European Journal of Operational Research 000 (2015) 1–8

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

Use of queue modelling in the analysis of elective patient treatment governed by a maximum waiting time policy

Dawid Kozlowski [a,*], Dave Worthington [b]

[a] Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[b] Department of Management Science, Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, UK

ARTICLE INFO

Article history:
Received 2 May 2014
Accepted 11 January 2015
Available online xxxx

Keywords:
Queueing
Simulation
Waiting lists
Waiting time guarantee

ABSTRACT

Many public healthcare systems struggle with excessive waiting lists for elective patient treatment. Different countries address this problem in different ways, and one interesting method entails a maximum waiting time guarantee. Introduced in Denmark in 2002, it entitles patients to treatment at a private hospital in Denmark or at a hospital abroad if the public healthcare system is unable to provide treatment within the stated maximum waiting time guarantee. Although clearly very attractive in some respects, many stakeholders have been very concerned about the negative consequences of the policy on the utilization of public hospital resources. This paper illustrates the use of a queue modelling approach in the analysis of elective patient treatment governed by the maximum waiting time policy. Drawing upon the combined strengths of analytic and simulation approaches we develop both Continuous-Time Markov Chain and Discrete Event Simulation models, to provide an insightful analysis of the public hospital performance under the policy rules. The aim of this paper is to support the enhancement of the quality of elective patient care, to be brought about by better understanding of the policy implications by hospital planners and strategic decision makers.

1. Introduction

Identifying methods for ameliorating excessive waiting lists for elective patient treatment have been high on the agenda for many countries with public healthcare systems. As shorter waiting times usually result in higher degrees of patient and citizen satisfaction, considerable funds have been invested in their reduction over the years. Different countries have tried to address this problem in various ways, and one interesting method entails a Maximum Waiting Time Guarantee (MWTG). A number of papers reporting the implementation of the policy in different countries and providing general descriptions of its content are available in the literature (Ashton, 2009; Hanning, 1996; Hanning & Spångberg, 2000; Johannesson, Johansson, & Söderqvist, 1998; Karlberg & Brinkmo, 2009; Siciliani & Hurst, 2005; Vrangbæk, Østergren, Birk, & Winblad, 2007; Wiley, 2005).

Introduced in Denmark in 2002, this guarantee entitles patients to treatment at a private hospital in Denmark, or at a hospital abroad, if the local public healthcare system is unable to provide treatment within the stated maximum waiting time guarantee. In such cases, the Danish region where the patient resides (and which is responsible for providing the local public hospital care), now has to cover the cost of treatment either in private for-profit hospitals in Denmark or in hospitals in neighbouring countries (Madsen, 2010). As noted by Pedersen, Christiansen, and Bech (2005), the position of the patient has been significantly strengthened through the implementation of the policy. Patients are no longer simply a passive party waiting for capacity to become available; they have now more options to choose from, and hence some decision power, which should lead to a reduction in waiting lists and waiting times.

After the introduction of a two-month maximum waiting time guarantee for the entire course of treatment (i.e. examination and treatment) in 2002, it was decreased to one month in 2007. Although clearly very attractive in some respects, many stakeholders have been concerned about the possible negative consequences of such policies, if taken too far, on the performance of public hospitals. In particular, there are claims that a one month guarantee would cause increased overflows of patients to the private sector, with a consequent inefficient use of public hospital resources. As a result, the Danish government has recently extended the maximum waiting time guarantee into one month for examination procedures and two months for the treatment.

Clearly, there exists a trade-off between short waiting times and high resource utilization in public hospitals. A maximum waiting time guarantee should limit waiting times of elective patients by offering the possibility of faster treatment in the private sector when

* Corresponding author. Tel.: +45 65503966.
E-mail addresses: kozlowskidawid86@gmail.com, dko@sam.sdu.dk (D. Kozlowski), d.worthington@lancaster.ac.uk (D. Worthington).

http://dx.doi.org/10.1016/j.ejor.2015.01.024
0377-2217/© 2015 Elsevier B.V. All rights reserved.

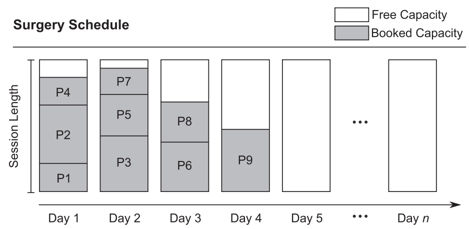

Fig. 1. Example of a rolling schedule under 'first slot available' rule.

the queues in public system are too long. However, given that the whole system is stochastic, very short queues for examination and treatment can result in some inefficiencies as shown in Worthington (1987). Specifically, resources can become under-utilised more often, because there will be a higher chance of having no patients available to fill empty timeslots in a schedule. The negative impact of the policy would be magnified if all patients use the possibility of private treatment.

However, there are other potential sources of inefficiency which can result from short waiting lists. The first one relates to a mismatch between available patients and examination or treatment session lengths. Specifically, the policy rules state that patients should be scheduled for examination and operation based on the FCFS rule. Moreover, they should receive an appointment for a particular day in the future in advance, so the hospital planner has to reserve the capacity required for the procedure based on his best knowledge about its expected process time. As overbooking is typically not allowed, this can result in sessions that are not fully booked, and hence underused capacity. It is simply difficult to match the process times of upcoming patients in a way that they cover 100 percent of available time. An example of a patient case-mix scheduled for a given day with use of this policy is presented in Fig. 1. To increase the utilization of resources exemptions from the FCFS rule may occur when the expected process time of an upcoming patient fits into an empty gap in the schedule left by the preceding patients (e.g. patients 4 and 7). Hence, the longer the queue gets the better is the chance of a good fit, as there are simply more patients to choose from. The goodness of fit is likewise affected by the patient case-mix, for example it is easier to fit short procedures into a schedule than long ones.

Another way in which utilization may be lost is due to patients withdrawing from the system or rescheduling their appointments. The first event relates to the patients who have made an appointment but leave the system before the due date, not requiring examination or treatment any more. The latter event describes the situation in which the patients change their previous appointments (due to medical or personal reasons) but still need to be examined or treated on another day. In both cases, the potential loss of utilization is due to the difficulty of finding a replacement for the cancelled appointment. In practice the replacement can be found amongst newly arriving patients or patients that are already placed on a waiting list. The earlier a patient gives notice about the cancellation, the greater is the chance of finding an arriving patient to take his place. If this is not possible, the hospital planners typically contact other patients on the waiting list offering them an earlier examination or treatment. Clearly, the longer the waiting list is the higher is the chance of finding a replacement. However, if a patient gives very short notice or does not inform the hospital at all, it may be impossible to find a replacement and the capacity will not be utilized.

Given these sources of utilization losses it is easy to see that short queues can lead to inefficiencies. However, increasing the size of queues that are already long increases patients' waiting times for ever decreasing improvements in utilization. (In fact, in practice longer waiting lists can sometimes reduce utilization due to greater no-shows or other problems associated with managing long waiting lists.) It is therefore important to investigate and compare the trade-offs between waiting times and utilization levels prior to making any changes in the policy.

Not surprisingly, waiting list management and hospital admissions have been given some attention in the literature. Whilst economists perceive waiting lists as a rationing mechanism in the absence of price rationing (Lindsay & Feigenbaum, 1984), operations researchers tend to see the waiting time as a result of stochastic demand for health services. Consequently, the latter often use stochastic modelling techniques to tackle waiting list problems. For example, Worthington (1987) investigated the implications of certain waiting list and demand management strategies using a single station queueing model. Vissers, Adan, and Dellaert (2007) used simulation to examine different admission policies in a simplified hospital setting and proposed a platform for comparing alternative hospital admission systems. VanBerkel and Blake (2007) also used a simulation to analyze the waiting list for surgery procedures and to aid capacity planning decisions at a surgical department. Goddard and Tavakoli (2008) evaluated economic aspects of managed public sector waiting lists using a single station queueing model with a simple withdrawal mechanism. They evaluated a guaranteed maximum duration of wait, pointing out that it may not always serve a welfare maximizing policy. Bekker and Koeleman (2011) combined Quadratic Programing with a time-dependent queueing model to determine hospital admissions quota for scheduled admissions that results in a stable bed demand. Knight, Williams, and Reynolds (2012) developed a simulation tool for modelling patient choice in healthcare systems where individual patient's decision to join a waiting list at a given hospital is made upon observation of the entire system (e.g. travel distance and length of waiting list). This work is extended in Knight and Harper (2013) who used a game theoretical approach where the choice of public hospitals by individual patients is interpreted as a routing game. Griffiths, Williams, and Wood (2013) constructed a queuing model of a neurological rehabilitation unit which incorporated balking to model the dissuading effect that long waiting lists may have on arrival rates. Queueing theory approach was also used by Stanford, Lee, Chandok, and McAlister (2014) to address waiting time inconsistency in solid organ transplantation.

The process of elective treatment usually involves a specialist consultation at an outpatient clinic for which patients will normally have to wait a period of time. Some of these patients will require treatment as inpatients, for which they will again usually join a waiting list before receiving treatment. Consequently, a more detailed queueing representation of this process is likely to involve two stations. Analysis of tandem configurations consisting of service facilities with finite queues linked in series has a long history in the literature. Blocking caused by finite capacities makes exact analytical analysis of open queueing networks intractable, except for very small networks, and hence approximation approaches have been developed for these systems (see e.g. Altiok, 1982, 1989; Bierbooms, Adan, & van Vuuren, 2013; Jun & Perros, 1990; Perros, 1986). Other authors have provided useful performance bounds for such systems (e.g. So & Chin, 1992; Van Dijk & Lamond, 1988). Unfortunately, none of these models includes the waiting list mechanisms needed for the MWTG problem, such as the various sources of efficiency losses, and the waiting time guarantee which refers to total waiting time in both queues.

The primary aim of this paper is to tackle and offer insights into the MWTG problem using an innovative combination of Continuous-Time Markov Chain (CTMC) and Discrete Event Simulation (DES) modelling approaches. The combined use of the two approaches allows us to incorporate all the important waiting list mechanisms in a tandem queue formulation. In particular, we develop a CTMC in which the additional sources of inefficiency are introduced into the queue formulation through state-dependent processing rates. In addition, we have built an associated DES model to help investigate features such as patient booking in more detail. Simulation results are used to calibrate state-dependent processing rates in the CTMC model so