Computers & Industrial Engineering 59 (2010) 378–386

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Developing nonlinear queuing regressions to increase emergency department patient safety: Approximating reneging with balking ☆

Jeffery K. Cochran [a], James R. Broyles [b,*]

[a] Air Force Institute of Technology, Department of Operational Sciences, 2950 Hobson Way Wright-Patterson AFB, OH 45433-7765, USA
[b] Department of Industrial, Systems, and Operations Engineering, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

Article history:
Received 24 March 2009
Received in revised form 22 March 2010
Accepted 12 May 2010
Available online 16 May 2010

Keywords:
Reneging
Balking
Queuing
Nonlinear regression
Emergency departments

ABSTRACT

Administrators know when Emergency Department (ED) overcrowding is a problem in their hospital. Lead times to change ED capacity are long and require strategic tools. ED patients who Leave WithOut Treatment (LWOT) before seeing a physician are, in queuing nomenclature, 'reneging' from an overcrowded situation and are an important measure of ED patient safety. We propose to enable strategic decision making on future ED capacity on the basis of patient safety (rather than congestion measures). We hypothesize that the LWOT reneging percentage is captured by the balking probability (p_K) relationship of an M/M/1/K queue. If true, this relationship is superior to the typical ad hoc regression relationships commonly found. Since it is based on a physical scientific mechanism, the sample size requirements and extrapolation power are improved. We derive the form of a binomial response nonlinear weighted regression model that best fits p_K for predicting LWOT to long-term ED performance by means of Gauss–Newton linearization. Our results include asymptotic Wald confidence intervals on prediction, specific Pearson and Deviance model goodness-of-fit tests, and residual analysis that facilitate identification of outlying data points. None of these features exist for reneging (or balking) models previously presented in the literature.

1. Introduction

Emergency Department (ED) overcrowding is a great concern to hospital administrators. ED sizing decisions are strategic in nature. Lead times to add or remove capacity are long. The data supporting such decisions covers years. Consequences of overcrowding are extreme in terms of fatalities and lawsuits. Further, community access to healthcare depends in significant part on ED capacity. In the context of long-term strategic capacity planning in the ED, management needs to be able to convert estimates (including forecasts) of ED arrival rate into estimates of ED patient safety.

An important performance measure of ED overcrowding is LWOT. LWOT is the percentage of patients who arrived to an ED, but then Left WithOut Treatment. In this paper, we argue that the safety of these patients is the foundation of a new approach to managing ED capacity. Not all of them get served; that is unsafe.

There are significant challenges to managing ED capacity using patient safety. First, the decision to LWOT has a behavioral component. The same capacity offered to one societal group may not produce the same percentage of LWOT. Second, an ED does not operate in a static environment. Volume of business has seasonality by hour, weekday, and month, and corresponding service provided does not always match business volume well. Third, the physical layout of a facility (or the psychology of the staff) may induce or reduce LWOT with all other factors held constant. Fourth, EDs have 'outlier events' when, for example, changes in computer technology or clinical care are introduced. Fifth, the ED does not operate in isolation; rather it is part of a patient flow pattern throughout the hospital. Finally, even if one were able to screen for all these effects, there is a stochastic component to arrival and service that historically has required the use of simulation or queuing.

Our goal is to capture patient safety (LWOT) response to business demand in a universal manner. A universal assessment would have the property that a particular ED could benefit from the functional form of the assessment on the basis of small sample sizes. A universal assessment would not be based upon an ad hoc description of patient LWOT behavior, but rather on a scientific mechanistic description. We hypothesize that the reneging mechanism of queuing theory will capture patient safety response for ED capacity management decision making.

The remainder of this paper is organized as follows. Section 2 discusses pertinent research that precedes this work. Section 3 presents the assumptions of the reneging approach chosen and

☆ The views expressed in this article are those of the authors and do not reflect the official policy of the United States Air Force, Department of Defense, or the United States Government.

* Corresponding author.
E-mail addresses: Jeffery.Cochran@afit.edu (J.K. Cochran), James.Broyles@asu.edu (J.R. Broyles).

0360-8352/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cie.2010.05.010

derives in complete detail a binomial regression approach to fit it to standard ED monthly data. Section 4 presents two methods of strategic deployment of the LWOT regression curve. Section 5 concludes with important findings and provides avenues for future research. Throughout, validation of our work is demonstrated using a system of seven hospital EDs in the United States.

2. Literature review

To quantify the importance of the problem, let us look at statistics from one country, the United States of America. In March 2003, The United States General Accounting Office (2003) reported that EDs are under increasing pressure with the rise in national ED visits from 95 million in 1997 to 108 million in 2000. Urgent Matters, the George Washington University Medical Center, and the Department of Health Policy (2004) describe American EDs as being in a "crisis and bursting at the seams." They also state that many hospitals have done little to address the patient flow obstacles that lead to overcrowded EDs. The US Department of Human and Health Services (2006) report that in 2004 American emergency departments were so crowded that the national mean waiting time to see a physician is 47.4 min. On average, 1.9% of all ED visits result in the patient leaving without being seen by a nurse or physician. Individual hospitals, approaching capacity, experience dramatically worse behavior than the average. This research is intended for those hospitals.

There are two schools of thought on how best to measure ED effectiveness and consequently patient safety. One measure is some form of patient waiting time. Garcia, Centeno, Rivera, and DeCario (1995) and Mahapatra et al. (2003) analyze W_q using queuing analysis of a fast-track ED. Jacobson, Hall, and Swisher (2006), Kraitsik and Bossmeyer (1993), Blake, Carter, and Richardson (1996) simulate to provide process suggestions to reduce ED waiting time. McGuire (1994), Miller, Ferrin, and Szymanski (2003), Samaha, Armel, and Starks (2003) use several patient lengths of stay (LOS denoting the service portion of stay) as well as W and W_q in the ED as measures of ED effectiveness. They simulate to quantify allowable throughput increase in the presence of decreasing LOS. All these authors state that they use some form of patient delay because an increase in time spent in an ED decreases patient safety.

Polevoi, Quinn, and Kramer (2005) is part of an extensive literature that shows that LWOT possesses critical clinical patient safety concerns. They develop a regression model for LWOT based on percentage of used ED bed capacity, acuity of ED patients, length of stay of discharged ED patients, ED patients awaiting an admission bed, inpatient floor capacity, intensive care unit capacity, and the characteristics of the attending physicians. McMullan and Veser (2004) use LWOT to determine maximum patient-to-physician ratios. Green, Soares, Giglio, and Green (2006) use queuing analysis to identify staffing patterns that reduce LWOT. They create a multivariate logistic regression model to determine if a patient will renege as a function of the new staffing profile, daily mean total ED LOS, daily mean total visits, and mode of arrival. Goodacre and Webster (2006) construct a multivariate regression model to predict the patients who LWOT as a function of patient age, sex, triage priority, postal code, initiator of attendance, mode of arrival, time, day, and month of presentation. Macdonald et al. (2005) build a simulation model in which patients who wait longer than a threshold time are randomly selected to LWOT. For this group of authors, LWOT is the preferred patient safety measure because it is a direct measure of patient safety.

We propose to support strategic decision making by developing a response of LWOT to the single factor volume. No one knows at what level of busyness patients waiting in a busy ED leave. However, there is clearly some relationship between busyness and LWOT (we will explore the mathematics of the relationship in Section 3). Leaving a waiting line after entering it is called reneging in queuing terminology. We argue that the proper regression relationship is that LWOT is some function of delay and therefore we need to explore reneging to understand LWOT.

Reneging outside of healthcare is well developed mathematically. Gross and Harris (1998) provide system performance measures in an M/M/1 system where reneging and balking occur. Abou-el-ata and Hariri (1992) and Montazer-Haghichi et al. (1986) quantify both balking and reneging for an M/M/c/K queuing system. Park (1989) exemplifies how reneging probabilities can affect warehouse restocking decisions. Ward and Glynn (2005) analyze reneging and balking in a generally distributed arrival and service rate queue. Although these methods are mathematically sound, they assume that the reneging rate distribution is known (in these examples, it is either Poisson or general distribution). This is not the case in hospital EDs. Also, these studies assume that the mean and standard deviation of reneging, arrival, and service rate are constant, in other words, distributional homogeneity over time of the reneging, arrival, and service processes is assumed. This is also not the case in hospital EDs.

In this paper, we explore the hypothesis that there is a universal relationship between LWOT and demand for service. We present a methodology based solely on physical queuing phenomena as an explanation and prediction of LWOT. Specifically, the balking probability of a single server Markovian (Poisson arrival and service rate) queue with truncation (M/M/1/K in Kendall's notation) is proposed as a strategically actionable surrogate for ED patient reneging. A preliminary formulation of this research is in Broyles and Cochran (2007) and Cochran and Roche (2009). They do not address model fitting techniques, do not use an adequate nonlinear fit quality measure (they use R^2), and do not provide forecast errors. This paper does. Our original contribution is not in queuing theory, rather it lies in the derivation of regression models that use queuing formula rather than ad hoc expressions. The queuing foundation includes the stochastic component of ED patient flow. The regression adjusts the queuing model to capture an ED's unique characteristics of LWOT behavior, seasonality, physical layout, and effects patient flow throughout the entire hospital. The regression's statistical inference provides a test for outlying events such as changes in computer technology or clinical care.

3. Formulation

3.1. Approximations and assumptions

The modeling approximations and assumptions in this formulation are driven by the queuing theory foundation of this approach, the literature, the authors' experience over five years in seven Banner Health hospitals, and extensive interviews with, and presentations to, leadership and management engineers in those hospitals. We believe this model to be valid and representative of ED balking behaviors in the real world.

Table 1 displays a list of the regression model variables. Queuing relationships among the variables in the M/M/1/K queue are introduced as derivations proceed.

In Table 1, reneging behavior is captured in a single number, K^*. This is the main approximation involved in choosing the M/M/1/K model. Classically, reneging is a continuous-time Markov chain state-dependent phenomenon. An assessment of the rate of reneging (read LWOT) would need to be estimated for each value of the number in ED queue. No hospital has the capability to estimate this data. Further, such data would be facility-specific, which we desire to avoid. We hypothesize that the patients in an ED, while actually