European Journal of Operational Research 239 (2014) 427–436

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

A stochastic appointment scheduling system on multiple resources with dynamic call-in sequence and patient no-shows for an outpatient clinic

Pei-Fang Jennifer Tsai *, Guei-Yu Teng

Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei 10608, Taiwan

ARTICLE INFO

Article history:
Received 21 September 2013
Accepted 23 April 2014
Available online 4 May 2014

Keywords:
Stochastic processes
OR in health services
Appointment scheduling
Overbooking
Patient no-shows

ABSTRACT

This research focuses on the stochastic assignment system motivated by outpatient clinics, especially the physical therapy in rehabilitation service. The aim of this research is to develop a stochastic overbooking model to enhance the service quality as well as to increase the utilization of multiple resources, like therapy equipment in a physical therapy room, with the consideration of patients' call-in sequence. The schedule for a single-service period includes a fixed number of blocks of equal length. When patients call, they are assigned to an appointment time for that block, and an existing appointment is not allowed to be changed. In each visit, a patient might require more than one resource and a probability of no-show. Two estimation methods were proposed for the expected waiting and overtime cost with multiple resources: Convolution Estimation Method and Joint Cumulative Estimation Method for the upper and lower bound value; respectively. A numerical example based on a physical therapy room was used to show that this stochastic model was able to schedule patients for better profitability compared with traditional appointment systems based on four prioritization rules. The workload in each appointment slot was more balanced albeit more patients were assigned to the first slot to fill up the empty room.

1. Introduction

In a healthcare delivery system, efficient appointment scheduling can serve as a connection between two important quality-related criteria: timely access to services and efficient healthcare services, especially for the outpatient clinics (Cayirli & Veral, 2003). Appointment delay, or the time interval between the need for service and the actual use of healthcare service, can be effectively measured by the timeliness in the service accessibility (Liu, Ziya, & Kulkarni, 2010; Robinson & Chen, 2010). On the other hand, the efficiency of healthcare service was usually measured by patients' waiting time in the clinic and physicians' idle time or overtime in sessions (Bailey, 1954).

Many studies have scrutinized the accessibility issue of appointment systems and found a strong relationship between appointment delay and the rate of cancellations and no-shows. A 1997 study in the UK analyzed 1927 referrals from 26 general practitioners and found a longer interval between referral and appointment as one of the five factors significantly associated with non-attendance (Hamilton, Round, & Sharp, 2002). A study of referrals to the Johns Hopkins Bayview Medical Center community psychiatry outpatient programs from 1995 to 2000 found that the rate of cancellations and no-shows increased to 23 from 12% for patients who were given a next-day appointment rather than on the same day as their initial contact (Gallucci, Swartz, & Hackerman, 2005). A study analyzing data for patients in outpatient additional treatment program units from 2003 to 2006 demonstrated a strong negative correlation in the probability of completing four sessions of treatment with increasing time between the clinical assessment and first treatment session (Hoffman, Ford, Tillotson, Choi, & McCarty, 2011). To address the issue of no-shows, a new advanced access model, or "open access policy, was introduced to ensure patients called in the morning to obtain same-day appointments (Murray & Tantau, 2000).

The efficiency of an appointment scheduling system can be represented as a combination of patient satisfaction and hospital revenue gains. A good appointment system should also reduce the patient's waiting time and the hospital's operational cost at the same time (Gupta & Denton, 2008). Early researches observed excessive waiting for patients due to their being called to the clinic in large batches at a volume greater than the physicians' capacity to see them (Bailey, 1954; Welch, 1964). Even when patient was scheduled individually for a specific treatment time, Vissers (1979) observed that punctual patients might reduce idle time for the physician but result in unacceptable waiting time for the

* Corresponding author. Tel.: +886 2 2771 2171x2354; fax: +886 2 2731 7168.
E-mail address: ptsai@ntut.edu.tw (P.-F.J. Tsai).

http://dx.doi.org/10.1016/j.ejor.2014.04.032
0377-2217/

patients. Patients might end up undermining the appointment time by either arriving too early to receive treatment earlier or by arriving later to avoid waiting.

Research on the appointment problem pertaining to physical therapy has been rather limited. Chien, Tseng, and Chen (2008) had formulated the rehabilitation patient scheduling problem as a hybrid shop scheduling problem. Their solution approach involved two phases: to find the optimal sequencing of patients using genetic algorithm, and then to construct the feasible therapy sequence for all patients using a timetabling algorithm based on the sequence from the first phase. The numerical results showed that, after considering the patients' preference, their approach was more effective than the hospital's empirical method in terms of reducing patient's waiting time as half of the patients were without an appointment. Ogulata, Koyuncu, and Karakas (2008) also formulated the scheduling problem in a physiotherapy clinic as deterministic mathematical models. Since a high demand for the service was assumed, the entire problem was decomposed to three hierarchical models: selection of patients, assignment of patients to therapists, and scheduling of patients. None of these studies considered walk-in patients, rate of no-shows, and uncertainty concerning the actual service times.

This research is motived by the assignment scheduling problem in outpatient departments for special clinics, especially from the observation for a physical therapy room in an academic medical center in Taipei (Rau et al., 2013). Most patients' treatment plans include two to five physical treatments with no predetermined sequence. As discussed in Rau et al. (2013) about therapist pooling to reduce patient wait times, a major concern in adopting this strategy from therapists' viewpoint was the additional work in synchronizing patient appointment schedules. Further, with the consideration of patient promptness and no shows, the study hinted a need to develop an advanced appointment scheduling to further improve the patient flow in this room. Focusing on real-world settings, this research aims to develop a stochastic model for the appointment scheduling problem to achieve better service quality and profitability for outpatient clinics. Each service session is divided into a fixed number of appointment blocks with equal length. With the consideration of existing appointments, this appointment system optimally assigns patients to an appointment block with given probabilities of patient no-shows. The remainder of the paper is organized as follows. In the next section, literature on appointment scheduling is briefly reviewed. The introduction of mathematical formulation of the stochastic process and the implementation results are depicted in Sections 3 and 4, respectively. Then, discussion and concluding remarks are summarized in Section 5.

2. Literature on appointment scheduling problems

The development of an appointment system depends on the design to access medical services among three options: traditional model, carve-out model, and advanced-access model (Murray & Berwick, 2003). In a traditional model, appointments are completely arranged and booked in advance. Same-day urgent care is piled on top of an already full schedule or deflected to the emergency department. This model presents a significant obstacle to obtaining care due to the system's inability to offer appointments soon, which leads to high no-show rates. In a carve-out model, some of the appointment slots are reserved for urgent demands and the rest are booked for non-urgent requests in advance. The caveat of this model is the fairness in the triage decisions in addition to the same impediment of delaying more non-urgent demands as in the traditional model. The advanced-access system, also known as a "same-day access" or "open access" system, tries to eliminate appointment delay entirely, with the clinic having to offer an appointment to patients on the same day as they call about any problem, urgent or not (Murray & Berwick, 2003). Successful implementation showed significant reduction in patients' appointment delay from a range of 26 to 55 days to just one to two days, along with an increase in satisfaction of patients, staff, and even service providers (Murray & Tantau, 2000). Although benefits are substantial, this advance-access policy is more vulnerable to the variations in arrival demand, and is sensitive to the imbalance between demands and capacities (Liu et al., 2010; Robinson & Chen, 2010).

The design of an appointment system has three components: an appointment rule, patient classification or preference, and an adjustment policy for walk-ins and no-shows. The design of the appointment rule also has three components: appointment block sizes, appointment intervals, and the initial block design (Cayirli & Veral, 2003). The appointment interval can be fixed or variable, and the initial block design can be present or absent. The "single-block" rule assigns all patients to arrive as a block at the beginning of the clinic session, and the "individual-block" rule gives patients unique appointment times distributed evenly over the clinic session. The "multiple-block" rule assigns a fixed number of patients to each appointment block (Blanco White & Pike, 1964; Fetter & Thompson, 1966).

Consideration of patient preference can be helpful in deciding the patients' appointment schedule at the time of booking. Also, patient classification can make the resulting schedule more practical if the appointment intervals are adjusted based on patients' acuity levels. Other classification includes new/return patients, variability in the service times, and type of procedure (Cayirli & Veral, 2003). Walter (1973) found that, for the same chest examination on an ordinary X-ray machine, patient age had a significant impact on the average examination length, and the variances of the times for patients age 60 or older and those under 20 were both significantly higher. Vanden Bosch and Dietz (2000) adopted a classification scheme to sequence assignments for a specific primary clinic based on patients' appointment history or type of procedure. Their simulation results concluded that there was no easy rule for the optimal assignment scheduling.

The dynamic case is to allocate the appointment block/time specifically to a patient depending on the state of the system represented by the number of appointed patients in each session. This appointment system tends to have a variable block size, and the next patient is allocated optimally by considering the possible number of remaining patients at the end of each session. Fries and Marathe (1981) and Liu and Liu (1998) proposed a dynamic programming approach to find the optimal allocation for new patients with the consideration of the expected number served in each block. Gupta and Wang (2008) extended the dynamic appointment case so as to include patient choices and proposed a threshold-type optimal policy for a primary clinic with a single physician. Green, Savin, and Wang (2006) considered the appointment problem for diagnostic facilities to find the optimal capacity allocation for inpatient and outpatient demands. Day, Dean, Garfinkel, and Thompson (2010) applied the appointment system design to both single and multiple testing stages as patients might need the same type of test room more than once.

To access various appointment systems under realistic environment settings, simulation modeling has prevailed. Klassen and Rohleder (1996) used simulation to evaluation designs in sequencing patients with "high" or "low" service time variability while the appointment intervals were constant. They concluded that better results can be obtained if patients with higher service-time standard deviations are scheduled toward the end of the appointment session. Cayirli, Veral, and Rosen (2006) tested similar appointment rules as in Klassen and Rohleder (1996) in primary care clinics in New York. Su and Shih (2003) evaluated four appointment policies