Simulation Modelling Practice and Theory 77 (2017) 317–337

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

A reusable simulation model to evaluate the effects of walk-in for diagnostic examinations

A. Braaksma[a,b,c,*], N. Kortbeek[a,b,c], K. Smid[b,d], M.E.S. Sprengers[e]

[a] Center for Healthcare Operations Improvement and Research (CHOIR), University of Twente, Enschede, Netherlands
[b] Department of Quality and Process Innovation, Academic Medical Center, Amsterdam, Netherlands
[c] Stochastic Operations Research, Department of Applied Mathematics, University of Twente, Netherlands
[d] Industrial Engineering and Business Information Systems, School of Management and Governance, University of Twente, Netherlands
[e] Department of Radiology, Academic Medical Center, Netherlands

ARTICLE INFO

Article history:
Received 23 June 2016
Revised 26 June 2017
Accepted 28 July 2017
Available online 2 August 2017

Keywords:
Health services accessibility
Diagnostic services
Discrete event simulation
Appointments and schedules
Computed tomography scans

ABSTRACT

Enabling patients to walk in for their diagnostic examination without an appointment has considerable potential in terms of quality of care, patient service, and system efficiency. We present a model to evaluate the effect of implementing a combined walk-in and appointment system, offering appointments to all patients preferring or strictly requiring these, while enabling all other patients to walk in. In a combined system, appointments can be scheduled in periods with low walk-in demand to counterbalance the possible high variability in walk-in arrival rates. We develop a discrete event simulation model, combined with an intelligent algorithmic methodology for appointment schedule optimization, for evaluating the implementation of a combined walk-in and appointment system for diagnostic examinations. Our simulation model is reusable: its component-based structure and generic underlying logic enable it to automatically represent any type of diagnostic facility, for which it can then evaluate the effect of implementing a combined walk-in and appointment system. Applying this approach, we quantitatively investigate the impact of implementing a combined walk-in and appointment system for CT-scans, performing a case study at the Academic Medical Center (AMC) Amsterdam. Inspired by the results, the AMC CT-facility has implemented a combined walk-in and appointment system, thereby shortening patients' diagnostic trajectories, and decreasing the number of required hospital visits for many patients.

1. Introduction

Diagnostic examinations play an important role in determining a patient's condition and deciding on the course of treatment. Enabling patients to walk in for their diagnostic examination without an appointment has considerable potential in terms of quality of care, patient service, and system efficiency. Access time, defined as the number of days between the request and the actual examination, may have a negative impact on the patient's condition and health outcomes when prolonged [1]. Walk-in eliminates access times completely. Consequently, the diagnostic trajectory is shortened which in turn shortens the period the patient and his relatives are in a state of suspense about the patient's condition. Further, walk-in

* Corresponding author.
E-mail address: braaksma@mit.edu (A. Braaksma).

http://dx.doi.org/10.1016/j.simpat.2017.07.004
1569-190X/

provides patients with autonomy to choose their preferred moment for the examination. When walking in right after the consultation in which the necessity of the examination has been revealed, the patient saves one hospital visit. In terms of system efficiency, walk-in eliminates three downsides of an appointment system. First, no-shows are absent in a walk-in system, while these cause server idle time in an appointment system. Second, stochasticity of examination durations forces the inclusion of slack time in an appointment system, to guarantee acceptable patient waiting times (the time between a patient's arrival at the facility and the start of the examination). There is no need for inclusion of slack time in a walk-in system. A walk-in system thus facilitates a potentially higher system utilization. Third, a walk-in system eliminates the administrative work related to scheduling and possibly rescheduling appointments as well as reminding patients of their appointment times.

A disadvantage of a walk-in system is the possible high variability in patient arrivals, which results in a highly variable system utilization and high patient waiting times during busy periods. An appointment system, on the other hand, is ideal for spreading workload over time. Implementing a full walk-in system is not possible for most diagnostic examinations, because appointments are inevitable for certain patients (e.g., patients with high preparation times, examinations requiring the presence of a medical specialist from another department, or patients preferring an appointment over walking in). By employing a combined walk-in and appointment system, a diagnostic facility can have the best of both worlds: appointments can be scheduled in periods with low walk-in demand such that the overall daily arrivals are smoothed [2,3].

While walk-in is common practice for X-ray examinations, other diagnostic examinations, such as computed tomography (CT) and magnetic resonance imaging (MRI) scans, are appointment-based in most hospitals. In this paper, we present a computer simulation model to evaluate the effect of implementing a combined walk-in and appointment system for diagnostic examinations. The model is reusable: its component-based structure and generic underlying logic enable it to automatically represent any type of diagnostic facility (see Section 3.1 for a detailed explanation of reusable simulation modeling). Thus, our simulation model can be used to investigate the consequences of implementing a combined walk-in and appointment system for any type of diagnostic examination. In this paper, we apply it to evaluate the impact of implementing such a system for CT-scans.

This paper is organized as follows. Section 2 discusses relevant literature. Section 3 describes the reusable simulation model. Section 4 introduces our case study and presents the results, followed by a discussion and conclusions in Section 5.

2. Literature

The amount of literature on using computer simulation to study healthcare delivery is substantial [4–6], and also contains studies on CT-scan delivery (e.g., [7,8]). To the best of our knowledge, the impact of a walk-in system has only been quantitatively investigated before for nurse-led NHS walk-in centers, providing community-based ambulatory care [9], and for diagnostic clinics, treating non-emergent cases that would otherwise report to the emergency department [10]. Ashton et al. [9] developed a simulation model to advise a health center in Liverpool, consisting of an NHS walk-in center, a general practitioner's practice, and various primary and community healthcare services, on how to operate their services upon relocation to new premises. Because the various services were to share waiting areas and treatment rooms in the new situation, analyzing the interaction between the services and providing suggestions for limiting the number of patients simultaneously present in the central waiting room were important aspects of the simulation study. Reilly et al. [10] aimed to enhance performance of a diagnostic clinic employing a complete walk-in system, by adjusting physician staffing patterns and introducing so-called *delay scheduling*. In periods of congestion in the clinic, walk-in patients were assigned a *delay time*. Patients could decide to temporarily leave the clinic and return after the delay time (i.e., later on the same day), in which case the delay time was not counted as clinic-accountable waiting time. Upon returning to the clinic, such patients were treated with priority. Reilly et al. [10] employed simulation modeling to evaluate the delay scheduling and staffing changes, and concluded that these resulted in reduction of manpower by 10% while significantly reducing the clinic-accountable waiting time. While both these studies aimed at improving the performance of existing complete walk-in systems, we develop a methodology for evaluating the impact of implementing a combined walk-in and appointment system.

In combined walk-in and appointment systems, the appointment schedule naturally affects system performance. Although a review on outpatient appointment scheduling in 2003 concluded that the presence of walk-in was neglected in most studies [11], several authors have recently studied the question which slots to reserve for appointments and which slots to leave open for walk-in in combined systems. This has been investigated for a single consultation session [12–14] or for a multi-day cycle [3,15]. Aiming to find the optimal moments at which to schedule appointments, authors either assume the number of appointments to schedule to be pre-specified [12,15], or optimize this number simultaneously [3,13,14]. All studies focus on schedule performance from a walk-in patient's and a system perspective by optimizing some combination of patient waiting times, provider idle time, provider overtime, and a revenue for each patient seen. [3] is the only study additionally considering access times for scheduled patients. Because access time is an essential performance indicator in a combined walk-in and appointment system, we combine our simulation study with optimization of the appointment schedule based on the methodology by Kortbeek et al. [3].

While scheduling appointments in periods with low walk-in demand has been demonstrated to effectively counterbalance high variability in patient arrivals [2,3], the concept of call-in has been shown to achieve similar effects [16]. The literature defines *call-in* patients as those who (i) do not require immediate service – as emergency patients do, (ii) have a medical urgency that prohibits them from being scheduled as regular elective patients, and (iii) are available to be served at the provider's earliest convenience. Examples are inpatients requiring diagnostic examinations [17–19], patients requiring expedited inpatient admission [16], and semi-urgent surgical patients [20]. With emergency, call-in, and elective patients competing for access to the same resource, studies focus on the optimal design of appointment schedules [19,20], and on the decision how many patients of each type to start serving at each point in time [16–18,20]. The objective is to minimize costs or maximize profits, accounting for type-specific treatment revenues, waiting costs, and deferral penalties, as well as opportunity costs for resource idling. Note the fundamental difference between call-in and walk-in patients: while call-in patients have a medical necessity for expedited service, may be readily available as inpatients [17–19], and may benefit from increased quality of service through call-in [16], the walk-in patients considered in this paper are outpatients and thus receive the best quality of service when receiving their examination on the day of their arrival to the diagnostic facility, within an acceptable waiting time.

Our contribution is threefold. First, we develop a reusable simulation model for evaluating the effects of a combined walk-in and appointment system for diagnostic examinations, providing any type of diagnostic facility with quantitative support in decision making on (i) the potential implementation of a combined walk-in and appointment system, and (ii) configuring such a system. Second, using our model, we quantitatively investigate the impact of implementing a combined walk-in and appointment system for CT-scans. Thereby, we provide CT-facilities in general, and the Academic Medical Center (AMC) Amsterdam's CT-facility in particular, with quantitative insight into this alternative system that has considerable potential in terms of quality of care and service as well as system efficiency. To the best of our knowledge, we are the first to quantitatively investigate the impact of a combined walk-in and appointment system. Third, we incorporate an intelligent algorithmic methodology for appointment schedule optimization into this quantitative investigation. Not only are we the first to propose this incorporation, but we also develop a procedure that connects the strengths of a generic mathematical methodology for generating appointment schedules [3] to the specifics required of appointment schedules that are to be used at diagnostic facilities. Leveraging simulation modeling to adequately capture those real-world specifics, our contribution is a methodology that combines the best of both worlds: optimization and simulation.

3. Methods

This section presents our generic methodology for evaluating the impact of implementing a combined walk-in and appointment system at a diagnostic facility. We introduce the concept of reusable simulation modeling, followed by a description of the reusable simulation model developed in this study.

3.1. Reusable simulation modeling

We develop a reusable discrete-event computer simulation model to evaluate the effects of implementing a combined walk-in and appointment system for diagnostic examinations. Although simulation modeling is a suitable technique for this study because of its favorable comparison both to experimenting with the real system and to other modeling techniques, developing a simulation model is a time-consuming process [21]. This disadvantage can be overcome by reusable simulation modeling, i.e., developing a simulation model in such a way that (parts of) it can be used again by others in their simulation study of another, possibly similar type of system. We develop a reusable simulation model, such that the modeling effort from this study can be exploited to evaluate the effects of implementing a combined walk-in and appointment system for various types of diagnostic examinations, and possibly also for other types of healthcare facilities. For reusable simulation modeling, [22] presents a reuse spectrum, consisting of code scavenging, function reuse, component reuse, and full model reuse. The four different types of reuse are listed here in decreasing order of application frequency, and in increasing order of complexity. We apply the component reuse approach, leading to reduced model development time and cost during reuse, and extensibility and evolvability, which make the model more adaptable and easier to modify [23].

3.2. Simulating a combined walk-in and appointment system

Fig. 1 depicts the core patient flow in a combined walk-in and appointment system. Upon emergence of a patient's service request, the patient's eligibility for walk-in is assessed. If the patient is eligible, he can directly proceed to the waiting room and is then served. In case the patient is not eligible, the appointment scheduler schedules an appointment, the patient comes to the facility on the day of the appointment, waits in the waiting room, and is then served. Notice that there are two main arrival modes in a combined walk-in and appointment system: appointment patients and walk-in patients. Also, there are two service modes: scheduled and unscheduled. While all appointment patients are served as scheduled patients and the majority of walk-in patients are served as unscheduled patients, certain walk-in patients join the category of scheduled patients. The latter happens if, due to either a medical or a logistical reason, a walk-in patient cannot receive unscheduled service. Throughout this paper, we use the terminology of appointment and walk-in patients to refer to patients' arrival modes, and the terminology of scheduled and unscheduled patients to refer to patients' service modes.

There may be several additions to the core process described above: patients may need to undergo tests or preparation steps before they can be served, and there may be multiple, possibly distinct servers. The core process makes up the heart of our simulation model, that is, only systems adhering to this basic structure can be evaluated by means of our reusable model. To incorporate the possible additions to the core process, we define three types of components: preparation steps,

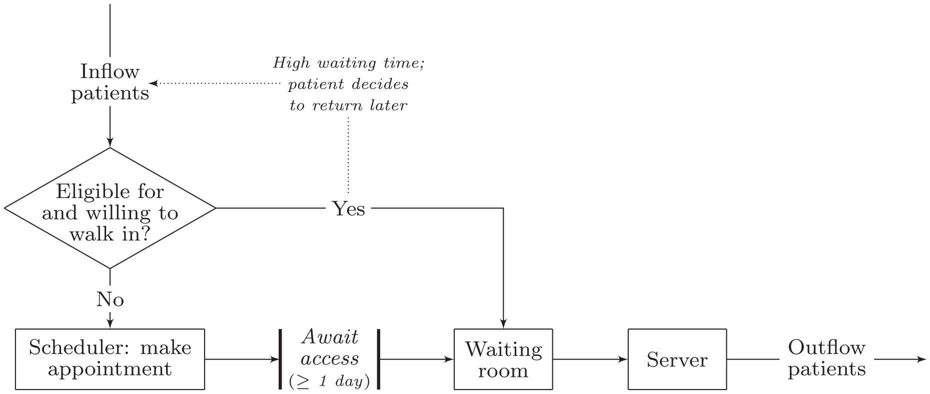

Fig. 1. The core patient flow in a combined walk-in and appointment system.

tests, and servers. Our model has the flexibility that any number of tests and preparation steps can be added to the core process before the patient's service in any order, to maximize the model's ability for representing different types of diagnostic facilities. Also, our model can accommodate any number of servers, and these can be of distinct types. Servers are always the final process step.

Our simulation model in itself does not represent any type of diagnostic facility, but contains the logic of the core process, as depicted in Fig. 1, and the logic of the three types of components that can be added to that. In order to make the simulation model represent a given diagnostic facility, the user has to specify a number of general diagnostic facility characteristics, patient type characteristics, and characteristics of the various components. All these characteristics can be specified in the model's Excel (Microsoft, Redmond, WA) front end, which comes with a user guide. A specific instance of the simulation model, representing an actual diagnostic facility, is constructed automatically based on the information it reads from this front end. Each element of our reusable simulation model contains a generic initialization procedure that automatically generates the instance-specific content for that model-element based on the front end information, as exemplified in Fig. 2. Thus, the user does not need to access the simulation software, which makes the model usable for people with minimal simulation experience. As an example of one out of many diagnostic facility processes our simulation model can thus represent, Fig. 3 displays the conceptual simulation model for our case study.

In what follows, we present our simulation model in detail by first describing the general model representation of a diagnostic facility, the representation of patient types, and the three types of components, followed by a description of the execution of the model and an overview of the performance indicators. Finally, we present our approach for combining simulation with appointment schedule optimization. Fig. 4 supports our model description by giving a schematic overview of the reusable simulation model.

3.2.1. The diagnostic facility

From a logistical perspective, a diagnostic facility is characterized by its demand, its process, and a set of policy decisions relating to capacity allocation. First, the simulation model characterizes a facility's demand as follows:

– The number of patient types. Section 3.2.2 elaborates on the simulation model's characterization of a patient type.
– The overall patient arrival rate in each time slot of each day in the cycle. We assume that patient arrivals follow a cyclic (e.g., weekly repeating) pattern, which is a non-stationary Poisson process in the current model setup. In the simulation model, all patients arrive according to this cyclic pattern; upon arrival a patient's type is drawn according to the percentage of patients that is of each type.
– The no-show probability of scheduled patients.

Second, the simulation model contains the following generic logic characterizing a diagnostic facility's process:

– The number of preparation steps.
– The number of tests.
– The number of servers.

Section 3.2.3 provides a detailed description of the model's characterization of these three types of components. Finally, the model contains a subset of a diagnostic facility's policy decisions relating to capacity allocation, namely those decisions that affect the performance of a combined walk-in and appointment system:

– The facility's business hours.
– The time of day at which the facility stops accepting walk-in patients (walk-in 'closing' time). A facility may choose to stop accepting walk-in patients some time before closing time, in order to reduce the risk of overtime.

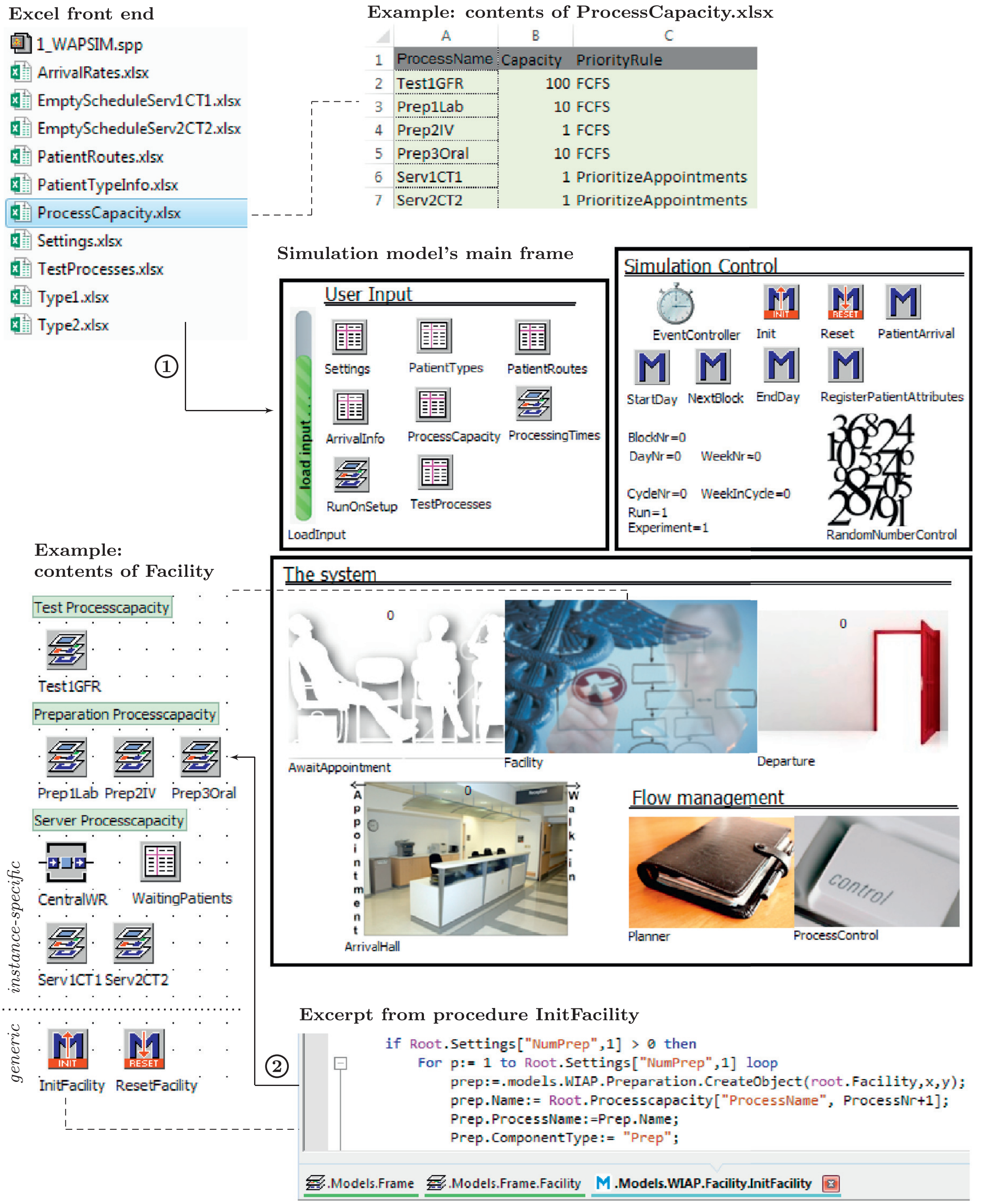

Fig. 2. Screen shots exemplifying how our reusable simulation model automatically generates a model instance representing a specific diagnostic facility. The Excel front end feeds the model's User Input tables (1). Generic initialization procedures use that information to build the instance-specific content of the model's elements (e.g., 2).

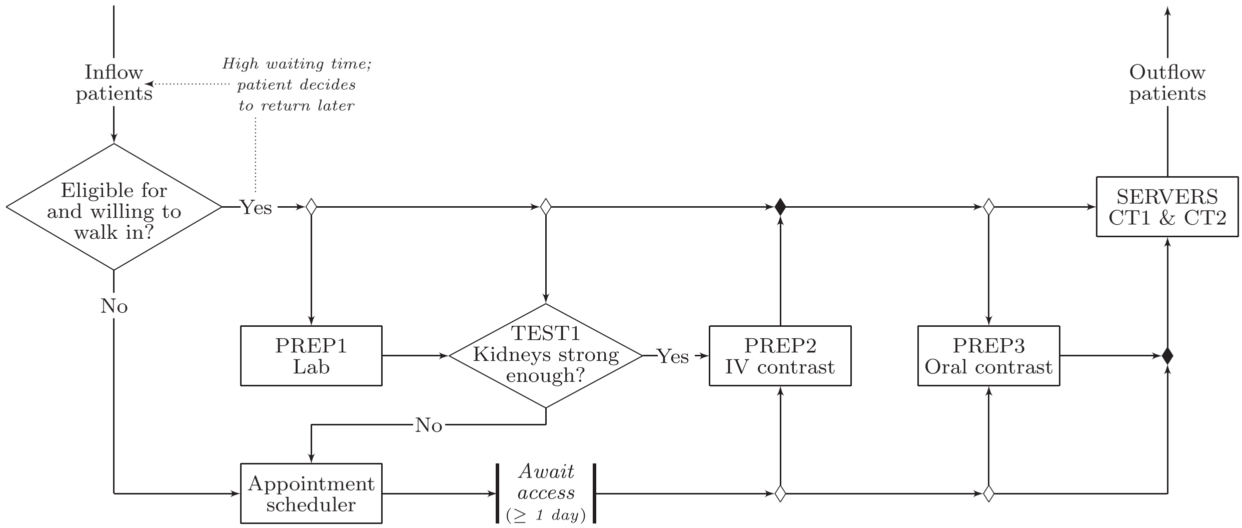

Fig. 3. Conceptual simulation model for our case study. (At ◇ (✦), the routes of different patient types through the system split (merge).)

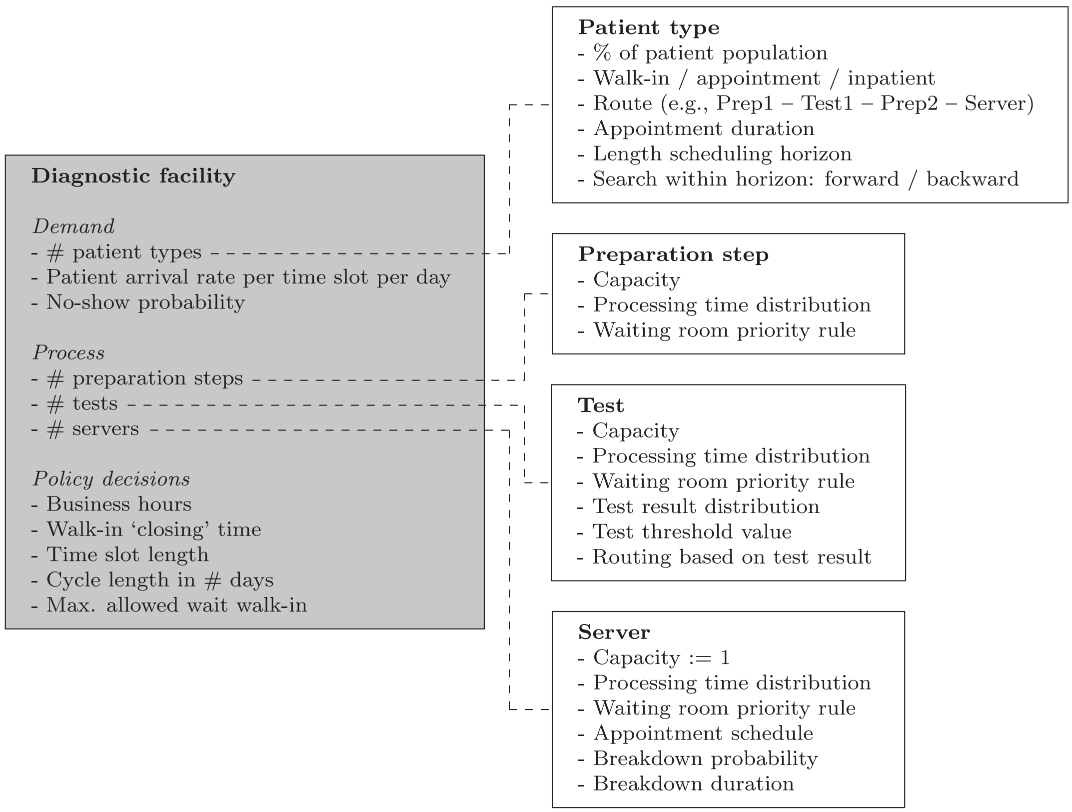

Fig. 4. Schematic overview of our reusable simulation model.

– The length of a time slot (in minutes).
– The cycle length in number of days. Diagnostic facilities typically employ cyclic appointment schedules, where the cycle is a repeating sequence of days.
– The number of time slots a walk-in patient is allowed or willing to wait for service. If, upon a walk-in patient's arrival at the facility the expected waiting time is higher than this number of slots, the patient is offered an appointment on a later day in order to prevent patients from excessive waiting times, and the facility from overcrowding. We refer to such patients, who arrive as a walk-in patient but are served as a scheduled patient, as *deferred walk-in patients*.

3.2.2. Patient types

The simulation model characterizes a patient type as follows:

- The percentage of the total patient population that is of this type.
- Whether patients of this type arrive as walk-in patients, as appointment patients, or as inpatients (i.e., hospitalized patients). Appointment patients receive an appointment on a future day, whereas inpatients are preferably offered a same-day appointment.
- This patient type's pre-specified route through the system. The user specifies this route by assigning a 1 to the first preparation step, test, or server the patient should visit, and consecutive integers to consecutive process steps. Components not belonging to the route of a patient type are not assigned any number, and thus omitted from the route. In specifying the route, the user also indicates which servers are appropriate for serving this patient type.
- The number of time slots that should be reserved when scheduling an appointment for this patient type.
- The horizon (in number of days) within which an appointment should be scheduled for this patient type, and whether the appointment scheduler should search for the first appointment slot from the start of the horizon forwards, or from the end of the horizon backwards. The start of the horizon is always the next day, except for inpatients, who are preferably offered a same-day appointment.

3.2.3. Three types of components

Recall that our simulation model has the flexibility that it can represent the process flow of any diagnostic facility consisting of preparation steps, tests, and servers. While servers are always the final process step, service can be preceded by any number of test and preparation steps in any configuration. Here, we describe these components in detail.

3.2.3.1. Preparation step. A preparation step takes a certain amount of time, after which the patient can proceed to the next step in the process. For example, before undergoing a CT-scan, certain patients need to be administered contrast fluid, either orally, intravenously (IV), or both. A preparation step has the following characteristics:

- The capacity, that is, the maximum number of patients that can be served simultaneously.
- The distribution of the processing time, specified separately for each patient type. Possible processing time distributions in the current model setup are: beta, binomial, deterministic, empirical, Erlang, gamma, normal, log-normal, triangle, uniform, or Weibull.
- The priority rule used when a new patient can be taken into service and there are multiple patients waiting. Possible priority rules in the current model setup are: first come first served (FCFS), prioritize scheduled patients, or prioritize scheduled patients unless there is an unscheduled patient who has been waiting longer than a certain threshold time.

3.2.3.2. Test. Although each patient type has a pre-specified route through the system, certain patients might need to deviate from that due to their medical condition. A test represents a point in the process at which, based on a patient's medical condition, it is decided whether the patient can continue on the pre-specified route or should be rerouted. For example, patients with weak kidney performance, indicated by a high creatinine level, may not be administered IV contrast fluid. Such patients require additional preparation before their CT-scan that cannot be performed on a walk-in basis, and therefore have to be served as scheduled patients. Thus, in this example, the simulation model contains a test applicable to all patients requiring IV contrast fluid. It represents the decision, based on the patient's creatinine level, to either let the patient continue to IV preparation or reroute the patient to the appointment scheduler. A test has all the characteristics of a preparation step, along with the following additional characteristics:

- The specification of the probability with which a patient can continue on his pre-specified route versus requiring rerouting. This is specified by a test result distribution and a threshold value θ. Let X denote the random variable drawn from the test result distribution, then the patient continues on his pre-specified route if $X \leq \theta$, and is rerouted otherwise. Possible test result distributions in the current model setup are: beta, binomial, empirical, Erlang, gamma, normal, log-normal, triangle, uniform, or Weibull.
- The specification of the patient's further route through the system in case of rerouting. In the current model setup, this route consists of one component. Therefore, currently, patients can be rerouted either to a server or to the appointment scheduler.

3.2.3.3. Server. Service is the main and final process step, and the only step for which appointments can be scheduled. After service, the patient leaves the system. A server has all the characteristics of a preparation step, along with the following additional characteristics:

- The cyclic appointment schedule, specifying for which type(s) of patients the server is available in each time slot. As opposed to tests and preparation steps, the capacity of a server is always one, such that each server can have its own appointment schedule. This gives our simulation model the flexibility to represent both facilities that have distinct servers, where certain patient types can only be served by particular servers, and facilities that have interchangeable servers. Patients who can be served by multiple servers are served by the first appropriate server that becomes available for them, irrespective of the specific server they had an appointment with. Section 3.2.6 provides details on how we generate appointment schedules for the servers.

– The breakdown probability: the probability that the server will unexpectedly and suddenly be unavailable, requiring repair in order to become operable again.
– The duration of breakdown, which is deterministic and counted in number of days in the current model setup. When a server breaks down, appointments scheduled on the days of breakdown are rescheduled as soon as possible. We assume that breakdown does not apply to tests and preparation steps.

3.2.4. Simulation execution

When all the required information (as described in Sections 3.2.1–3.2.3) has been specified and a simulation model representing a given diagnostic facility has automatically been constructed based on that (as exemplified in Figs. 2 and 3), the execution of the simulation is as follows. Patients arrive according to the specified non-stationary Poisson process. Upon a patient's arrival, his type is drawn in accordance with the percentage of patients that is of each type. If the patient is an appointment patient or an inpatient, an appointment is scheduled immediately and the patient will physically arrive at the facility on the day of the appointment. Scheduled patients are assumed to be punctual, that is, they arrive exactly on time for their appointment. Once a patient has arrived at the facility, the patient follows his pre-specified route through the process, unless rerouted due to a test. The patient may incur waiting time at each process step. When a new patient can be taken into service at a given component and there are multiple patients in its waiting room, selection occurs according to the priority rule specified for that particular component. We assume that patients do not leave the system prematurely, that is, patients stay until having been served or having received an appointment. During a patient's stay in the system, several performance indicators are recorded, which we discuss now.

3.2.5. Performance indicators

We formulate the following performance indicators for our simulation study:

– *Fraction walk-in directly served.* The fraction of walk-in patients that is served on the day of arrival, i.e., as an unscheduled patient. (Recall that walk-in patients are offered an appointment, and thus join the category of scheduled patients, if upon their arrival the expected waiting time is higher than a pre-specified norm.)
– *Access time.* The number of days from a patient's request until the examination.
– *Waiting time.* The number of minutes from the physical arrival of the patient in the waiting room for the servers until the start of the examination. (The model measures waiting times for all process steps. In this article, we only present the waiting time for the servers.)
– *Overtime.* The number of minutes examination time performed after the facility's closing time per day.
– *Server utilization.* Total examination time performed in the facility's business hours as a percentage of total examination time available from the servers during the facility's business hours.

3.2.6. Combining simulation with schedule optimization

When simulating a combined walk-in and appointment system, an appointment schedule is required for the servers, which prescribes when appointments can be scheduled for patients requiring or preferring an appointment, and for deferred walk-in patients. As pointed out in Section 2, we apply the approach developed by Kortbeek et al. [3] to generate appointment schedules. This methodology, which from now on we refer to as the walk-in schedule generator, consists of two mathematical models, linked by an iterative procedure. The first model calculates scheduled patients' access time, given the number of slots reserved for appointments on each day in the cycle. It thereby decides how many appointment slots to reserve in order to guarantee a pre-specified access time service level norm. On each day, these reservations are then assigned to specific time slots, whereupon the second model calculates the expected number of deferred walk-in patients. Because these deferred walk-in patients also require appointments, a second iteration starts, with increased scheduled patient arrival rates for the first model. This iterative procedure continues until the expected number of deferred walk-in patients, resulting from the second model, matches the number of scheduled patients anticipated for in the first model. The walk-in schedule generator's output is a vector specifying for each time slot in the cycle how many appointment reservations should be made in that time slot (i.e., each element of this vector is an integer between 0 and the number of servers). The output's generic format warrants its applicability to both facilities with distinct servers, where certain appointments may be assigned to specific servers only, and facilities with interchangeable servers.

Because the walk-in schedule generator is a mathematical methodology, several assumptions are imposed to allow for analytical tractability. Some of the assumptions and model outcomes do not exactly match actual practice at diagnostic facilities, while simulation modeling allows for a more detailed representation of reality. First, the walk-in schedule generator assumes that all patients require an examination with a deterministic length of one time slot, while in the simulation model examination durations are stochastic and may require multiple slots for certain patient types. Second, the walk-in schedule generator assumes that there is only one type of scheduled patients, while in the simulation model there may be multiple types of scheduled patients, some of which require specific time slots (e.g., due to the required presence of a medical specialist from another department). Third, the walk-in schedule generator does not take into account that patients may require preparation before their examination, and, as described in [3], it has the tendency to reserve time slots for appointments early on the day. However, because preparation cannot start before the diagnostic facility's opening time, examinations for patients requiring preparation cannot be scheduled too early on the day.

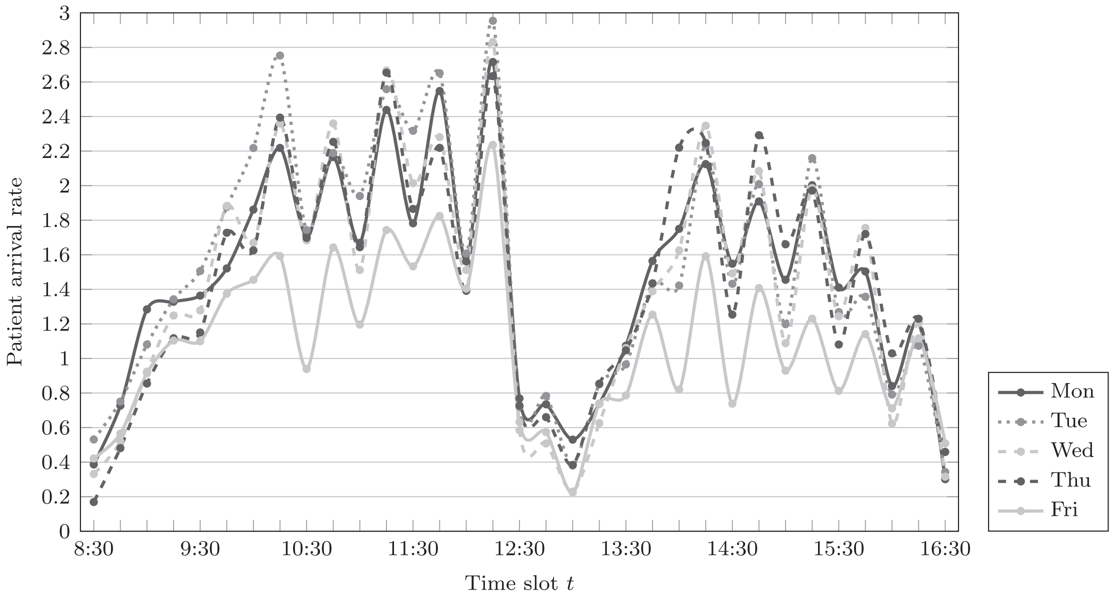

Fig. 5. Overall patient arrival rates per slot per day.

Appendix A describes our approach for dealing with these three inconsistencies between the walk-in schedule generator and the simulation model in order to obtain feasible appointment schedules for our simulation study.

4. Results – case study

In this section, we evaluate the effects of implementing a combined walk-in and appointment system for CT-scans by applying our simulation model to the CT-scan facility at the radiology department of the Academic Medical Center (AMC) in Amsterdam, a Dutch university hospital. The facility has two scanners that jointly perform approximately 12,000 examinations per year for outpatients and inpatients. Emergency patients are scanned on a dedicated CT-scanner at the emergency department and are therefore excluded from this study. The business hours of the CT-scan facility are from 8:30 to 16:45 on each weekday, with time divided in 15-minute slots. When this study started, the facility employed an appointment system, while management considered the implementation of a combined walk-in and appointment system.

We deduce the input data for our model from the radiology information system (RIS/PACS, Agfa HealthCare, Mortsel, Belgium), which contains detailed information about all CT-scans performed, and the hospital-wide electronic calendar system X/Care (McKesson, San Francisco, CA), which contains the appointments scheduled for CT-scans. We collected data over the period May 2012 until April 2013. For data not accurately registered in these systems, we performed measurements in September 2013.

The structure of this section is as follows. Section 4.1 provides additional information on the case study. We describe our experimental design in Section 4.2. Before presenting the numerical results in Section 4.4, we elaborate on the verification and validation of our simulation model in Section 4.3. The simulation model was implemented in Tecnomatix Plant Simulation 12 (Siemens, Plano, TX) and tested on an Intel 2.6 GHz PC with 16.0 GB of RAM.

4.1. Case study description

Fig. 5 displays the joint patient arrival rates per time slot and per day, as used for the AMC CT-facility in our experiments. Because the facility employed a complete appointment system when this study started, there was data on total patient arrival rates, but not distinguished by day and time slot. Therefore, we estimate the arrival rates per day and time slot as follows: for the 11 most-referring outpatient clinics with respect to CT-scans (jointly accounting for 81.44% of all outpatient CT-scan requests), we calculate the average number of patient consultations in each time slot on each weekday and multiply this by the specialty's CT-referral probability. We incorporate a 30-minute delay between the outpatient consultation and the patient's arrival at the CT-facility to account for patient transfer times, and extrapolate the resulting arrival pattern to match the total number of yearly CT-scan requests. Note that, as a natural consequence of the cyclic weekly pattern of the outpatient clinics' appointment schedules, the estimated arrival rates follow a cyclic weekly pattern. The no-show probability of scheduled patients is 2.60%. Based on a patient preference survey conducted at the AMC CT-facility with respect to implementing a combined walk-in and appointment system, walk-in patients are offered an appointment, and thus join the category of scheduled patients, when their expected waiting time is over 1.5 hours.

Table 1
Input data patient types.

Patient type		%	Arrival mode	Route					# CT slots	CT processing time (in min.)		
				Prep1Lab*	Test1GFR	Prep2IV	Prep3Oral	Server (CT)		Distribution	Min	Max
1	IV-Oral60	19.09	Walk-in	1	2	3	4	5	1	$\ln\mathcal{N}(2.35, 0.23)$	4.5	40.0
2	IV-Oral30	5.04	Walk-in	1	2	3	4	5	1	$\ln\mathcal{N}(2.35, 0.23)$	4.5	40.0
3	IV	15.71	Walk-in	1	2	3		4	1	$\ln\mathcal{N}(2.35, 0.23)$	4.5	40.0
4	Oral30	0.34	Walk-in				1	2	1	$\ln\mathcal{N}(2.35, 0.23)$	4.5	40.0
5	Outpatient	37.30	Walk-in					1	1	$\ln\mathcal{N}(2.35, 0.23)$	4.5	40.0
6	Colonography	1.23	Appointment					1	2	$\mathcal{N}(26.80, 56.75)$	3.0	60.0
7	Puncture	1.43	Appointment					1	3	Erl(0.12,8.00)	25.0	150.0
8	Cardiac	7.30	Appointment					1	1	$\ln\mathcal{N}(2.41, 0.33)$	3.0	30.0
9	Inpatient	12.56	Inpatient					1	1	$\ln\mathcal{N}(2.35, 0.23)$	4.5	40.0

*A lab visit is only required for the 3.00% patients with unknown creatinine level.

We distinguish nine patient types (see Table 1), differentiated based on patients' arrival modes, flows through the process, and specific appointment slot requirements for certain patients. Three different arrival modes can be distinguished: patients being eligible for walk-in (77.48% of patients), outpatients requiring specific appointment slots (9.96%), and inpatients (12.56%). The outpatients requiring specific appointment slots are subdivided into three distinct categories: colonography examinations requiring preparation starting three days in advance (patient type 6), puncture examinations, for which patients need to be anesthetized (patient type 7), and cardiac examinations requiring the presence of a cardiologist (patient type 8). Further, inpatients are patient type 9, and we distinguish five types of patients eligible for walk-in by different flows through the process (patient types 1–5). Note that the process flows for patient types 1 and 2 differ due to different oral contrast fluid preparation durations. Patients of type 1 have to start drinking contrast fluid at least 60 minutes before their CT-scan (referred to by 'Oral60' in Table 1), whereas patients of type 2 have to start drinking water at least 30 minutes before their CT-scan ('Oral30'). For each patient type, Table 1 displays its route through the system, as well as the number of time slots reserved in practice when scheduling an appointment for a patient of that type. Additionally, the table displays the CT processing time distribution per patient type, which is either log-normal $\left(\ln\mathcal{N}(\mu, \sigma^2)\right)$, normal $\left(\mathcal{N}(\mu, \sigma^2)\right)$, or Erlang (Erl($\lambda$, n)), and equal for both servers. We use truncated distributions with the minimum and maximum as listed. We used the Datafit package that is part of Tecnomatix Plant Simulation 10.1 (Siemens, Plano, TX) to determine the processing time distributions for our experiments, selecting the distribution with the best quality of fit according to the Kolmogorov-Smirnov (KS) test, while ensuring that the Anderson-Darling test also supported the selection of that distribution (both at significance level $\alpha = 0.05$). All walk-in patients and inpatients share the same CT processing time distribution (KS-test statistic: $D = 0.66$; critical value: 1.358), whereas the distributions for outpatients requiring specific appointment slots (patient types 6 to 8) are different due to the additional complexity involved in scanning these patients ($D_6 = 0.65, D_7 = 0.66, D_8 = 0.80$). While certain inpatients may also require tests or preparation steps before their CT-scan, these are not included in the table since inpatients receive their preparation at their inpatient unit and thus arrive at the CT-facility prepared. When scheduling an appointment for a patient, recall that the appointment scheduler could search for the first appointment slot from the start of the horizon forwards, or from the end of the horizon backwards. In our case, the appointment scheduler always searches from the start of the horizon forwards.

The CT-scan process at the AMC, as depicted in Fig. 3, consists of one test, three preparation steps, and two servers. The first preparation step is a lab visit, to determine the creatinine level of patients requiring IV contrast fluid but lacking an up-to-date creatinine measurement. (Creatinine measurements are considered up-to-date when these are less than one year old at the time of the CT-scan; 3.00% of the patients requiring IV contrast lack up-to-date creatinine measurements.) This preparation step has capacity 10 (i.e., at most 10 patients can undergo this preparation step simultaneously) and the processing time is deterministic at 60 minutes. The next process step, applying to all patients requiring IV contrast fluid, is a test of the patient's creatinine level. If the creatinine level is below a given maximum, the patient can continue to IV preparation. Otherwise, the patient is rerouted to the appointment scheduler. Because this test is merely a simple check that determines the patient's further route, it effectively has a processing time of 0 minutes and infinite capacity. Its result is uniformly distributed, $U(0, 1)$, with threshold value $\theta = 0.94$ (see Section 3.2.3 for the definition of a test's threshold value). The second preparation step is the placement of an IV line, with capacity 1 and empirically distributed processing time with mean 8.85 minutes. Drinking oral contrast fluid is the third preparation step, with capacity 10 and a deterministic processing time, as displayed in Table 1 (using 'Oral60' and 'Oral30'). Patients are instructed to distribute the intake of contrast fluid evenly over the specified time interval and to hold on to their last cup until being called for their CT-scan. These instructions ensure that contrast fluid is adequately distributed in the bowels, also if patients experience waiting time between this preparation step and their CT-scan. For the test and all three preparation steps, FCFS is used as the priority rule for selecting which patient from the waiting room to serve next. The final process step consists of the two servers, CT1 and CT2, both with capacity 1. Patients requiring a CT-scan without contrast fluid can proceed to the servers immediately after arriving at the CT-facility. Recall that the processing time distribution for the servers varies for different patient types (see Table 1). To be able to evaluate the effect of implementing a combined walk-in and appointment system purely, uncontaminated by other

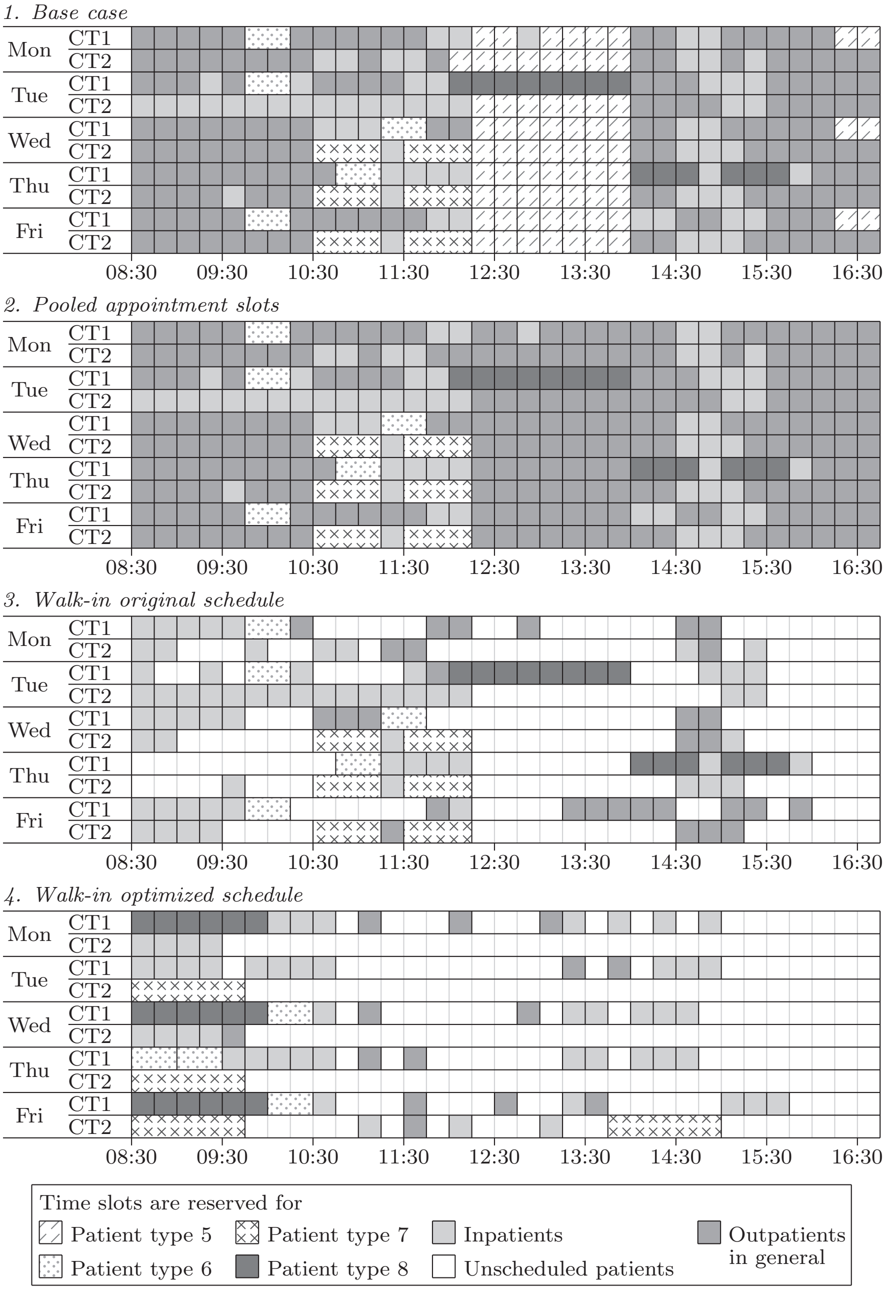

Fig. 6. CT-scanner appointment schedules for Experiments 1–4.

disruptive effects, we set the breakdown probability for the servers to 0. Because we use different appointment schedules for the servers in different experiments, we present the appointment schedules along with our experimental design (in Section 4.2). When a server becomes idle and the next patient has to be selected from the waiting room, scheduled patients receive priority over unscheduled patients.

4.2. Experimental design

This section describes the twelve experiments we perform to evaluate the effects of implementing a combined walk-in and appointment system for CT-scans. Figs. 6 to 8 display the appointment schedules used for the CT-scanners in Experiments 1 to 4, 5 to 8, and 9 to 12, respectively.

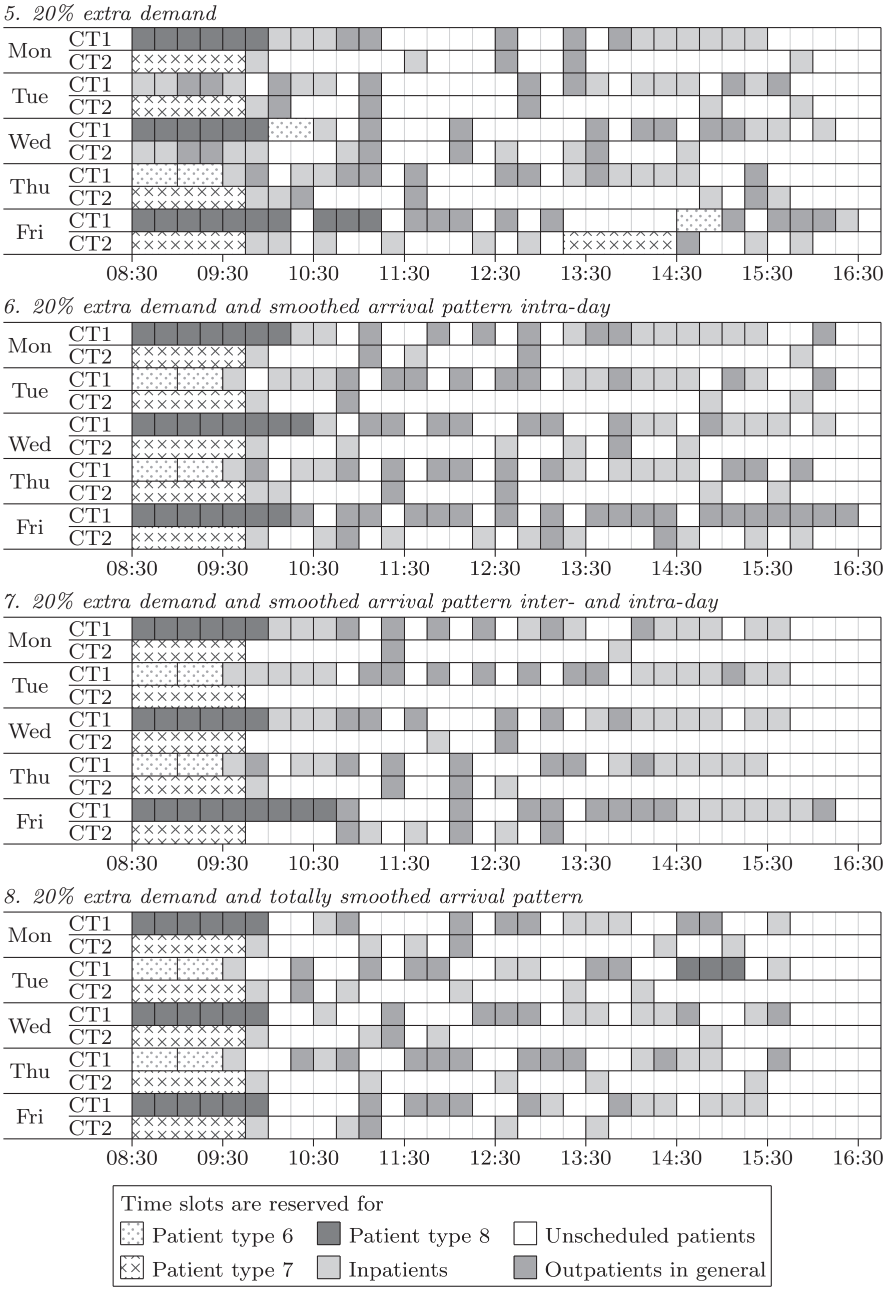

Fig. 7. CT-scanner appointment schedules for Experiments 5–8.

(1) Base case. We first evaluate the performance of the base case scenario, as described in Section 4.1, which resembles current practice at the time this study started. In the base case scenario, all patients are scheduled; there are dedicated appointment slots for inpatients and for patient types 5 to 8, and general appointment slots for outpatient types 1 to 4.

(2) Pooled appointment slots. Reserving dedicated appointment slots for certain patient types, as in Experiment 1, entails efficiency loss and performance decrease compared to a pooled situation which benefits from the portfolio effect. For example, dedicated appointment slots for a patient type A might all be filled in a given week, thereby increasing access time for type A patients, whereas dedicated appointment slots for a patient type B in the same week might remain empty, thereby decreasing utilization. In a pooled situation, type A patients could have used the 'type B' free slots, resulting in an improvement in both access time and utilization. To enable a fair comparison between a full appointment system and a combined walk-in and appointment system, we experiment with a full appointment system

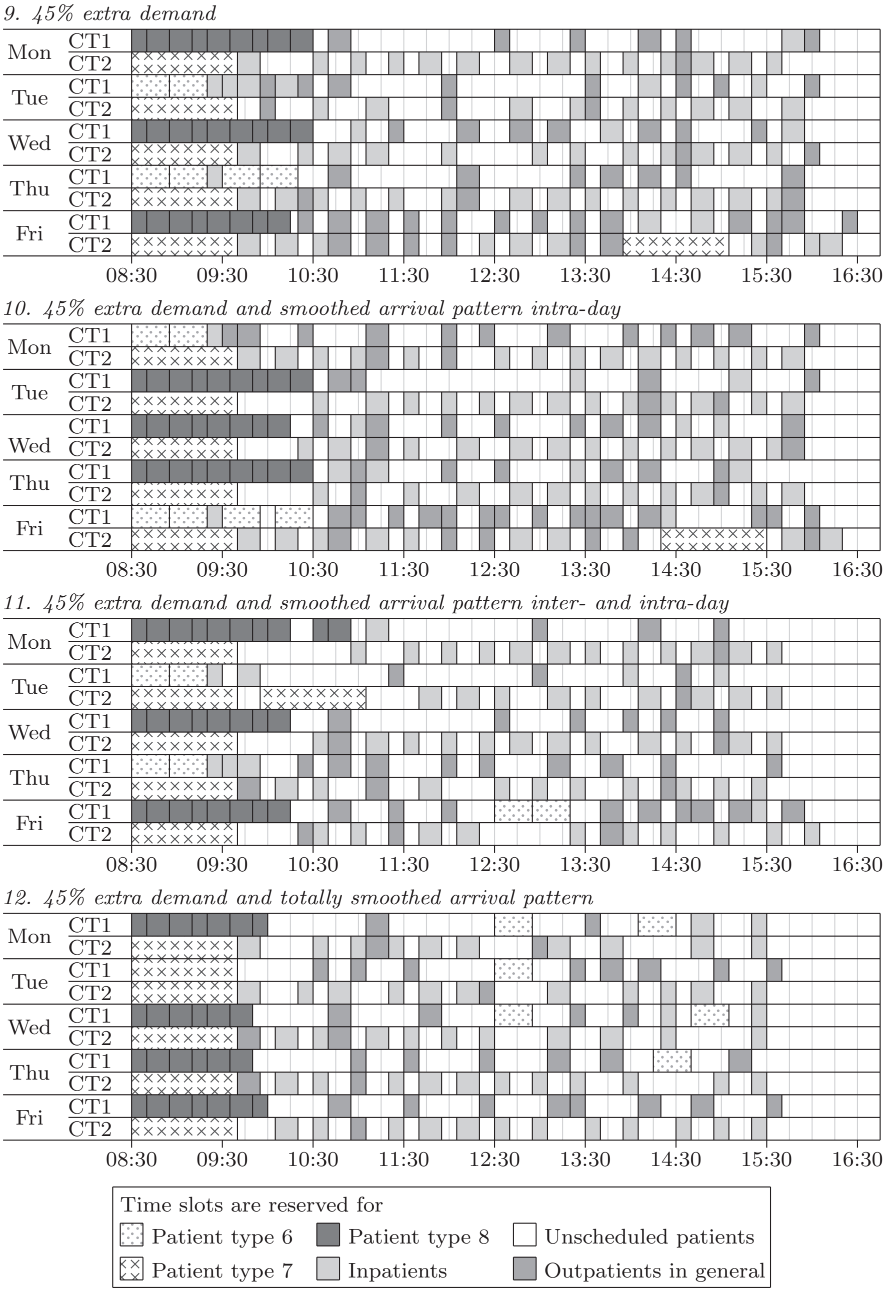

Fig. 8. CT-scanner appointment schedules for Experiments 9–12.

in which all patient types eventually eligible for walk-in share general appointment slots, while specific reservations are only kept for patient types strictly requiring appointments (i.e., types 6 to 8, and inpatients).

(3) *Walk-in original schedule.* In this experiment, we enable all eligible patients to walk in. Patient types strictly requiring appointments keep the dedicated slots they had when this study started. A small number of general appointment slots is required for walk-in patients who turn out to be ineligible for unscheduled service. We use the methodology by Kortbeek et al. [3] to determine how many and which slots to reserve as general outpatient appointment slots. The tendency of this methodology to reserve slots early on the day, as described in Section 3.2.6, conflicts with scheduling appointments for patients with lengthy preparations, which is indeed a problem in this particular experiment. We resolve this issue by exchanging each early general outpatient appointment slot for an inpatient slot later on the same day.

(4) Walk-in optimized schedule. In the first three experiments, we adhere to some form of the appointment schedule that was employed by the AMC CT-facility when this study started. Now, to evaluate the full potential of jointly implementing a combined walk-in and appointment system and optimizing the appointment schedule, we use the methodology by Kortbeek et al. [3] to redesign the entire appointment schedule. As this methodology includes adherence to a pre-specified access norm for scheduled patients, both the number of slots reserved for each patient type and the timing of these slots differ from Experiment 3. In line with AMC policy, we set the access time service level norm such that 95% of the scheduled patients have their examination within two weeks (i.e., ten days). For inpatients, who are preferably offered a same-day appointment, we require that 95% have their examination within one day. In addition, note the mismatch between the 3 time slots (i.e., 45 minutes) reserved for a puncture patient (patient type 7) in the CT-facility's original schedule, and this patient type's CT processing time distribution, which has a mean of 67.6 minutes. Reserving too few slots for a puncture patient involves the risks of significant waiting times for subsequent patients, and underestimating the total workload on days with puncture appointments, resulting in walk-in patient deferrals that could have been avoided. Therefore, along with replacing the CT-facility's original schedule by an optimized schedule, from this experiment on we extend the number of time slots reserved for a puncture patient from 3 to 5 slots.

(5) 20% extra demand. The improved accessibility and patient service inherent to a combined walk-in and appointment system may result in an increase in patient arrival rates [24]. Therefore, we evaluate the ability of our combined walk-in and appointment system to cope with a 20% increase in arrival rates, by multiplying the patient arrival rate in each time slot (Fig. 5) by 1.2 and again redesigning the entire appointment schedule.

(6) 20% extra demand and smoothed arrival pattern intra-day. The efficiency of a walk-in system depends on the variability in patient arrivals, because highly variable arrivals result in highly variable system utilization and high patient waiting times during busy periods. In Experiments 6 to 8, we investigate how much a hospital could gain by balancing the various outpatient clinics' consultation hours, thereby smoothing walk-in arrival rates. In this experiment, starting from the arrival rates of Experiment 5, we smooth the arrival pattern on each day as follows. Let χ_t^d be the patient arrival rate in time slot t on day d, and $\bar{\chi}^d$ the average arrival rate per time slot on day d. Now we define the new arrival rate per time slot as $\hat{\chi}_t^d = (1 - 0.5344)\chi_t^d + 0.5344\bar{\chi}^d$, such that one in every ten patients originally arriving in a relatively busy time slot will now arrive in a relatively quiet time slot, where 0.5344 is the multiplication factor that corresponds to achieving this effect. (Instead of adjusting the planning of outpatient clinics' consultation hours, a similar smoothing effect might also be achievable through informing patients on busy and quiet walk-in time slots, and encouraging them to choose a quiet time slot.)

(7) 20% extra demand and smoothed arrival pattern inter- and intra-day. In addition to smoothing the arrival pattern on each day as in Experiment 6, in this experiment we also smooth the arrival pattern in each week. Let ξ^d be the total patient arrival rate on day d, and $\bar{\xi}$ the average daily arrival rate over all days. We move two in every 100 patients from a relatively busy to a relatively quiet day by defining the new arrival rate per day as $\hat{\xi}^d = (1 - 0.4659)\xi^d + 0.4659\bar{\xi}$.

(8) 20% extra demand and totally smoothed arrival pattern. For the purpose of insight, in this experiment we explore the effect of a totally smoothed arrival pattern. We define the arrival rate in each time slot on each day to be the overall average arrival rate (i.e., $\hat{\chi}_t^d = \frac{1}{d \cdot t} \cdot \sum_{d,t} \chi_t^d$).

(9)-(12) 45% extra demand. To investigate the sensitivity of the system's performance to its utilization, we repeat Experiments 5 to 8 with a 45% increase in arrival rates compared to the base case. Notice that in our case study, 97.34% of patients have an expected CT processing time of 12 minutes (see Table 1). The AMC CT-facility's 15-minute slots thus introduce on average 20% slack time for the majority of scheduled appointments. To accommodate the 45% increase in arrival rates, we allow appointments to start at each 5-minute interval in Experiments 9 to 12. We reserve 25 and 70 minutes for type 6 and 7 appointments, respectively, and make alternating 10- and 15-minute reservations for all other patient types, as displayed in Fig. 8.

4.3. Verification and validation

We verified our simulation model using the eight techniques suggested by Law and Kelton [25]: debugging in modules, having multiple authors review the computer program, running the simulation under a variety of settings of the input parameters and checking that the output is reasonable, tracing the state of the simulated system and comparing it with hand calculations, running the model under simplifying assumptions for which its true characteristics can easily be computed, observing animations of the simulation output, computing the sample mean and variance for each simulation input probability distribution and comparing these with the historical mean and variance, and using a commercial simulation package to reduce the amount of programming required.

Based on an analysis of all performance indicators (see Section 3.2.5) for 20 test runs, we set the warm-up period and the run length for our experimentation. The warm-up period is determined by applying Welch's procedure [26] and is set to 26 weeks. Using Robinson's graphical method [27], we set the run length (including the warm-up period) to 10.5 years (546 weeks). The number of replications is set based on applying the confidence interval method [21] with a desired half-width of 5% for the 95% confidence intervals (aiming for sufficiently narrow confidence intervals, having a range no larger than

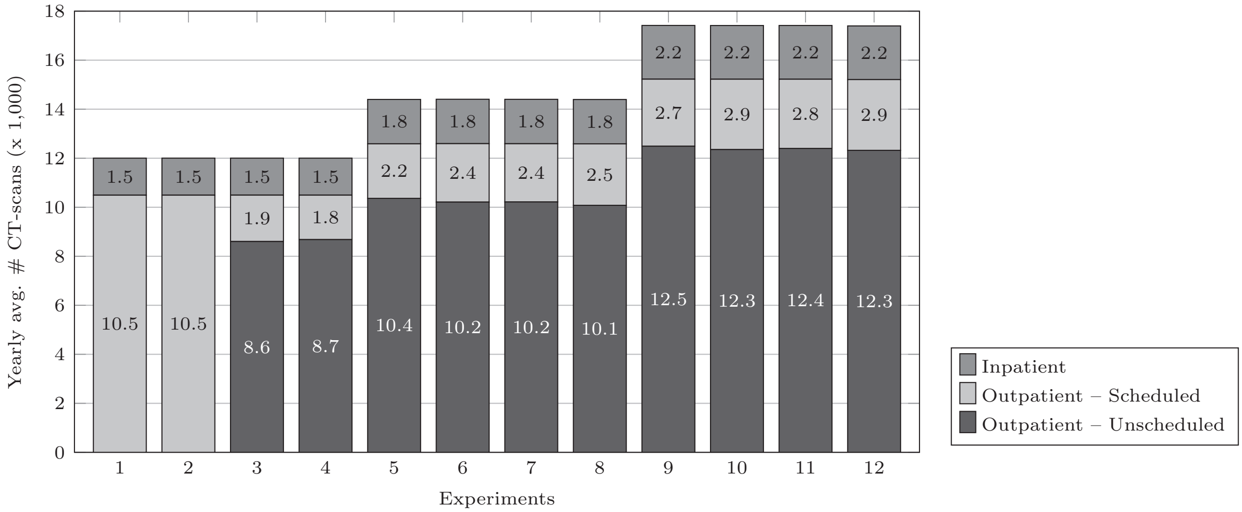

Fig. 9. For each experiment: average number of CT-scans performed yearly for inpatients, scheduled outpatients, and unscheduled outpatients. Every unscheduled patient can save at least one hospital visit.

10% of their average). In addition to all performance indicators, we include the total realized processing time in hours into our analysis for determining the number of replications. Because this is a large number (approximately 25,000, 30,000, and 37,000 for Experiments 1–4, 5–8, and 9–12, respectively), we require a half-width less than 0.15% for the 95% confidence interval of the processing time. Based on this analysis, we set the number of replications at 40.

We validate our simulation model by statistically comparing current practice performance, as measured in September 2013, to simulation model performance, using 95% Welch confidence intervals [25]. The average waiting time is 4.32 and 6.20 minutes for practice and simulation, respectively, with confidence interval for the difference of $[-2.47, -1.30]$. Because $0 \notin [-2.47, -1.30]$, we conclude that the difference between the average waiting times is statistically significant; our simulation model overestimates waiting times, and is thus conservative with respect to this performance indicator. The average daily overtime is 4.40 and 4.61 minutes for practice and simulation, respectively, with confidence interval for the difference of $[-4.70, 4.28]$. Average server utilization is 55.10% and 58.12% for practice and simulation, respectively, with confidence interval for the difference of $[-9.30, 3.26]$. As 0 is contained in both these confidence intervals, we conclude that there are no statistically significant differences in the average overtime and server utilization. The simulation's average access time of 5.37 days cannot be validated statistically, because in practice data it is unclear whether a relatively long access time is the result of the patient's or medical preferences, or caused by a full appointment schedule. However, according to expert opinions of the appointment schedulers at the AMC CT-facility, average access time was indeed around five days in September 2013. We exclude the fraction of walk-in patients that is served on the day of arrival (i.e., as unscheduled patients) from our validation, because in our historical comparison month (September 2013), the CT-facility of the AMC employed a complete appointment system.

Note that there will be differences between the simulation output reported here for validation, and the output of the base case scenario (Experiment 1) in the next section. In September 2013, when practice measurements were performed, the CT-facility at the AMC had two scanners: a Philips Brilliance and a Philips MX8000, the first being newer and having a lower radiation dosage and higher quality images, such that this was the only acceptable machine for certain patient types. The simulation experiments we performed for validation resemble this situation. Before implementing a combined walk-in and appointment system, however, facility management intended to replace the Philips MX8000 by a new scanner, such that all patient types could be scanned on both machines, which is thus the underlying assumption for all other experiments.

4.4. Experimental results

Table 2 displays the results of the twelve experiments described in Section 4.2. For definitions of the performance indicators, the reader is referred to Section 3.2.5.

The implementation of a combined walk-in and appointment system significantly contributes to patient-centeredness, as it provides patients with a one-stop-shop. Over 90% of walk-in patients are served as unscheduled patients in all experiments, and thus save a hospital visit. Taken one step further, specific types of patients, for whom the evaluation of imaging results and the follow-up consultation with the referring medical specialist take place on the same day, save two hospital visits. At the AMC, for example, this is currently the case for patients undergoing breast or gastrointestinal cancer screening. Fig. 9 displays the number of patients scanned yearly as unscheduled outpatients, scheduled outpatients, or inpatients. It shows that implementing a combined walk-in and appointment system at the AMC CT-facility enables 8600 patients (Ex-

Table 2
Numerical results of the twelve experiments. (In addition to the confidence intervals presented in this table, Fig. 10 presents the distributions for access time and waiting time, and Fig. 11 presents the distribution for overtime.)

Performance indicators	Experiments											
	1	2	3	4	5	6	7	8	9	10	11	12
Walk-in directly served (%)												
incl. medical deferrals	-	-	92.47 ± .03	93.31 ± .02	92.89 ± .03	91.46 ± .03	91.54 ± .03	90.35 ± .04	92.55 ± .03	91.56 ± .03	91.89 ± .03	91.39 ± .03
excl. medical deferrals	-	-	98.61 ± .02	99.53 ± .01	99.05 ± .02	98.48 ± .02	98.59 ± .02	98.09 ± .01	98.38 ± .03	98.07 ± .03	98.42 ± .03	98.44 ± .03
Access time (in days)	4.70 ± .02	2.27 ± .00	1.36 ± .00	0.64 ± .01	0.50 ± .01	0.51 ± .01	0.66 ± .02	0.70 ± .02	0.48 ± .01	0.51 ± .01	0.51 ± .01	0.50 ± .01
Per patient category:												
Scheduled outpatients	5.34 ± .02	2.56 ± .00	8.04 ± .02	3.91 ± .08	3.01 ± .05	2.90 ± .04	3.83 ± .13	3.70 ± .13	2.92 ± .07	2.92 ± .05	2.93 ± .05	2.75 ± .06
Inpatients	0.27 ± .00	0.27 ± .00	0.71 ± .00	0.38 ± .00	0.27 ± .00	0.26 ± .00	0.25 ± .00	0.42 ± .00	0.19 ± .00	0.23 ± .00	0.28 ± .00	0.33 ± .00
Unscheduled outpatients	-	-	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00
Waiting time (in min.)	2.51 ± .02	2.43 ± .02	9.04 ± .06	3.95 ± .02	7.59 ± .04	8.00 ± .04	7.71 ± .04	7.72 ± .04	18.10 ± .10	18.28 ± .09	16.96 ± .10	17.48 ± .09
Per patient category:												
Scheduled outpatients	2.60 ± .02	2.53 ± .02	5.74 ± .03	2.16 ± .02	3.91 ± .03	4.45 ± .03	3.99 ± .03	4.18 ± .02	6.21 ± .04	6.57 ± .04	6.23 ± .03	6.80 ± .04
Inpatients	1.85 ± .02	1.77 ± .02	4.95 ± .04	1.85 ± .01	3.55 ± .02	3.58 ± .02	3.41 ± .02	3.44 ± .01	6.82 ± .03	6.41 ± .03	5.99 ± .02	6.49 ± .03
Unscheduled outpatients	-	-	10.49 ± .07	4.69 ± .03	9.08 ± .05	9.61 ± .05	9.33 ± .06	9.37 ± .05	22.68 ± .13	23.10 ± .12	21.35 ± .13	21.93 ± .13
Overtime (in min.)	4.47 ± .04	3.54 ± .05	4.03 ± .04	3.15 ± .05	5.35 ± .06	5.04 ± .06	4.68 ± .06	4.53 ± .04	16.99 ± .13	17.31 ± .14	15.91 ± .12	14.97 ± .12
Utilization (%)	58.13 ± .07	58.25 ± .07	59.20 ± .07	59.30 ± .07	70.98 ± .08	71.03 ± .07	71.04 ± .07	71.01 ± .07	84.78 ± .08	84.72 ± .08	84.86 ± .07	84.90 ± .08

Names of the experiments

1 - Base case	5 – 20% extra demand	9 – 45% extra demand
2 - Pooled appointment slots	6 – (5) & smoothed arrival pattern intra-day	10 – (9) & smoothed arrival pattern intra-day
3 - Walk-in original schedule	7 – (5) & smoothed arrival pattern inter- and intra-day	11 – (9) & smoothed arrival pattern inter- and intra-day
4 - Walk-in optimized schedule	8 – (5) & totally smoothed arrival pattern	12 – (9) & totally smoothed arrival pattern

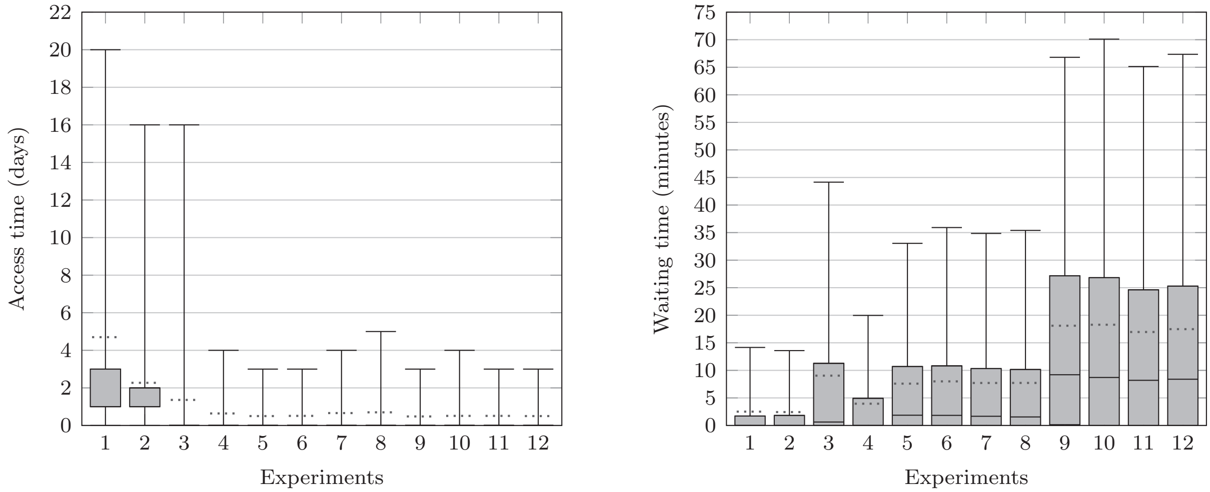

Fig. 10. Distribution of patients' access times (left) and waiting times (right) in the different experiments. The box plots display the 5th percentile (lower whisker), 25th percentile (box bottom), median (solid line), mean (dotted line), 75th percentile (box top), and 95th percentile (upper whisker). In several cases the box is not visible because the 75th percentile is 0.

periment 3) to save atleast one hospital visit, while schedule optimization increases this number to 8700 (Experiment 4). Table 2 displays two performance metrics for the percentage of walk-in patients that can be served on the day of arrival (i.e., as unscheduled patients). A walk-in patient can be offered an appointment, and thus become a scheduled patient, either for a medical or for a logistical reason. Medical reasons in our case are: the patient's creatinine level is too high, or a patient walks in so close to the facility's closing time that he is expected to finish preparation after the facility's business hours. Such 'medical deferrals' are unavoidable for the CT-facility. In the first performance metric displayed in Table 2, we do count these patients as deferred walk-in patients. In the second metric, these patients are not taken into account, because they are not regarded as 'true' walk-in patients. From the experiments we conclude that over 98% of true walk-in patients can indeed be served as unscheduled patients (Experiment 3), while schedule optimization yields an increase to over 99% (Experiment 4). Enhancing both quality of care and patient-centeredness, a combined walk-in and appointment system reduces access times. Table 2 and Fig. 10 display the access time confidence interval and distribution, respectively. For the AMC CT-facility, the average access time reduces from 4.70 days in the base case (Experiment 1) to less than one day in Experiment 4. A reduction from 4.70 to 2.27 days can be achieved by reducing the amount of dedicated appointment slots (Experiment 2), and the additional access time reduction to 0.64 days is the result of implementing a combined walk-in and appointment system. While the access time is eliminated completely for unscheduled patients, the average access time for scheduled patients is 3.91 days in Experiment 4, slightly lower than in the base case. This decrease is the result of utilizing the walk-in schedule generator to determine the number of appointment slots per patient type that guarantees an access time service level of 95% within ten days. Note, from comparing the results of Experiments 3 and 4, that schedule optimization is required in order to prevent the access time for scheduled patients from increasing when implementing a combined walk-in and appointment system. Additionally, schedule optimization significantly improves the tail of the access time distribution (see Fig. 10) and enables adherence to the AMC's access time targets. Although average inpatient access time is slightly higher in the combined walk-in and appointment setting (Experiments 3 and 4) compared to the base case, when employing schedule optimization the majority of inpatients can still be offered a same-day appointment, as preferred (Experiment 4). Patients' waiting times, for which also both the confidence interval (Table 2) and the distribution (Fig. 10) are displayed, are very reasonable in a combined walk-in and appointment system at the AMC CT-facility. Again, schedule optimization improves the system's performance significantly, given the comparison between the results of Experiments 3 and 4. By employing schedule optimization, average waiting time performance similar to the base case can be achieved for inpatients and scheduled outpatients. Observe from Fig. 10 that the tail of the waiting time distribution does increase, but still 75% of all patients experience a waiting time less than 5 minutes, and the waiting time is below 20 minutes for at least 95% of all patients. For unscheduled patients, an average waiting time below five minutes can be realized under the basic demand (Experiment 4).

The daily overtime confidence interval and distribution are displayed in Table 2 and Fig. 11, respectively. For the AMC CT-facility, daily overtime stays similar to the base case when implementing a combined walk-in and appointment system (Experiment 3), while schedule optimization improves the average as well as the tail of the distribution (Experiment 4).

While Experiments 1 to 4 all have the exact same demand, Table 2 displays that the realized system load is slightly higher in Experiments 3 and 4 compared to Experiments 1 and 2. (Where realized system load consists of all work done during business hours – displayed via utilization – and all work done in overtime.) The reason why a higher system load can be realized in Experiments 3 and 4, is that walk-in eliminates no-shows. In other words, the lower the percentage of

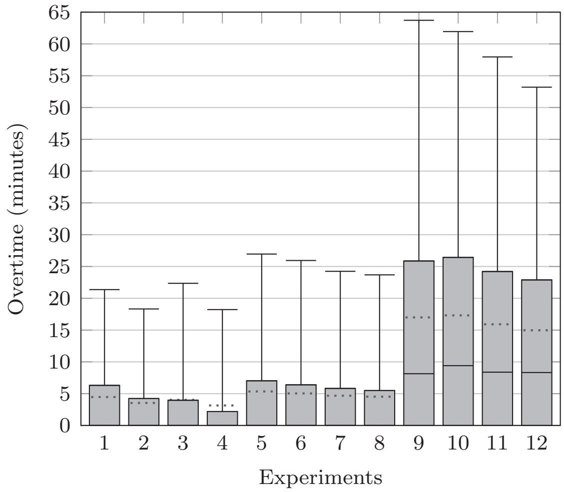

Fig. 11. Distribution of daily overtime in the different experiments.

scheduled patients, the lower the loss of demand due to no-shows, and the higher the system load that can be realized. The assumption to have no-show patients leave our simulation model instead of regenerating a patient arrival was established together with the AMC CT-facility.

From the results of Experiments 5 and 9, with 20% and 45% extra demand, respectively, we conclude that implementing a combined walk-in and appointment system while applying schedule optimization enables an increase in productivity, while patients can still be provided with high-quality service. Note that Experiment 5, with 20% extra demand, outperforms Experiment 3, with regular demand but lacking schedule optimization, on nearly all performance indicators. Only the median waiting time (Fig. 10) and the overtime average (Table 2) and tail (Fig. 11) are slightly worse in Experiment 5. The access time averages and distributions of Experiments 5 and 9 even outperform those of Experiment 4 – Experiments 4, 5, and 9 are equivalent except for demand, which is regular, 20% extra, and 45% extra, respectively. This is due to the fact that for the more heavily utilized cases of Experiments 5 and 9, the walk-in schedule generator reserves more buffer capacity than for Experiment 4, to ensure that the access time service level norm can always be met. Thus, not only is the total number of appointment slots greater in Experiments 5 and 9, but the average number of appointment slots available per arriving appointment request is also higher. Notice that despite this increase in access time performance combined with the higher demand, the percentage of walk-in patients served on the day of arrival (i.e., as unscheduled patients) decreases by no more than 1.2 percentage point. The increased demand does result in longer waiting times, with considerable impact on unscheduled patients, while, rightly, average waiting times for scheduled patients increase only slightly, to 3.91 and 6.21 minutes in Experiments 5 and 9, respectively (see Table 2). Given that unscheduled patients can save a hospital visit, thus eliminating the time of a commute to the hospital, waiting times less than 11 and 28 minutes for 75% of patients and less than 34 and 67 minutes for 95% of patients in Experiments 5 and 9, respectively, argue in favor of a combined walk-in and appointment system. Recall from Section 4.1 that patients indicated a willingness to wait up to 1.5 hours when being enabled to walk in. Daily overtime also increases with demand. While the 75th and 95th overtime percentiles of Experiment 5 exceed the base case (Experiment 1) slightly, the 45% extra demand of Experiment 9 results in considerable increases in overtime (see Fig. 11). Setting a certain time before closing time at which the facility stops accepting walk-in patients – to be specified by facility management – could decrease overtime, but at the cost of decreasing the percentage of walk-in patients served on the day of arrival, and deteriorating access time results. To put the increases in overtime and waiting time into perspective, observe from Fig. 9 that only the number of unscheduled patients in Experiment 9 exceeds the total number of patients in Experiment 4.

For the experiments with smoothed walk-in arrival patterns (Experiments 6 to 8 and 10 to 12), Fig. 9 displays the counter-intuitive effect of fewer patients receiving unscheduled service than in Experiments 5 and 9, respectively. Table 2 shows that this is due to an increased number of medical deferrals in Experiments 6 to 8 and 10 to 12. Smoothing the arrival pattern increases patient arrival rates in the relatively quiet time slots towards the end of each day, resulting in a slightly higher number of patients not being able to finish their preparation before the facility's closing time. As displayed by Fig. 11, smoothing the arrival pattern has a positive effect on the tail of the overtime distribution. Note that apart from overtime, neither Experiments 6 and 7 nor the totally smoothed case of Experiment 8 yield notable performance improvements compared to Experiment 5. The same holds true when comparing Experiments 10, 11, and 12 to Experiment 9. We conclude that the walk-in schedule generator from [3] very adequately counterbalances the variability in walk-in arrivals. For our case study, smoothing walk-in arrival patterns does not yield additional benefits on top of that. However, the effectiveness of smoothing is highly case specific, because it depends on the combination of system utilization, the original shape of walk-in arrival patterns (including the degree of variability), and the number of time slots walk-in patients are allowed or willing to wait for service. The slight differences in access time and waiting time performance between Experiments 5 to 8 and between Experiments 9 to 12 are the result of a complex interaction between patient arrival patterns, access time service level norms, and the appointment schedules generated by the walk-in schedule generator. Observe from Figs. 7 and 8 that despite the total demand being exactly equal in Experiments 5 to 8 and Experiments 9 to 12, respectively, CT-scanner appointment schedules are notably different, naturally affecting the access time and waiting time performance.

5. Discussion

In this study, we have developed a reusable computer simulation model for evaluating the effects of a combined walk-in and appointment system for diagnostic examinations, combined that with an intelligent algorithmic methodology for appointment schedule optimization, and applied our approach to quantitatively investigate the impact of implementing a combined walk-in and appointment system for CT-scans.

By setting up the simulation model according to the component reuse approach, and defining generic preparation, test, and server components, the developed simulation model has the flexibility to represent any diagnostic examination process, consisting of any number of preparation steps and tests in any configuration, with the servers being the final process step. To maximize user-friendliness, no computer programming is required to make the simulation model represent a particular diagnostic examination process. All a user has to do is specify the required input data in the Excel front end. Based on this information, the developed simulation model automatically generates a model instance representing the user-specified diagnostic facility, thus minimizing the time and effort required for reuse. While the simulation model has been completely verified in this study, validation should be performed each time the model is applied [22,23]. Although the model's credibility increases with each successful validation, the user always has to ensure it is also valid for his particular purpose.

Our quantitative investigation reveals that for CT-scans, a combined walk-in and appointment system does indeed achieve its potential in terms of quality of care, patient service, and system efficiency. For the CT-facility in the AMC, average access times can be reduced from around five days to less than one, thus shortening patients' diagnostic trajectories by four days on average. Almost all patients eligible for walk-in can be served on the day of arrival, while experiencing very acceptable waiting times, such that patients can save one, and possibly even two hospital visits. As an academic medical center, the AMC treats patients from all over the Netherlands and abroad, with considerable travel distances, making this one-stop-shop principle very attractive. The attractiveness of a combined walk-in and appointment system is that this principle is not imposed; patients who are not able or willing to have their CT-scan performed same-day have the autonomy to choose to walk in on another day or have an appointment scheduled. A combined walk-in and appointment system also enhances system efficiency, as our experiments demonstrate that with the given capacity, the AMC CT-facility can examine an increased number of patients, while providing them with high-quality service and average access times less than one day.

The best performance of a combined walk-in and appointment system is achieved when employing appointment schedule optimization. Our experiments indicate that the methodology developed in [3] very adequately counterbalances the high variability in walk-in arrival rates. A combined system with schedule optimization clearly outperforms a combined system without schedule optimization, even when the first faces a significant increase in demand while the latter stays at basis demand. Although the methodology from [3] thus has considerable potential, this study has also revealed a number of shortcomings in its applicability to practice. Future research should improve this methodology by incorporating (i) appointments with a duration of multiple slots, (ii) patient types requiring specific appointment slots, (iii) preparation time that may be required before a patient's examination, and (iv) random server breakdowns.

Based on the results of this study, the CT-facility at the AMC has started employing a combined walk-in and appointment system as of April 4, 2016. Key factors leading to the decision to implement were (i) the close collaboration between the operations researchers and the CT-facility's clinicians and administrative staff via weekly project team meetings and frequent one-on-one contact, (ii) the project team's inspiring tour of the CT-facility at Rijnstate hospital in Arnhem, that had implemented a combined walk-in and appointment system in 2010, and (iii) having a radiologist (the fourth author of this paper) championing the implementation. While we started this project in 2008 and simulation results were available in 2009, it was not until the occurrence of (ii) and (iii) that management of the CT-facility decided to start the implementation. Several other major changes at the CT-facility, such as the implementation of a new radiology information system in 2011, the replacement of one of the facility's two CT-scanners in 2015, and the implementation of a hospital-wide electronic medical record system in 2015, also kept management from deciding to start the implementation of a combined walk-in and appointment system earlier. While the CT-facility's clinicians were initially skeptical towards changing a well-functioning process, the increasing emphasis on patient-centered healthcare motivated them to pursue implementation. The CT-facility chose to employ a phased approach for their implementation, in each phase allowing a subset of the hospital's outpatient clinics to start sending their patients to the CT-facility on a walk-in basis. This approach enabled the implementation team to thoroughly educate outpatient clinic staff, who in turn informed patients. As the CT-facility's clinicians started receiving positive feedback from patients and outpatient physicians, their motivation increased as implementation progressed. In future work, once the fully implemented system has been operable for a period of time, we intend to report in detail on the implementation and its results.

The reusable simulation model developed in this study can also be used to investigate the impact of implementing a combined walk-in and appointment system at other CT-facilities or for other types of diagnostic examinations. Note that the performance of a combined walk-in and appointment system depends on the number of servers at a given facility. The higher the number of servers, the greater the flexibility to schedule slots reserved for appointments in a way that minimizes walk-in waiting times, for example by always keeping at least one server free for walk-in. Thus, facilities having one server should not expect to achieve performance similar to our case study, while facilities with more than two servers could most likely improve upon our results. Next to investigating the impact of switching from a complete appointment system to a combined walk-in and appointment system, our methodology could also be used by facilities already employing a combined system, for investigating improvements to their appointment schedule, thereby possibly enhancing their system performance. Additionally, our methodology enables facility management to quantify the impact of other potential improvements, such as smoothing walk-in arrival patterns, changing the facility's business hours, or adjusting resource capacities, thereby supporting logistical decision making. Thus, while already being beneficial to the AMC CT-facility and its patients, the methodology developed in this paper has the potential to serve many diagnostic facilities and their patients.

Appendix A. Generating feasible schedules using the walk-in schedule generator

This appendix describes our approach for using the walk-in schedule generator [3] to generate feasible appointment schedules, given the three inconsistencies between the walk-in schedule generator and the simulation model described in Section 3.2.6. Fig. A.1 summarizes this approach. Initially, we perform three steps for each patient type separately:

1. To address the walk-in schedule generator's assumption that all examinations have a duration of one time slot, we multiply a patient type j's daily appointment request arrival rates λ_j^d by its examination duration H_j, to obtain the correct load patient type j places on the system.
2. The fact that the walk-in schedule generator contains only one type of scheduled patients necessitates merging of the arrival processes of all patient types requiring appointments. However, as a result of the portfolio effect, the walk-in schedule generator will reserve fewer appointment slots for this joint appointment request arrival process than actually required by the J patient types requiring specific appointment slot reservations. This is due to the fact that buffer capacity, required for meeting access time targets while there is variability in appointment request arrivals, can be shared in a joint arrival process, but not among disjoint arrival processes. Therefore, in a second step we determine the number of appointment slots required for each patient type separately, M_j, using the first model from [3].
3. The total number of appointment slots reserved for patient type j has to be a multiple of its examination duration H_j. Thus, we round M_j up to the nearest integer multiple of the examination duration, denote the result by $\hat{M}_j$, and then recover daily appointment request arrival rates by distributing $\hat{M}_j$ over the days in the cycle proportional to the original appointment request arrival rates λ_j^d (see the formula in Fig. A.1).

Finally, we perform two steps for all patient types jointly:

For each patient type separately

Step 1: *multiply appointment request arrival rates by # slots per appointment:*
$\hat{\lambda}_j^d := \lambda_j^d \cdot H_j$

Step 2: $M_j :=$ *minimum # slots required as determined by first model from [3]*

Step 3: $\hat{M}_j := M_j$ *rounded up to nearest integer multiple of appointment length* (H_j)*;*
translate result back to daily appointment request arrival rates by
distributing $\hat{M}_j$ *over the days* d *proportional to* λ_j^d*:*
$\bar{\lambda}_j^d := \frac{\lambda_j^d}{\sum_d \lambda_j^d} \cdot \hat{M}_j$

For all patient types jointly

Step 4: *(i)* *merge appointment request arrival processes*
(ii) *run walk-in schedule generator*
(iii) *distribute reserved slots over patient types*

Step 5: *(i)* *determine # slots required for outpatient appointments in general*
(ii) *combine schedules from 4iii and 5i to obtain final schedule*

Fig. A1. The approach for applying the methodology from [3] to generate feasible schedules for an actual diagnostic facility.

4. We merge all appointment request arrival processes, run the walk-in schedule generator, and distribute the slots reserved for appointments over the different appointment patient types. Each patient type requiring specific reservations is given a number of slots according to its previously determined requirement, $\hat{M}_j$. All remaining slots are reserved for outpatient appointments in general, and can be used both by appointment patients not requiring specific reservations, and by deferred walk-in patients (who arrive as a walk-in patient but need to be served as a scheduled patient). In assigning specific slots to specific patient types, we strive to avoid the issue mentioned in Section 3.2.6 concerning reservations early on the day, by assigning early slots to patient types not requiring preparation before their examination.
5. The number of slots that have thus far been reserved for outpatient appointments in general is probably too small, because the walk-in schedule generator assumes that deferred walk-in patients may utilize all slots reserved for appointments, while in the simulation model these patients may only utilize the slots reserved for outpatient appointments in general (and not the slots reserved for specific patient types). Therefore, we determine the number of slots required for outpatient appointments in general by running the walk-in schedule generator again, now excluding all patient types requiring specific appointments, and also excluding the slots reserved for them by considering these as slots in which the servers are not available. Ultimately, we combine the slots now reserved for outpatient appointments in general and the slots previously reserved for specific patient types to obtain our final schedule.

References

[1] M. Murray, D.M. Berwick, Advanced access: reducing waiting and delays in primary care, JAMA 289 (8) (2003) 1035–1040.
[2] E.J. Rising, R. Baron, B. Averill, A systems analysis of a university-health-service outpatient clinic, Oper. Res. 21 (5) (1973) 1030–1047.
[3] N. Kortbeek, M.E. Zonderland, A. Braaksma, I.M.H. Vliegen, R.J. Boucherie, N. Litvak, E.W. Hans, Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals, Perform. Eval. 80 (2014) 5–26.
[4] D. Fone, S. Hollinghurst, M. Temple, A. Round, N. Lester, A. Weightman, K. Roberts, E. Coyle, G. Bevan, S. Palmer, Systematic review of the use and value of computer simulation modelling in population health and health care delivery, J. Public Health 25 (4) (2003) 325–335.
[5] M.M. Gunal, M. Pidd, Discrete event simulation for performance modelling in health care: a review of the literature, J. Simul. 4 (1) (2010) 42–51.
[6] J.B. Jun, S.H. Jacobson, J.R. Swisher, Application of discrete-event simulation in health care clinics: a survey, J. Oper. Res. Soc. 50 (2) (1999) 109–123.
[7] P.H.P. Fung Kon Jin, M.G.W. Dijkgraaf, C.L. Alons, C. van Kuijk, L.F.M. Beenen, G.M. Koole, J.C. Goslings, Improving CT scan capabilities with a new trauma workflow concept: simulation of hospital logistics using different CT scanner scenarios, Eur. J. Radiol. 80 (2) (2011) 504–509.
[8] W.A.M. van Lent, J.W. Deetman, H.J. Teertstra, S.H. Muller, E.W. Hans, W.H. van Harten, Reducing the throughput time of the diagnostic track involving CT scanning with computer simulation, Eur. J. Radiol. 81 (11) (2012) 3131–3140.
[9] R. Ashton, L. Hague, M. Brandreth, D. Worthington, S. Cropper, A simulation-based study of a NHS walk-in centre, J Oper Res Soc 56 (2) (2004) 153–161.
[10] T.A. Reilly, V.P. Marathe, B.E. Fries, A delay-scheduling model for patients using a walk-in clinic, J. Med. Syst. 2 (4) (1978) 303–313.
[11] T. Cayirli, E. Veral, Outpatient scheduling in health care: a review of literature, Prod. Oper. Manag. 12 (4) (2003) 519–549.
[12] T. Cayirli, K.K. Yang, S.A. Quek, A universal appointment rule in the presence of no-shows and walk-ins, Prod. Oper. Manag. 21 (4) (2012) 682–697.
[13] Y. Peng, X. Qu, J. Shi, A hybrid simulation and genetic algorithm approach to determine the optimal scheduling templates for open access clinics admitting walk-in patients, Comput. Ind. Eng. 72 (2014) 282–296.
[14] C. Yan, J. Tang, B. Jiang, Sequential appointment scheduling considering walk-in patients, Mathl. Probl. Eng (2014). Article ID 564832, 12 pages
[15] T. Cayirli, E.D. Gunes, Outpatient appointment scheduling in presence of seasonal walk-ins, J. Oper. Res. Soc. 65 (4) (2014) 512–531.
[16] J.E. Helm, S. AhmadBeygi, M.P. Van Oyen, Design and analysis of hospital admission control for operational effectiveness, Prod. Oper. Manag. 20 (3) (2011) 359–374.
[17] L.V. Green, S. Savin, B. Wang, Managing patient service in a diagnostic medical facility, Oper. Res. 54 (1) (2006) 11–25.
[18] R. Kolisch, S. Sickinger, Providing radiology health care services to stochastic demand of different customer classes, OR Spect. 30 (2) (2008) 375–395.
[19] S. Sickinger, R. Kolisch, The performance of a generalized Bailey–Welch rule for outpatient appointment scheduling under inpatient and emergency demand, Health Care Manag. Sci. 12 (2009) 408–419.
[20] M.E. Zonderland, R.J. Boucherie, N. Litvak, C.L.A.M. Vleggeert-Lankamp, Planning and scheduling of semi-urgent surgeries, Health Care Manag Sci 13 (2010) 256–267.
[21] S. Robinson, Simulation: the practice of model development and use, John Wiley & Sons Ltd, Chichester, England, 2004.
[22] S. Robinson, R.E. Nance, R.J. Paul, M. Pidd, S.J.E. Taylor, Simulation model reuse: definitions, benefits and obstacles, Simul. Model Pract. Theory 12 (7) (2004) 479–494.
[23] N. Oses, M. Pidd, R.J. Brooks, Critical issues in the development of component-based discrete simulation, Simul. Model Pract. Theory 12 (7–8) (2004) 495–514.
[24] R.L. Tideman, M.K. Pitts, C.K. Fairley, Effects of a change from an appointment service to a walk-in triage service at a sexual health centre, Int. J. STD AIDS 14 (12) (2003) 793–795.
[25] A.M. Law, W.D. Kelton, Simulation modeling and analysis, third edition, McGraw-Hill, New York, NY, 2000.
[26] P.D. Welch, The Statistical Analysis of Simulation Results, in: S.S. Lavenberg (Ed.), The computer performance modeling handbook, Notes and Reports in Computer Science and Applied Mathematics, 4, Academic Press, New York, NY, 1983, pp. 268–328.
[27] S. Robinson, An heuristic technique for selecting the run-length of non-terminating steady-state simulations, Simul. 65 (3) (1995) 170–179.