritization approaches including DPQ. We find that a simple DPQ approach can achieve a better balance of patient wait time across ESI priority levels than both the existing setting at Liberty ED (with ESI prioritization) and versions of FT, and the results are robust to changes in the proportion of patients in each priority level. This paper also highlights that the proposed DPQ is a general and powerful approach to managing patient wait times in that different dynamic prioritization approaches for implementing DPQ can be designed to mimic the behavior of various systems. For example, DPQ in the existing system uses a static ESI priority (with First-Come-First-Served within each ESI level), and DPQ in the FT system uses two prioritization schemes: one based only on ESI 4 and 5, for the resources allocated to low priority patients and another based only on ESI 1, 2, and 3, for the other resources. Thus, the DPQ concept is more general than prior work on the accumulating priority queue (APQ) approach (e.g. Stanford et al., 2014; Li et al., 2017), as DPQ can incorporate changes in arrival patterns, ESI level proportions, resource allocation, and availability of alternative resources. Also, DPQ can be implemented in a variety of ways, including special consideration of ED patients returning to see a specific physician after blood work or radiology.

This paper makes the following contributions to the OM literature:

1. We compare different resource allocation and patient prioritization approaches as a means to managing ED patient wait times including the existing system at the pediatric Liberty ED, Fast Track (FT), FT with improved service times, FT with virtual streaming, and versions of the proposed DPQ using simple dynamic prioritization. We illustrate that, conceptually, the DPQ approach dominates the other approaches.

2. We show that FT by itself is insufficient to achieve the goal of significantly reducing LOS for the lower priority patients without increasing LOS for the higher priority patients. In our experiments, FT results in better performance and achieves the afore-mentioned patient LOS goal if the physician's mean service time is reduced by 25% and standard deviation of service time is reduced by 50%. It is likely that this degree of improvement may be facilitated by the process focus that is afforded by FT (e.g., see Song et al., 2015), though a number of factors can mediate the relationship between process focus and performance improvement (Douglas and Fredendall, 2004; Ding, 2014; Goldstein and Iossifova, 2012; Dobrzykowski et al., 2016).

3. We find that simple approaches to implementing DPQ, for instance, using accumulated wait time or accumulated flow time with the appropriate ESI-based weight, dominates the other approaches tested, including FT with process improvement in the form of lower mean and variance of service time. Thus, expanding decision making about patient prioritization from only considering the patient's clinical condition to also including operational data can improve performance (i.e. patient wait times) dramatically, even without service time improvements. Further, in a multi-step process where the physician sees a patient more than once, the physician should consistently prioritize a patient at the head of the queue specifically waiting for him/her over a new patient, who can be seen by any physician.

For practitioners, this study presents an alternative solution to the challenging problem of effectively managing the wait time of low priority patients. More broadly, it represents a paradigm shift from the Fast Track-dominated landscape in the emergency medicine literature, where FT has thus far emerged as the natural way to handle patients that are prioritized using ESI levels. By reducing patient wait times, our findings have meaningful implications for improved patient access, reduced congestion for ancillary services and ambulance diversion, increased hospital revenues, and potential improvements in the patients' perception of quality and satisfaction ratings. These findings can be transposed to various service settings where customers have complex and varying service needs, and different priority levels.

2. Related literature

There is a large body of literature on ED operations and patient flow, spanning from patient entry into the ED, to patient flow within and out of the ED (cf. Saghafian et al. (2015) for a comprehensive review). Our work on patient prioritization fits in the category of patient flow within the ED, and the two pertinent streams of literature are server capacity partitioning for different customer classes, and dynamic prioritization of multi-class customers.

2.1. Capacity partitioning

To provide timely access and care delivery to low-priority patients in the ED setting, the traditional approach has been to dedicate resources by reassignment of rooms, supplies, and staff. In practice, this approach is often referred to as Fast Track (FT), and is a form of server capacity partitioning that has been extensively studied in the emergency medicine literature (e.g., Meislin et al., 1988; Simon et al., 1996; Cooke et al., 2002; O'Brien et al., 2006; Sanchez et al., 2006; Rodi et al., 2006; Williams, 2006; AHRQ, 2008), and in the operations management literature (e.g. Badri and Hollingsworth, 1993; Blake et al., 1996; Samaha et al., 2003; Mahapatra et al., 2003). Collectively, these studies have reported benefits of the FT approach in terms of reduced length of stay (LOS) for patients in the FT, shorter wait time to be seen by a physician, lower rate of patients leaving without being seen (LWBS), no changes of the rates of mortality or revisits, and improved patient satisfaction. Yet from an operations management perspective, key operational drivers of improvements have not been clearly identified. First, it is not always clear whether additional resources are brought on