Operations Research for Health Care 26 (2020) 100267

Contents lists available at ScienceDirect

Operations Research for Health Care

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

An optimization model for multi-appointment scheduling in an outpatient cardiology setting

Lida Anna Apergi [a,*,1], John S. Baras [b,1], Bruce L. Golden [a], Kenneth E. Wood [c]

[a] Robert H. Smith School of Business, University of Maryland, College Park, MD, USA
[b] Institute for Systems Research, University of Maryland, College Park, MD, USA
[c] R Adams Cowley Shock Trauma Center, University of Maryland School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 14 March 2019
Received in revised form 10 July 2020
Accepted 21 July 2020
Available online 3 August 2020

Keywords:
Outpatient scheduling
Integer programming
Cardiology

ABSTRACT

In this paper, we tackle the problem of outpatient scheduling in the cardiology department of a large medical center. The outpatients have to go through a number of diagnostic tests and treatments before they are able to complete the final interventional procedure or surgery. We develop an integer programming (IP) formulation to ensure that the outpatients will go through the necessary procedures on time, that they will have enough time to recover after each step, and that their availability will be taken into account. Our goal is to schedule appointments that are convenient for the outpatients, by minimizing the number of visits that the patients have to make to the hospital and the time they spend waiting in the hospital. We propose formulation improvements and introduce valid inequalities to the IP, which help the running times to decrease significantly. Furthermore, we investigate whether scheduling outpatients in groups can lead to better schedules for the patients. This would require coordination between the different members of the scheduling staff within the cardiology department. The results show improvements in the total objective value over a period of one month, ranging from 0.45% to 2.33% on average, depending on the scenario taken into account.

1. Introduction

1.1. Motivation

Approximately 11% of the adult population in the U.S. is diagnosed with heart disease [1], which is also ranked first as a cause of death [2]. Receiving treatment on time is crucial in increasing the chances for survival [3]. It is critical to investigate ways of making access to care as effortless as possible for patients with a heart condition. This is expected to increase the willingness of patients to visit their physicians and go through the necessary treatment [4].

1.2. Background

We focus on appointment scheduling for outpatient interventional procedures (also known as outpatient programs) and elective surgery in cardiology. The outpatient programs that we study are the transcatheter aortic valve replacement (TAVR), the transcatheter mitral valve repair (TMVR), the patent foramen ovale (PFO) closure, valvuloplasty, and the Watchman. Examples of elective surgery that an outpatient could go through include the coronary artery bypass grafting (CABG), the ventricular assist device (VAD) implanting, heart valve surgery, heart transplant, and others.

Patients have to go through a number of steps including consultations, diagnostic tests, and treatments before they are able to go through one of the outpatient programs or elective surgeries described above. The steps that patients are required to go through depend on their history and condition. Fig. 1 includes the steps that TAVR patients have to complete. More details about the steps included in each outpatient program or surgery can be found in e-component 1. We generated the procedure diagrams based on discussions we had with the scheduling staff and nurse practitioners in the University of Maryland Medical Center (UMMC) in Baltimore. We were not able to find information about the exact sequence of steps that outpatients have to follow in other hospitals. Nevertheless, the requirements in other hospitals were similar to those of UMMC [5–10]. Thus, the model that we are studying is commonly used across different hospitals.

From Fig. 1 we observe that not all TAVR patients will necessarily go through all of the steps appearing in the diagram. The steps that the patients have to visit depend on their health

* Corresponding author.
E-mail addresses: lapergi@rhsmith.umd.edu (L.A. Apergi), baras@isr.umd.edu (J.S. Baras), bgolden@rhsmith.umd.edu (B.L. Golden), kenneth.wood@umm.edu (K.E. Wood).
1 The work of Lida Anna Apergi and John S. Baras was partially supported by the Lockheed Martin Chair funds with the Institute for Systems Research, USA.

https://doi.org/10.1016/j.orhc.2020.100267
2211-6923/

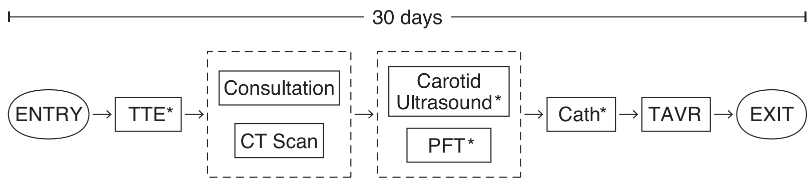


Fig. 1. Procedure diagram for TAVR patients.

and the procedures they have recently undergone. The dashed rectangles in the figure indicate that the procedures contained in them can occur in any order. The arrows indicate precedence between procedures or groups of procedures. Also note that no more than 30 days should elapse from start to finish, i.e., from the day that the patient gets referred to the hospital to the day that the patient goes through the TAVR procedure. This is important both for the health of the patient and because some of the tests have to be repeated if more than 30 days pass (e.g. history & physical, and blood tests [11]), which leads to unnecessary use of resources. Each step depicted in Fig. 1 requires numerous resources to be available in order for the patient to pass through them.

Outpatients are patients who get treatment in the hospital without staying overnight or getting admitted. The patients we are studying get admitted only after going through the final procedure. This means that the patients have to travel to the hospital on multiple occasions. The UMMC treats patients throughout Maryland and the surrounding region. Thus, the patients may have to travel long distances to get to the hospital. Furthermore, they usually depend on someone else to drive them to and from the hospital. This complicates the scheduling procedure, and in some cases the patients may not show up for their appointments or they arrive late. Therefore, it is important to propose a scheduling approach that makes the visits to the hospital as convenient as possible for the patients.

1.3. Problem statement

Currently, appointments are scheduled manually without the support of a decision model. While the scheduling staff tries to generate schedules that will not cause unnecessary burdens on the patients, it can be time consuming to generate such a schedule, and the outcome may not be optimal. The resources required for each appointment are in many cases shared with other departments in the hospital, which means that the various departments are competing for the same resources. Each outpatient is scheduled separately. Different outpatient programs and surgeons have their own scheduling staff, who coordinate with the outpatients to book the appointments. The scheduling staff first talks to the patients to learn when they can come to the hospital. Then they call the corresponding labs to find out the availability of appointments. After booking all the required appointments, they inform the patient about the days and times of the appointments.

In this study, we propose an optimization model to help the staff decide how to schedule the appointments. The objective is to generate schedules that are more convenient for the outpatients. Furthermore, we investigate the effect that scheduling patients in groups would have in the quality of the resulting appointments. This could help distribute the available resources better since more information is available when additional patients are included in the optimization. A number of hospital settings schedule patients in groups instead of one at a time. Examples include scheduling appointments for destination medical centers [12], for radiotherapy pre-treatment [13], and for care of neuromuscular diseases [14]. Scheduling patients in groups requires coordination between the scheduling staff of the various outpatient programs and surgeons. There is a big overlap in the resources required across the procedures that we study. If combining the scheduling of the outpatients leads to better outcomes for the patients, the hospital wants to encourage this collaboration. An additional advantage of grouping the patients is that the communication with the labs will become easier. The scheduling staff will not have to book appointments as often, since patients will be scheduled in batches. An additional derivative benefit, resulting from this scheme, is cost savings on the side of the hospital treating the patients. It can be time consuming every time that the scheduling staff has to find out the availability of each resource, given that there is no centralized way of doing this. Each lab and personnel has to be contacted separately. For the reasons discussed above, this is a direction that the hospital is interested in exploring.

1.4. Constraints

Our objective is to generate schedules that are convenient for the patients. The patients should not have to travel to the hospital more times than necessary. Furthermore, the patients should not have to spend too much time waiting in between appointments, because this may tire or stress them. The scheduling process must satisfy the following constraints:

- Once an appointment for a patient is set, it must not get canceled in order to accommodate another patient.
- The preference and the availability of the patient with respect to the different days of the week should be taken into consideration when scheduling the appointments.
- The patients have to complete all the required steps within the allowed time limit.
- The patient should not go through more steps than those required.
- Enough time must be provided for the patient to prepare (or recover) before (or after) a procedure.
- No more than the available resources can be used.

1.5. Contributions

The contribution of this research is threefold. First, to the best of our knowledge this is the first study that looks into multi-appointment scheduling in outpatient cardiology. In this work, we discuss the parameters and the constraints that need to be taken into account in such a problem and develop an IP formulation. We also provide the procedure diagrams including the steps that the outpatients have to go through. Second, we develop formulation improvements which help solve the IP significantly faster. Those types of improvements could be applied to other scheduling problems with a similar formulation approach. Third, we examine whether outpatients should be scheduled in groups, which requires the collaboration of different scheduling staff in outpatient cardiology. We investigate what is the appropriate size of the group based on different levels of resource availability and external demand. Taking external demand into account is important in a hospital setting, since the various departments do not have complete control over all resources used by their patients.

The rest of the paper is organized as follows. Section 2 provides a review of the existing literature relevant to this problem. Section 3 includes a description of the multi-appointment scheduling problem in outpatient cardiology. The formulation of the scheduling problem is presented in Section 4, which also includes a number of formulation improvements and discusses running times. Section 5 provides the computational experiments and the discussion of the results. Section 6 discusses the managerial insights. Finally, Section 7 states the main conclusions and provides directions for future research.

2. Literature review

This section discusses some of the relevant literature in outpatient scheduling. A thorough review of the existing outpatient appointment scheduling literature and suggested research opportunities can be found in [15–17]. We study scheduling outpatients to go through multiple procedures, which, due to resource or recovery constraints, have to take place on more than one days. While the majority of the outpatient scheduling research examines single appointments, a number of studies have looked at scheduling multiple appointments for the same patient [17]. In particular, the existing literature can be classified into one of the following groups: single appointments, combination appointments, or appointment series [17]. Combination appointments refer to multiple appointments booked for the same day, while appointment series refers to appointments booked on more than one days for the same patient. However, appointment series usually refer to the same types of appointments scheduled on multiple days, since as part of the treatment, the patients have to make repeated visits. Examples include scheduling of chemotherapy [18,19] or radiotherapy patients [20,21], where a similar type of treatment is repeated on many days. Thus, our problem is better classified as a combination appointment even though the appointments can be scheduled on different days. An extensive literature review of multi-appointment scheduling for both inpatients and outpatients is provided in [22]. Our research considers the multidisciplinary scheduling group of problems, since the appointments scheduled include staff and resources from various parts of the hospital. A literature review on multidisciplinary planning can be found in [23], which discusses research on a number of settings, including outpatient clinics, emergency care, blood collection sites, etc.

A mixed integer program (MIP) formulation is presented in [24] for scheduling a set of patients to go through their path of required procedures in an ophthalmology clinic on a given day, where the order of the procedures is predetermined. The objective is to minimize the weighted sum of patient waiting time, resource use overtime, and congestion. Adaptive scheduling heuristics are used to solve the problem, where patients are transferred from more to less busy slots. Patients who have to go through radiotherapy pre-treatment are scheduled in [13]. A MIP formulation with multiple hierarchical objectives is proposed, in an effort to minimize the number of patients waiting longer than the waiting time targets. Each objective is added to the model separately, and its optimal value is passed as a constraint to the next phase which includes the subsequent objective. Due to complexity, the problem is solved by passing initial solutions and not optimal solutions from one phase to the next. It is assumed that the sequence of procedures that the patients go through is known. Scheduling procedures in nuclear medicine is considered in [25], where there are strict constraints on the time that each step of the medical procedure takes place. Four different scheduling algorithms are taken into consideration including scheduling the patients as soon as possible, scheduling based on the day that the patient prefers to go to the clinic, and two algorithms that take into account the preference of the patient up to a point and then schedule them as soon as possible. A MIP formulation is developed in [26] for scheduling in a pathology laboratory as an effort to maximize the patient satisfaction, which is achieved by minimizing the completion time.

Scheduling procedures in nuclear medicine are also considered in [27], where an online scheduling algorithm and a stochastic online scheduling algorithm are proposed. The stochastic online scheduling algorithm consists of a two-stage stochastic program that also takes into account future patient arrivals when scheduling each new request for appointment. The future arrivals include patients that might request an appointment on the same day as the patient being scheduled requested. Scheduling chemotherapy appointments is studied in [28], during which each patient goes through three stages. The uncertainty of the type of appointments requested from new patients during a day and possible cancellations are dealt with by developing a template schedule based on historical data and the deterministic version of the problem. It is possible to update the template dynamically when a new patient arrives who does not fit in the existing template. The authors in [29] study appointment scheduling in a cancer clinic, in an effort to minimize the weighted sum of the waiting time of the patients, the time that the resources are being idle, and the overtime. A stochastic IP is proposed, since the routes that each patient will follow are uncertain. The sample average approximation method is used in order to obtain the upper and lower bounds of the objective. The length of treatment and the course of the treatment of inpatients is predicted through a discrete time Markov model in [30]. A two-stage stochastic IP is proposed in [31] for scheduling patients in primary care. The patients go through a sequence of predetermined steps with stochastic duration. In the first stage the appointment times of all patients are scheduled, and in the second stage the service times in each stage are realized. The objective is to minimize the idle time of the physician and the waiting time of the patients. Scheduling outpatient surgeries is studied in [32], where the patients go through the pre-operation, the surgery, and the recovery stage. Each stage has stochastic duration. The authors propose simulation-based TABU search methods to solve the scheduling problem.

The objective of our study has similarities with research on scheduling appointments in rehabilitation and destination medical centers. Both types of problems aim to generate schedules that are convenient for the patients. Scheduling patients in destination medical centers means that the patients have to travel to the location of the hospital and stay in this location for the duration of their treatment. Thus, in [12] patients are scheduled in an effort to have them start their treatment as close to their start day as possible. Furthermore, the time that the patients wait until all steps of their treatment are completed is also minimized, in order to avoid having patients spend too much time in a hotel. The problem is formulated as a MIP, and is solved using a hybrid particle swarm optimization algorithm. An IP for scheduling rehabilitation patients is proposed in [33]. Each appointment request is satisfied as it arrives. The problem is formulated as a multi-objective minimization problem, where the weight of each criterion is decided ahead of time. The eleven criteria taken into account in the objective include the number of extra visits the patient has to make to the clinic (in addition to the minimum number necessary), the time it takes until the first appointment of the patient takes place, and the number of unscheduled appointments. The formulation allows for a percentage of appointments in each discipline not to get scheduled. In the majority of the cases, the optimal solution of the problem is obtained within a few seconds. The number of visits that a child with a neuromuscular disease has to make to the hospital for rehabilitation is limited to one per year in [14]. Based on this constraint, an IP is proposed to schedule the appointments as well as decide on which children will visit the hospital on each day. The formulation allows for patients to have only a subset of the required appointments scheduled. The objective consists of five criteria, including the time the patients spend being idle, and the number of patients with complete or partial visits. The patients are scheduled in groups.

In this work, we aim to generate schedules that are convenient for the outpatients. In particular, due to the specifics of the outpatient cardiology setting we do not want to have patients travel to the hospital too many times or wait too long in between appointments. The majority of the research in outpatient multi-appointment literature has different objectives, with most studies focusing on decreasing the waiting time for the patients [13,24,27,29,31]. However, it is crucial to investigate this type of objective because it is more suitable to the case of non-urgent outpatients. Generating convenient schedules for those patients will help increase their willingness to go through the necessary treatment. Research on rehabilitation outpatient scheduling [14,33] has the most similarities to our problem compared to other studies in the literature. Both of the cited studies formulate the problem using a multi-objective function. It can be difficult to determine the necessary weights of the objective function in a way that provides the hospital with the best schedule in every case. In our study, we propose providing the hospital with Pareto optimal solutions and allow them to decide which schedule is preferable. Thus, the weights do not have to be determined ahead of time, which gives more flexibility to the hospital to decide on the appropriate schedule. Furthermore, our method of finding the Pareto optimal solutions is much more general and complete than the weighted sum approach, because the fundamental problem here is that of trade-off analysis between the various objectives, and our solution provides precisely all the Pareto points corresponding to such trade-off analysis. The weighted sum approach is the so-called "scalarization method" for trade-off analysis that provides only one Pareto point, the one corresponding to the weights used [34]. Deciding on the weights ahead of time is not very easy or intuitive for the scheduling staff and may not lead to the desirable balance between the two components of the objective. In addition, the constraints of our problem are different from [14,33] due to the characteristics of our system. For example, we allow for recovery and preparation time when necessary, and we have a strict target day for completion. Finally, we investigate whether patients should be scheduled in groups by having different teams collaborate. This could help generate appointments of better quality in a setting where there are other departments in the hospital competing for the same resources. This is crucial, because in many hospital settings the resources are shared across multiple departments. To the best of our knowledge this has not been studied before in outpatient multi-appointment scheduling problems. In general, we see that only a small portion of the outpatient literature focuses on multi-appointment scheduling [17], which highlights the importance of investigating different aspects of this problem.

3. Problem description

This section discusses the assumptions made for solving the problem of multi-appointment scheduling and includes the description of the different elements of the formulation and the corresponding notation.

3.1. Assumptions

In order to model the multi-appointment scheduling problem we make the following assumptions.

1. *The duration of each process is deterministic.*
   Some process steps may last for a longer or shorter time period than the duration for which they were booked. Nevertheless, in this work, we are assuming that the duration of the appointments for each type of diagnostic test or treatment is fixed. This is the approach currently followed in the hospital when scheduling, since appointments are booked based on slots of a predetermined length. Since we want to provide a support tool for the scheduling staff in the hospital, we decided to solve the problem following the same approach. A deterministic duration is also assumed in [12,14,33]. The risk of assuming deterministic durations is that the patient may not finish a step in the expected time, and therefore not have enough time to go through a subsequent step. More complicated procedures (e.g. cardiac catheterization), which can have higher variability in duration due to their complexity, also require a long recovery time. This ensures that there will not be another procedure booked right after. Therefore, we assume that the duration is deterministic. In the experiments run, we use the maximum duration of a diagnostic test or procedure as the duration of the appointment.
2. *The steps that patients have to go through are determined as soon as they arrive.*
   In reality, some diagnostic test might be added later, after finding out more about the patient's health, or some future test might get canceled because it was deemed that the particular procedure was not a good fit for the patient. Similarly, an entire series of steps might have to get canceled if it is deemed that the patient is not a good match for the final procedure at that point in time or in general. However, those cases are less frequent, and the combinations of all possible paths that the patient might end up following in this case are many. Therefore, in order to get a simpler optimization model which is easier to solve, we make this assumption.
3. *The scheduling staff has enough time to book an appointment before the resource availability changes.*
   In other words, when the scheduling staff obtains the available appointments they will have enough time to solve the optimization problem and call back each resource to book the appointments, without some other department being able to book these resources for their own patients in the meantime. Depending on the number of patients that are getting scheduled at the same time, solving the problem could take from a few seconds to a few minutes. We are assuming that the probability of another department booking appointments that overlap with the ones proposed by our model is very low, and thus there is no need to take this case into consideration.

3.2. Definitions

3.2.1. Time

Let $\mathcal{D}$ denote the set of days and $\mathcal{H}$ be the set of time slots or time units in the planning horizon. For example, if each slot corresponds to one hour, we would have $|\mathcal{H}| = 24 * |\mathcal{D}|$ because there are 24 h in a day. The set of time slots $\mathcal{H}^d$ denotes the time slots $h \in \mathcal{H}$ included in a day $d \in \mathcal{D}$, so in our example we would have $|\mathcal{H}^d| = 24, \forall d \in \mathcal{D}$. The length of each slot is set to be equal to the greatest common divisor of the appointment durations of

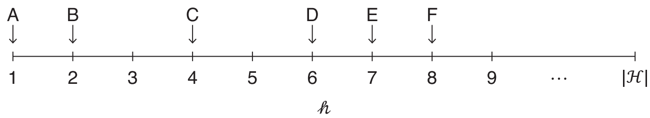


Fig. 2. Example of patient arrivals during the time horizon.

all the procedures that patients could possibly go through in the system. Thus, the duration of each procedure can be expressed as an integer multiple of the defined time unit. We can schedule patients either one at a time, or in groups. Let $\mathcal{T}$ denote the set of decision epochs in the planning horizon. In a decision epoch t, it is decided how to schedule any patients that arrived between decision epoch $t-1$ and t, who have not been scheduled yet. Decision epochs do not take place at predetermined or fixed intervals. They depend on the arrival times of the last patient in the group. Function $o : (\mathcal{T}) \rightarrow (\mathcal{H})$ shows the time in the planning horizon that a specific decision epoch corresponds to. Fig. 2 includes an example of the times that patients A through F were referred to the hospital. If patients are scheduled one at a time, then $o(1) = 1$. The first decision epoch corresponds to the arrival time of the first patient. On the other hand, if the patients are scheduled in groups of two, we have $o(1) = 2$. In other words, the decision epoch equals to the arrival time of the second patient. If patients are scheduled in groups of three, we get $o(1) = 4$, since the decision epoch equals to the arrival time of the third patient.

3.2.2. Patients

Let $\mathcal{I}$ denote the set of patients that will be scheduled. There exists an amount of time within which a patient $i \in \mathcal{I}$ has to complete the program, which is denoted by $\mathcal{L}^i$. In other words, this is the maximum time that a patient is allowed to spend in the system. It is assumed that as soon as a patient is referred to the hospital and the history of the patient becomes available, the patient enters the system. The entry time is denoted by a_i. The planning horizon is set based on the latest it can take for a patient in the group being scheduled to exit the system without violating the allowed period of completion. A patient exits the system once the last procedure is completed, which is normally the main reason that the patient was referred to the hospital. For example, in the TAVR program, the final procedure is the TAVR being performed on the patient in a hybrid operating room.

3.2.3. Positions in system

Patients take various positions throughout their stay in the system. These positions can be categorized into two different types: *procedure* and *waiting positions*. A patient that is in a *procedure position* is a patient who is at a hospital going through a diagnostic test or a treatment. Patients who are in a *waiting position* might be in the hospital or not. A patient, who is in a *waiting position*, has already gone through a procedure and is expected to go through another one on the same day, is assumed to be waiting in the hospital. These types of positions are occupied by patients who are either waiting to recover from a previous procedure or are waiting to go through their next appointment which might not be available right away due to capacity constraints. A patient who is in a *waiting position* is not using any hospital resources. Let $\tilde{\mathcal{P}}$ denote the set of *procedure positions* available in the hospital, and $\tilde{\mathcal{Q}}$ be the set of *waiting positions*. Note that for every element $\tilde{p} \in \tilde{\mathcal{P}}$ there exists an element $\tilde{q} \in \tilde{\mathcal{Q}}$ which is the *waiting position* that the patients will take after they complete procedure $\tilde{p}$. Also note that after the patients go through their last position in the system they go to a *waiting position* where they stay for the entire horizon, which is the *exit* position denoted by l. All patients that have reached this position are assumed to have exited the system. Set $\tilde{\mathcal{Q}}$ has one additional element ($|\tilde{\mathcal{Q}}| = |\tilde{\mathcal{P}}| + 1$), which is the waiting position *entry* which all patients take when they first enter the system before they go through any procedures and is denoted by b.

A patient entering the system will only visit a subset of the procedures (and their corresponding waiting positions) offered in the hospital. Let $\tilde{\mathcal{P}}_i$ and $\tilde{\mathcal{Q}}_i$ denote the sets of *procedure positions* and *waiting positions*, respectively, that a specific patient i has to visit, where $\tilde{\mathcal{P}}_i \subseteq \tilde{\mathcal{P}}$ and $\tilde{\mathcal{Q}}_i \subseteq \tilde{\mathcal{Q}}$. An example of the procedures that a specific TAVR patient i may have to go through can be found in Fig. 3(a). In this case the elements appearing in the rectangles will be included in set $\tilde{\mathcal{P}}_i$, since they refer to procedures that patient i will go through, while those in the ovals will be part of the $\tilde{\mathcal{Q}}_i$ set, since the patient is waiting at those positions. For this example, the patient will follow one of the two paths presented in Fig. 3(b). The patient arrives in the system at position *entry*. Then the patient will either follow the first path in order to get a consultation and then do a CT scan or the second path in which the patient will first do a CT scan and then get a consultation. Finally, the two paths will merge into one and the patient will go through the TAVR procedure and then exit the system. It is important to point out that a patient who followed the first path and is going through a CT scan is in a different state in the system than a patient that followed the second path and is going through the CT scan. Therefore, multiple types of the same position are created, one for each path that a patient could follow. Let $\mathcal{P}_i$ denote the set of all *procedures* in all possible paths that patient i could go through, and correspondingly let $\mathcal{Q}_i$ be the set of all *waiting positions* in all possible paths that patient i could go through. Finally, let $\mathcal{S}_i$ be the set of positions that a patient i may potentially visit, so $\mathcal{S}_i = \mathcal{P}_i \cup \mathcal{Q}_i$. Let $g : (\mathcal{S}_i) \rightarrow (\tilde{\mathcal{P}}_i \cup \tilde{\mathcal{Q}}_i)$, denote a function that returns an element $\tilde{s} \in (\tilde{\mathcal{P}}_i \cup \tilde{\mathcal{Q}}_i)$ for each argument $s \in \mathcal{S}_i$. Table 1 includes the elements that each of the sets defined above consists of, in the example presented in Fig. 3. Finally, let $\mathcal{F}_{i,s}$ and $\mathcal{G}_{i,s}$ denote the set of positions right before and right after a position $s \in \mathcal{S}_i$ for patient i, respectively.

3.2.4. Availability, resources, and durations

For a patient to go through a procedure at a given time, the patient must be available to come to the hospital on that day, and the necessary resources must be available for the procedure to take place. Let A_i^d denote the availability of patient i to come to the hospital on day d. In particular, A_i^d is a binary parameter taking the value 1 if the patient can come to the hospital on day d, and 0 otherwise. Let $\mathcal{R}$ be the set of all resources in the hospital that are required for the procedures to take place. These resources include equipment, labs, and staff. Let $P_{\tilde{p},r}$ denote the number of each resource type $r \in \mathcal{R}$ required for a procedure of type $\tilde{p} \in \tilde{\mathcal{P}}$ to take place. Let R_r^h denote the number of resources of type r that are available to be assigned to patients at time h. This number will change from one decision epoch to the next, since it will be updated based on the patients that got scheduled in the previous decision epoch and the demand from other departments in the hospital. Note that on weekends or during time in a weekday outside the 8 am–5 pm range, the availability of all resources is 0.

Each procedure $\tilde{p} \in \tilde{\mathcal{P}}$ that a patient will go through has a duration $d_{\tilde{p}}$. Any resources that are allocated to the patient for a procedure $\tilde{p}$ will be allocated to the same patient for the entire duration of the procedure. Every *procedure position* $\tilde{p}$ is followed by a *waiting position* $\tilde{q}$, which is necessary to ensure that the patient has enough time to rest and recover from the procedure. Let $\omega_{\tilde{q}}$ be the time that a patient needs to recover from a preceding procedure $\tilde{p}$, or, in other words, the minimum time that the patient will spend in the *waiting position* $\tilde{q}$. Depending on the procedure, the recovery time could last up to

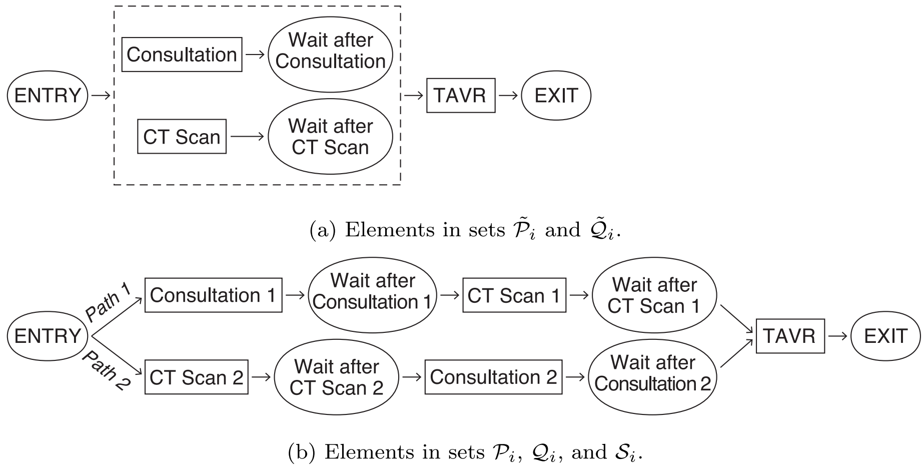


Fig. 3. Example of paths followed by a TAVR patient i.

Table 1
Elements included in each set in the example illustrated in Fig. 3.

Set	Elements
$\tilde{\mathcal{P}}_i$	CT scan, Consultation, TAVR
$\mathcal{P}_i$	CT scan 1, CT scan 2, Consultation 1, Consultation 2, TAVR
$\tilde{\mathcal{Q}}_i$	Entry, Wait after CT scan, Wait after Consultation, Exit
$\mathcal{Q}_i$	Entry, Wait after CT scan 1, Wait after CT scan 2, Wait after Consultation 1, Wait after Consultation 2, Exit
$\mathcal{S}_i$	CT scan 1, CT scan 2, Consultation 1, Consultation 2, TAVR, Entry, Wait after CT scan 1, Wait after CT scan 2, Wait after Consultation 1, Wait after Consultation 2, Exit

Table 2
Scheduling problem sets, function, and parameters.

	Sets
$\mathcal{D}$	Set of days in the planning horizon, indexed d
$\mathcal{H}$	Set of time slots in the planning horizon, indexed h
$\mathcal{H}^d$	Set of time slots belonging to the same day d, indexed h
$\mathcal{T}$	Set of decision epochs, indexed t
$\mathcal{E}_i^d$	Set of available appointment times for patient i on day d, indexed e
$\mathcal{I}$	Set of patients, indexed i
$\tilde{\mathcal{P}}$	Set of procedures, indexed $\tilde{p}$
$\tilde{\mathcal{P}}_i$	Set of procedures that patient i will visit, indexed $\tilde{p}$
$\mathcal{P}_i$	Set of procedure position states over all paths that patient i may visit, indexed p
$\tilde{\mathcal{Q}}$	Set of waiting positions, indexed $\tilde{q}$
$\tilde{\mathcal{Q}}_i$	Set of waiting positions that patient i will visit, indexed $\tilde{q}$
$\mathcal{Q}_i$	Set of waiting position states over all paths that patient i may visit, indexed q
$\mathcal{S}_i$	Set of position states over all paths that patient i may visit, indexed s, s'
$\mathcal{F}_{i,s}$	Set of positions preceding position s for patient i, indexed s, s'
$\mathcal{G}_{i,s}$	Set of positions subsequent to position s for patient i, indexed s, s'
$\mathcal{C}$	Set of steps required to reach the positions, indexed c
$\mathcal{K}_{i,c}$	Set of positions that can be reached by patient i after exactly c steps, indexed s, s'
$\mathcal{P}_i^*$	Set of procedure position states that will solely take place on a day, indexed p^*
$\mathcal{R}$	Set of resources, indexed r
	Functions
$o()$	Function mapping elements in $\mathcal{T}$ to elements in $\mathcal{H}$
$f()$	Function mapping elements in $\mathcal{H}$ to elements in $\mathcal{E}_i^d$
$g()$	Function mapping elements belonging in $\mathcal{S}_i$ to elements in $\tilde{\mathcal{P}} \cup \tilde{\mathcal{Q}}$
	Parameters
R_r^h	Number of resources of type r available at time h
$P_{\tilde{p},r}$	Number of resources of type r required for a process $\tilde{p}$ to take place
A_i^d	Binary indicating availability of patient i to visit the hospital on day d
a_i	Arrival time of the patient to the system, $a_i \in \mathcal{H}$
L_i	Number of time units within which patient i should complete the program
b	First position that patients take in the system, $b \in \mathcal{S}_i$
l	Last position that patients take in the system, $l \in \mathcal{S}_i$
$d_{\tilde{p}}$	Duration in time units of procedure $\tilde{p}$
$\omega_{\tilde{q}}$	Time units of recovery needed after a procedure, spent in the subsequent waiting position $\tilde{q}$
$\gamma_{\tilde{p}}$	Time units of rest required before procedure $\tilde{p}$
M	Upper bound for the time in a day that a patient can be in the hospital
N_i^j	Normalizing constant for patient i used in component j of the objective, $j = 1, 2$
λ	Trade-off parameter in the objective function, $\lambda \in [0, 1]$

a few days. Finally, let $\gamma_{\tilde{p}}$ denote the additional time required before a specific procedure $\tilde{p}$ takes place. This is the case for procedures that require some preparation. For example, if the patient is not supposed to eat or drink anything for a few hours before the procedure. In those cases, in order to avoid stressing the patients further, we impose some additional rest time before the procedure.

4. Scheduling problem

This section describes the formulation used to solve the problem of multi-appointment outpatient scheduling. In Sections 4.1 and 4.2, we present the decision variables, the objective function, and the constraints of the scheduling problem. The problem is formulated based on the notation discussed in Section 3 and included in Table 2. Section 4.3 describes the approach used to normalize the components of the objective function, and Section 4.4 includes an example for obtaining the Pareto optimal solutions. In Section 4.5 we include improvements of the initial formulation in order to decrease the time it takes to obtain an optimal solution.

4.1. Decision variables

The decision variables for the IP are presented below.

$$w_{i,s}^h = \begin{cases} 1, & \text{if patient } i \in \mathcal{I} \text{ arrives at position } s \in \mathcal{S}_i \\ & \text{by time } h \in \mathcal{H} \\ 0, & \text{otherwise} \end{cases}$$

$$y_{i,s}^h = \begin{cases} 1, & \text{if patient } i \in \mathcal{I} \text{ is at position } s \in \mathcal{S}_i \text{ at time } h \in \mathcal{H} \\ 0, & \text{otherwise} \end{cases}$$

$$x_i^d = \begin{cases} 1, & \text{if patient } i \in \mathcal{I} \text{ visited the hospital on day } d \in \mathcal{D} \\ 0, & \text{otherwise} \end{cases}$$

$$x_{i,p}^d = \begin{cases} 1, & \text{if patient } i \in \mathcal{I} \text{ went through procedure } p \in \mathcal{P}_i \\ & \text{on day } d \in \mathcal{D} \\ 0, & \text{otherwise} \end{cases}$$

$u_i^d \equiv$ time that patient $i \in \mathcal{I}$ arrived at the hospital on day $d \in \mathcal{D}$

$v_i^d \equiv$ time that patient $i \in \mathcal{I}$ left the hospital on day $d \in \mathcal{D}$

Note that the decision variable $w_{i,s}^h$ shows whether a patient has arrived at a position by a specific time h. Once this variable takes the value 1 for a specific patient and position, it will take the value 1 for all subsequent times, since the patient has reached this position by all future times. This decision variable was first introduced in [35]. On the other hand, the decision variable $y_{i,s}^h$ only takes the value 1 if a patient is in position s at a specific point in time h. Fig. 4 shows the values for these two types of variables. A patient is scheduled to go through a specific position from time h_1 to time h_2. We can see from the figure that decision variable $w_{i,s}^h$ takes the value 1 for all times after h_1. On the other hand, decision variable $y_{i,s}^h$ equals to 1 only for the times between h_1 and h_2. E-component 2 includes an example which illustrates the difference between the decision variables and the values they take.

4.2. The integer program

Below is the mathematical formulation of the IP. The hospital would have to run the following IP at every decision epoch $t \in \mathcal{T}$, in order to schedule all patients that have arrived since the last scheduling took place. The patients that have already been scheduled are not taken into account again in the decision making since it is not allowed to reschedule already scheduled patients. Nevertheless, the resources they occupy are updated to represent the availability at time t. However, when previously scheduled patients do not show up for an appointment, they might be taken into account again in the decision making. In this case their appointments have to get rescheduled. In particular, if the appointment that the patient did not show up for can be rescheduled without affecting the timing of future appointments it can be rescheduled manually, since it is considered as a single appointment. If the no-show affects a combination of steps, then the patient is scheduled again through the IP. The IP follows:

$$\min \lambda \sum_{d\in\mathcal{D}} \sum_{i\in\mathcal{I}} \frac{1}{N_i^1} x_i^d + (1-\lambda) \sum_{d\in\mathcal{D}} \sum_{i\in\mathcal{I}} \frac{1}{N_i^2} [(v_i^d - u_i^d) - \sum_{p\in\mathcal{P}_i} (d_{g(p)} - 1) x_{i,p}^d] \tag{1}$$

subject to:

$$w_{i,s}^h = 0, \forall h \in \mathcal{H} : h \le a_i - 1, i \in \mathcal{I}, s \in \mathcal{S}_i \tag{2}$$

$$w_{i,s}^h = 1, \forall h \in \mathcal{H} : h \ge a_i, i \in \mathcal{I}, s = b \tag{3}$$

$$w_{i,s}^h = 1, \forall h \in \mathcal{H} : h \ge a_i + L_i, i \in \mathcal{I}, s = l \tag{4}$$

$$w_{i,s}^h = 0, \forall h \in \mathcal{H} : h \le o(t), i \in \mathcal{I}, s \ne b \tag{5}$$

$$w_{i,s}^{h-1} \le w_{i,s}^h, \forall \{h, h-1\} \in \mathcal{H}, i \in \mathcal{I}, s \in \mathcal{S}_i \tag{6}$$

$$w_{i,s}^h \le \sum_{\substack{s'\in\mathcal{F}_{i,s}: \\ \{h-max\{\omega_{g(s')},\gamma_{g(s)}\}\}\in\mathcal{H}}} w_{h-max\{\omega_{g(s')},\gamma_{g(s)}\}}^{i,s'}, \forall h \in \mathcal{H}, i \in \mathcal{I}, s \in \mathcal{S}_i \cap \mathcal{P}_i \tag{7}$$

$$w_{i,s}^h = \sum_{\substack{s'\in\mathcal{F}_{i,s}: \\ \{h-d_{g(s')}\}\in\mathcal{H}}} w_{i,s'}^{h-d_{g(s')}}, \forall h \in \mathcal{H}, i \in \mathcal{I}, s \in \mathcal{S}_i \cap \mathcal{Q}_i \backslash \{b\} \tag{8}$$

$$w_{i,s}^h - \sum_{s'\in\mathcal{G}_{i,s}} w_{i,s'}^h \le y_{i,s}^h, \forall h \in \mathcal{H}, i \in \mathcal{I}, s \in \mathcal{S}_i \tag{9}$$

$$\sum_{s\in\mathcal{S}_i} y_{i,s}^h \le 1, \forall h \in \mathcal{H}, i \in \mathcal{I} \tag{10}$$

$$\sum_{i\in\mathcal{I}} \sum_{s\in\mathcal{S}_i\cap\mathcal{P}_i} y_{i,s}^h P_{g(s),r} \le R_r^h, \forall h \in \mathcal{H}, r \in \mathcal{R} \tag{11}$$

$$x_i^d \le A_i^d, \forall d \in \mathcal{D}, i \in \mathcal{I} \tag{12}$$

$$y_{i,s}^h \le x_i^d, \forall d \in \mathcal{D}, h \in \mathcal{H}^d, i \in \mathcal{I}, s \in \mathcal{S}_i \cap \mathcal{P}_i \tag{13}$$

$$x_i^d \le \sum_{h\in\mathcal{H}^d} \sum_{s\in\mathcal{S}_i\cap\mathcal{P}_i} y_{i,s}^h, \forall d \in \mathcal{D}, i \in \mathcal{I} \tag{14}$$

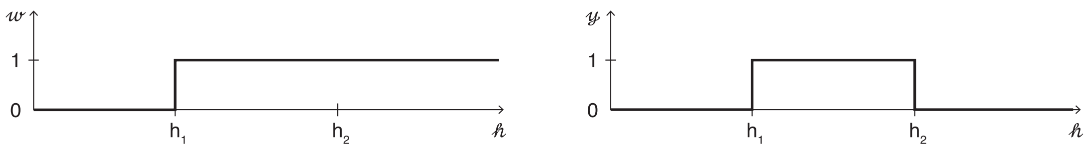


Fig. 4. Difference between decision variables $w^h_{i,s}$ and $y^h_{i,s}$.

$$y^h_{i,s} \leq x^d_{i,p}, \forall d \in \mathcal{D}, h \in \mathcal{H}^d, i \in \mathcal{I}, p \in \mathcal{P}_i, s = p \quad (15)$$

$$x^d_{i,p} \leq \sum_{h \in \mathcal{H}^d} y^h_{i,s}, \forall d \in \mathcal{D}, i \in \mathcal{I}, p \in \mathcal{P}_i, s = p \quad (16)$$

$$u^d_i \leq (h - |\mathcal{H}^d| * (d-1))y^h_{i,s} + M(1 - y^h_{i,s}) - M(1 - x^d_i), \forall d \in \mathcal{D}, h \in \mathcal{H}^d,$$

$$i \in \mathcal{I}, s \in \mathcal{S}_i \cap \mathcal{P}_i \quad (17)$$

$$v^d_i \geq (h - |\mathcal{H}^d| * (d-1))y^h_{i,s}, \forall d \in \mathcal{D}, h \in \mathcal{H}^d, i \in \mathcal{I}, s \in \mathcal{S}_i \cap \mathcal{P}_i \quad (18)$$

$$w^h_{i,s}, y^h_{i,s}, x^d_i, x^d_{i,p} \in \{0, 1\}, \forall d \in \mathcal{D}, h \in \mathcal{H}, i \in \mathcal{I}, p \in \mathcal{P}_i, s \in \mathcal{S}_i \quad (19)$$

$$u^d_i, v^d_i \in \mathbb{Z}^+, \forall d \in \mathcal{D}, i \in \mathcal{I} \quad (20)$$

The objective function (1) minimizes the linear combination of the number of times that the patients have to visit the hospital, and the time that the patients spend in the hospital waiting for their next appointment. The first component of the objective has weight λ, and is estimated based on the total number of visits to the hospital over all patients. For each patient, the number of visits is normalized to take a value between 0 and 1. This is achieved by dividing the number of visits of each patient by a constant N^1_i. More details about how the value of the normalizing constant is chosen are included in Section 4.3. The second component of the objective has weight $(1 - \lambda)$, and corresponds to the total time that the patients spent in the hospital not going through a procedure, or in other words the time spent in the hospital in-between procedures. Similarly to the first component, the total idle time for each patient is normalized by dividing by a constant N^2_i. We plan to provide the hospital with multiple schedules to choose from. Therefore, in each problem instance we estimate the Pareto optimal solutions based on various values of λ. Varying λ in this way allows us to obtain the convexification of the Pareto points boundary (see [34]), which is much better than providing the solution for a single value of λ. Such an example is discussed in Section 4.4.

Constraint (2) makes sure that the patients will not reach any position in the system before they get referred to the hospital. Constraints (3), and (4) enforce that patients will visit the first and the last positions in their path by their entry time and at most after spending the maximum allowed time in the system respectively. Constraint (5) states that patients who are referred to the hospital cannot go through any procedure before the decision epoch following their arrival. Constraint (6) guarantees that once a patient has reached a position, this position will have been reached in all future times by this patient. Constraint (7) enforces a recovery and preparation time before the patient can move to the next procedure. Similarly, constraint (8) imposes that a patient will move on from a procedure to the subsequent *waiting position* exactly after time equal to the duration of the procedure. Constraint (9) ensures that a patient is in a specific position, when the patient has reached the current position but none of the subsequent positions. Constraint (10) makes sure that a patient is at most at one position in the system at all times. Constraint (11) verifies that no more than the available resources are used at each point in time. Constraint (12) enforces that all appointments are scheduled on days that the patient is able to come to the hospital. Based on the discussions we had with the hospital, we decided to consider availability based on the day and not the time. Since the hospital tries to fit as many procedures as possible on a single day, we decided that patients are available if they are available all day. However, if necessary, availability per time can be easily introduced to the IP by changing constraint (12) to $y^h_{i,s} \leq A^h_i$, where A^h_i is the availability of the patient in each time slot h.

Constraints (13) and (14) capture whether a patient is scheduled to be in the hospital on a specific day. Similarly, (15) and (16) capture whether a patient is scheduled to go through a specific procedure on a specific day. Constraints (17) and (18) capture the arrival and departure times of each patient on each day. Since this is a minimization problem and variable u^d_i has a negative sign, the variable will take the largest value possible. Thus, $u^d_i \leq (h - |\mathcal{H}^d| * (d-1))y^h_{i,s}$ ensures that variable u^d_i will equal the time of day d that the patient is scheduled for the first procedure of the day. If a patient is not scheduled for procedure p on day d, but is scheduled for a different procedure, constraint (17) becomes $u^d_i \leq M$ for procedure p. This ensures that u^d_i will equal to the start time of a procedure that the patient is scheduled to go through on that day. This requires a large enough M. On the other hand, if the patient is not scheduled for any procedure on day d, constraint (17) becomes $u^d_i \leq M - M \Rightarrow u^d_i \leq 0$ for all procedures. Since v^d_i has a positive sign in the objective, it will take the smallest value possible. This corresponds to the time that the patient finished with the last procedure for the day. Constraints (19) and (20) define the binary and the nonnegative integer decision variables of the problem respectively.

Some of the constraints introduced above either come from, or are based on, constraints from [35,36]. Constraints (6) and (7) were used in [35,36], and ensure connectivity between the positions and connectivity in time. Constraint (8) is similar to (7) but with an equal instead of a less than or equal sign. We adapted this type of constraint to ensure that a patient will spend time in a procedure exactly equal to the duration of the appointment. Constraint (11) has been extended to include the number of resources required for a procedure to take place. The remaining constraints and objective functions were created for the purpose of this problem.

4.3. Choice of normalizing constants

The two components of the objective are estimated based on different metric units. The first component refers to the number of visits to the hospital, and the second to the total idle time during a patient's visits to the hospital. The idle time component tends to get larger values, which effectively gives this component a larger weight in the objective. Therefore, it is important to normalize the two components. For every patient getting scheduled, each component of the objective will take values between 0 and

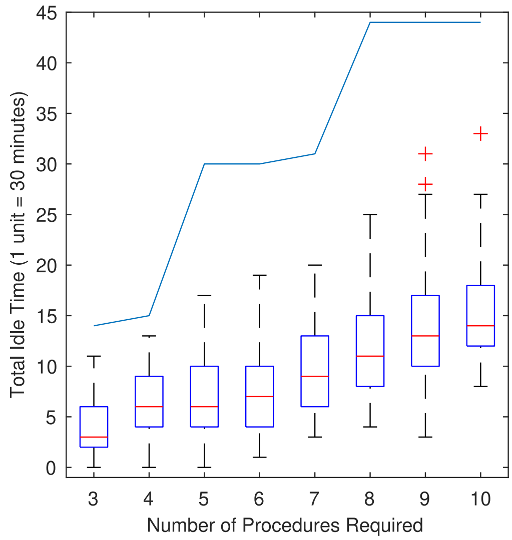


Fig. 5. Idle times observed per number of procedures required.

1 after the normalization. This helps achieve a better balance between the two components, and avoids the well know pitfall of different dynamic ranges (i.e. ranges of values) between two or more objectives in trade-off analysis.

In the case of the number of visits to the hospital we achieve this by dividing the total number of visits that a patient has to make to the hospital with the maximum number of visits that the patient could make to the hospital (i.e., the number of procedures required). With the second component of the objective we want to use a similar approach when normalizing. However, in the worst case scenario, where the patients have to wait in the hospital for the longest time possible, the total wait time for each patient becomes a very large number. If we divide the total idle time for each patient by the maximum possible wait time that this type of patient could spend in the hospital, we obtain a very small number. The reason is that the worst case scenario never occurs in reality. Therefore, the first component of the objective would effectively have a larger weight. In an effort to decrease the value of the normalizing constant in the second component of the objective, we ran a number of experiments where the objective only focused on minimizing the number of visits to the hospital (i.e. $\lambda = 1$). For each patient, we estimated the total idle time in the hospital. The patients were grouped based on the number of required procedures they have to go through. The results are presented in Fig. 5. The line at the top of the figure shows the maximum idle time for each group in the worst case scenario. From the box plots we can see that the actual idle times are considerably lower than the longest possible idle time. Therefore, the normalizing constant for each type of patient is set equal to the longest idle time observed in the experiments.

4.4. Pareto solutions

Our objective consists of two components, which are incompatible with each other. A Pareto optimal solution ensures that in order to improve one of the components, the other will have to become worse. The hospital should be the one to choose between schedules that provide Pareto optimal solutions. Thus, the scheduling staff will select the schedule that provides the desirable balance (or compromise) between the two components.

An example of the different solutions provided when solving a specific scheduling problem is presented in Fig. 6. In this example, three patients are getting scheduled simultaneously. In order to obtain the different solutions, we solved the IP using different values of λ (i.e., $\lambda \in \{0, 0.1, \dots, 1\}$). Furthermore, we used different time limits, after which the IP provided the best solution found up to that point. We also allowed enough time for the IP to reach optimality in some of the experiments. Note that in cases where the solver is not allowed enough time to prove that a proposed solution is optimal (i.e., to reach optimality), we subsequently cannot prove that a solution is Pareto optimal. Nevertheless, since enough time was allowed in this particular example, we have generated Pareto optimal solutions.

The filled red points in Fig. 6 are the Pareto optimal solutions. The gray areas in both graphs indicate values that cannot be part of the solution, since there is a minimum number of visits that a patient has to make to the hospital. Fig. 6(a) shows the values for each component of the objective. The y and the x axis of the graph range from 0 to 3, since for each of the 3 patients getting scheduled both components of the objective take at most the value 1. Fig. 6(b) includes the same solution without normalizing the two components of the objective.

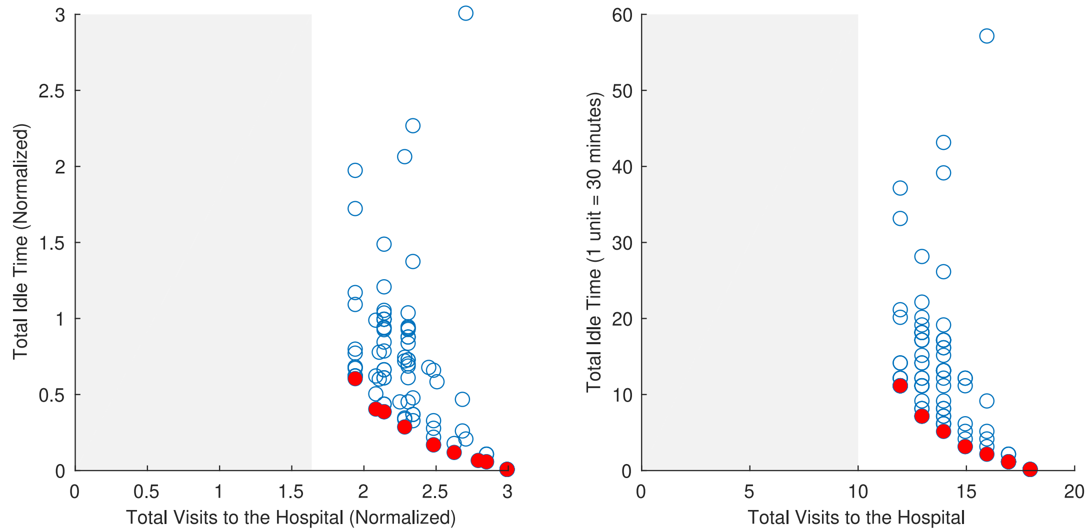
It can be observed from Fig. 6 that there exists a trade-off between the two components of the objective. As it was expected, low values of idle time are achieved when the patient makes many visits to the hospital. On the other hand, the time spent at the hospital while waiting increases as the number of visits to the hospital decreases. If the hospital is more interested in obtaining schedules where the patients visit the hospital few times, they could choose a schedule that corresponds to 12 visits and 11 time units of wait in total for the three patients being scheduled. Another option would be to choose a schedule that corresponds to 13 visits and 7 time units waiting time in the hospital in total. The second schedule proposed might be preferable, since the hospital would require one additional visit from one of the three patients, but save four time units of wait, which corresponds to two hours. Nevertheless, the hospital will have to decide whether this is indeed preferable.

4.5. Formulation improvements

We make improvements to the IP presented in Section 4.2 in order to help in solving it faster. The improvements guarantee that the resulting IP will provide the same solutions as the initial formulation. First, we ensure that the big M will take the smallest possible value, which helps the IP to be solved faster. Second, we introduce valid inequalities. These are constraints that eliminate some non-integer solutions, which are feasible in the initial formulation. However, none of the integer solutions are eliminated. The improvements that we introduce are based on the structure of this particular problem.

Valid inequalities have been used before in the outpatient scheduling literature [37]. In [37], the valid inequalities have to do with the fact that the total resource demand on a given day cannot be higher than the available capacity, and that a procedure cannot be scheduled if none of the resources are used by this procedure on a specific day. The authors use Benders' decomposition, and their objective function only depends on the day that a specific procedure is scheduled. The formulation approach and constraints are different to the IP proposed here, which means that we use different valid inequalities.

As we described in Section 4.2, parameter M has to be large enough to ensure that it will not be smaller than a potential appointment time. At the same time, for computational efficiency M has to take the smallest possible value. One option is to set M equal to $|\mathcal{H}^d|$, since all appointments are completed by the end of the day, and there are $|\mathcal{H}^d|$ time units in a day. Nevertheless, we achieve an even stricter bound by setting different M values for different days and different patients. So in fact M becomes $M_i^d, \forall d \in \mathcal{D}, i \in \mathcal{I}$.



(a) Normalized values of the two objective function components. (b) Values of the two objective components.

Fig. 6. Pareto optimal solutions (in red) in example of scheduling 3 patients simultaneously. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We know the available resources and the procedures that each patient is going to go through. Thus, we can estimate the earliest and the latest time on each day that each patient could potentially be scheduled to go through a procedure. We generate set $\mathcal{E}_i^d = \{0, 1, \ldots, E\}$ for each patient on each day. We define the earliest time that the patient could be in the hospital as $e = 1$, and the latest time as $e = E$. For times in a day that a patient will definitely not be in the hospital, we have $e = 0$. If the patient cannot be scheduled for any procedure on a day, we get $\mathcal{E}_i^d = \{0\}$. Let $f : (\mathcal{H}) \to (\mathcal{E}_i^d)$ denote a function that returns the order in a day for each argument $h \in \mathcal{H}$. If h corresponds to a time outside the appointment range it returns 0. For example, consider the case where a patient could only be at the hospital for any appointment from 8 am to 1 pm on a specific day. This corresponds to $h = 8, 9, \ldots, 13$, assuming that this is the first day in the planning horizon and the time unit is in hours. The corresponding set $\mathcal{E}$ for the patient on this specific day, would include elements $\{0, 1, 2, \ldots, 6\}$, where $e = 1$ corresponds to $h = 8$, $e = 2$ corresponds to $h = 9$, etc. In addition, $e = 0$ corresponds to all h on that day that the patient could not possibly be at the hospital (e.g. 4 am). Using this new information obtained from preprocessing, constraints (17) and (18) can be replaced by constraints (21) and (22), where $M_i^d = |\mathcal{E}_i^d| - 1$. Through this change in the formulation, we obtain much smaller M values.

$$u_i^d \leq f(h) y_{i,p}^h + M_i^d (1 - y_{i,p}^h) - M_i^d (1 - x_i^d), \quad \forall d \in \mathcal{D}, h \in \mathcal{H}^d, i \in \mathcal{I}, p \in \mathcal{P}_i \tag{21}$$

$$v_i^d \geq f(h) y_{i,p}^h, \forall d \in \mathcal{D}, h \in \mathcal{H}^d, i \in \mathcal{I}, p \in \mathcal{P}_i \tag{22}$$

$$\sum_{s \in \mathcal{K}_{i,c}} w_{i,s}^h \leq 1, \forall h \in \mathcal{H}, i \in \mathcal{I}, c \in \mathcal{C} \tag{23}$$

$$\sum_{h \in \mathcal{H}} \sum_{\substack{s \in \mathcal{S}_i \cap \mathcal{P}_i: \\ g(s) = \tilde{p}}} y_{i,s}^h = d_{\tilde{p}}, \forall i \in \mathcal{I}, \tilde{p} \in \tilde{\mathcal{P}}_i \tag{24}$$

$$(v_i^d - u_i^d) \geq \sum_{p \in \mathcal{P}_i} (d_{g(p)} - 1) x_{i,p}^d, \forall d \in \mathcal{D}, i \in \mathcal{I} \tag{25}$$

$$x_{i,p^*}^d + x_{i,p}^d \leq 1, \forall d \in \mathcal{D}, i \in \mathcal{I}, p^* \in \mathcal{P}_i^*, p \in \mathcal{P}_i : p \neq p^* \tag{26}$$

The time it takes to solve the IP can be decreased further by introducing valid inequalities (23) through (26). Let $\mathcal{C}$ denote the set including the number of steps required for patients to reach any position in their path. Let $\mathcal{K}_{i,c}$ denote the set including all positions that could be possibly reached by patient i after exactly c steps. Valid inequality (23) states that at most one position will be reached among those requiring exactly c steps to be reached by patient i. This is similar to the antichain inequality used in [36]. Valid inequality (24) states that the time spent over all procedures of the same type will equal to the duration of this specific procedure. Valid inequality (25) has to do with the fact that the time that patients spend in the hospital on a given day is greater or equal to the duration of the procedures that they go through. Let $\mathcal{P}_i^*$ denote the set of *procedure positions* for a patient i, where each *procedure position* included in the set will solely take place on a day. In other words, the patient will not go through other procedures on the same day if he goes through a procedure in $\mathcal{P}_i^*$. These are procedures that might require long preparation and/or recovery times. Valid inequality (26) states that all patients can either go through a procedure in $\mathcal{P}_i^*$ or any other procedure on a given day, but not both.

4.5.1. *Resulting running times*

We ran experiments of scheduling 60 patients, where one patient was scheduled at a time, in order to see how the proposed formulation improvements affected the running times. The maximum time that the solver was allowed to run was set to 1500 s. If by that time no optimal solution was found the solver stops and returns the best solution obtained up to that point in time. This limit was selected due to our 3rd assumption, which states that "*The scheduling staff has enough time to book an appointment before the resource availability changes*". Therefore, we need to impose a limit to the time that we allow the IP to run. We notice that the solver finds the optimal solution relatively fast, but it takes a long time to prove optimality. In other words, by limiting the time to 1500 s we might still get an optimal solution but it is not proven that it is optimal. In Section 5 we ran the simulation experiments where we schedule patients in groups of 2 up to 5 patients. This complicates the problem and it takes longer for the

Table 3
Results of the Wilcoxon rank sum test (comparing each model with the IF).

Model	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1$
IF + (21), (22)	0.3853	0.0597	0.4917	0.4988	0.1975	0.4893
IF + (24)	0.2381	**0.0241**	0.2193	0.3127	**0.0168**	0.4768
IF + (25)	**0.0000**	**0.0077**	**0.0208**	**0.0344**	**0.0110**	0.2491

IP to run. Nevertheless, within the 1500 s limit all IPs solved in the experiments had enough time to go through the presolve section of Gurobi and generate a solution, even through it was not proven to be optimal. Therefore, 1500 s seems to be reasonable time limit for our problem.

All experiments were coded in Java 8. We used the Gurobi 7.5 [38] optimization solver through Java in order to solve the IP. Experiments were run on a computer with an 2.9 GHz Intel Core i5 processor and 8 GB 1867 MHz memory.

Fig. 7 includes the running times of the initial formulation (IF) and the final formulation (FF), after introducing improvements (21) through (26). For all values of λ the final formulation, including all of the proposed improvements, performed considerably better than the initial formulation. We use the Wilcoxon rank sum test, which is a nonparametric test, in order to test whether the initial and the final formulations have statistically different running times. The null hypothesis states that they are samples from distributions with equal medians. The resulting p-values reject the null hypothesis at the 5% level, with the exception of $\lambda = 1$. The scheduling problem solved already fast for $\lambda = 1$, so adding the valid inequalities did not significantly change the running time.

Fig. 8 presents the results of experiments ran, where proposed improvements in the formulation were added separately in the initial formulation (IF). While all formulation improvements seem to provide faster running times in all cases, valid inequality (25) is the one with the largest proportion of gains in running time. Table 3 includes the resulting p-values of the Wilcoxon rank sum tests. Note that no multiple testing procedure has been used. We see that valid inequalities (24) and (25) rejected the null hypothesis of equal medians for at least some values of λ. However, introducing inequalities (21), (22) to the initial formulation did not result in statistically different running times. Nevertheless, because we see an improvement in the running times even though is not significant at the 5% level, we decided to keep the aforementioned improvements in the formulation.

5. Numerical experiments

In this section, we include the results of the simulation experiments. First, we conduct benchmark analysis. Second, we investigate under which circumstances it is better to group outpatients together when scheduling, and the gains obtained in those cases.

5.1. Simulation

Due to the lack of historic data including the availability of resources over time and information about the outpatients, we make a number of assumptions regarding the distributions of the data.

5.1.1. Outpatients

The outpatients usually get referred to the hospital from different physicians, so the referral times are independent from one another. Therefore, we assume that outpatients get referred to the hospital following a Poisson distribution [23,39]. In particular, outpatients get referred to the hospital only on days Monday through Friday, and during the 8 am–5 pm time range. Based on discussions we had with the hospital personnel we know that on average 60 outpatients per month go through cardiology programs or elective surgery. About 10 patients per month go through the TAVR program, 4 through the TMVR, 4 through the Watchman, 1 through the PFO closure, 1 through the valvuloplasty, and 40 through elective surgery. Each new outpatient that arrives gets assigned to one of the aforementioned outpatient programs or to elective surgery based on the probabilities generated from the corresponding average numbers discussed above. Then, the combination of required steps is determined. Each combination is equally likely to be required for an outpatient of a specific condition. Finally, the outpatients are assigned the days that they are able to come for an appointment to the hospital within the planning horizon. We are assuming that with probability 0.1 a patient will not be able to come to the hospital on a given day. We were not able to find information about the real availability of the patients. We chose a non-zero probability to illustrate that our formulation takes into account the availability of the patient. However, the number is low enough to avoid infeasible solutions. We assume that the outpatients will make the effort to come to the hospital unless there are exceptional circumstances that makes them unavailable. This rare case is captured by this low probability.

Table 4
Procedure durations and recovery times.

Type	Procedure	Duration	Recovery time
Common	TTE [40]	30 min	0 min
	CT scan [41]	30 min	0 min
	Carotid ultrasound [42]	30 min	0 min
	PFT [43]	30 min	30 min
	PREP [44]	2 h	1 h
	TEE [45]	1.5 h	24 h
	Cath [46]	1.5 h	72 h
TAVR	Consultation	1.5 h	30 min
	Procedure [47]	2.5 h	(admitted)
TMVR	Consultation I	1 h	30 min
	Consultation II	30 min	0 min
	Procedure [48]	2.5 h	(admitted)
PFO closure	Consultation	1 h	30 min
	Procedure [49]	2 h	(admitted)
Valvuloplasty	Consultation	1 h	30 min
	Procedure [50]	1.5 h	(admitted)
Watchman	Consultation	1 h	30 min
	Procedure [51]	1.5 h	(admitted)
Surgery	Consultation	1 h	30 min
	Procedure [52]	4 h	(admitted)

5.1.2. Durations

Table 4 includes the duration of each procedure and the minimum required recovery time after a procedure. Those times are maximum estimations of the durations, and are based on information found online and discussions we had with the staff in the hospital. These durations are used to illustrate the performance of the system. However, they can be easily adjusted if the real numbers from the hospital are different. Based on these durations we determine the length of the time unit in the scheduling problem. A time unit is equal to 30 min, since all durations and recovery times are multiples of 30 min. We choose 30-minute slots for this

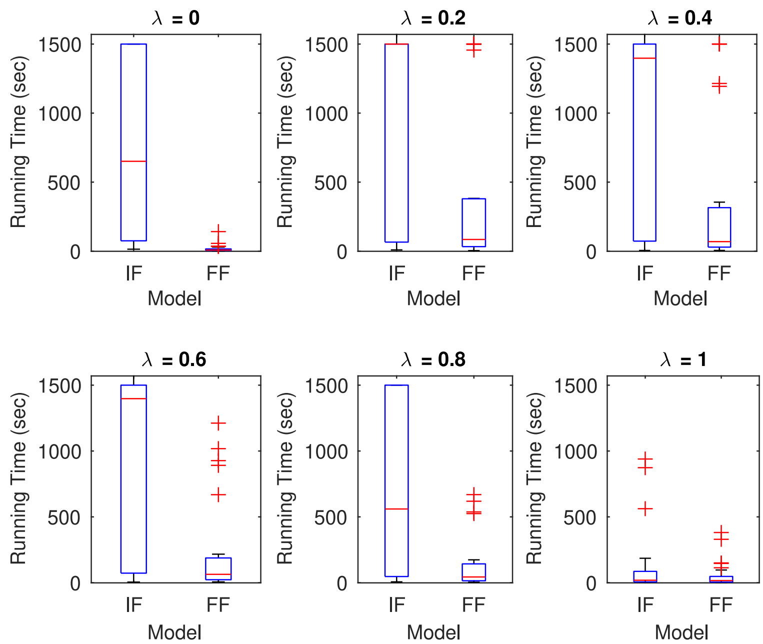


Fig. 7. Running times of initial and final formulation.

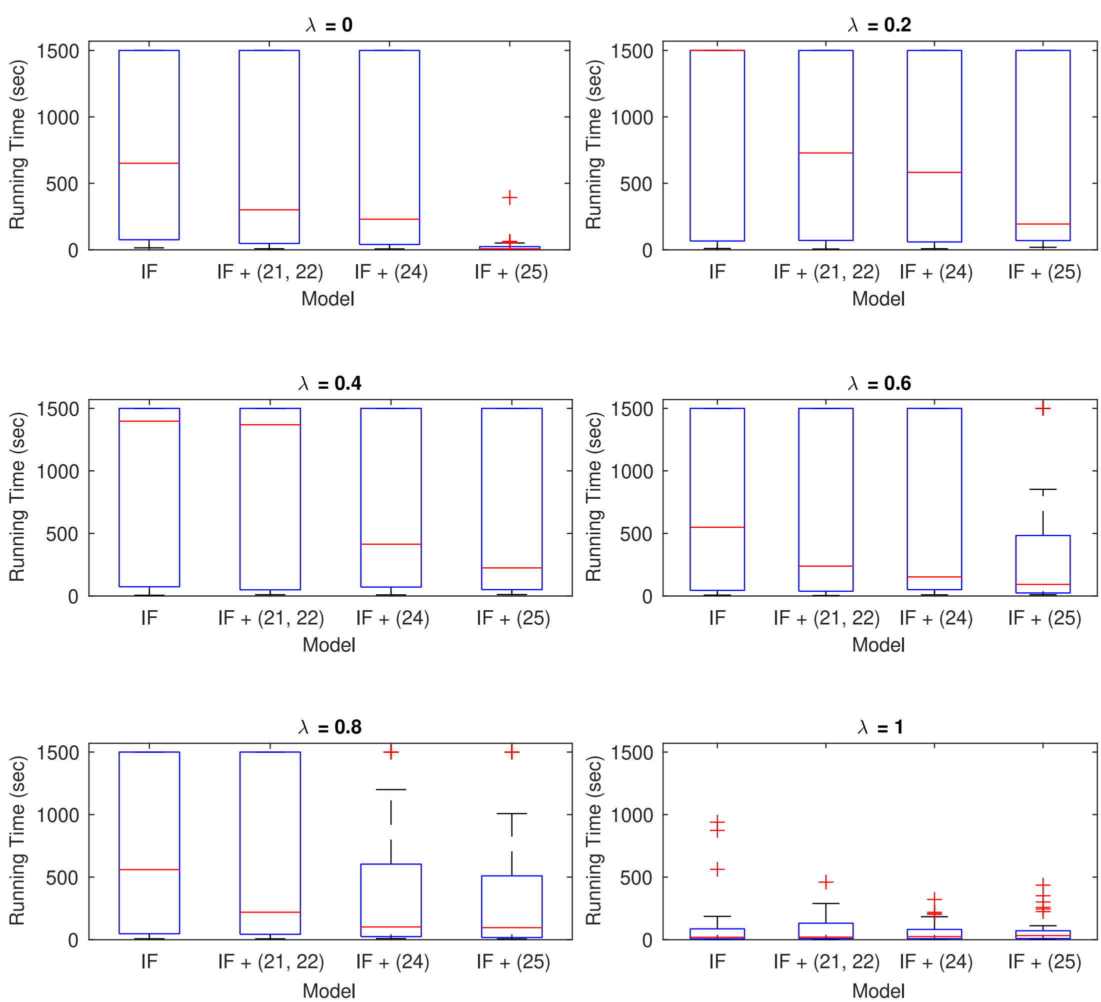


Fig. 8. Running times after separately introducing improvements to the formulation.

simulation, because we have no data available showing the length of the slots used in the hospital. The parameters of the IP can be easily adjusted to include slots of different length, as for example 15 min slots. Thus, the scheduling staff can change the slot length if necessary in the future. In addition to the times presented in Table 4, we assume that the patients will spend some time after they enter the system but before they go through any procedures. This time is set equal to 48 h and it represents the time it takes for the hospital to learn about the patients and decide on the procedures that they will follow. Some procedures require some preparation time, where the patient usually is not allowed to eat or drink for some time. Those procedures are the TEE, the Cath, and all the procedures where the patient gets admitted afterwards.

5.1.3. Scenarios

In the beginning of each simulation we generate the initial resources available. We consider scenarios of different initial availability. In particular, low initial availability corresponds to 30% of the appointments of each resource being free. High availability corresponds to 50% of the appointments being free. The slots of each physical resource are grouped in blocks that equal to the maximum duration of all the procedures that this resource may participate in. In the case of personnel, the slots are grouped in blocks equal to half-days. This way we avoid having fragmented availability of resources. Each block is free with probability 0.5 or 0.3, depending on the scenario. The total number of each type of resource present in the hospital can be found in e-component 3. The available resources are updated during the simulation, as *external demand* and new cardiology outpatients arrive.

The *external demand* arrives following a Poisson distribution during working hours on weekdays. We consider different arrival rates (low and high) of *external demand*. The arrival rates depend on the number of blocks of slots present in the hospital for each resource. In the case of high levels of *external demand* about 14% of the blocks will be occupied in the entire planning horizon. When there are low levels of *external demand* 7% of the blocks present in the hospital will be occupied. Based on these numbers we produce the interarrival times of *external demand* for each type of resource. Thus, resources that can treat higher number of patients during a fixed time also have higher rates of *external demand*. In other words, we assume that a resource is present in higher numbers in the hospital because there is also higher demand for this resource.

We investigate four different scenarios based on the initial resource availability and *external demand*. Scenario 1 considers an environment with low availability of resources and high external demand. Scenarios 2 and 3 consider the cases of low resource availability and low external demand, and high resource availability and high external demand respectively. Scenario 4 corresponds to a setting with high availability of resources and low external demand.

5.1.4. Simulation components

The parameters described in Sections 5.1.1 through 5.1.3 are used as an input in the simulation. Fig. 9 shows the steps followed in our simulation.

1. Based on the scenario we are in, we generate the availability of each type of resource during the planning horizon. Furthermore, we generate the list of cardiology outpatients that are going to get referred to the hospital, which includes the referral time, the steps that they have to complete, and their availability (as discussed in Section 5.1.1). Similarly, depending on the scenario, we generate the external demand, which is a list of resource types and the times that they are requested. Finally, we set the group size (i.e. how many outpatients are scheduled together in each decision epoch).
2. Next we run the simulation, where we schedule appointments for cardiology outpatients and external demand. Based on the referral times of the last outpatient in each group (i.e., if the group size is four, the referral times of the 4th, 8th, 12th outpatient, etc.), and the arriving times of external demand we generate *events*. During an *event* we either solve the IP, if the last outpatient in a group has arrived, or an abbreviated version of the IP, which simply schedules the external demand, if an external demand has arrived. Thus, the IP presented in Section 4.2 is solved $\lceil \frac{\text{\# of outpatients}}{\text{\# of outpatients in each group}} \rceil$ times in each simulation. After each *event* we update the available resources based on the schedule resulting from the corresponding IP. Once the last *event* is executed, the simulation terminates.

As we see from the above, all of the random factors in the system are generated in the first step. This allows us to rerun the second step multiple times using the same input, but with a different group size or λ value each time. Thus, we are able to compare the change in the objective value and the resulting schedules in the same setting. To get different input we use different seeds, which allows us to run multiple simulations and estimate the average effects.

5.2. Numerical results

This section presents the numerical results. We use five different seeds in the simulations. For each seed, we run the simulation once for each combination of: scenario, $\lambda = \{0, 0.2, \ldots, 1\}$, and group size (we consider sizes 1 through 5).

5.2.1. Benchmark analysis

Based on the results from the simulations we conduct a benchmark analysis. First, we estimate the number of visits and the time spent waiting in between appointments for different types of patients in the ideal case, where there are no resource restrictions. In total, we have 88 types of patients in our setting. Each type has a unique set of steps they have to complete. For example, TAVR patients can be classified into 16 different types. However, a TAVR patient that has to go through 3 procedures is expected to behave differently from a TAVR patient that has to go though 7 procedures. For that reason we categorize patients based on the number of steps they are required to go through and not the condition they have. Table 5 includes the average number of visits and the units of time spent waiting when there are no resource restrictions. We use $\lambda = 0.8$ to estimate the results. We can see that patients with more requirements make more visits to the hospital and spend more time waiting in between appointments, even when there are no resource restrictions.

Next, for each scenario, we estimate the average number of visits and the idle time based on the optimal schedules generated from the IP. We assume that $\lambda = 0.8$ in order to compare the results with the best case scenario presented in Table 5. Fig. 10 includes the average change in the values included in Table 5. Note that the values presented the figure are not integer because they correspond to the average over many patients of each type. We see that in most cases Scenario 1 has the largest increase in the number of visits and the idle time. This was expected, since Scenario 1 has the fewest available resources. The opposite effect is observed for Scenario 4. We also see that as the number of requirements increases the patients are more likely to visit the hospital more times than the minimum required. The same is true for the time they spend waiting. This was expected,

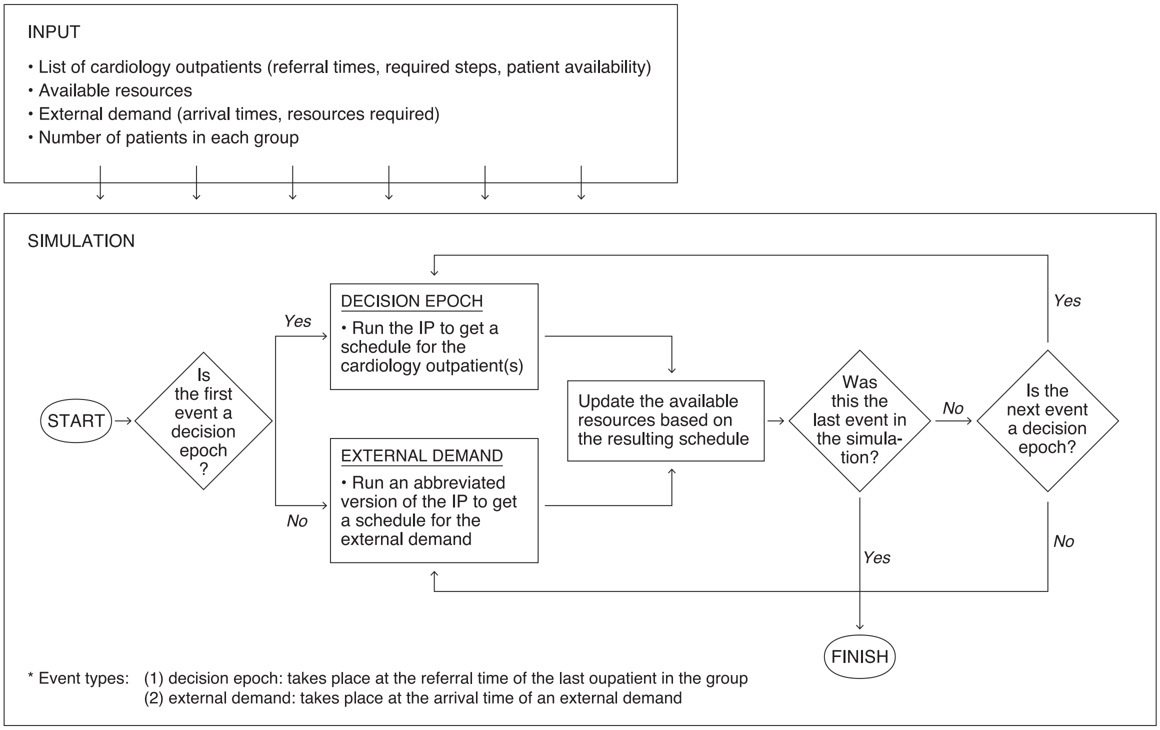


Fig. 9. Simulation diagram and corresponding input.

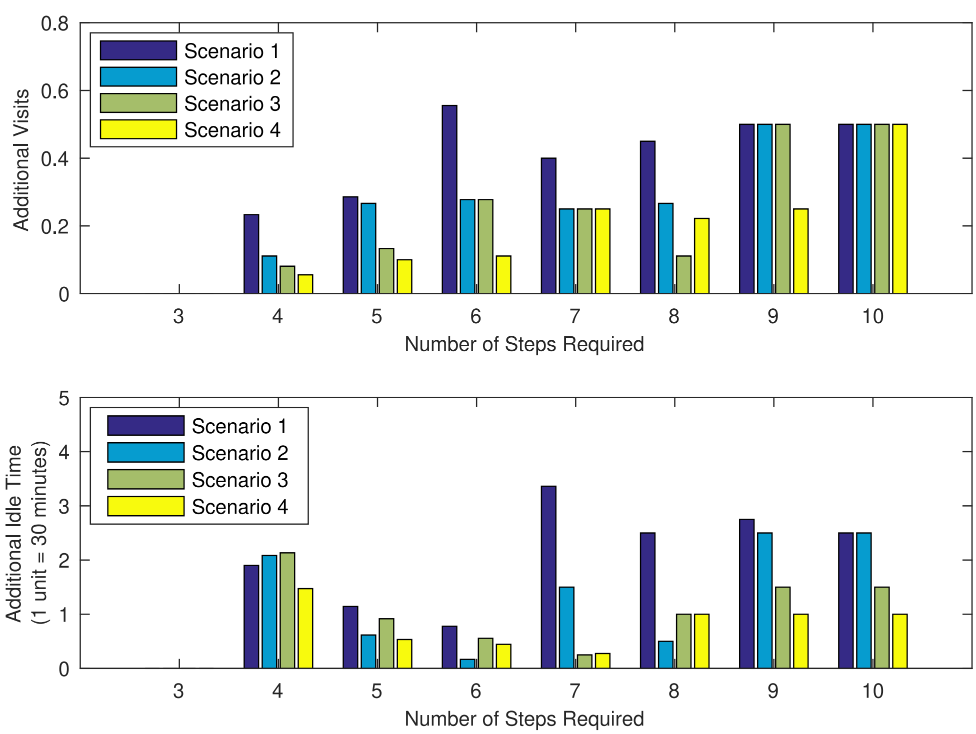


Fig. 10. Additional visits and additional idle time spent in the hospital on average, compared to the *no resource restrictions* setting.

Table 5
Average visits and time spent waiting when there are no resource restrictions.

Number of steps required:	3	4	5	6	7	8	9	10
Average visits	2	2.34	3.01	3.56	4.25	5	5	5
Average idle time	0.67	2.5	2.67	3.56	3.5	4.5	4.5	6

Table 6
Type of change observed in the objective value after scheduling patients in groups.

Outcome	2 per group	3 per group	4 per group	5 per group
No change	28.33%	29.17%	21.67%	27.5%
Increase	8.33%	15.83%	25%	34.17%
Decrease	63.33%	51.67%	45.83%	35%
Infeasible	0%	3.33%	7.5%	3.33%

because when there are limited resources it is harder to create convenient schedules for patients that have more requirements. Nevertheless, there is no indication that the proposed IP benefits some patients more than others. On average, the extra visits that patients have to make are less than 1 for all types of patients. The additional waiting time is less than 2 h on average. Thus, we do not observe a large variation between the different types of patients.

5.2.2. Scheduling in groups

Initially we schedule one patient at a time and then we schedule patients in groups of two to five patients. Patients are grouped together based on the order that they were referred to the hospital. Thus, we are able to compare the total objective value at the end of the month for each case.

Table 6 includes the percentage of group scheduling cases where the objective value increased, decreased, or there was no change observed compared to the baseline case of scheduling one patient at a time. It also includes the percentage of cases where an infeasible solution was obtained. Each column is based on the experiments of a particular group size, which includes the 6 λ values taken into account, and the 4 Scenarios, using 5 different seeds. In other words, in each column we have 120 objective values to compare with the case of grouping 1 patient at a time. Each objective value is compared to the corresponding objective value (same λ, Scenario, and seed) in the setting of scheduling one patient at a time. Thus, the percentages of each column show in which portion of the 120 objective values there was an increase, decrease, no change, or we got an infeasible solution. We see that over 60% of the cases performed better when scheduling patients in groups of two instead of one at a time. However, in about 8% of the cases there was an increase in the objective value. The increase in the objective value when scheduling in groups can be the result of missed opportunities, since we have to wait before scheduling a patient. One situation is that the day that a convenient combination of resources was available for the patient has passed when we finally scheduled the patient. Another case is that the external demand occupied the resources that would have otherwise been occupied by the patient. We observe that as the number of patients getting scheduled simultaneously increases, there is a smaller percentage of cases where a decrease in the objective value was observed. On the other hand, there were more cases of increase in the objective value. The cause of this is that patients getting scheduled in groups of a larger size, usually have to wait longer to get scheduled and therefore have a higher chance of missing a good opportunity. Thus, while scheduling patients in groups provides us more information about patients, waiting too long can have a negative effect. Finally, in the cases of groups of three patients or more, we can observe that there is a chance of getting an infeasible solution. This means that of all the patients scheduled during the month we are taking into consideration, at least one patient was not able to complete all steps in the required time. For those patients it would take more than 30 days to go through all the required steps.

Table 7 presents the results included in Table 6 separately for each value of λ. The chance of getting an infeasible solution does not seem to depend on the value of λ, based on the results included in Table 7. Nevertheless, we can observe that the chance of increase or decrease in the objective value can differ considerably for different values of λ. For example, for $\lambda = 0$ we can see that for all group sizes in the majority of the cases the objective value does not change, excluding the few cases with infeasible solutions. However, for $\lambda = 0.2$ or $\lambda = 0.4$ it seems to make sense to schedule patients in groups of two or three patients, since in most cases the objective value decreases. For $\lambda = 0.8$ the results suggest that we could use even larger groups of patients when scheduling. Nevertheless, as we discussed in Section 4.4, the hospital will be provided with Pareto optimal solutions to choose from. This requires using multiple values for λ, when scheduling the same group of patients, in order to obtain multiple solutions. Therefore, since we have to decide on the size of the group ahead of time, we should select a size that provides good enough results for all values of λ. Therefore, based on Table 7 a good choice seems to be scheduling patients in groups of two or three in order to avoid the increases in the objective value cancel out any potential decreases.

Table 7
Type of change observed in the objective value for different values of λ when scheduling in groups.

λ	Outcome	2 per group	3 per group	4 per group	5 per group
0	No change	100%	95%	90%	95%
	Increase	0%	0%	0%	0%
	Decrease	0%	0%	0%	0%
	Infeasible	0%	5%	10%	5%
0.2	No change	65%	65%	35%	35%
	Increase	0%	0%	20%	25%
	Decrease	35%	35%	40%	35%
	Infeasible	0%	0%	5%	5%
0.4	No change	5%	0%	0%	0%
	Increase	10%	20%	50%	50%
	Decrease	85%	80%	45%	45%
	Infeasible	0%	0%	5%	5%
0.6	No change	0%	0%	0%	0%
	Increase	5%	35%	40%	75%
	Decrease	95%	60%	55%	25%
	Infeasible	0%	5%	5%	0%
0.8	No change	0%	0%	0%	20%
	Increase	10%	20%	20%	10%
	Decrease	90%	75%	70%	65%
	Infeasible	0%	5%	10%	5%
1	No change	0%	15%	5%	15%
	Increase	25%	20%	20%	45%
	Decrease	75%	60%	65%	40%
	Infeasible	0%	5%	10%	0%

When scheduling one patient at a time, the patients get informed quickly about their appointments. However, when grouping the patients together, the patients have to wait until the arrival of the last patient in the group in order to get scheduled. This means that patients may have to wait up to a few days in order to find out about their appointments. The patients prefer to learn their schedule as early as possible, in order to be able to plan ahead. Fig. 11 presents an example of the hours that patients arriving over a month had to wait for each group size. As expected, for smaller group sizes the patients are more likely to wait for a shorter period of time. Also it appears that when scheduling patients in groups of two or in groups of three patients, the wait time is approximately the same. On the other hand, for groups with size four and five we can observe a considerable increase in the wait time.

The level of change in the objective value for each Scenario is presented in Table 8. The results were estimated after excluding the cases where an infeasible solution was obtained. In particular, all of the infeasible cases occurred under Scenario 1, where there

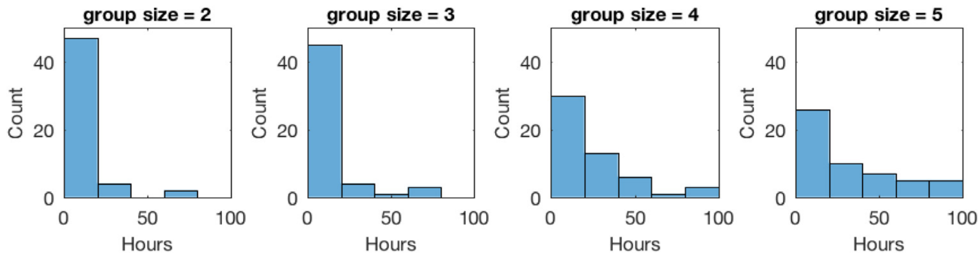


Fig. 11. Wait time in hours from the time the patients got referred to the time that the appointments were scheduled.

Table 8
Amount of change observed in the objective value for each Scenario.

Scenario	Measure	2 per group	3 per group	4 per group	5 per group
1	Mean change	−0.43	−0.21	−0.1	−0.07
	Mean pct. change	**−2.33%**	−1.57%	−0.93%	−0.93%
2	Mean change	−0.14	−0.13	−0.03	−0.03
	Mean pct. change	−0.44%	**−0.45%**	−0.05%	−0.15%
3	Mean change	−0.12	−0.13	−0.04	−0.04
	Mean pct. change	−0.73%	**−0.74%**	−0.32%	1.44%
4	Mean change	−0.05	−0.12	−0.16	−0.10
	Mean pct. change	−0.15%	−0.39%	**−0.59%**	−0.24%

is low availability of resources and high external demand. Infeasible solutions appeared in two out of the five seeds used in the simulation. In each case there was one infeasible group across all the groups getting scheduled. The bold entries in Table 8 indicate the largest percentage improvements across all group sizes for each scenario. In Scenario 1 the highest percentage of decrease was observed when scheduling in groups of two patients. Since Scenario 1 also included infeasible solutions, it seems that having groups of two patients is a reasonable choice in a setting with low initial availability and high external demand. In an environment with big competition for resources, it is best not to risk waiting too long to schedule the appointments of the patients. Scenarios 2 and 3 seem to behave in a similar manner with each other. They both appear to have the highest percentage of decrease when the patients are scheduled in groups of three, but they only perform slightly better compared to scheduling patients in groups of two. Finally, Scenario 4 gives better results when scheduling in groups of four patients. As was expected, when there are many resources available, and not a lot competition caused by the external demand, it is possible to wait for more patients to arrive before scheduling without worrying that the appointments are going to get booked. The results in Table 8 show that the scheduling staff could adjust the size of the groups if they have information about the availability of resources and the expected rate of external demand in the next few weeks. Nevertheless, if is not possible to get an estimate of the state of the system regarding the resources and the demand, the hospital could schedule the patients in groups of two or three based on the results obtained from the simulations. The exact group size should be determined by the hospital depending on the likelihood of getting each Scenario and on their willingness to risk.

Table 9 shows the increase observed in the average time it takes for the outpatients to find out about their appointments, and the percentage of change observed in the objective. We consider the different group sizes separately, but in each case the size of the group increases by 1 patient. We see that the smaller increase in the time it takes to find out the appointments is observed when going from a group size of 1 to a size of 2, and from a group size of 4 to a size of 5. Furthermore, the only transition to a larger group size that leads to a decrease to the objective value in all of the scenarios studied is increasing the group size from 1 patient to 2. Thus, from the above we see that having a group of two patients seems to improve the objective value, while at the same time not considerably increasing the time it takes for outpatients to get informed about their appointments.

While the results show that there are gains in grouping patients when scheduling, the percentages of decrease in the objective value are small. We use an example from the simulations to illustrate. The example is in a Scenario 2 setting, where $\lambda = 0.8$. During the month of the simulation 54 patients were scheduled. When scheduling each patient separately, there were 196 visits to the hospital with a total wait time of 248 time units (124 h). If we scheduled in groups of two patients, the total objective value in this example decreased by 0.29%. The resulting schedule corresponded to 195 visits to the hospital and a wait of 248 time units. In other words, a decrease of 0.29% corresponded to one less visit to the hospital across all patients. On the other hand, if we scheduled in groups of three patients there was a decrease of 1.45% in the objective value, which corresponded to a schedule of 193 visits and 251 time units of waiting. Thus, compared to scheduling one patient at a time, there were three fewer visits to the hospital but the wait time increased by one and a half hours. We can see from the examples above that the improvements to the final schedule are relatively small, and are expected to affect only a few patients among those getting scheduled.

5.2.3. *Scheduling in a larger department*

Based on the results presented in Section 5.2.2, we see that only a small number of patients will get an improved schedule when scheduling in groups in our outpatient cardiology setting. In this section, we investigate the effect of scheduling in groups in a department treating a larger number of outpatients each month. In order to study this, we increased the average number of outpatients referred to the cardiology department by 50%. All other input parameters remain the same based on the description of Section 5.1. This allows us to investigate the effect of scheduling in groups in a larger department.

We ran experiments for $\lambda = 0.8$. The change observed in the objective value compared to scheduling one patient at a time is included in Table 10. The results show that the largest decrease is observed when scheduling in groups of 3 patients. Furthermore, the patients had to wait for a shorter time period in order to find out about their scheduled appointments. For example, the average time was around 10 h in the original setting (smaller

Table 9
Change in the objective value and the wait time to find out about the appointments.

From size	To size	Wait increase (in hours)	Pct. change in objective value			
			Scenario 1	Scenario 2	Scenario 3	Scenario 4
1	2	4.77	−2.33%	−0.44%	−0.73%	−0.15%
2	3	6.19	0.78%	−0.01%	−0.01%	−0.24%
3	4	8.5	0.65%	0.40%	0.42%	−0.20%
4	5	4.45	0.00%	−0.10%	1.77%	0.35%

Table 10
Amount of change observed in the objective value.

Measure	2 per group	3 per group	4 per group	5 per group
Mean percentage change	−1.11%	−4.01%	−1.85%	−1.37%

department) when scheduling in groups of 3 patients. In this setting, where we consider a larger department, the average time decreased to about 6 h. This was expected, because patients arrive more often, and therefore the groups are completed in higher rates.

Let us consider a specific example in this setting. In total, 88 cardiology outpatients were referred to the hospital during the course of one month. In this case, scheduling one patient at a time leads to 316 total visits and 165 h of waiting in the hospital between appointments. Scheduling in groups of three patients at a time leads to 303 visits to the hospital and 164 h of waiting in the hospital. In other words, there were 13 fewer visits in the hospital in this case. Thus, the experiments show us that in larger departments there are greater gains obtained when scheduling in groups, since more patients are affected by the improved schedules. Furthermore, in a larger department the patients do not have to wait as long on average to find out about their schedule.

6. Implementation and managerial insights

This study resulted from discussions we had with the procedural director of the heart and vascular center, who identified the outpatient procedures as an area of interest. However, the proposed IP was not used in the hospital. Nevertheless, in this paper we provide the decision tool and an initial analysis of the expected outcomes. This information can be used in the future if the hospital decides to go forward with this scheduling approach. In this section we discuss the implementation requirements for using such a tool and provide some managerial insights.

The implementation requirements refer to (a) the deployment environment, and (b) information exchange.

With regards to (a), the hospital must have access to a solver (e.g. Gurobi, CPLEX, etc.) coupled with a suitable platform to allow the scheduling staff to provide information about outpatient availability, required steps, and resource availability. With regards to (b), information sharing between different members of scheduling staff will facilitate scheduling patients in groups by gradually inputting information about patients, until the group is complete. In addition, shared information between scheduling and lab staff facilitates the procedure of determining the availability of resources in advance of each appointment and scheduling accordingly while at the same time considering patient convenience.

The proposed IP can help generate schedules that are convenient for the patients. Especially as the number of procedures that the patient has to go through increases it becomes very complicated to schedule the appointments manually. By providing multiple schedules to choose from, the scheduling staff will be able to choose the one that best fits the needs of each specific patient. Finally, we see that grouping patients when scheduling provides better schedules. The improvements only affect a small number of patients in the case of our outpatient cardiology setting. However, in the case of a slightly larger department, which treats more outpatients each month, the results showed that an increased number of outpatients gets positively affected by scheduling in groups. Furthermore, in a larger department the outpatients would not have to wait as long to find out about their appointments. Thus, having groups within the same department collaborate when scheduling can lead to improvements that benefit the outpatients. Additionally, scheduling in groups means that the staff responsible for scheduling the appointments will not have to contact the various resources as often to find out about the availability, since there are fewer decision epochs in each month. Considering that this will also be done collaboratively with other members of the scheduling staff means that this option will safe them time during the course of the month.

7. Conclusions

In this paper we discussed the problem of multi-appointment scheduling in outpatient cardiology. We generated the procedure diagrams including the steps that the patients have to go through in order to be able to complete the procedure they were referred to the hospital to have. Each step was linked to the corresponding resources that need to be available for the patient to be able to go through the step. We proposed an IP formulation, in order to support the staff in making appointment scheduling decisions, which are currently done manually. We identified the objective of the problem as minimizing the combination of the number of visits that the patients make to the hospital and the time the patients spend in the hospital in-between appointments. The formulation allows for patients to be scheduled either one at a time or in groups. The scheduling staff will be provided with Pareto optimal solutions to choose from when deciding on which appointment to book for the patients. We discussed improvements to the initial formulation and added valid inequalities to the IP in order to decrease the time it takes for the solver to find an optimal solution and prove that it is optimal. Through the suggested improvements the running time of the IP decreased significantly. Finally, we investigated the advantages of scheduling patients in groups. We considered various hospital settings based on the initial availability of resources and the rate of external demand. Since many of the resources used by the cardiology outpatients are shared across many departments in the hospital, it is crucial to take into account the external demand generated by the patients in those departments. The results varied for different levels of initial resources and external demand. In particular, when more resources were available there were better results obtained for larger sizes of groups. On the other hand, when fewer resources were available there was a risk of obtaining infeasible solutions when scheduling in large group sizes. Also, there was a larger decrease in the number of visits and the wait time for smaller group sizes. Nevertheless, the improvements observed across the various settings were on the level of few visits or time units. Thus, while on average there are improvements in scheduling in groups, those improvements were not major and expected to affect only a small portion of the patients getting scheduled in our cardiology outpatient setting. Additional experiments showed that in departments that treat an increased number of outpatients, scheduling in groups positively affects a larger portion of patients.

7.1. Limitations and future research

One limitation of this work is that we do not have access to data from the hospital. This non-availability of real data from a hospital is typical. It is due either to privacy concerns by the hospital, or to current processes in many hospitals of not collecting and preserving systematically such data. For that reason, we were not able to generate a simulation model that can be validated to match data from the current system. This did not allow us to make comparisons between the schedules that the scheduling staff currently generates and the schedules resulting from the proposed IP. Furthermore, while we were able to conduct some initial analysis about scheduling patients in groups, this could be examined further. Future directions of this research could investigate ways of getting a good estimate of the state of the system. This refers to the levels of available resources and the rate of external demand. Based on historical data, it could be possible to make predictions about how the availability of resources will change within the scheduling horizon. Thus, the scheduling staff would be able to decide on the size of the groups and adjust them if necessary.

The second limitation of this work is that we assume a number of parameters to be deterministic. We assume that each appointment that the outpatient goes through will start on time, and will not take longer than the duration that it was booked for. However, in reality this might not be the case. Having considerable delays in the system may not allow the patient to go through all necessary steps booked for a specific day. Therefore, future research could investigate how to schedule the outpatients after taking into account this variability in the system. A robust schedule to uncertainty can be achieved by simulating multiple scenarios of the duration of the individual procedures and formulating a stochastic programming model. In this case, the objective of the problem is to minimize not a deterministic value of the number of visits or idle time, rather than a stochastic measure thereof, such as their average or worst case values according to the underlying statistical distribution.

CRediT authorship contribution statement

Lida Anna Apergi: Methodology, Software, Formal analysis, Writing - original draft. **John S. Baras:** Supervision, Methodology, Writing - review & editing. **Bruce L. Golden:** Supervision, Writing - review & editing. **Kenneth E. Wood:** Resources.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.orhc.2020.100267.

References

[1] J.S. Schiller, J.W. Lucas, J.A. Peregoy, Summary Health Statistics for US Adults: National Health Interview Survey, 2011, 2012.
[2] M.P. Heron, Deaths: leading causes for 2017, Natl. Vital Stat. Rep. 68 (6) (2019).
[3] M. Giuseppe De Luca, H. Suryapranata, J.P. Ottervanger, E.M. Antman, Time delay to treatment and mortality in primary angioplasty for acute myocardial infarction, Diabetes 248195 (211135) (2004) 0–002.
[4] K. Dracup, D.K. Moser, M. Eisenberg, H. Meischke, A.A. Alonzo, A. Braslow, Causes of delay in seeking treatment for heart attack symptoms, Soc. Sci. Med. 40 (3) (1995) 379–392.
[5] Johns Hopkins Medicine, News & Publications, TAVR at Johns Hopkins, https://www.hopkinsmedicine.org/news/articles/tavr-at-johns-hopkins, (Accessed 2 September 2019).
[6] Westchester Medical Center, Medical Services, Cardiothoracic Surgery, Transcatheter Valve Program, (TMVR) Transcatheter Mitral Valve Repair, https://www.westchestermedicalcenter.com/transcatheter-valve-program, (Accessed 2 September 2019).
[7] Mayo Clinic, Diseases & Conditions, Patent foramen ovale, https://www.mayoclinic.org/diseases-conditions/patent-foramen-ovale/diagnosis-treatment/drc-20353491, (Accessed 2 September 2019).
[8] T.R. Keeble, A. Khokhar, M.M. Akhtar, A. Mathur, R. Weerackody, S. Kennon, Percutaneous balloon aortic valvuloplasty in the era of transcatheter aortic valve implantation: a narrative review, Open Heart 3 (2) (2016) e000421.
[9] myheart.net, Watchman Device, https://myheart.net/articles/watchman-device-explained-and-faqs-answered-by-a-cardiologist/, (Accessed 2 September 2019).
[10] Michigan Medicine, Cardiac Surgery, Tests Prior to Surgery, https://medicine.umich.edu/dept/cardiac-surgery/patient-information/pre-post-operation-information/tests-prior-surgery, (Accessed 2 September 2019).
[11] S.S. Naidu, H.D. Aronow, L.C. Box, P.L. Duffy, D.M. Kolansky, J.M. Kupfer, F. Latif, S.R. Mulukutla, S.V. Rao, R.V. Swaminathan, et al., SCAI Expert consensus statement: 2016 best practices in the cardiac catheterization laboratory:(Endorsed by the cardiological society of india, and sociedad Latino Americana de Cardiologia intervencionista; Affirmation of value by the Canadian Association of interventional cardiology–Association canadienne de cardiologie d'intervention), Catheter. Cardiovasc. Interv. 88 (3) (2016) 407–423.
[12] M. Rezaeiahari, M.T. Khasawneh, An optimization model for scheduling patients in destination medical centers, Oper. Res. Health Care 15 (2017) 68–81.
[13] E. Castro, S. Petrovic, Combined mathematical programming and heuristics for a radiotherapy pre-treatment scheduling problem, J. Sched. 15 (3) (2012) 333–346.
[14] N. Kortbeek, M. van der Velde, N. Litvak, Organizing multidisciplinary care for children with neuromuscular diseases at the Academic Medical Center, Amsterdam, Health Syst. 6 (3) (2017) 209–225.
[15] T. Cayirli, E. Veral, Outpatient scheduling in health care: a review of literature, Prod. Oper. Manage. 12 (4) (2003) 519–549.
[16] D. Gupta, B. Denton, Appointment scheduling in health care: Challenges and opportunities, IIE Trans. 40 (9) (2008) 800–819.
[17] A. Ahmadi-Javid, Z. Jalali, K.J. Klassen, Outpatient appointment systems in healthcare: A review of optimization studies, European J. Oper. Res. 258 (1) (2017) 3–34.
[18] S. Sevinc, U.A. Sanli, E. Goker, Algorithms for scheduling of chemotherapy plans, Comput. Biol. Med. 43 (12) (2013) 2103–2109.
[19] A. Turkcan, B. Zeng, M. Lawley, Chemotherapy operations planning and scheduling, IIE Trans. Healthcare Syst. Eng. 2 (1) (2012) 31–49.
[20] D. Conforti, F. Guerriero, R. Guido, Optimization models for radiotherapy patient scheduling, 4OR 6 (3) (2008) 263–278.
[21] D. Conforti, F. Guerriero, R. Guido, M. Veltri, An optimal decision-making approach for the management of radiotherapy patients, OR Spectrum 33 (1) (2011) 123–148.
[22] J. Marynissen, E. Demeulemeester, Literature review on multi-appointment scheduling problems in hospitals, European J. Oper. Res. (2018).
[23] A. Leeftink, I. Bikker, I. Vliegen, R. Boucherie, Multi-disciplinary planning in health care: a review, Health Syst. (2018) 1–24.
[24] C.K.Y. Lin, An adaptive scheduling heuristic with memory for the block appointment system of an outpatient specialty clinic, Int. J. Prod. Res. 53 (24) (2015) 7488–7516.
[25] E. Pérez, L. Ntaimo, W.E. Wilhelm, C. Bailey, P. McCormack, Patient and resource scheduling of multi-step medical procedures in nuclear medicine, IIE Trans. Healthcare Syst. Eng. 1 (3) (2011) 168–184.
[26] A. Azadeh, M. Baghersad, M.H. Farahani, M. Zarrin, Semi-online patient scheduling in pathology laboratories, Artif. Intell. Med. 64 (3) (2015) 217–226.
[27] E. Pérez, L. Ntaimo, C.O. Malavé, C. Bailey, P. McCormack, Stochastic online appointment scheduling of multi-step sequential procedures in nuclear medicine, Health Care Manag. Sci. 16 (4) (2013) 281–299.
[28] S. Hahn-Goldberg, M.W. Carter, J.C. Beck, M. Trudeau, P. Sousa, K. Beattie, Dynamic optimization of chemotherapy outpatient scheduling with uncertainty, Health Care Manag. Sci. 17 (4) (2014) 379–392.
[29] A. Leeftink, I. Vliegen, E. Hans, Stochastic integer programming for multi-disciplinary outpatient clinic planning, Health Care Manag. Sci. (2017) 1–15.
[30] A.S. Kapadia, S.E. Vineberg, C. Rossi, Predicting course of treatment in a rehabilitation hospital: a Markovian model, Comput. Oper. Res. 12 (5) (1985) 459–469.

[31] H.-J. Oh, A. Muriel, H. Balasubramanian, K. Atkinson, T. Ptaszkiewicz, Guidelines for scheduling in primary care under different patient types and stochastic nurse and provider service times, IIE Trans. Healthcare Syst. Eng. 3 (4) (2013) 263–279.
[32] A. Saremi, P. Jula, T. ElMekkawy, G.G. Wang, Appointment scheduling of outpatient surgical services in a multistage operating room department, Int. J. Prod. Econ. 141 (2) (2013) 646–658.
[33] A. Braaksma, N. Kortbeek, G.F. Post, F. Nollet, Integral multidisciplinary rehabilitation treatment planning, Oper. Res. Health Care 3 (3) (2014) 145–159.
[34] M. Ehrgott, Multicriteria Optimization, Springer, 2005.
[35] D. Bertsimas, S.S. Patterson, The air traffic flow management problem with enroute capacities, Oper. Res. 46 (3) (1998) 406–422.
[36] D. Bertsimas, G. Lulli, A. Odoni, An integer optimization approach to large-scale air traffic flow management, Oper. Res. 59 (1) (2011) 211–227.
[37] A. Riise, C. Mannino, L. Lamorgese, Recursive logic-based Benders' decomposition for multi-mode outpatient scheduling, European J. Oper. Res. 255 (3) (2016) 719–728.
[38] Gurobi Optimization LLC, Gurobi Optimizer Reference Manual, 2018, http://www.gurobi.com.
[39] L. Green, Queueing analysis in healthcare, in: Patient Flow: Reducing Delay in Healthcare Delivery, Springer, 2006, pp. 281–307.
[40] HEARTe, Imaging, https://www.heartelearning.org/labyrinths?id=47886, (Accessed 2 September 2019).
[41] Mayo Clinic, Patient Care & Health Information, Tests & Procedures, CT Scan, https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675, (Accessed 2 September 2019).
[42] Mayo Clinic, Patient Care & Health Information, Tests & Procedures, Carotid Ultrasound, https://www.mayoclinic.org/tests-procedures/carotid-ultrasound/about/pac-20393399, (Accessed 2 September 2019).
[43] Lung Medicine, Pulmonary Function Tests, http://www.lungmedicine.com/Pulmonary-Function-Test.pdf, (Accessed 2 September 2019).
[44] University of Maryland Medical System, Perioperative Services, PREP Center Appointments, https://www.umms.org/ummc/health-services/perioperative-services/prep-center-appointments, (Accessed 2 September 2019).
[45] Cedars-Sinal, Transesophageal Echocardiography, https://www.cedars-sinai.org/programs/heart/diagnostics/transesophageal-echocardiography.html, (Accessed 2 September 2019).
[46] Heart and Stroke, Heart, Cardiac Catheterization, https://www.heartandstroke.ca/heart/tests/cardiac-catheterization, (Accessed 2 September 2019).
[47] Texas Heart Institute, Women & Transcatheter Aortic Valve Replacement (TAVR), https://www.texasheart.org/heart-health/womens-heart-health/straight-talk-newsletter/women-transcatheter-aortic-valve-replacement-tavr/, (Accessed 2 September 2019).
[48] The University of Vermont Medical Center, Conditions & Treatments, TMVR, https://www.uvmhealth.org/medcenter/Pages/Conditions-and-Treatments/TMVR.aspx, (Accessed 2 September 2019).
[49] Cleveland Clinic, Health Library, Treatments & Procedures, Patent Foramen Ovale (PFO) & Catheter-based Procedures, https://my.clevelandclinic.org/health/treatments/11627-patent-foramen-ovale-pfo--catheter-based-procedures, (Accessed 2 September 2019).
[50] Lahey Health, Percutaneous Balloon Valvuloplasty, https://www.lahey.org/lhmc/department/thoracic-cardiothoracic-surgery/treatments/percutaneous-balloon-valvuloplasty/, (Accessed 2 September 2019).
[51] Franciscan Health, WATCHMAN Device Reduces Stroke Risk; Eliminates Need for Blood Thinners in Patients with Atrial Fibrillation, https://www.franciscanhealth.org/news-and-events/news/new-device-reduces-stroke-risk-eliminates-need-blood-thinners-patients-atrial-fibrillation, (Accessed 2 September 2019).
[52] Healthline, Heart Transplant Surgery, https://www.healthline.com/health/heart-disease/transplants, (Accessed 2 September 2019).