A simulation model to support implementation of a combined walk-in and appointment system for CT scanning in a radiotherapy centre

Catarina Pereira
Master of science Degree Student
Instituto Superior Técnico
Lisboa, Portugal

Abstract

Hospitals often use appointment systems to assign patients to their imaging facilities. This policy avoids having crowded waiting rooms and allows them to organize the capacity of the involved resources more easily by balancing the workload on each working day. Although, allowing patients to walk into imaging facilities right after the first consultation, without an appointment, also lodge several benefits for both patients and hospitals. From a logistics' viewpoint, machines' utilization rates are increased; there is no waste of time after completing a service (thus avoiding slack and surplus times), fewer resources are needed for scheduling appointments and it is easier to deal with no-shows.

In this work, we aim to quantitatively assess the improvement of having walk-in patients in a capacity allocation solution for imaging devices through the analysis of an appointment system combining walk-ins and appointments that maximizes the number of walk-in patients. We analyse the case study of the CT-scanning at the radiotherapy department of the AVL. To this end, we use a 'Walk-in Generator' model [1] to obtain a capacity allocation solution for the two types of patients, which was quantitatively evaluated through a discrete-event simulation model.

The best capacity allocation solution obtained with the 'Walk-in Generator' model was the one using a literature benchmark with spread time slots through the work day. This solution showed a reduction of 1,5 days in the access times, an increase of 49% in the fraction of walk-in patients served and satisfied the service level target of serving 95% within 2 days. We build upon previously validated research works and develop a model that, for the first time, quantitatively assess the advantage of having walk-in patients in the managing of resources' capacity allocation.

Keywords— Resource allocation, appointment schedule, walk-ins, appointments, discrete-event-simulation, access time, waiting time

I. MOTIVATION AND PROBLEM DESCRIPTION

On a worldwide scale, healthcare institutions are trying to reduce costs while maintaining the same quality of care. It is known that costs in health systems are mainly associated with the need for highly qualified staff and the use of very expensive resources.

While health systems face increasing waiting times for several medical services, it is known that long waiting time can have a negative impact on patient's health outcome [2]. Therefore, there has been an increasing pressure to reduce patients' waiting times in hospitals without incurring additional costs.

The imaging track is a critical part of health systems' management. It is a necessary path for many treatment modalities and its resources are usually expensive and very limited in capacity. Thus, an efficient capacity management of the devices used in imaging facilities is a task of increasing importance in healthcare management that has shown solid improvement in reducing waiting times for treatment [2].

Hospitals often use appointment systems to assign patients to their imaging facilities. This policy avoids having crowded waiting rooms and allows them to organize the capacity of the involved resources more easily by balancing the workload on each working day. Allowing patients to walk into imaging facilities right after the first consultation, without an appointment, may lodge several benefits for both patients and hospitals. We define this type of accessibility as "walk-ins". From a patient's perspective, the number of visits to the hospital is reduced; the access time (time between an appointment request and the actual appointment) is eliminated and consequently, the psychologic distress of having to wait for a cancer treatment is minimized. From a logistics' viewpoint, machines' utilization rates are increased; there is no waste of time after completing a service (thus avoiding slack and surplus times), fewer resources are needed for scheduling appointments and it is easier to deal with no-shows. The downsides of such a working principle may be the longer waiting times for patients in periods of peak demand and difficulties in planning the capacity of human resources, as working overtime may often be needed. Besides, not all patients are eligible to walk into an imaging facility right after a given examination is prescribed, as other appointments in between may be needed. Therefore, it is advantageous having both scheduled patients and walk-ins, contributing for speeding up the process and, consequently, increased patient satisfaction.

The efficient utilization of the CT-scanners belonging to the RT department of the AVL (RT-AVL) is directly affected by the allocation of its capacity to different patient types. As a physical resource used by several patient types, it has a limited capacity defined by the time that it is available for utilization. The allocation of that capacity means the division of this time to be reserved to each patient type – appointment or walk-in - that needs to use it. This capacity allocation is represented by a schedule where the working hours on each work day are divided in time slots. Each time slot has a patient type associated, meaning that the scanner can be only be booked by the appointment office if the time slot is previously allocated to a patient of that type.

As in many RT centres, capacity allocation decisions at the RT-AVL are performed in an intuitive way, relying on the expertise of the booking staff, without any clear procedure or decision support system behind. Therefore, the capacity allocation solution developed and used on a daily basis in each of the 2 CT-scanners may be far from optimal when considering the minimization of patients' waiting times. An analysis of the current way of working showed that the CT scanners of the RT-AVL had only between 2 and 3 walk-in patients per day. Given the fact that most patients were eligible to walk-in, this number was considered low by the managers of the department. Thus, they were seeking a new solution that could maximize the number of walk-in patients in the clinic. In this work, we aim to find new capacity allocation solutions for the CT scanners of the RT-AVL that maximizes the number of walk-in patients in the clinic without increasing the access times for patients that need an appointment.

II. RESEARCH OBJECTIVES

Our proposed methodology encompasses three major steps:

- Use a walk-in generator (Kortbeek et al.[1]) to obtain a solution for the clinical data gathered in the RT-AVL;
- Adapt the solution, together with clinicians, taking into account the medical and technological constraints of the RT-AVL;
- Quantitatively assess the impact of implementing such a solution for the CT-scanners on the RT-AVL through a discrete-event simulation model.

We aim to find efficient capacity allocation solutions that are able to guide appointment scheduling combining appointments and walk-ins in the RT department of the AVL, allowing as much walk-in patients as possible. We conclude that we are operating in the managerial area of resource capacity planning. Besides, since we aim to generate decisions that affect the mid-term operation of the centre and constrain the operational decisions, we can see that we are working at the tactical level of decision-making.

III. LITERATURE REVIEW

Exact approaches work well for small instances of the problems, but for bigger instances the computational running time is not practical, thus a solution found in a completely mathematical way is not commonly successful in the medical practice. Heuristic approaches do not reach the optimal solution, and can be stuck in a local optimum, however are easier to develop and good solutions are obtained in a faster way than exact approaches. For instance, Littman et al. [3] show that queuing theory works well for small instances, but for large instances it is not solvable in a practical running time. Besides this, exact approaches cause medical acceptance difficulties because it requires dealing with computer-based decision rules [4]. This opens space to explore heuristic approaches [5], namely local and constructive heuristics. All papers report improvement after applying local search on their constructive heuristic method. Computer simulation approaches provide big advantages in finding bottlenecks, visually representing the real system and evaluating solutions generated by other methods. These methods usually verify a better acceptance, and consequently more chances of implementation of new interventions by health care practitioners.

Existing models from other sectors cannot readily be adapted to health systems. Besides, we can see that patient categorization and their aggregation in patient-types is a common approach in healthcare planning. This arises from their distinct service requirements and priority levels, which also leads to patient prioritization approaches.

IV. WALK-IN SCHEDULE GENERATION

The work of Kortbeek et al. [1] is used to design schedules combining walk-ins (unscheduled jobs) and appointments (scheduled jobs), with the goal of maximizing the fraction of unscheduled jobs served on the day of the arrival, while satisfying a pre-specified access time norm for scheduled jobs. Scheduled jobs are given priority in case of conflicting access, with the possibility of no-shows. Both unscheduled and scheduled jobs have a cyclic pattern of arrival. Thus, Kortbeek et al. develop a method to obtain a cyclic appointment schedule (CAS), i.e. the maximum number of jobs that may be scheduled on each day. The best CAS is attained by employing an iterative algorithm that generates CASs using a First Come First Served (FCFS) principle with two different approaches: complete enumeration, and heuristic, in which the expected fraction of unscheduled jobs served on the day of arrival is maximized, while for scheduled jobs the access time service level is minimized. Each CAS is evaluated with two models in different time scales. The first model -Model I - provides an evaluation of the access time for scheduled jobs, having days as the time scale. The second model - Model II - evaluates the performance of a single day in the CAS, with a time scale of hours. This analytical methodology gives the best CAS using model I to find the capacity cycle, the number of time slots for appointments per day which minimizes the access times. The first model is combined with model II, which finds the time of the day for previously defined number of time slots for scheduled jobs, minimizing the number of deferred patients. Deferred patients are patients that were initially walk-in, but because they had to wait above a certain user-defined threshold, they are given an appointment for the next day.

The iterative procedure

The iterative procedure, shown in Figure 1, links access and day processes in order to maximize the unscheduled jobs within the pre-specified access time service level norm established by the user. In model II unscheduled jobs are not willing to wait more than g time slots in the day of arrival to the facility. If the time needed to wait is superior to g, they are offered an appointment for the next day. Therefore, the number of deferred jobs from CAS is accounted for in the appointment request arrival distribution γ. This is done by recalculating the distribution of the appointment request arrival by adding the deferred jobs from the previous iteration, having as an objective the approximation of the optimal F, the fraction of walk-in patients. There are two ways of doing this iterative procedure, according to Kortbeek *et al.* [1] work: 1) a complete enumeration, with the disadvantage of longer running times; and 2) a heuristic approach, faster and more practical, but not optimal as with the complete enumeration. Both follow the pseudo-code depicted in Figure 1.

Step 1: *specify input*	*Specify:* $R, T, D, g, q, S^{\text{norm}}(y), \epsilon$; $\forall d : \lambda^d; \forall d, t : \chi_t^d$.
Step 2: *initialize iterative procedure*	$n := 1; \forall d : \nu^d(1) := 0, \gamma^d(1) := \text{Poisson}(\lambda^d)$.
Step 3: *find candidate CAS*	*Execute complete enumeration (see Section 6.2) or heuristic procedure (see Section 6.3).*
Step 4: *assess current solution*	*If* $\forall d : \lvert\nu^d(n) - \nu^d(n-1)\rvert < \epsilon$, *then stop, else proceed to Step 5.*
Step 5: *adjust deferrals*	$\forall d : \nu^d(n+1) := \nu^d(n), \phi^d(n+1) := \phi^d(n)$, $\gamma^d(n+1) := \text{Poisson}(\lambda^d) * \phi^d(n+1)$; $n := n+1$ *and return to Step 3.*

Figure 1: Pseudo code of the iterative procedure

Solutions obtainned

The best results using the Walk-in Generator model in [6] were obtained with 'Heuristic 1' for capacity cycle and 'Heuristic 3' for the day schedule. The difference in the fraction of walk-in patients between the solutions created with local search techniques and the ones without local search was 0.001% [6]. Due to this small difference, we did not use local search techniques, helping us to reduce the solution space. Besides the 'Heuristic 1' and 'Heuristic 3', we also used the literature benchmark, 'Benchmark 1', which spread the time slots through the appointment schedule, avoiding concentration of time slots in a certain time frame of the day. Table 1 shows the configurations we had in the Walk-in Generator Model, we analyse the different heuristics and the different values for the g parameter, obtaining eight configurations.

Table 1: Capacity allocation configurations tested

Number of resources	1: CT08		2: CT08 and CT04	
Capacity cycle heuristics	H1	H1	H1	H1
Day schedule heuristics	H3	Spread	H3	Spread
g parameter	4/6	4/6	4/6	4/6
	1h	1b	2h	2b

Table 2: Capacity allocation solutions configurations' results

	g=4			g=6	
	One resource			One resource	
	1h	1b		1h	1b
SL	97,42	96,63	SL	95,4974	95,7311
AT	5,725455	5,879091	AT	6,011727	6,173045
F	0,8256	0,82076	F	0,858624	0,85359
	Two resources			Two resources	
	2h	2b		2h	2b
SL	97,28	95,44	SL	95,3616	95,5768
AT	5,856818	6,163636	AT	6,149659	6,471818
F	0,8551	0,8645	F	0,889304	0,89908

Table 3: Difference between solutions with g=6 and g=4, for capacity allocation '2b'

SL	0,001
AT	0,050
F	0,040

SL: Service Level; AT: Access Time; F: Fraction of walk-in patients

From Table 2, considering that the service level of serving 95% of the patients within 2 days is always obtained, we select the best configuration based on the fraction of the walk-in patients served. The solution '2b', for two resources and using the spreading procedure has the biggest value for F. In Table 3 we compare the service level, access time and fraction of walk-in patients for the solution '2b', with g=4 and g=6. The results are the difference of each KPI when g goes from 4 to 6. All the KPIs increase, F increases 4%, AT increases 5% and SL increases 0,1%. For an increase of 4% in the fraction of walk-in patients against an increase of one hour in the waiting time in the waiting room, the managers of RT-AVL decided that it would not worth. For this reason, we perform the discrete-event-simulation only for the four capacity allocation solutions configurations, with g=4. These four solutions are shown in the following pictures and have already been modified by the managers.

Repairing of solutions to be evaluated

To the four capacity allocation solutions, the team of experts performed some rearrangements having in mind the specific RT-AVL constraints. These rearrangements consisted in changing some time slots for walk-ins in the morning to appointment type time slots. The reason for this to happen was the constrains related with the staff availability and with the consultations hours.

Staff availability

A CT-scan may require that an IV contrast is administered to a patient before the scan. In these situations, there must be a physician to administrate the contrast to the patient. In RT-AVL, the physicians are only available after 8h45 AM, and thus every time slot for appointments before 8h45 A.M in the new appointment schedule as to be a non-IV-contrast time slot, following the current doctors' schedule in the hospital.

Consultation hours

The walk-in patients, as well as the appointment patients, are referred to the CT-scanner after consultation. Because walk-in patients can have the CT-scan done in the same day of consultation, before the end of the first consultation hour there are no walk-in patients available for the scanner. Knowing that the first consultation of a work day ends at 9h40 A.M, every slot for walk-ins patients before that hour will never be fulfilled, and represent a waste of capacity. Thus, in the output solutions, every time slot for walk-in before 9h40 A.M was adapted to an appointment-type time slot.

All changes have been performed maintaining the number of time slots for appointments (capacity cycle), and only changing the order of certain time slots. In this new capacity allocation solutions, there are time slots with the number 5, referring to the appointment time slots without IV contrast needed. The walk-in time slots are marked as “0” and the appointments’ time slots as “1”. The black time slots are the breaks.

All the figures presented are already rearranged because of the constrains, ‘2hr’ and 2br’, count for both CT scanners, weighting the results for each performance measurement with 0.113 for CT04 and 0.887 for CT08. We only show the appointment schedule for CT04 in ‘2hr’, because the one for ‘2br’ is the same.

	Monday	Tuesday	Wednesday	Thursday	Friday
7:45	5	5	5	5	5
8:10	5	5	5	5	5
8:45	1	1	1	1	1
9:10	1	1	1	1	1
9:35	1	1	1	1	1
10:00	0				
10:25	0	0	0	0	0
10:50	0	1	0	0	1
11:15	0	1	0	0	0
11:40	0	0	0	0	0
12:05	0	0	0	0	0
12:30	1	1	0	1	1
12:55	1	1	0	1	1
13:20	1	1	1	1	1
13:45	1	1	1	1	1
14:10	1	1	1	1	1
14:35	1	0	0	1	1
15:00					
15:25	0	0	0	0	0
15:50	0	0	0	1	0
16:15	0	0	0	0	0
16:40	0	0	0	0	0

Figure 2: Capacity allocation solution ‘1hr’ (rearranged)

	Monday	Tuesday	Wednesday	Thursday	Friday
7:45	5	5	5	5	5
8:10	5	5	5	5	5
8:45	1	1	1	1	1
9:10	1	1	1	1	1
9:35	1	1	1	1	1
10:00	0				
10:25	0	0	0	0	0
10:50	0	0	0	0	0
11:15	0	0	0	0	1
11:40	0	0	0	0	0
12:05	0	0	0	0	0
12:30	1	1	0	1	1
12:55	1	1	0	1	1
13:20	1	1	0	1	1
13:45	1	1	0	1	1
14:10	1	1	1	1	1
14:35	0	0	0	1	1
15:00					
15:25	0	0	0	0	0
15:50	0	0	0	1	0
16:15	0	0	0	0	0
16:40	0	0	0	0	0

	Monday	Tuesday	Wednesday	Thursday	Friday
7:45					
8:10					
8:45		1	1	1	1
9:10					
9:35		0		0	0
10:00					
10:25		0		1	1
10:50		1	1	1	1

Figure 3: Capacity allocation solution ‘2hr’ (rearranged), CT08 (left) and CT04 (right)

	Monday	Tuesday	Wednesday	Thursday	Friday
7:45	5	5	5	5	5
8:10	5	5	5	5	5
8:45	1	1	1	1	1
9:10	1	1	1	1	1
9:35	1	1	1	1	1
10:00	0				
10:25	0	0	0	0	0
10:50	0	1	0	0	1
11:15	1	1	1	1	1
11:40	0	0	0	0	0
12:05	1	1	0	1	1
12:30	0	0	0	0	0
12:55	1	1	0	1	1
13:20	1	1	0	0	1
13:45	1	1	0	1	1
14:10	1	1	0	1	1
14:35	0	0	1	1	1
15:00					
15:25	0	0	0	0	0
15:50	0	0	0	0	0
16:15	0	0	0	1	0
16:40	0	0	0	0	0

Figure 4: Capacity allocation solution ‘1br’ (rearranged)

	Monday	Tuesday	Wednesday	Thursday	Friday
7:45	5	5	5	5	5
8:10	5	5	5	5	5
8:45	1	1	1	1	1
9:10	1	1	1	1	1
9:35	1	1	1	1	1
10:00	0				
10:25	0	0	0	0	0
10:50	0	0	0	0	1
11:15	1	1	1	0	1
11:40	0	0	0	0	0
12:05	1	1	1	0	1
12:30	0	0	0	0	0
12:55	1	1	1	0	1
13:20	1	1	0	0	1
13:45	1	1	0	1	1
14:10	1	0	0	0	1
14:35	0	0	1	0	0
15:00					
15:25	0	0	0	0	0
15:50	0	0	0	0	0
16:15	0	0	0	1	0
16:40	0	0	0	0	0

Figure 5: Capacity allocation solution ‘2br’ (rearranged)

V. SIMULATION MODEL

The conceptual model can be generally represented by the Figure 2. Upon the patient’s arrival, an evaluation whether he/she is eligible for walk-in takes place. If the patient is of appointment type, then he goes to the planner, if not, he can walk into the server right away.

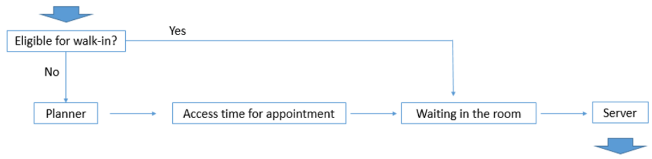

Figure 2: Overview of the simulation model

In a more detailed view of the system four components can be identified: tests, preparations, planner and servers. There are two components which are common for every patient: the initial test, identified as ‘Eligible for walk-in?’, and the destination, identified as ‘Server’ in Figure 3. Only appointment-type patients pass by the ‘Planner’. In the destination, independently of the care plan assigned to the patient, she/he will be assigned to one of the three types of time slots: walk-in time slot (0) , appointments time slot (1) (with or without contrast) and appointment with no-contrast time slot (5).

We gathered information about the patient population during the years of 2014, 2015 and 2016, and clustered the patients in three patient types, corresponding to the three time slot types to which they are assigned: walk-ins; appointments with IV contrast; appointments without IV contrast.

Patient route

The route a patient follows in the simulation model depends on the care plan of the patient, i.e. depends on the medical needs of the patient. Figure 3 represents all the possible routes a patient may experience in the model.

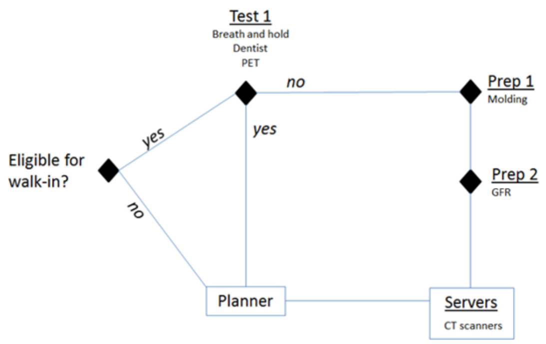

Figure 3: Simulation model routes

Each patient is attributed with a care plan, corresponding to a care pathway, following a distribution based on the data analysed. If the patient is eligible for walk-in follows the route to the Test 1, if not, follows the route to the planner.

In the planner, the model needs to know if the patient needs IV contrast, to search an appointment time slot suitable for the patient. This route end in the server after finding the time slot.

In the route to the Test 1, the patient has three binary tests to perform. Only if the three results are negative, the patient follows the route as a walk-in, ending in the server without passing by the planner. After Test 1 the walk-in patient may need one procedure, two or none. The number of procedures needed does not change his or hers final destination to the server as a walk-in patient. If one of the tests is positive, the patient goes to the planner as an appointment patient type.

Test 1 and Preparation procedures

Within the same care plan exist patients of appointment type and walk-in type. To distinguish them we perform the Test 1. According to each care plan, patients can perform one, two or all the test components. The test is modelled by a binomial distribution, with a success probability (p) and number of samples (n) for each component:

Table 4: Components' distribution parameters

	n	p
1- Breathe and Hold	1356	0,346
2- Dentist Consultation	346	0,287
3- PET-CT	543	0,178

The test is assumed to have a zero-processing time and an infinity capacity, it serves the patients in a First Come First Served (FCFS) fashion. For a positive result the patient goes to the planner as an appointment type of patient, if the result is negative he/she goes to the walk-in route.

The result of the Test 1 does not change only the route, but also the planning horizon. For instance, if the Test 1 is positive for the type breathe–and-hold technique (ID equal to 1), the patient has a planning horizon greater or equal than one day, this way the patient can practice the technique at home. If the Test 1 is positive with an ID equal to 2 (dentist consultation), the patient has a planning horizon greater or equal than 8 days (one week and a half), to align with the dentist's schedule. If the Test 1 is positive with an ID equal to 3 (PET-CT needed), the planning horizon is extended until the next Monday or Friday, so the PET-CT scan be performed in the same day the MRI scan is performed. This shift in the plan horizon does not count for access time, just the days after this shift count.

Besides Test 1, the model incorporates two preparation procedures. These procedures are the moulding and GFR (blood analysis). If both procedures are needed, they always happen in sequence, as shown in Figure 3. According to the care plan of the patient, it can happen that both procedures are needed, only one of them is needed, or none. The processing times for preparation procedures are deterministic and equal to one time slot (25 minutes). Besides, it is assumed that there is sufficient capacity to get zero waiting times for these procedures. If only one procedure is needed, the patient is counted as able to walk into the server after one time slot from the time of his arrival. If both procedures are needed, he is only counted after two time slots. This way, the time spent to undertake these procedures are not counted for the waiting time of the patient.

Patient allocation

The simulation model receives as input a capacity allocation solution: an empty schedule to allocate the patients while they arrive to the facility. Each patient, according to his care plan, is assign to one of the three types of time slots in a FCFS way. The three time slots are: appointment; walk-ins and appointments with no contrast. All with the same time length of 25 minutes.

Servers

There are two servers with different capacity. We assume that there is no preference of one over another, i.e. the patient is allocated to the one that is available with the smallest access time or waiting time. If there are slots available in both servers, with the same access time or waiting time, the choice between the servers is randomly chosen.

Test 1 and procedures parameters

The concept of a test for deciding the route of some care plans was a representation of the real system approved by the hospital experts. We fitted the data into these tests and estimated their distribution. The procedures' processing times are not recorded by the hospital, however it was agreed that these processing times would equal one time slot. With the help of several interviews and after some site visits we were able to define a patient route for each care pathway.

Arrival process

Our arrival rate of patients is a considered a stochastic process, it takes the uncertainty into account, which is very important, specially in healthcare, due to the high variability verified in the system. Theoretically, the patients' arrival rate holds the property of "lack of memory", i.e. the arrival of a patient at a certain time slot is independent from the arrival of the next patient, and so on. This can be represented by means exponential distribution, with the mean inter-arrival time being $E[X]=1/\lambda$, where λ is the mean arrival rate per time unit (time slot of 25 minutes) in a Poisson distribution. The arrival rate of all patients was calculated with data recorded from 1 of January 2015 to 31 of December 2015, i.e. for the whole year 2015. A Chi-squared test was done to compare the observed data with the expected data generated using the Poisson distribution. We used a $p > 0{,}05$ to accept the hypothesis that the arrival rate fits a Poisson distribution.

VI. Results

We divided the results into 4 categories: for one resource (CT); for two resources (CTs); the solution chosen versus the traditional solution; the solutions with the managers 'modifications; and the results for the different arrival rates.

The access time (AT) referred in the tables 5 to 12, is the access time for 95% of the patient population. This way we know that when this value exceeds the 2 days, the service level norm of having 95% of the patients served within 2 days is not fulfilled anymore. The waiting time (WT) in these tables has an upper bound of 4 time slots, 1hour and 40minutes, we can analyse its behaviour due to the increase in workload, however it is not a decisive performance measurement.

One resource

In these experiments, the performance measurements were done considering only one CT scanner, analysing the best capacity allocation solution obtained with Walk-in Generator model ('1h') and with the literature benchmark with the spreading procedure ('1b'). In both cases, we analyse the difference between having the time slots rearranged or not, according to the RT-AVL managers' recommendations.

Table 5: Results of the one resource capacity allocation solutions

	1h	1hr	1b	1br
AT (95%)	1,606054	2,645021	1,624851	1,918818
WT	2,778001	2,711051	3,40749	3,020713
FWI	0,605206	0,69122	0,651725	0,704645
UT	0,388101	0,434978	0,414547	0,443108

In this set of solutions, we chose the one with the lowest waiting time (WT) and access time (AT), a higher utilization rate (UT) and fraction of walk-in patients (FWI). The solution '1br' has 70% of fraction of walk-in patients and 44% of utilization rate, being the solution with the highest values in these performance measurements in Table 5.

Two resources

We also performed the analysis having two CT-scanners, closer to the real situation in RT-AVL centre. Solution '2br' has 76% of FWI and 62% of UT, being the one with the highest values in these performance measurements from the Table 6.

Table 6: Results of the two resources capacity allocation solutions

	2h	2hr	2b	2br
AT (95%)	2,189792	1,480905	1,460386	1,547265
WT	0,352261	2,37398	2,839733	2,562651
FWI	0,647739	0,729378	0,713114	0,761848
UT	0,62758	0,635664	0,623228	0,623156

Choosing the solution with best performance from Table 5 and 6, we can compare the solution '1br' and '2br', to see the effect of having one or more resources in RT-AVL department in Table 7.

Table 7: Comparison between capacity allocation solution 1br and 2br

	1br	2br	Difference
WT	3,020713	2,562651	-0,458062
FWI	0,704645	0,761848	0,057203
UT	0,743108	0,623156	-0,119952

With the second CT-scanner, the RT-AVL has less 15% (half day) in waiting time (WT), more 6% in the fraction of walk-in patients, and a decrease of 12% in the utilization rate (UT).

The managers' modifications

As shown in Table 8, there is no statistical significance, with $p>0.05$, in the values of access time (AT) and utilization rate (UT) for the solutions '2b' and '2br'. There is a difference in the waiting time values, however as explained in the beginning of this section, this is a bounded value. The fraction of walk-in patients served with '2br' is 5% bigger than with '2b', with 76% of the walk-.in patients served as a walk-in.

Table 8: Comparison between capacity allocation solutions '2b' and '2br'

	2b	2br	Statistical difference (p>0.05)
AT (95%)	1,460386	1,547265	No
WT	2,839733	2,562651	Yes
FWI	0,713114	0,761848	yes
UT	0,623228	0,623156	No

The solution '2br' shows the biggest values for the performance measurements FWI and UT, indicated in Table 8, the analysis performed in the next section will be with this capacity allocation solution.

The traditional versos the new solution

Concluding that the capacity allocation solution marked as '2br' is the one with the best overall performance, we compared it with the current practice capacity allocation solution being used at the RT-AVL.

Table 9: Comparison between capacity allocation solution in practice and '2br'

	Practice	2br	Gains
AT (95%)	2,880759	1,547265	1,5 days
WT	2,344007	2,562651	
FWI	27%	76%	49%
UT	0,73684	0,623156	

As shown in Table 9, we have less 1,5 days in access time for 95% of the patient population, fulfilling the service level norm with half day of margin. The capacity allocation solution '2br' includes the combination of appointments and walk-ins, serving 76% of the walk-in patient population, with an increase of 49% of the fraction of walk-in patients served.

The increased workload

The improved accessibility and patient service may occur into an increase of patient's arrival. This way, we obtained the results for the performance measurements of capacity solution 2br with several increases in the workload.

In Table 10, we can compare the performance measurements of the base case, with the normal arrival rate expected for RT-AVL, with an increase of 20%. The capacity allocation solution shows an increase in the fraction of walk-in patients (FWI) from 76% to 91% and in the utilization rate from 62% to 75%, keeping the service level norm. Due to these good results, we also perform an extreme case with an increase of 40% in the workload. Comparing with the 20% increase scenario, in this case the utilization rate has an increase of 2%, but the fraction of walk-in patients goes to 67%, with a decrease of 24%, and the service level norm is not fulfilled, although it only has a small increase of 0.2 days.

We perform an itineration in the simulation runs to find the critical increment in the workload in terms of fulfilling the service level norm. That value founded is shown in Table 10 as 'Critical increase of 29%'. This value of 29% represent the biggest increment in the workload supported by the capacity allocation while preserving the service level norm. Comparing the performance measurements of the critical situation with the base situation, the capacity allocation solution shows a decrease of 7% in the fraction of walk-in patients and an increase of 18% in the utilization rate.

The capacity allocation solution 2br shows a top performance of 91% in the fraction of walk-in patients served and a utilization rate of 74%, when the workload is increased by 20%, having the flexibility of fulfilling the service level norm until the incensement of 29% in the workload.

Table 10: Different workloads increments (base, 20%,40%,29%) for capacity allocation solution '2br'

	Base arrival rate	Increase of 20%	Increase of 40%	Critical increase of 29%
AT (95%)	1,547	1,857	2,166	1,996
WT	2,563	3,075	3,588	3,306
FWI	0,762	0,914	0,673	0,690
UT	0,623	0,748	0,772	0,804

.In Table 11 we show the results of incrementing the workload just in the morning period, until 11h40 AM. Comparing the values with an increase of 20% between Table 10 and Table 11, where the increment is equal to all time slots during the day versus only in the morning time slots, respectively, the fraction of walk-in patients has a difference of 2% and the utilization rate is 10% bigger when the increase is just in the morning. The critical value for the increment in the utilization rate is bigger when the increment is just in the morning than when it is during all day. The critical increment founded for this situation was 33%, with a difference of 4% between Table 10 and 111.

Table 11: Different workloads morning increments (20%,29%,33%) for capacity allocation solution '2br'

	Increase of 20%	Increase of 29%	Critical increase of 33%
AT (95%)	1,779	1,934	1,988
WT	2,947	3,203	3,408
FWI	0,740	0,777	0,744
UT	0,717	0,733	0,855

When the increment in workload is just in the morning we obtain better results, referred above, then when it is during the whole day. To check if this was due to the fact that the total daily increment is lower if it is just in the morning than in the all day, we also analyse afternoon increments in Table 12.

In Table 12 for an increase of 20% in the workload in the afternoon, the fraction of walk-in patients is 70%, when for the same increase but for the whole day (Table 10), the fraction is 91%. The utilization rate is 78%, when for the same increase but for the whole day (Table 10), is 75%. The critical value for increment in the workload was 23%, 10% smaller than for the same amount of increment but in the morning (Table 11), and 6% smaller than for the same amount of increment but for the whole day (Table 10).

Table 12: Different workloads afternoon increments (20%,29%,33%) for capacity allocation solution '2br'

	Increase of 20%	Critical increase of 23%
AT (95%)	1,934	1,994
WT	3,203	3,331
FWI	0,702	0,699
UT	0,779	0,810

VII. Discussion

This Section is divided in four sub sections, the first reflects about the number of resources needed by the RT-AVL center, the second about the effect of the rearrangements, the third compares our solution with the current one and the last one analyses the effect of having different workloads.

Number of resources

With the capacity allocation solution '1br', counting only with one resource and using the 'b' algorithm (time slots spread during the day with the benchmark number one in Walk-in Generator Model) the performance measurements show satisfactory results. These results are 74% for the utilization rate and 70% for the fraction of walk-in patients, with the service level norm fulfilled. This tell us that only one CT-scanner would fulfill the managers' requirements for RT-AVL. However, when comparing these values with the ones from '2br' solution, for both CT-scanners, we realize that just by adding scanner 'CT04' with four work days and only eight time slots per working day, there is a bigger space for improvement and the facility can serve more patients. We conclude this by looking to the difference in the utilization rate, 12% of the capacity in the scanner becomes available for new patients. Because of the increase in capacity, the fraction of walk-in patients also increases by 6%.

After the analysis made considering the number of resources, we chose the capacity allocation solution ‘2br’ as the one to perform following analyses.

Managers’ rearrangements

In Table 8 we compared the performance measurements for the solution ‘2b’ with (‘2br’) and without (‘2b’) rearrangements. There is no statistical difference between the results for access time and utilization rate, for the solution with and without rearrangements. This is an optimistic result for the implementation of the solution, once the constrains we had to add do not cause loss of performance.

There is an increase in the fraction of walk-in patients served of 5%, from 71% without rearrangements to 76% with the rearrangements. The improvement in this performance measurement is expected because the rearrangement done was the shifting of appointment time slots to the early morning walk-in time slots, where do not exist walk-in patients yet. In the Walk-in Generator Model was not possible to input this type of constrain, therefore we did the rearrangements afterwards, and the results were positive.

Traditional versus the new solution

Because the solution ‘2br’ has the higher FWI and UT, we compare it with the traditional solution, the one in current practice at RT-AVL. The combination of walk-in and appointment time slots in the capacity allocation solution showed a reduction of access times of one and half day, 46% reduction, and a reduction of 11% in the utilization rate. This lead to more time slots available to new patients, and show that a new capacity allocation solution could optimize the scanners capacity.

The outliers found in the access time results represent 7,5% of the runs in the simulation model. Because they are less than 10%, they do not influence our results. The waiting time showed in Table 9 for the practice has a standard deviation of 6,12, this value reflects the lack of a walk-in culture in RT-AVL, and for this reason is not a god performance measurement to compare.

Increase of the workload

As stated in the previous paragraphs, the utilization rate reduced when we applied the ‘2br’ capacity allocation solution. This optimization in capacity allowed to receive more patients, and because of that when the arrival rate is increased by 20% the results for the performance measurements still under the managers’ requirements. As it is shown in Table 10, 11 and 12, the higher utilization rate and fraction of walk-in patient is attained when the arrival rate is increased by 20%, obtaining 75% in UT and 91% in FWI. We can conclude that with the implementation of the ‘2br’ solution the RT-AVL CT-scanners could serve 20% more patients, which would be 5444 patients instead of 4537, more 907 patients per year. This would be, in average, more 18 patients per week.

From Table 10 we can observe that the ‘2br’ solution is flexible in terms of keeping the performance measurements with the increasing arrival rate. The solution can get an 80% in UT and a 69% in FWI with a 30% increase in workload. This means that the resources can serve until 5853 patients per year, which means 1326 more patients per year or 25more patients per week.

We analyze the flexibility during the work day, comparing between an increase in the morning, until 11h40, and in the afternoon, after 13h45. The capacity allocation solution is more flexible during the morning, with a maximum value of increase in the workload, fulfilling the service level norm, of 33% against a maximum value of 23% in the afternoon. This shows the importance of the CT04 in the capacity management of the scanners. With CT04 available in the morning slots, the system became more responsive to an increase in the workload in the morning.

VIII. CONCLUSIONS

In the final chapter we look at the work done in this thesis and its contribute for literature and for the cancer center, next we explore some limitations and possible improvements and lastly we give some ideas for future research in literature and in the cancer center.

We started this research with the following objective: optimization of capacity allocation of the imaging resources in diagnostic facilities, combining walk-in and appointment patients, in the cancer center of NKI-AVL, with the main goal of increasing the fraction of walk-in patients. To reach this goal we divided the work in two parts. The first one was the search for a state-of-the-art model, theoretical consistent and case study tested, to produce capacity allocation solutions. The second part was the development of a simulation model, to evaluate those solutions in RT-AVL. We started with an extensive review of theory, which gave us the concepts, definitions and terminology that we needed to develop a simulation model that can be used to analyze the combination of walk-in and appointments. Through a process analysis of the CT- scan casus of the RT-AVL and by studying the processes described by research conducted at other diagnostic facilities we identified the elements we need to simulate the service process of diagnostic facilities.

By combining the theory and the process analysis we developed the discrete event simulation model, which has of four elements: (i) the initial decision about a patient being eligible to walk-in to the CT-scanner, (ii) a planner for appointments, (iii) a test component and (iv) preparation components (see Figure 36). All these elements can be combined to match real world systems, such as that of the radiotherapy department of the AVL. By adding design choices corresponding to when walk-in is possible, how appointments are scheduled and how many time patients are willing to wait in the waiting room, there is a broad range of functionalities to experiment with.

The CT-scans at the AMC are currently organized through a 100% appointment system. Strictly speaking only 10% of all patients can get an appointment on the same day, being a walk-in patient. The relatively small number of time slots in a CT-scan for walk-ins is not aligned with the patient population type, where 63% can be served as a walk-in. In the current situation is an indication that the CT-scans could be organized through a combined walk-in and appointment system. We used the discrete-event-simulation model to simulate this new situation and the results show that:

- Most walk-in patients can be served on the day of the request;
- Access time of appointments decreases compared to the current situation;
- More patients incur waiting time than in the current situation;
- Number of served patients increase.

The patient clustering was decisive to attain the research goal. Because of this we attribute a major importance to the part where we study the RT-AVL system. Interviews to hospital's experts and visits on site were crucial. The two-patient cluster used in Walk-in Generator model was a good simplification to start the patient population study. We ended with three clusters of patients after the modifications by the managers, which was still a low number. We consider this an advantage to simplify the capacity allocation solution. If there are some patients that do not fit in any of the three groups, the staff in the planner have the ability, and resources, to choose the best time slot for those patients. These kinds of patients should represent 5% of the whole patient population.

Furthermore, we experimented with the patient's patience factor, allowing the decision makers to make trade-offs between the performance indicators:

- Increasing the allowed waiting time for walk-in patients reduces the fraction of walk-in patients who is deferred, but also leads to higher waiting time for walk-in patients;

The sensitivity analysis showed that the system remains stable with an increased patient arrival rate, by workload during all day, or just during the morning or in the afternoon, being more robust to increased workloads in the morning.

Some of the results, such as the response to the increased workload, can be used to give practical advice. However, there is not a best practice that can be advised. The power of the simulation model is that all effects can be mapped to aid the decision makers. To make full use of this power we suggest a method in which the main effects from the simulation are used to start a discussion between members of the project group, and that the interaction effects between different factors are used as catalyst to come to a consensus. We also note that the experimental factors in the model are only one side of the decisions corresponding to a combined walk-in and appointment system. There are also internal and external organizational changes, such as changes in information requirements and responsibilities. These changes need to be mapped if a department decides to implement a combined walk-in and appointment system. Based on this research the radiotherapy department of the AVL decided to continue with the analysis of these internal and external organizational changes, with the goal to implement a combined walk- in and appointment system for the CT-scans.

In our work we used some improvements stated in the literature and build upon that. Bailey and Welch, through their research work, showed that heuristic approaches work better for capacity allocation solutions with a deterministic service time, consequently a defined time slot length, equal through all the schedule/solution. Although Cayiril *et al.* [7] have work with a dome pattern for time slots, the approach of Bailey and Welch is still the most used and successful in the latest research works. Both Lin et al. [8] and Freville and Plateau [9], support solution space reduction through dynamic programming and myopic heuristic, respectively. We also did solution space reduction, but with a simpler way: we did not apply local search to the heuristics used, because we considered we had already too much entropy in our solutions. In other way, Denton et al. [10] used simulated annealing to improve the initial solution, we did not do this once we got good results with the initial solutions we got.

Klassen and Rohleden [11] show that empty time slots in the beginning of the day, for urgent patients, reduce waiting times, while at the end of the day such time slots can be able to serve a bigger number of patients that are willing to wait. To obtain both advantages, the time slots for urgent patients should be spread equally during the day. In the same way, Su and Shih [12] show that alternating sequences of appointment schedules with walk-ins works better. This supports that the best results we obtained were with the '2br' solution that spreads the time slots through the day. Therefore, some of our results were validated by the literature, while others add new insights from which future research works can build upon that, namely the relationship between the decrease of access time and increase of utilization rate when combining appointments and walk-ins patients, through the almost direct application of the Kortbeek et al. model.

IX. Discussion and Future Research

In this section we will discuss three topics that relate to the development of the simulation model and its further use: (i) generalizability of results, (ii) validation through more extensive cases, and (iii) the link between theory and practice.

For the first topic we look at the possible implications of attaining the same improvements of combining walk-ins and appointments, increasing to the maximum the fraction of walk-in patients, for other diagnostic facilities. The results for the CT-case of the RT-AVL are promising, however we have also seen that case-dependent factors were very determining for the performance of the system, for example the effect of having the first time slots adjusted to the staff availability. To avoid erroneous conclusions, it is important to note that the model only allows the user to understand the direct effect of having walk-ins in the optimization of the capacity allocation. A future user would have to translate the new real system to the components in the model, tests and procedures, and to do the process analysis, data collection and input preparation. Even more so, this simulation model is really simple, meaning that there are few assumptions, which could be too restrictive for some cases.

The second topic is closely related to this issue. The model was validated by simulating the current situation of the CT-scans at the RT-AVL, a system with only appointments. This validation does not in any way ensure that this model can be used for experiments on other cases that also have an appointment system. It is a limitation of the research that we only reviewed one case extensively. We think that the model can be used for a broad range of cases, but cannot claim that the model is truly generic.

The last discussion topic is related to the development of the model and the discrepancy between theory and practice. Two suggestions we distilled from the theory review were:

- Define independent components, this allows other modelers to work in a flexible manner. A model to which extra components/elements can easily be added/removed has the most potential when it comes to component based simulation.

- Complete model reuse is so complex that it is considered the holy grail of simulation modeling.

Taking a component based perspective was very useful to think of a workable structure that allows reuse. However, we found that it was not possible to develop independent components that are easily combined to form a model that was useable for experimentation with a specific purpose. In other words, some of the components we defined are dependent on the general structure of the model and the availability of information provided by other components. The result is that we developed a model that can be completely reused, is very flexible, but only within the defined limits. The consequence is that if a case requires additional functionalities these can be added, but for most functionalities this will mean that the structure of related components will have to be adapted to this change.

The ideas for future research concern the analytical model used, the simulation that was build and possibility of implementation. When using the Walk-in Generator model for two resources, the total arrival rate is split for each resource and the model runs independently for each one. This goes against what happens in practice since the RT department verifies a single arrival rate for both scanners, and the patients walk into one of the available resources.

For the discrete event simulation model, we took some simplifications and assumptions into account, that could be improved in a future work. The day and time of the breaks in the new capacity allocation solutions were assumed to be the same that existed in practice. A study about the best time slots for the breaks in each of the scanners, as well as the length of those time slots, could be an interesting line for further research. In terms of implementation, we believe that a valuable contribution would be a deeper study in the field of health sciences about the patient population and possible clusters based on care content, which is crucial for research works in resource allocation problems.

ACKNOWLEDGMENT

This reaserach was performed in RT-AVL hospital center, in Amsterdam, The Netherlands, with the support of Dr. Jeroen van de Kamer, clinical physics, and Dr. Corine van Vliet-Vroegindeweij, head of Clinic Physics Department. The research also had the support of Dr. Erwin Hans and Phd candidate Bruno Vieira from University of Twente.

REFERENCES

[1] N. Kortbeek et al., “Designing cyclic appointment schedules for outpatient clinics with scheduled and unscheduled patient arrivals,” Perform. Eval., vol. 80, no. C, pp. 5–26, 2014.

[2] E. Hans, M. Van Houdenhoven, and P. J. H. Hulshof, “A Framework for Health Care Planning and Control,” Chem. …, pp. 1–23, 2004.

[3] L. V. Green, S. Savin, and B. Wang, “Managing Patient Service in a Diagnostic Medical Facility,” Oper. Res., vol. 54, no. 1, pp. 11–25, 2006.

[4] R. Kolisch and S. Sickinger, “Providing radiology health care services to stochastic demand of different customer classes,” OR Spectr., vol. 30, no. 2, pp. 375–395, 2008.

[5] P. T. VanBerkel and J. T. Blake, “A comprehensive simulation for wait time reduction and capacity planning applied in general surgery,” Health Care Manag. Sci., vol. 10, no. 4, pp. 373–385, Oct. 2007.

[6] K. Smid, “A simulation model to support implementation of a combined walk-in and appointment system at diagnostic facilities,” 2013.

[7] S. A. Q. T. Cayirli, K.K. Yang, “A universal appointment rule in the presence of no-shows and walk-ins,” Prod. Oper., vol. 21, no. 4, pp. 682–697, 2012.

[8] J. Lin, K. Muthuraman, and M. Lawley, “Optimal and approximate algorithms for sequential clinical scheduling with no-shows,” IIE Trans. Healthc. Syst. Eng., vol. 1, no. 1, pp. 20–36, 2011.

[9] A. Freville and G. Plateau, “An efficient preprocessing procedure for the multidimensional 0-1 knapsack problem,” Discret. Appl. Math., vol. 49, no. 1–3, pp. 189–212, 1994.

[10] T. R. Rohleder, P. Lewkonia, D. P. Bischak, P. Duffy, and R. Hendijani, “Using simulation modeling to improve patient flow at an outpatient orthopedic clinic,” Health Care Manag. Sci., vol. 14, no. 2, pp. 135–145, 2011.

[11] T. R. R. K.J. Klassen, “Scheduling outpatient appointments in a dynamic environment,” J. Oper., vol. 14, no. 2, pp. 83–101, 1996.

[12] C. L. S. S. Su, “Managing a mixed-registration-type appointment system in outpatient clinics,” Int. J. ofMedical Informatics, vol. 70, no. 1, pp. 31–40, 2003.