Operations Research for Health Care 35 (2022) 100358

Contents lists available at ScienceDirect

Operations Research for Health Care

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

Integrated master surgery and outpatient clinic scheduling

Thomas Reiten Bovim [a,b,*], Anita Abdullahu [a], Henrik Andersson [a], Anders N. Gullhav [a,b]

[a] Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, NO 7491, Trondheim, Norway
[b] Center for Health Care Improvement, St. Olav's Hospital, Trondheim, Norway

ARTICLE INFO

Article history:
Received 21 September 2021
Received in revised form 31 August 2022
Accepted 9 September 2022
Available online 13 September 2022

Keywords:
Master surgery scheduling
Outpatient scheduling

ABSTRACT

In this paper, we study an integrated master surgery and outpatient clinic scheduling problem, motivated by the situation at the Orthopaedic Department at St. Olav's Hospital, Trondheim. During a treatment process, the patients require one or several consultations at the outpatient clinic, and potentially a surgery in one of the operating rooms. The physicians perform both consultations and surgeries, and coordinating the two facilities is challenging. The surgeons are trained to handle different surgical specialties, and they differ in experience. The overall goal is to schedule the specialties, and a number of qualified surgeons, to time slots in the outpatient clinic and operating rooms through the week, to efficiently handle the patient demand. Our main contribution is an optimisation model for solving the integrated master surgery and outpatient clinic scheduling problem. In addition to allocating specialties and a number of surgeons, the model also schedules activity types (surgery categories and outpatient clinic consultation types) to the time slots. These can guide the operational scheduling of individual patients at a later stage. A computational study is performed, demonstrating the use of the optimisation model to provide a set of master schedules, based on a set of different resource capacity cases. We develop a simulation model for evaluating the master schedules in an operational setting, and three different operational scheduling policies are compared. We conclude that scheduling patients to activities governed primarily by the optimisation model solution outperforms a FIFO scheduling policy based only on specialty.

1. Introduction

Surgical costs account for approximately 40% of the total hospital costs [1], and Freeman et al. [2] state that 60%–70% of all patients admitted to a hospital require some surgical intervention. However, surgeries are not performed in isolation, and by neglecting aspects of coordination when performing capacity planning, we may arrive at suboptimal solutions. Surgical patients typically require a consultation at the outpatient clinic (OC) both prior to, and following surgery. In addition, many patients require a stay in a hospital ward to recover from surgery. The surgeons serve both the OC rooms and the operating rooms (ORs), and their time must be carefully divided between the two facilities to provide a coordinated and efficient service.

The Master Surgery Scheduling Problem (MSSP) is a frequently studied problem within operations research. The Master Surgery Schedule (MSS), is a cyclic schedule where surgical specialties are assigned to OR slots through the week, such that the demand for surgery is covered. The time horizon considered in the MSSP is typically in the range of months, and the MSS is repeated through the time horizon. An equivalent problem can be formulated for the OC.

In this paper, we study the integrated master surgery and outpatient clinic scheduling problem. The problem is motivated by the situation at the Orthopaedic Department as St. Olav's Hospital, Trondheim. Vik et al. [3] report poor coordination of key health care activities as one of the major challenges within the department. From our point of view, a lack of coordination between the OC and the ORs can lead to unwanted variations in demand for activities, and we believe that the work presented in this paper can help to increase coordination.

During the treatment process, the patients require one or several consultations in the OC rooms, and potentially a surgery in one of the ORs. The surgeons are trained to handle different surgical specialties, and they differ in experience. The goal is to handle patient demand by allocating specialties, and a number of qualified surgeons, to time slots in the OC rooms and the ORs through the planning horizon. In addition, we also schedule activity types (surgery categories and outpatient clinic consultation types) to the available time slots, which can be used for scheduling individual patients at a later stage. To our knowledge this problem has not been studied in literature. It differs both from

* Corresponding author at: Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, NO 7491, Trondheim, Norway.
E-mail address: thomas.bovim@ntnu.no (T.R. Bovim).

https://doi.org/10.1016/j.orhc.2022.100358

multi-appointment scheduling, which mainly considers outpatient services, and from the MSSP, that seldom includes upstream units.

The main contribution in this paper is an optimisation model for solving the integrated master surgery and outpatient clinic scheduling problem. The model produces two cyclic master schedules; one for the OC rooms and one for the ORs. To evaluate the performance of the master schedules, they are implemented in a discrete-event simulation (DES) model, where patient arrivals and the paths of individual patients are modelled as stochastic processes. We compare three different operational scheduling policies. This allows us to investigate the value of scheduling activity types in the master schedules.

A computational study is performed based on data from the Orthopaedic Department at St. Olav's hospital. In this study, five alternative resource capacity cases are considered, each representing a strategy which the department can implement to increase patient throughput. Even though we study the situation at the orthopaedic department, the problem under study is rather generic and can be found in many hospital departments that perform surgical activities. It is common for such departments to serve patients that require multiple services, and where a set of resources are involved in providing more than one service. With some adjustments, the problem under study is relevant to departments that face similar problems.

The rest of the paper is structured as follows. In Section 2, a literature review is provided to position our contribution. Then, in Section 3 the problem under consideration is described. The mathematical model is presented in Section 4, while the simulation model is provided in Section 5. In Section 6, the computational study is reported on, before presenting managerial insights in Section 7. Finally, in Section 8, the paper is concluded.

2. Literature review

Tactical outpatient and surgery scheduling, that considers multiple resources, constitutes the field of interest. First we present selected literature on surgery scheduling, and the MSSP in particular, before turning to the tactical OC planning.

2.1. Tactical OR planning

Within health care planning, the tactical decision level addresses the organisation of the execution of the health care process [4]. At this level, the available resource capacities, settled at the strategic level, are divided among patient groups. Blueprints for the operational planning are created that allocate resources to different tasks, specialties and patient categories. According to Hulshof et al. [4], the MSSP is considered a tactical planning problem within surgical care services.

Cardoen et al. [5] review the literature on OR planning and scheduling, and find that about half of the recent contributions limit their scope to an isolated OR. Among the contributions that regard additional facilities, the wards, the Intensive Care Unit and the Post Anesthesia Care Unit are most frequently included. However, the modelling of ORs in interaction with other hospital facilities remains a main topic for further research [5].

When considering the MSSP, the downstream facilities, and the wards in particular, are frequently considered. Li et al. [6], Moosavi and Ebrahimnejad [7] and Adan et al. [8] include the Intensive Care Unit when analysing the MSSP, while Schneider et al. [9] and Fügener et al. [10] consider multiple downstream units. The upstream activities are seldom regarded in the MSSP literature, and Schneider et al. [9] propose the inclusion of upstream units, such as the OC, as a topic for future research. Moosavi and Ebrahimnejad [7] regard the wards as both up- and downstream capacities, acknowledging the fact that patients might need a bed both prior to, and following surgery.

Schneider et al. [9] cluster surgery types into surgery groups, and instead of scheduling surgery specialties, they schedule surgery groups to the OR blocks. This eases the operational scheduling of individual patients.

Unlike the majority of literature on the MSSP, we consider both up- and downstream activities in our problem. By including the OC we can schedule more of the surgeons' activities, and instead of considering the derived demand for surgery we can include the demand for new referrals. This allows us to control the number of patients that are sent to surgery based on the overall capacity of the system. Furthermore, like Schneider et al. [9], we schedule surgery groups for the OR slots, and through simulation we evaluate the value of using these when performing the operational scheduling of surgeries.

2.2. Tactical OC planning

According to Hulshof et al. [4], OC planning is categorised as ambulatory care services, and the existing literature is mainly focusing on the operational appointment scheduling. At the tactical level, the allocation of capacity to patient groups is the most frequently studied problem in the OC planning literature [11].

Historically, the majority of research has considered patients requiring a single appointment. However, in recent years, an increasing number of researchers has considered several resources and the fact that patients may require multiple consultations [12]. According to Marynissen and Demeulemeester [12], the multi-appointment scheduling problem is designed to act as an umbrella for both combination appointments, in which patients require multiple appointments on the same day, and appointment series, in which patients need to revisit the same set of resources several times. Furthermore, the authors define the multi-appointment scheduling problem as an operational problem, but emphasise the importance of reserving capacity at a tactical level. Examples of tactical multi-appointment scheduling problems can be found in Bikker et al. [13], Nguyen et al. [14] and Hahn-Goldberg et al. [15]. The first two represent an appointment series problem, while the last is a combination appointment problem for scheduling patients for chemotherapy.

Care processes can be analysed as a multi-appointment scheduling problem, as the patients typically require several visits to the hospital. Hulshof et al. [16] consider the tactical resource allocation for elective patient admission planning in care processes. The authors analyse a care process comprising of a visit to the OC, followed by surgery and a revisit to the OC, which is similar to the one studied in this paper. In the modelling framework presented by Hulshof et al. [16], each care process is represented by a set of consecutive queues, and patients are routed between queues to represent the demand for each care process. In each time period, a number of patients are served in each queue, while the patients remain in the queue to the next time period. To serve a patient in a given queue, a set of resources are required throughout the time period. All resources have a given capacity, restricting the flow of patients between queues. The decision variables represent the number of patients treated from each queue, in each time period, and a solution represents the resource capacity devoted to each queue in each time period.

Another branch of hospital planning that relates to multi-appointment scheduling is multi-disciplinary scheduling. Leeftink et al. [17] define a multi-disciplinary care system as a care system in which multiple interrelated appointments per patient are scheduled, where health care professionals from various facilities, or with different skills are involved. The authors categorise the literature according to a hierarchical planning structure, and within capacity planning the generation of blueprint schedules, patient admission planning and temporary capacity changes are typical outcomes. Like for multi-appointment scheduling, the consultations in multi-disciplinary scheduling can be performed within a day (see Liang et al. [18]), or as multiple revisits (see Braaksma et al. [19]).

To the best of our knowledge, the existing papers on tactical multi-appointment scheduling (and similar problems) in hospitals only consider outpatients. Although Hulshof et al. [16] encounter a setting which has similarities to ours, the problems differ. In Hulshof et al. [16], the resource requirements to perform an activity (serve patients from a queue) is given. In our problem, surgeons of different experience levels can perform OC consultations, allowing for flexibility. Furthermore, Hulshof et al. [16] apply discrete time steps in their model, requiring that each activity is started and finished within one time period. If a time step resembles one day, coordination of resources within a day is problematic as surgeons cannot serve one activity in the morning and another in the afternoon. If the time steps represent shorter time periods, coordination is possible. However, requiring that each activity spans one time period makes short time steps problematic. Finally, beds cannot be analysed as resources, as patients cannot wait for a bed following surgery. Therefore, their framework is most suitable for an outpatient setting.

Our model is suitable for handling inpatients, and we consider the demanding task of coordinating and scheduling resources to serve patients that require multiple services.

3. Problem description

In the integrated outpatient and master surgery scheduling problem, the aim is to generate two cyclic schedules; one for the OC rooms, and one for the ORs. The time available during the working day is divided into fixed time slots, and we define a slot within a room as a room slot. In both schedules, surgical specialties are assigned to the available room slots through the planning cycle (typically one week), such that the system can serve at least the expected demand of new referrals during the planning horizon (typically half a year). As an example, Fig. 1 illustrates two such schedules where two OC rooms and two ORs are shared between the orthopaedic and the surgical department. Here, two slots are available in each room, representing a morning and an afternoon slot. A scheduled slot in the operating theatre comprises both the medical specialty and the number of surgeons assigned to the slot. For the OC, only the specialty is given, as it is always sufficient with one surgeon to perform an OC consultation.

Four activity types can be performed; surgery and three different types of OC consultations. In an *initial consultation* (IC), the patient is examined and the surgeon decides on whether further intervention is needed, and if so, what kind of intervention. If a non-surgical intervention is required, the patient receives a *treatment consultation* (TC) in the OC. Finally, following either surgery or a treatment consultation, the patient is summoned for one or more *follow-up consultations* (FU) in the OC. The length of an OC consultation depends on the specialty of the patient, and what type of consultation that is performed. Surgeries are categorised depending on what procedures that are done during surgery. A surgery category is characterised by the planned surgery duration, the minimum number of surgeons that must be present during surgery, and the planned length of stay (LOS) in a ward following surgery.

There exists a set of surgical specialties, and patients are categorised based on the specialty they belong to. The activity types required by a patient depends on what specialty the patient belongs to. Following the initial consultation, a share of patients belonging to a specialty requires a treatment consultation, while another share requires a surgery. The remaining patients leave the system. Following a treatment consultation or surgery, the patients require one or multiple follow-up consultations. A number of patients within each specialty is referred straight for surgery, and the remaining paths of these patients are identical to the other patients within the same specialty.

Fig. 2 illustrates the activity types considered in the problem, and how they relate to each other for a specialty with two surgery categories available. In this example, x_{IC} initial consultations are scheduled. The expected demand for treatment consultations and surgery categories one and two are calculated as the probabilities that patients require these activities following an initial consultation multiplied by x_{IC}. Similar logic is used to calculate the expected demand of follow-up consultations based on the number of surgeries and treatment consultations that are scheduled.

There is a given number of OC rooms and ORs. The opening hours of the rooms are divided into consecutive time slots. The slot duration can differ between the OC rooms and the ORs, but they are of constant length within each facility. The slots are synchronised, such that a surgeon can serve one of the facilities in the morning, and the other one in the afternoon. As a consequence of these requirements, there can be at most two slots available during a day in each of the facilities. Each slot can be scheduled for one specialty, and the number of patients that can be scheduled within a slot is limited by the slot duration. If a specialty covers two consecutive slots in a room, an activity may begin in the first slot and end in the other. There are several wards available to serve the inpatients that require a bed following surgery. In every ward, a given number of beds are available each day, and each ward can serve patients from a subset of the surgery categories.

The surgeons are categorised according to what specialties they master, and their level of experience. Surgeons can be trained to master several specialties, and surgeons that master a given specialty may provide all activity types to patients belonging to that specialty. Based on the level of experience, surgeons are either consultants or residents, and each surgery requires the presence of at least one consultant. Both consultants and residents can perform activity types in the OC. Each surgeon type has a fixed number of surgeons available each day of the cycle.

To schedule a surgery of a given category there must be enough time available in an OR and a slot scheduled for the corresponding specialty. Furthermore, there must be enough surgeons scheduled to the same room and slot to perform the surgery, and the bed capacity must be sufficient to cover the entire LOS of the patient. To schedule an OC activity there must be enough time available in an OC room and a slot scheduled for the corresponding specialty.

The expected arrival rate of initial consultations during a cycle is constant and known for each specialty. In addition, there is a queue of patients that have not yet received an initial consultation at the beginning of the planning horizon. To maintain a stable waiting list for each specialty, a minimum throughput of initial consultations should be set such that the service rate is at least incrementally higher than the expected arrival rate of new referrals. Furthermore, to decrease a potential queue of patients waiting for an initial consultation, a reward is given for scheduling more than the minimum throughput. However, the maximum number of initial consultations that can be scheduled for each specialty in a cycle cannot exceed the expected arrival rate of new referrals, plus the length of the waiting list divided by the number of cycles in the planning horizon. To avoid queues from building up within the system, we must ensure that the system can handle the downstream demand for services generated.

Our goal is to serve at least the expected number of new referrals, and more if possible, while making sure that there is sufficient capacities to handle the derived demand for downstream services.

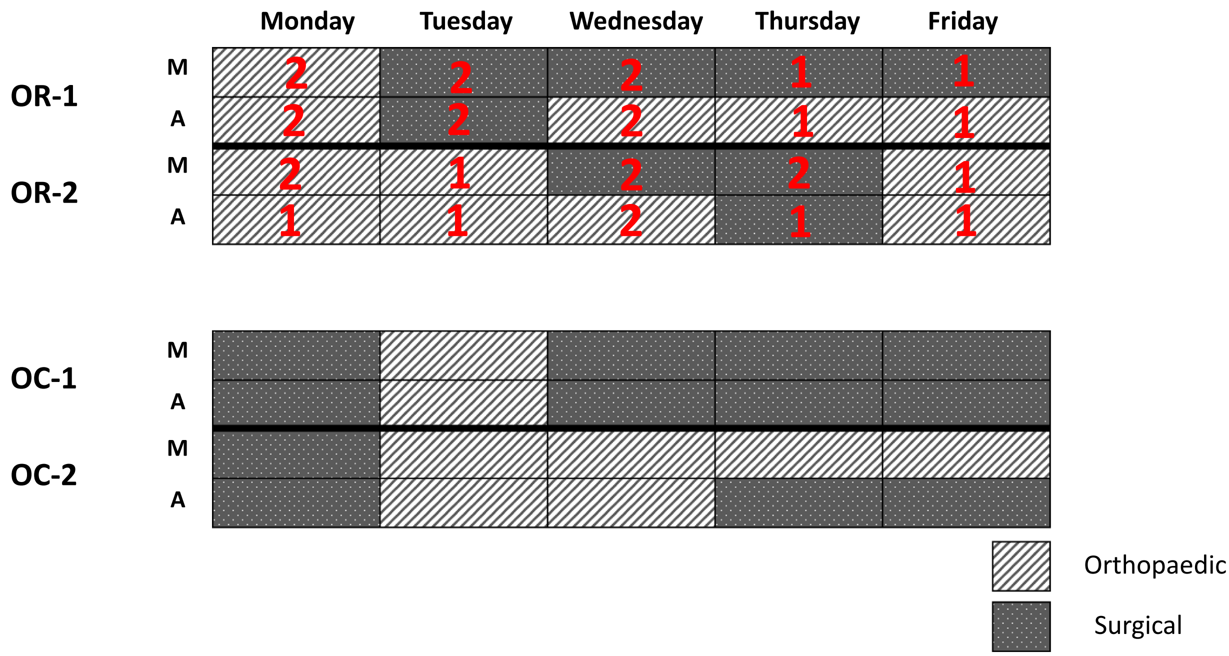

Fig. 1. Example of master schedules for the ORs and the OC rooms. Here, the capacity is shared between the orthopaedic and the surgical department. For the ORs, the number of surgeons assigned to each OR is given in addition to the specialty. In this example, two slots are available in each room each day. M is the morning slot, while A is the afternoon slot.

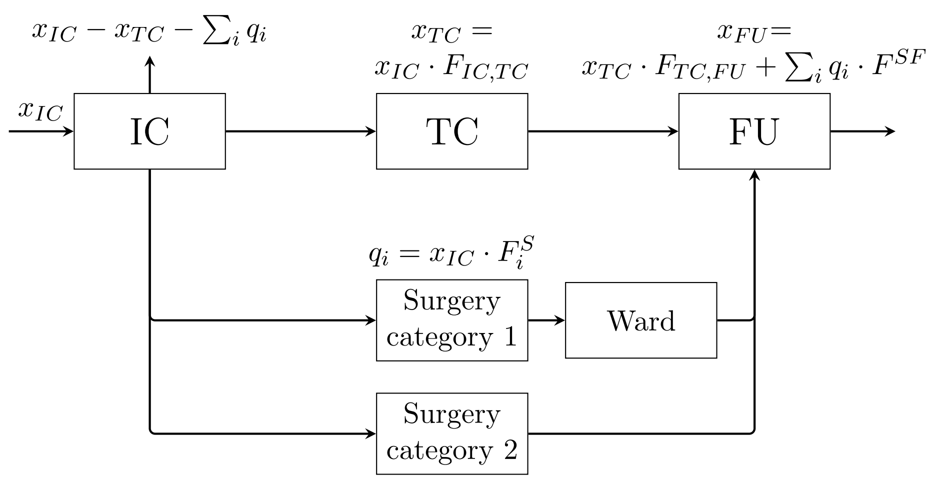

Fig. 2. The activity types considered in the problem, and how they relate to each other. x_{IC}: Number of initial consultations scheduled. x_{TC}: Derived demand for treatment consultations that must be scheduled. q_i: Derived demand for surgery category i that must be scheduled. x_{FU}: Derived demand for follow-up consultations that must be scheduled. $F_{IC,TC}$: Average fraction of initial consultations that yield a demand for treatment consultations. $F_{TC,FU}$: Average fraction of treatment consultations that yield a demand for follow-up consultations. F_i^S: Average fraction of initial consultations that yield a demand for surgery of category i. F^{SF}: Fraction of surgeries that yield a demand for follow-up consultations.

4. The mathematical model

In this chapter, the mathematical model for solving the problem is presented. Tables 1, 2, and 3 include all the notation used in the mathematical model. To enhance readability, the constraints are introduced in thematic groups.

Demand constraints

$$D_j \le \sum_{l\in\mathcal{L}} \sum_{d\in\mathcal{D}} x_{j,IC,ld} \le \overline{D}_j \qquad j \in \mathcal{J} \tag{1}$$

The demand constraints (1) ensure that we schedule between the planned demand and the upper limit of initial consultations for each specialty j. Here, $\overline{D}_j$ is the maximum number of initial consultations of specialty j that can be scheduled in a cycle, while D_j is the minimum throughput of new referrals for one cycle. $\overline{D}_j$ can be calculated by adding the number of patients from specialty j waiting for an initial consultation at the beginning of the planning horizon, Q_j^I, divided by the number of cycles in the planning horizon, T, to the minimum throughput:

$$\overline{D}_j = D_j + \left\lceil \frac{Q_j^I}{T} \right\rceil \qquad j \in \mathcal{J} \tag{2}$$

Slot constraints

$$\sum_{j\in\mathcal{J}} \beta_{jlsd} \le 1 \qquad l \in \mathcal{L}, s \in \mathcal{S}, d \in \mathcal{D} \tag{3}$$

$$\sum_{n\in\mathcal{N}} \sum_{j\in\mathcal{J}_k^K} \lambda_{njksd} \le 1 \qquad k \in \mathcal{K}, s \in \mathcal{S}, d \in \mathcal{D} \tag{4}$$

Table 1
Sets.

Symbol	Description	
$\mathcal{A}$	Set of consultation types performed at the OC	$a \in \mathcal{A} = \{IC, TC, FU\}$
$\mathcal{D}$	Set of days in a cycle	$d \in \mathcal{D}$
$\mathcal{I}$	Set of surgery categories	$i \in \mathcal{I}$
$\mathcal{J}$	Set of surgical specialties	$j \in \mathcal{J}$
$\mathcal{K}$	Set of ORs	$k \in \mathcal{K}$
$\mathcal{L}$	Set of OC rooms	$l \in \mathcal{L}$
$\mathcal{N}$	Set of number of surgeons that can be present during surgery	$n \in \mathcal{N}$
$\mathcal{P}$	Set of surgeon types	$p \in \mathcal{P}$
$\mathcal{S}$	Set of time slots	$s \in \mathcal{S}$
$\mathcal{W}$	Set of wards	$w \in \mathcal{W}$
$\mathcal{D}^{LOS}_{id}$	Set of days that a patient of surgery category i can have received surgery if the patient is still in a ward on day d	$d' \in \mathcal{D}^{LOS}_{id}$
$\mathcal{I}^J_j$	Set of surgery categories that can be handled by specialty j	$i \in \mathcal{I}^J_j \subseteq \mathcal{I}$
$\mathcal{I}^W_w$	Set of surgery categories that can rest in ward w following surgery	$i \in \mathcal{I}^W_w \subseteq \mathcal{I}$
$\mathcal{J}^K_k$	Set of specialties that can be scheduled to OR k	$j \in \mathcal{J}^K_k \subseteq \mathcal{J}$
$\mathcal{K}_j$	Set of ORs that can be utilised by specialty j	$k \in \mathcal{K}_j \subseteq \mathcal{K}$
$\mathcal{P}^C_j$	Set of surgeon types that are consultants and can cover specialty j	$p \in \mathcal{P}^C_j \subseteq \mathcal{P}$
$\mathcal{P}^R_j$	Set of surgeon types that are residents and can cover specialty j	$p \in \mathcal{P}^R_j \subseteq \mathcal{P}$
$\mathcal{W}^I_i$	Set of wards that can serve patients from surgery category i	$w \in \mathcal{W}^I_i \subseteq \mathcal{W}$

Table 2
Parameters.

Symbol	Description
A_{wd}	Number of staffed beds available for elective patients at ward w on day d
C_{pd}	Number of surgeons available of surgeon type p on day d
$\underline{D}_j$	Minimum number of initial consultations of specialty j that must be scheduled during one cycle
$\overline{D}_j$	Maximum number of initial consultations of specialty j that can be scheduled during one cycle
F^S_i	Fraction of initial consultations that yield a downstream demand for surgery category i
$F_{jaa'}$	Fraction of OC consultations of type a that yield a downstream demand for OC consultations of type a' for specialty j
F^{SF}_j	Fraction of surgeries of specialty j to be scheduled for a follow-up appointment at the OC
H^{OC}_{ja}	Planned time needed for consultations of type a of specialty j
$\underline{N}_i$	Minimum number of surgeons that must be present for a surgery of category i
$\overline{N}$	Maximum number of surgeons that can be present during surgery
Q_i	Number of patients from surgery category i that are referred straight to surgery during a cycle
Q^I_j	Number of patients from specialty j waiting for an initial consultation at the beginning of the planning horizon
$\overline{Q}_{nikd}$	Maximum number of surgeries of surgery category i that can be scheduled for surgery with n surgeons present in OR k on day d
R_j	Reward obtained from scheduling an initial consultation of specialty j
S_i	Planned surgery duration for patients of surgery category i
T	Number of cycles in the planning horizon
T^{OR}_{ksd}	Time available for surgeries in OR k, slot s on day d
T^{OC}_{lsd}	Time available for consultations in OC room l, slot s on day d
$\overline{X}_{jald}$	Maximum number of consultations of type a of specialty j that can be scheduled to OC room l on day d

Table 3
Variables.

Letter	Description
g^{OC}_{pjsd}	Number of surgeons from surgeon type p that cover specialty j allocated to slot s in the OC on day d
g^{OR}_{pksd}	Number of surgeons from surgeon type p allocated to OR k in slot s on day d
q_{nikd}	Number of surgeries of surgery category i scheduled in OR k with n surgeons present on day d
u_{iwd}	Number of beds occupied by patients of surgery category i in ward w on day d
x_{jald}	Number of OC consultations of type a of specialty j scheduled in OC room l on day d
β_{jlsd}	Indicates if specialty j is assigned OC room l in slot s on day d
λ_{njksd}	Indicates if specialty j is assigned OR k in slot s, with n surgeons present on day d

Constraints (3) make sure that at most one specialty j is assigned to each slot s in every OC room l on day d. Constraints (4) ensure that at most one specialty j with n surgeons is assigned to each slot s in OR k on day d.

Surgeon constraints

$$\sum_{l \in \mathcal{L}} \beta_{jlsd} \leq \sum_{p \in \mathcal{P}^C_j} g^{OC}_{pjsd} + \sum_{p \in \mathcal{P}^R_j} g^{OC}_{pjsd} \qquad j \in \mathcal{J}, s \in \mathcal{S}, d \in \mathcal{D} \tag{5}$$

$$\lambda_{njksd} \leq \sum_{p \in \mathcal{P}^C_j} g^{OR}_{pksd} \qquad n \in \mathcal{N}, j \in \mathcal{J}, k \in \mathcal{K}_j, s \in \mathcal{S}, d \in \mathcal{D} \tag{6}$$

$$n\lambda_{njksd} \leq \sum_{p \in \mathcal{P}^C_j} g^{OR}_{pksd} + \sum_{p \in \mathcal{P}^R_j} g^{OR}_{pksd} \qquad n \geq 2, j \in \mathcal{J}, k \in \mathcal{K}_j, s \in \mathcal{S}, d \in \mathcal{D} \tag{7}$$

$$\sum_{j \in \mathcal{J}} g^{OC}_{pjsd} + \sum_{k \in \mathcal{K}} g^{OR}_{pksd} \leq C_{pd} \qquad p \in \mathcal{P}, s \in \mathcal{S}, d \in \mathcal{D} \tag{8}$$

Constraints (5) ensure that all OC rooms that are scheduled for specialty j in slot s on day d, must be covered by at least one surgeon each. Constraints (6) require that, if specialty j is assigned to OR k in slot s on day d with n surgeons, there should be at least one consultant of the same specialty assigned to that OR, at that point in time. Constraints (7) state that, if specialty j is assigned to OR k in slot s on day d with two or more surgeons, this OR must be covered by enough surgeons from that specialty, at that point in time. Constraints (8) make sure that the number of surgeons allocated to slot s from surgeon type p on a given day d does not exceed the number of surgeons available from that surgeon type on that day.

Time capacity constraints in the OC

$$\sum_{a\in\mathcal{A}} H^{OC}_{ja} x_{jald} \leq \sum_{s\in\mathcal{S}} T^{OC}_{lsd}\beta_{jlsd} \qquad j\in\mathcal{J}, l\in\mathcal{L}, d\in\mathcal{D} \tag{9}$$

The time capacity constraints (9) make sure that the total time scheduled for initial consultations, treatment consultations, and follow-up consultations for specialty j, in OC room l on day d, cannot exceed the time scheduled for that specialty in that room on that day.

Patient flow constraints

$$\sum_{l\in\mathcal{L}}\sum_{d\in\mathcal{D}} F^S_i x_{j,IC,ld} + Q_i \leq \sum_{n=\underline{N}_i}^{\overline{N}}\sum_{k\in\mathcal{K}_j}\sum_{d\in\mathcal{D}} q_{nikd} \qquad j\in\mathcal{J}, i\in\mathcal{I}^J_j \tag{10}$$

Constraints (10) state that the number of scheduled surgeries of surgery category i, is at least the same as the sum of the planned demand for surgery of category i, derived from the initial consultations scheduled for specialty j, and the expected number of surgeries of surgery category i referred from other instances.

$$\sum_{l\in\mathcal{L}}\sum_{d\in\mathcal{D}} F_{j,IC,TC} x_{j,IC,ld} \leq \sum_{l\in\mathcal{L}}\sum_{d\in\mathcal{D}} x_{j,TC,ld} \qquad j\in\mathcal{J} \tag{11}$$

Constraints (11) ensure that we, for each specialty j, schedule enough treatment consultations in relation to initial consultations.

$$\sum_{n\in\mathcal{N}}\sum_{i\in\mathcal{I}^J_j}\sum_{k\in\mathcal{K}_j}\sum_{d\in\mathcal{D}} F^{SF}_j q_{nikd} + \sum_{l\in\mathcal{L}}\sum_{d\in\mathcal{D}} F_{j,TC,FU} x_{j,TC,ld} \leq \sum_{l\in\mathcal{L}}\sum_{d\in\mathcal{D}} x_{j,FU,ld} \qquad j\in\mathcal{J} \tag{12}$$

Constraints (12) make sure that we schedule at least the required fractions of follow-up appointments from specialty j after surgery or after a treatment consultation.

Time capacity constraints for surgery

$$\sum_{i\in\mathcal{I}^J_j} S_i q_{nikd} \leq \sum_{s\in\mathcal{S}} T^{OR}_{ksd}\lambda_{njksd} \qquad n\in\mathcal{N}, j\in\mathcal{J}, k\in\mathcal{K}_j, d\in\mathcal{D} \tag{13}$$

Constraints (13) ensure that time capacity is respected in the ORs. The total time scheduled for surgery categories belonging to specialty j within OR k on day d cannot exceed the time scheduled for that specialty, in that room, on that day.

Ward constraints

$$\sum_{i\in\mathcal{I}^W_w} u_{iwd} \leq A_{wd} \qquad w\in\mathcal{W}, d\in\mathcal{D} \tag{14}$$

$$\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}}\sum_{d'\in\mathcal{D}^{LOS}_{id}} q_{nikd'} = \sum_{w\in\mathcal{W}^I_i} u_{iwd} \qquad i\in\mathcal{I}, d\in\mathcal{D} \tag{15}$$

Constraints (14) state that the number of beds occupied at ward w on day d does not exceed the available number of beds at the ward. In constraints (15), we count the number of patients of surgery category i still present at a ward on a given day d.

Objective function

$$\max \sum_{j\in\mathcal{J}}\sum_{l\in\mathcal{L}}\sum_{d\in\mathcal{D}} R_j x_{j,IC,ld} \tag{16}$$

In the objective function we maximise the reward generated from covering more than the expected demand of initial consultations at the OC. This is an attempt to decrease the queue of referrals for a specialty during the planning horizon.

Variable domains

$$x_{jald} \in \{0,1,\ldots,\overline{X}_{jald}\} \qquad j\in\mathcal{J}, a\in\mathcal{A}, l\in\mathcal{L}, d\in\mathcal{D} \tag{17}$$

$$q_{nikd} \in \{0,1,\ldots,\overline{Q}_{nikd}\} \qquad i\in\mathcal{I}, n\geq\underline{N}_i, k\in\mathcal{K}, d\in\mathcal{D} \tag{18}$$

$$u_{iwd} \in \{0,1,\ldots,A_{wd}\} \qquad i\in\mathcal{I}, w\in\mathcal{W}, d\in\mathcal{D} \tag{19}$$

$$g^{OC}_{pjsd} \in \{0,1\} \qquad p\in\mathcal{P}, j\in\mathcal{J}, s\in\mathcal{S}, d\in\mathcal{D} \tag{20}$$

$$g^{OR}_{pksd} \in \{0,1,\ldots,\overline{N}\} \qquad p\in\mathcal{P}, k\in\mathcal{K}, s\in\mathcal{S}, d\in\mathcal{D} \tag{21}$$

$$\beta_{jlsd} \in \{0,1\} \qquad j\in\mathcal{J}, l\in\mathcal{L}, s\in\mathcal{S}, d\in\mathcal{D} \tag{22}$$

$$\lambda_{njksd} \in \{0,1\} \qquad n\in\mathcal{N}, j\in\mathcal{J}, k\in\mathcal{K}, s\in\mathcal{S}, d\in\mathcal{D} \tag{23}$$

Constraints (17) to (23) give the domains for the variables.

Prioritising among the specialties

Between the minimum demand and the upper limit for new referral consultations, it is possible to prioritise among the specialties. The majority of patient generated income comes from surgical activity. The income is correlated with the expected surgery duration, such that surgeries with long expected duration generates a high income. We can incorporate the surgery duration when prioritising the specialties in the reward set for each specialty. A way to calculate the reward for each specialty is as the product of the fraction of surgeries and the expected surgery duration, as shown in Eqs. (24).

$$R_j = \sum_{i\in\mathcal{I}_j} F^S_i S_i \qquad j\in\mathcal{J} \tag{24}$$

5. The simulation model

In this section, the discrete-event simulation (DES) model is introduced. To describe the simulation study and the DES model, the STRESS guidelines, introduced by Monks et al. [20], are used. First, we describe the objectives of the simulation study, before presenting the logic of the model. The data is introduced in Section 6, and described in details in Appendix B.

5.1. Objectives

The purpose of the simulation study is to evaluate the performance of the tactical schedules provided by the optimisation model when including random arrivals of new referrals, and random paths of patients through the system. The main input for the simulation model is the schedules generated by the optimisation model, while the main output is the development over time of the queues of patients waiting for both OC consultations and surgery, and the mean total service time of patients in the system. The length of the queues are recorded at the beginning of each simulated week. There are two main aims of experimentation: To evaluate the performance of the tactical schedules provided by the optimisation model, and to evaluate the effect of different operational scheduling policies.

We want to emphasise that the simulation model is not a replication of the complex system of the orthopaedic department. There are many events that are not considered, such as no-shows and absence of staff, and stochastic processes, such as surgery duration and patient length of stay, that are considered deterministic. To isolate the effects from implementing different schedules

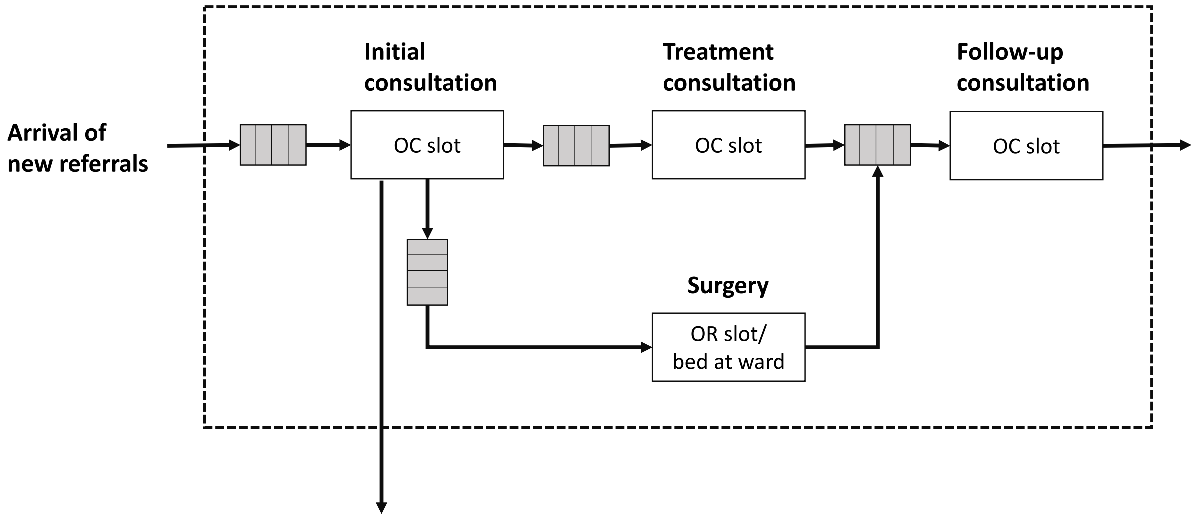

Fig. 3. The flow of patients in the DES model.

Table 4
The attributes of the patients.

Attribute number	Description
1	Specialty
2	Surgery category
3	Treatment at OC, {0, 1}
4	Surgery, {0, 1}
5	Ward
6	Number of follow-ups, {0, 1, 2}

and scheduling policies, we have chosen to keep the system under study rather simple. For this reason, a model validation is not applicable in our case.

5.2. Logic

The entities of the model are the patients, and the attributes of the patients are presented in Table 4. An illustration of the system considered can be seen in Fig. 3. The resources available are the OC rooms, and the combination of ORs and beds. The resource capacities available to the various specialties at different days are given by the schedules obtained from the solution of the optimisation model. The variables β_{jlsd} and λ_{njksd} indicate what specialty that has access to the different room slots during the week, while the variables x_{jald} and q_{nikd} indicate what activity types (type of OC consultation or surgery category respectively) that should be performed in the rooms. The daily bed capacity reserved for patients of surgery category i is given by the u_{iwd} variables. The activity types provided in the OC rooms are initial consultations, treatment consultations and follow-up consultations, while the activity types provided in the ORs are surgery categories and potentially a subsequent stay in a bed. There are four queues in the system, illustrated by the grey squares, one in front of each activity type. Each queue is split into subqueues, one for each specialty.

In Fig. 4, an overview of the DES model implementation can be seen. Fixed time increments of one week are applied, and we assume that no patients receive more than one consultation per week, and that patients are added to the queue for initial consultations the week following arrival. Although this may not always be the case in real life, it is a reasonable assumption when modelling elective patients whose waiting time limits are typically in the range of months. In the arrival algorithm, new referrals are generated each week and sent to the queue for initial consultations the week after. Then, the scheduling algorithm is initiated, which assigns patients to activities in the present week according to two scheduling policies, described below. When a patient is assigned for an activity, the post-scheduling algorithm is initiated. Here, patients are either sent to the queue for the subsequent activity, or, if no further activities are required, they leave the system. Patients do not join the queue for the subsequent activity before the following week. There can be a need for an additional delay between activities, and if so, a patient will not join the queue before the delay has passed.

We assume that the referrals arrive independently of each other, and model the arrivals as a Poisson process with expected arrival rates equal to the expected demand used in the optimisation model. The probabilities of generating patients of a given surgery category, and a given path, are set such that the flow of patients corresponds to the flow given by the F parameters in the optimisation model. In contrast to the arrival of patients and the paths required by patients, the planned service durations are deterministic. We acknowledge that there exist mechanisms that will impact the realised outcome of a schedule, and the importance of efficient rescheduling. However, these are not studied here.

In the scheduling algorithm, two different scheduling policies are used for assigning patients to activities. In both policies, all subqueues are sorted according to a FIFO principle, and when a patient of a given subqueue is to be scheduled, the first patient in the queue is chosen. Algorithms 1 and 2, in Appendix A, explain the two policies applied to the OC room activities. In the former, patients are scheduled in accordance with the optimisation model solution, that is the x_{jald} variables. If there are not enough patients present in the corresponding subqueue, the scheduled capacity is left idle. In the second scheduling policy, we schedule patients based on specialty and the remaining time available in an OC room. If a specialty is scheduled for a slot in an OC on a given day, and the remaining time in this room slot is sufficient to perform more OC consultations, the patient that has waited the longest for an OC consultation (either an initial, a treatment or a follow-up consultation) within the corresponding specialty is scheduled. If several patients have waited equally long, one is randomly chosen.

Corresponding policies are implemented for scheduling patients in the ORs. If, in the first scheduling policy, we cannot find a patient to schedule according to the solution of the optimisation model, the q_{nikd} variables, the corresponding OR capacity is left

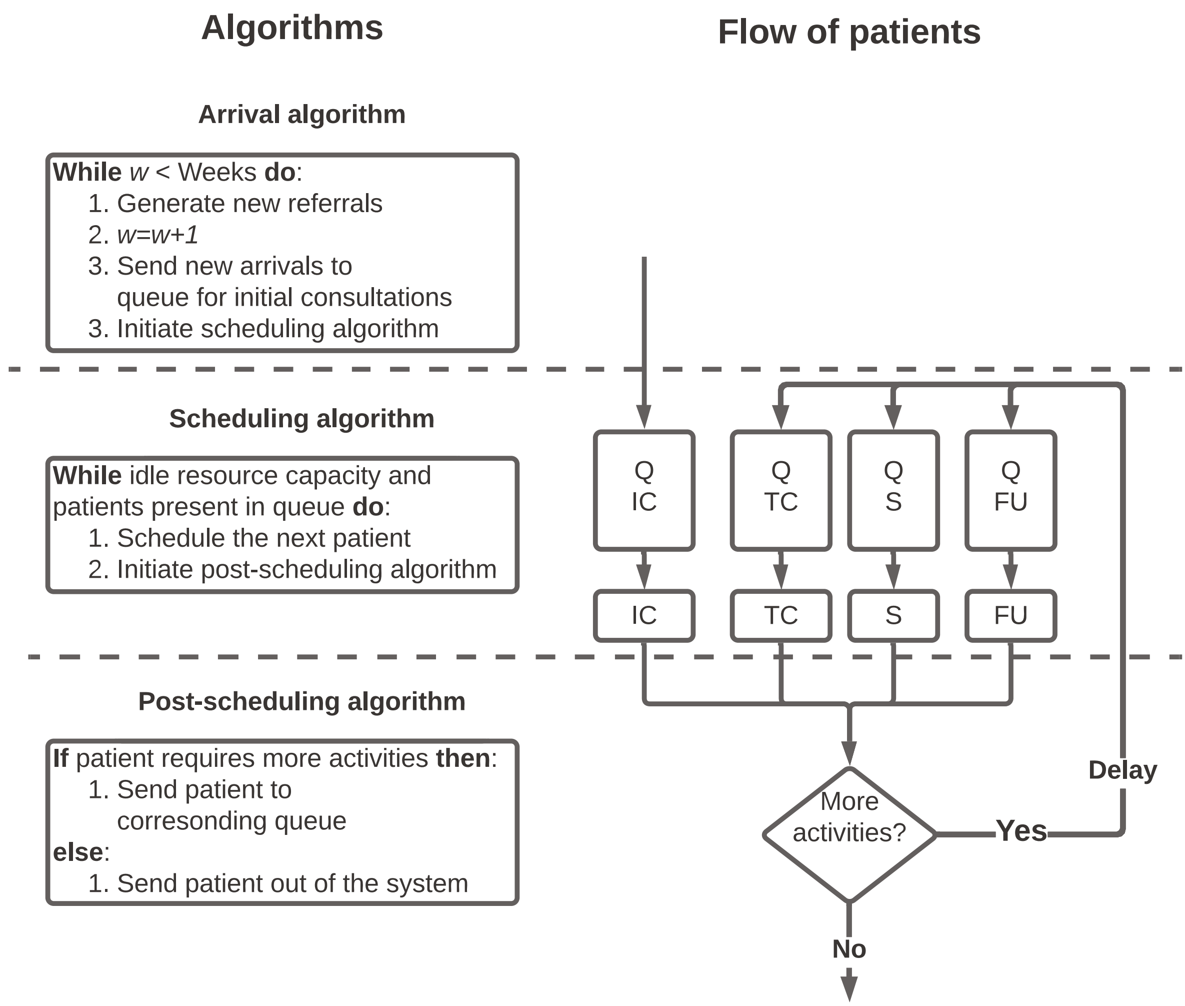

Fig. 4. Overview of the DES model implementation.

idle, and the bed capacity that was reserved for this patient is freed. In the second scheduling policy, we schedule patients based on specialty, the number of surgeons, and the remaining time available in an OR. For a patient to be scheduled, there must be enough time remaining in an OR scheduled for the right specialty, with enough surgeons available to perform the surgery, and there must be a bed available for the patient in a suitable ward for at least as many consecutive days as the LOS of the patient. The first patient (the one who has waited the longest) to fulfil these criteria is scheduled. Since the first scheduling policy applies information regarding the type of consultation (in the OC rooms) and the surgery category (in the ORs), we refer to this policy as the *Activity* (`Act`) policy. The second policy schedules patients based on specialty, and is therefore referred to as the *Specialty* (`Spec`) policy.

When applying the `Act` policy, we take advantage of the resource coordination provided by the optimisation model. If we compare the two scheduling policies, we can interpret the differences in outcome as the value of coordination. If the two scheduling policies are combined, and run successively, we may be able to utilise the capacity that is left idle after scheduling in accordance with the `Act` policy. In this case, the `Spec` policy may be thought of as having a list of patients that can be called and scheduled on short notice (the week before), if there is idle capacity. Performing activities on a short notice requires resource flexibility and responsiveness. Additional gains from combining the two policies can be interpreted as the value of flexibility.

When scheduling according to the `Act` policy, the solution from the optimisation model can be used to calculate the minimum possible queues that can be achieved. As a result, we may end up with queues that are shorter than the ones calculated based on the solution of the optimisation model.

6. Computational study

The aim of the computational study is to demonstrate how a department can apply the optimisation model to coordinate its OC and OR activities to decrease the queues of patients waiting to be served in either of the facilities. The study based on data from the Orthopaedic Department at St. Olav's Hospital. In addition to a base case, representing the present resource capacities at the Orthopaedic Department, we design multiple cases with slightly altered resource capacities to demonstrate how the department can temporarily alter its resources to enhance the system performance. Finally, we evaluate three operational scheduling policies to demonstrate the gains from scheduling activity types in addition to specialties when generating the master schedules. In accordance with the Orthopaedic Department, we use the term subspecialty instead of specialty throughout the computational study.

Table 5
The resource capacity cases.

Case name	Description
I_0	Base case
I_1	Homogeneous wards
I_2	Homogeneous ORs
I_3	All beds available during weekend
I_4	$I_1 + I_3$

An Intel(R) Core(TM) i7-8550U CPU @ 1.80 GHz, 16 GB RAM computer is used when performing the computational study. The optimisation model is implemented in IVE Xpress 8.6, while the simulation model is written in Python 3.7, and the package SimPy. To perform the random sampling, the algorithms included in Python are used.

6.1. Case descriptions

To establish the base case, data from the Orthopaedic Department at St. Olav's Hospital is used. All data necessary to define the base case are provided in Tables B.13–B.20, in Appendix B. To sum up, there are 19 surgery categories divided among seven subspecialties. There are eight OC rooms and seven ORs, and one slot per day is used. The OC rooms can be used by all specialties, while the ORs are heterogeneous and can only be accessed by a subset of the specialties. The slot length at the OC is 240 min, while a slot lasts for 480 min at the operating theatre. Four wards are available, with a total capacity of 28 beds from Monday to Friday, and seven beds during the weekend. The wards are specialised, and an inpatient that has received surgery can only access a subset of the wards. The initial queues of patients waiting for an initial consultation at the start of the planning period are specified in Table C.21, in Appendix C.

In addition to the base case, labelled I_0, four resource capacity cases are investigated, and these are presented in Table 5. In I_1, the wards are treated as homogeneous, implying that each inpatient can be assigned to all wards. In practice, this would imply that the nurses, who serve the wards, must gain a wider competence such that they can handle patients outside their main field of competence. In I_2, homogeneous ORs are applied, meaning that each subspecialty can be assigned to all ORs. Room size and location can inhibit complete homogeneity between ORs, but by introducing similar equipment in all ORs, they can be more or less homogeneous. In I_3 and I_4, the bed capacity is not decreased during the weekend, and in the latter the wards are homogeneous.

6.2. The results from the optimisation model

One alteration has been made to the mathematical formulation when performing the optimisation study. To avoid unnecessary opening of the OC rooms, we introduce a small penalty of 0.1 for assigning specialties to the OC room slots (that is, we penalise the β_{jlsd} variables in the objective function).

The main results from running the five cases for three hours are presented in Table 6. The problem has not been solved to optimality in any of the cases, so we provide the best objective function values found, together with the upper bound and the dual gap. In addition, the aggregated number of activity types scheduled per week is presented. In the base case, 129 initial consultations are scheduled every week. By imposing homogeneous wards, we are able to increase the activity type by seven initial consultations, which is three more compared with leaving the bed capacity constant all week. Combining the two yields one more initial consultation compared with only having homogeneous wards. Also note that the number of surgeries increases in these instances, providing a higher reward in the objective function. By introducing homogeneous ORs, we can expect to provide two more initial consultations each week, compared with the base case, and one additional surgery. Based on these results, we can conclude that the beds are a scarce resource, and that the separation of wards imposes considerable restrictions for the patient throughput.

Table 6
Results from running the optimisation model for three hours. UB: Upper bound on objective value. IC: Number of ICs scheduled. TC: Number of TCs scheduled. FU: Number of FUs scheduled. S: Number of surgeries scheduled.

Case	Obj func	UB	Gap	IC	TC	FU	S
I_0	7058.13	7211.40	2.13%	129	31	138	65
I_1	7680.15	7918.00	3.00%	136	30	143	71
I_2	7231.16	7385.67	2.09%	131	32	141	66
I_3	7469.96	7859.53	4.96%	133	30	141	69
I_4	7846.27	8100.66	3.14%	137	30	146	74

In Tables 7 and 8, the number of initial consultations and surgeries scheduled are provided respectively. Compared with the base case, all other cases schedule more initial consultations, resulting in more surgeries being scheduled as well. When introducing homogeneous wards, less initial consultations from the hand and tumour subspecialties are scheduled, while the capacity increases for the remaining subspecialties. The reason for this is that these specialties impose less reward in the objective function. To take advantage of the increased bed capacity, specialties that provide a higher reward are prioritised. In the base case, the ward that houses the arthroplasty patients is closed during weekend. If the bed capacity is not reduced on Saturday and Sunday, the activity related to the arthroplasty patients can be increased. Also when applying homogeneous ORs, the capacity is increased for arthroplasty patients. In the base case, the arthroplasty subspecialty has access only to ORs six and seven, and due to the long LOS of these patients, all surgeries must be performed on Monday and Tuesday. When allowing for homogeneous ORs, the arthroplasty subspecialty gains access to all ORs, enabling the surgery of one additional patient.

In Table 9, the total resource consumption for one week in the different cases is given. Increasing the resource flexibility results in a higher resource consumption, which increases the surgeon workload. The total surgeon capacity is calculated as the number of surgeons available multiplied by five days. However, the surgeons have other duties to fulfil, such as serving the wards and the emergency department, and conducting research, so having 220 surgeon days available for OC and OR activities is not realistic. As can be seen from Table 10, there can be days where all the available surgeon hours are utilised for OC and OR activities. The table presents the maximum daily utilisation of surgeon hours available for each subspecialty. The surgeon types that can cover several specialties are added to the capacity of all the corresponding specialties when performing the calculation.

The expected OC room and OR utilisation for the different cases can be seen in Table 11. Here, the resource utilisation is given relative to the scheduled resource capacity. When regarding the OC rooms, we see that the utilisation increases when allowing for a more flexible use of resources. For these rooms, a utilisation of 100% is possible as we have set the duration of OC consultations to be 30 min, which is a multiple of the slot duration. For the ORs, the utilisation is low compared with in the OC rooms. The reason for this is the combination of surgery durations not adding up to full slots, and the bed capacity restricting the possible combinations of surgeries on a day.

Table 7
The number of initial consultations scheduled in the different cases.

Specialty	Number of initial consultations					Exp. demand	Max demand	Obj. func. reward
	I_0	I_1	I_2	I_3	I_4			
Arthroscopy	17	19	17	17	17	17	20	42.4
Hand	23	20	23	23	20	19	23	35.1
Plastic	30	33	30	30	34	29	34	77.9
Arthroplasty	19	21	21	23	23	19	23	86.5
Reconstructive	18	20	18	20	20	18	22	46.1
Back	10	13	10	10	13	10	13	56.7
Tumour	12	10	12	10	10	10	12	13.2
Sum	**129**	**136**	**131**	**133**	**137**	**122**	**147**	

Table 8
The number of surgeries scheduled in the different cases.

Subspecialty	Surgery category	I_0	I_1	I_2	I_3	I_4
Arthroscopy	Arthroscopy (aggregated)	2	3	2	2	2
Arthroscopy	ACL	1	2	1	1	1
Arthroscopy	Meniscus	1	1	1	1	1
Arthroscopy	Patellae	1	1	1	1	1
Hand	Hand (aggregated)	7	6	7	7	6
Hand	CTS	2	2	2	2	2
Plastic	Plastic (aggregated)	12	13	12	12	14
Plastic	Carsinoma	1	1	1	1	1
Plastic	BCC	3	3	3	3	4
Plastic	Malignant melanoma	5	5	5	5	6
Plastic	Cancer mammae	3	4	3	3	4
Plastic	SCC	2	2	2	2	2
Arthroplasty	Hip (primary)	7	8	8	9	9
Arthroplasty	Hip (revision)	2	2	2	3	3
Arthroplasty	Knee (primary)	5	5	5	5	5
Arthroplasty	Knee (revision)	1	1	1	1	1
Reconstructive	Reconstructive (aggregated)	6	7	6	7	7
Back	Back (aggregated)	2	3	2	2	3
Tumour	Tumour (aggregated)	2	2	2	2	2
Sum		**65**	**71**	**66**	**69**	**74**

Table 9
The scheduled use of resources. For the beds, some numbers are underlined to indicate that more beds are available in these cases.

Resource type	Resource usage					Capacity per week
	I_0	I_1	I_2	I_3	I_4	
OC room slots	40	40	40	40	40	40
OR slots	24	29	30	28	30	35
Bed days	134	154	138	<u>149</u>	<u>161</u>	154/<u>196</u>
Surgeon days	85	96	97	93	98	220

Table 10
The maximum daily utilisation of surgeon hours.

Specialty	I_0	I_1	I_2	I_3	I_4
Arthroscopy	0.5	0.67	1.00	0.50	0.50
Hand	1.00	1.00	1.00	1.00	1.00
Plastic	1.00	1.00	1.00	1.00	1.00
Arthroplasty	0.64	0.55	0.64	0.55	0.55
Reconstructive	1.00	0.57	0.86	0.71	0.43
Back	1.00	0.60	1.00	0.60	0.60
Tumour	0.29	0.43	0.43	0.29	0.43

Table 11
The planned utilisation of OC rooms and ORs. The values represent the utilisation of the scheduled time.

Resource type	I_0	I_1	I_2	I_3	I_4
OC room slots	93.13%	96.56%	95.00%	95.00%	97.81%
OR slots	66.82%	62.90%	54.22%	61.12%	61.97%

6.3. The simulation study

The aim of the simulation study is to evaluate the performance of the tactical master schedules provided by the optimisation model, and to test different scheduling policies that can be implemented for the operational scheduling of patients. Furthermore, the arrival of patients, and whether the patients require surgery, a treatment consultation in the OC, or if they leave the system following an initial consultation are modelled as stochastic processes. Because the number of follow-ups after an orthopaedic treatment is rather standardised, we assume that the number of follow-up consultations is deterministic. Furthermore, there is a delay of minimum three weeks before a follow-up consultation.

6.3.1. Experimental setup

For each of the cases analysed in the previous section, the solution obtained from the optimisation model is implemented in the simulation model. Then, 100 simulated weeks are run with the simulation model, and a total of 100 replications are performed for each case. We only report on the 25 first weeks, but a cool-down period is added to ensure that all patients arriving within the first 25 weeks have left the system by the end of the simulation.

As the system under study is not a steady state system, no warm-up is applied. Furthermore, the system is not empty when starting the simulation, and the queues are pre-filled with patients to mimic a realistic situation. For the two first weeks, the expected number of the different activities are pre-assigned. Then, for the two following weeks, half of the expected number of activities are pre-assigned. There is also an initial queue of patients waiting to be scheduled for an initial consultation at the beginning of week 1. The queues of pre-assigned activities and unscheduled initial consultations are specified in Tables C.21 and C.22 in Appendix C.

Fig. 5, illustrates how the initial consultations performed in the first simulated week will cause a delayed downstream demand for the remaining services. From the first simulated week, the number of initial consultations performed is equal to the number of initial consultations performed in the solution of the optimisation model. The derived demand for treatment consultations and surgeries is evident from week two, while the derived demand for follow-up consultations is delayed to weeks five (first follow-up) and eight (second follow-up). As a consequence, the full impact of the altered scheduling regime will be evident from week eight. Other consequences of the delay related to follow-up consultations is that the queue of follow-ups will increase the first five weeks, and that the OC utilisation will not peak until week eight.

We do not make assumptions about how long the patients that are already in queue have been waiting, so we only report on the waiting time of patients arriving after the start of the simulated time. Furthermore, for patients that belong to subspecialties

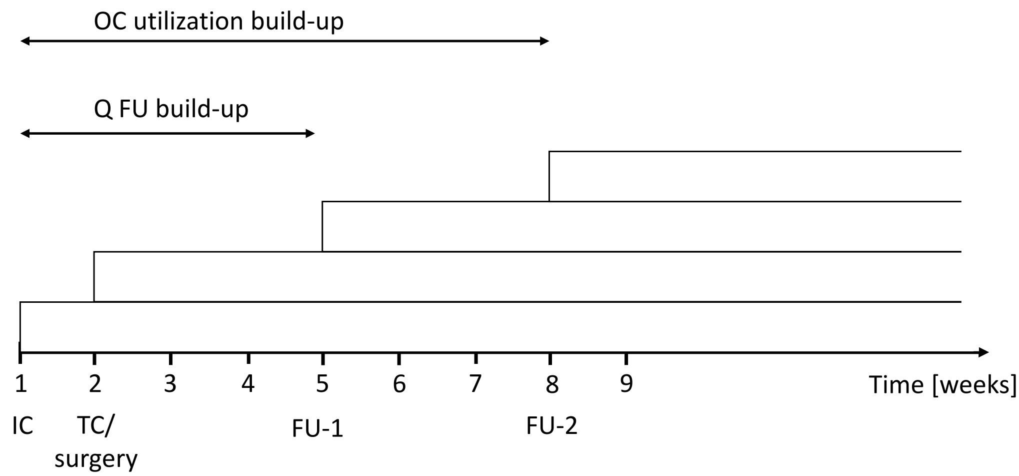

Fig. 5. Illustration of the initial development of activity in the simulation model. The derived demand for treatment consultations and surgeries is evident from week two, while the derived demand for follow-ups consultations is delayed to weeks five (first follow-up) and eight (second follow-up).

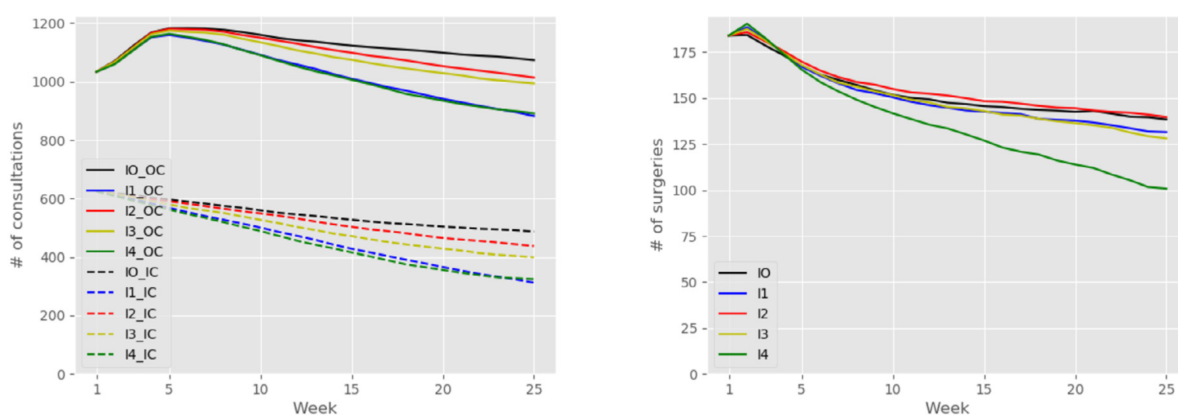

Fig. 6. The mean queue length for OC consultations (left, solid lines), initial consultations (left, dashed lines) at the OC, and surgeries (right) in the five different cases.

that require two follow-up consultations, we sample with a 50% probability for each outcome, whether a prescheduled patient is scheduled for his first or second follow-up consultation.

Before comparing the different scheduling policies, we first present results where the Act policy is applied, as these can be compared directly to the results of the optimisation model.

6.3.2. *The queues for OC consultations and surgery*

In Fig. 6, the mean queue length for OC consultations in total, initial consultations at the OC, and surgeries are provided for the five different cases. To calculate the total queue of OC consultations, the queue of initial, treatment and follow-up consultations are added together. Note that all follow-up consultations for a patient are added to the queue of follow-ups the week following either a treatment consultation or a surgery, even though there is a delay between these activities. Not surprisingly, the cases perform according to the rank of performance obtained from the solutions of the optimisation model. It is also evident that all queues decrease during the period. Initially, the queue of OC consultations increases, caused by the delay related to follow-up consultations. The number of initial consultations, and the relatively low throughput of surgeries in week one, leads to more patients being added to the queue for surgery than what is removed. Therefore, the queue for surgery increases from week one to two. However, from week two, the surgery rate increases and the queue decreases.

6.3.3. *The resource utilisation and overall efficiency*

The mean utilisation of the scheduled OC rooms, ORs, and the beds is displayed in Fig. 7. Since we model the arrivals and paths of patients as stochastic processes, the demand for activities will deviate from the expected demand, causing some of the prescheduled activities to be unused. This tendency becomes more evident as the queues decrease, leading to a decrease in resource utilisation towards the end of the planning horizon.

When regarding the OC room utilisation, the delayed demand for follow-up consultations yields a jump in utilisation in weeks five and eight. As previously shown, the derived demand for surgeries will be present from week two, causing the jump in OR utilisation from week one to two. The same effect causes similar behaviour for the bed utilisation.

To measure the system efficiency achieved in the different cases, we register the mean time that patients, who have received all necessary activities, stayed in the system. Fig. 8 illustrates the mean number of weeks that patients stay in the system as a function of when they arrive. In all cases, the patient waiting

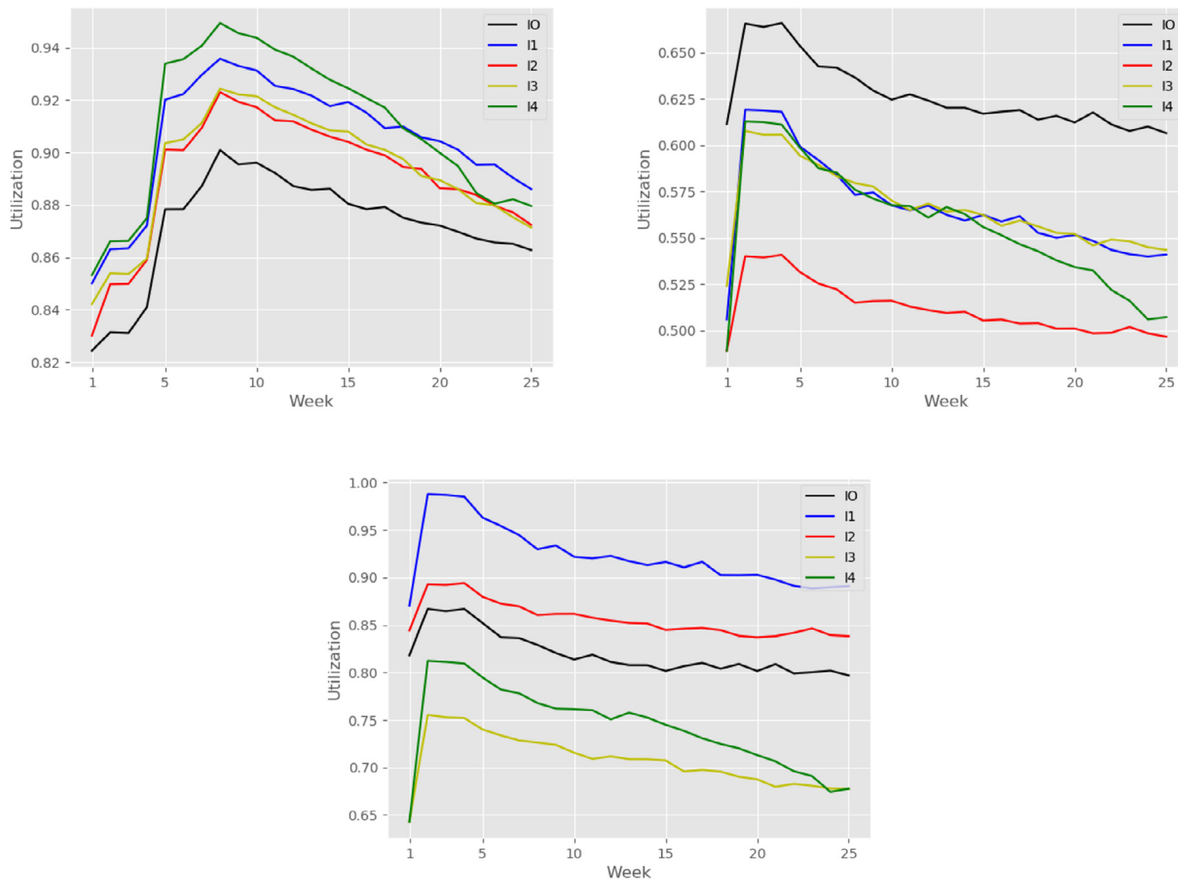

Fig. 7. The mean utilisation of the scheduled OC rooms (top left), the scheduled ORs (top right), and the beds (bottom) in the five cases.

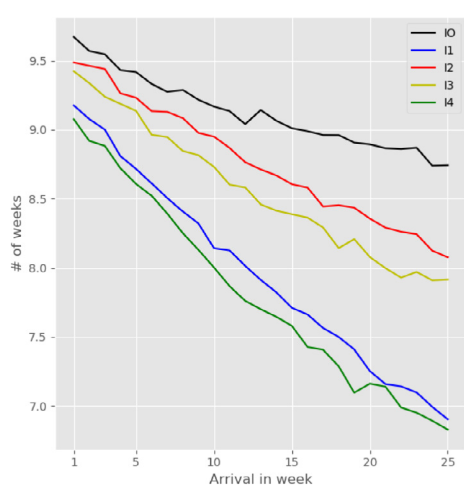

Fig. 8. The mean number of weeks that patients stay in the system as a function of when they arrive.

times decrease throughout the period, resulting in lower times spent in the system.

6.3.4. *Evaluating different patient scheduling policies*

In this section, three different scheduling policies are evaluated. The scheduling policies introduced in Section 5.2, the `Act` and the `Spec` policies, are evaluated individually. In the final scheduling policy, the two former policies are used successively, such that after having scheduled patients according to the optimisation model solution, we schedule additional activities if possible. This is equivalent to scheduling patients according to the `Act` policy, and then, if excess capacity is available, summon patients who can enter on a short notice. This can be achieved through establishing a calling list of patients who are willing and able to enter on a short notice. We refer to the final policy as the *Combined* (`Comb`) policy. When applying the `Act` policy, only the capacity that is pre-scheduled for activities can be utilised (see Table 11 to see how much of the scheduled capacity that can be utilised with this policy). However, when applying the two other policies, all the capacity scheduled for a specialty can be used, allowing for a higher resource utilisation in these policies. It is therefore not fair to compare the `Act` to the other policies, but we choose to include it to indicate the value of establishing a calling list.

The mean queue length for initial consultations, OC consultations, and surgery when applying the three different scheduling policies in I_1 can be seen in Fig. 9. If patients are scheduled according to the `Act` policy, the queues obtained in a deterministic reality can be calculated from the solution of the optimisation model. This queue has also been added to the figure. Not surprisingly, the optimisation model solution outperforms the `Act` policy for all queues. Furthermore, the `Comb` policy outperforms the `Act` policy, indicating the value of flexibility in terms of a calling list.

When regarding the queues for initial and OC consultations in total, the `Act` performs worse than the two other scheduling policies because it has access to less resource capacity. However,

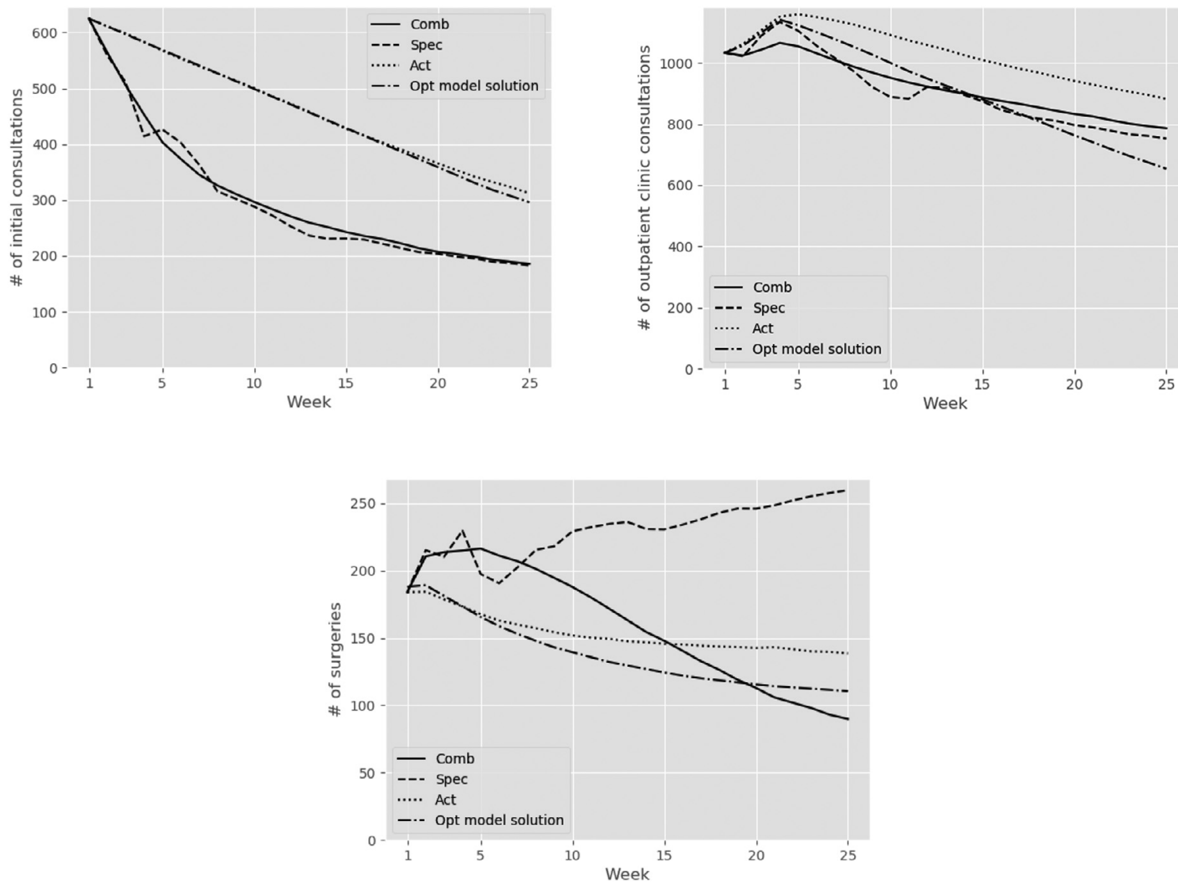

Fig. 9. The mean queue length for initial consultations (top left), OC consultations (top right), and surgery (bottom) for the three scheduling policies in I_1.

Table 12
Different measures in the last week of the planning horizon (week 25). The "Time in system" is the mean time [weeks] spent for patients arriving in week 25. "Throughput" is the mean number of patients that has left the system within week 25.

Case	Time in system			Queue OR			Queue OC			Throughput		
	Act	Spec	Comb	Act	Spec	Comb	Act	Spec	Comb	Act	Spec	Comb
I_0	8.74	10.58	6.33	138	288	201	1074	993	757	3175	3098	3421
I_1	6.90	7.46	5.72	132	260	90	882	751	778	3372	3382	3523
I_2	8.08	7.51	6.08	140	276	185	1013	781	738	3233	3335	3462
I_3	7.92	8.91	5.99	128	170	116	994	949	811	3269	3256	3470
I_4	6.83	5.91	5.82	101	64	60	891	825	812	3391	3505	3508

Table B.13
Base case: the sets.

Set	# of elements
Days	7
Subspecialties	7
Surgery categories	19
OC rooms	8
ORs	7
Wards	4
Activity types performed at the OC	3
Surgeons present	2
Surgeon types	17
Slots	1

when regarding the queue of surgeries, something interesting is observed. Due to the excessive scheduling of initial consultations during the first weeks, the queue of surgeries grows initially when applying either the Spec or the Comb policy. Since the Comb policy ensures a coordination between initial consultations and surgeries, the queue for surgeries eventually decreases below the level seen with the Act policy. However, this is not the case with the Spec policy, where the queue of surgeries keeps growing. This clearly indicates the value of coordination, especially when downstream capacities are scarce.

To evaluate the efficiency obtained from the three scheduling policies, the mean number of weeks that patients stay in the system as a function of when they arrive, is displayed in Fig. 10. The Comb policy clearly outperforms the Act policy, indicating the value of flexibility. As a consequence of poor coordination between initial consultations and surgeries, the Spec policy performs worse than the Act policy when approaching the end of the period.

In Table 12, we present different measures from the end of the planning horizon [week 25] in the simulation model. In general, we observe that the Comb policy outperforms the other scheduling policies, stating the value of coordination in combination

Table B.14
Base case: the subspecialties.

Specialty	OCRs available	ORs available	IC dur [min]	TC dur [min]	FC dur [min]	Reward for scheduling more than min
Arthroscopy	All	3 and 4	30	30	30	42.4
Hand	All	3	30	30	30	35.1
Plastic	All	1, 2 and 3	30	30	30	77.9
Arthroplasty	All	6 and 7	30	30	30	86.5
Reconstructive	All	1 and 5	30	30	30	46.1
Back	All	5	30	30	30	56.7
Tumour	All	5	30	30	30	13.2

Table B.15
Base case: the surgeons.

Surgeon type	Subspecialty	# of surgeons	Experience level
Arthroscopy 1	Arthroscopy	4	Consultants
Arthroscopy 2	Arthroscopy	2	Residents
Hand 1	Hand	2	Consultants
Hand 2	Hand	1	Residents
Plastic 1	Plastic	4	Consultants
Plastic 2	Plastic	4	Residents
Arthroplasty 1	Arthroplasty	6	Consultants
Arthroplasty 2	Arthroplasty	3	Residents
Reconstructive 1	Reconstructive	4	Consultants
Reconstructive 2	Reconstructive	2	Residents
Back 1	Back	4	Consultants
Back 2	Back	1	Residents
Tumour 1	Tumour	3	Consultants
Tumour 2	Tumour	1	Residents
Cons 1	Arthroplasty and tumour	1	Consultant
Cons 2	Arthroplasty and tumour	1	Consultant
Cons 3	Reconstructive and tumour	1	Consultant

Table B.16
Base case: the number of beds available.

Ward	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Trauma	4	4	4	4	4	2	2
Reconstructive	5	5	5	5	5	3	3
Elective	3	3	3	3	3	2	2
FT	16	16	16	16	16	0	0

Table B.17
Base case: the surgery categories.

Surgery category	Subspecialty	Surg dur [min]	# of surgeons	Ward available	LOS [days]
Arthroscopy (aggregated)	Arthroscopy	174	2	Elective	2
ACL	Arthroscopy	173	1	Elective	2
Meniscus	Arthroscopy	103	2	–	0
Patellae	Arthroscopy	176	2	Elective	1
Hand (aggregated)	Hand	107	2	–	0
CTS	Hand	54	2	Trauma	1
Plastic (aggregated)	Plastic	108	2	Trauma, reconstructive	2
Carsinoma	Plastic	52	1	Reconstructive	1
BCC	Plastic	59	2	Trauma, reconstructive, FT	1
Malignant melanoma	Plastic	85	1	–	0
Cancer mammae	Plastic	146	1	Reconstructive	1
SCC	Plastic	65	2	Reconstructive	1
Hip (primary)	Arthroplasty	110	2	FT	4
Hip (revision)	Arthroplasty	152	2	FT	4
Knee (primary)	Arthroplasty	122	2	FT	4
Knee (revision)	Arthroplasty	165	2	FT	4
Reconstructive (aggregated)	Reconstructive	145	2	Reconstructive	2
Back (aggregated)	Back	309	2	Elective	6
Tumour (aggregated)	Tumour	93	1	Reconstructive	1

with resource flexibility. Furthermore, the Spec policy yields a longer queue of surgeries at the end of the planning horizon compared with the two other policies (except for I_4). The patients waiting for surgery in week 25, will eventually require one or

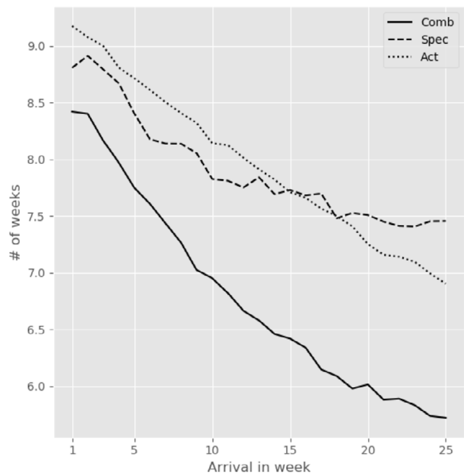

Fig. 10. The mean number of weeks that patients stay in the system as a function of when they arrive for the three scheduling policies in I_1.

Table B.18
Base case: the flow of patients at the OC.

Subspecialty	Expected demand	Max demand	Share to TC after I	Share to FU after T
Arthroscopy	17	20	0.01	1
Hand	19	23	0.55	2
Plastic	29	34	0.20	1
Arthroplasty	19	23	0.01	1
Reconstructive	18	22	0.06	2
Back	10	13	0.01	2
Tumour	10	12	0.53	2

Table B.19
Base case: the flow of patients at the operating theatre.

Surgery category	Share to surgery after I cons	Share to FU cons after surgery	Add surg demand
Arthroscopy (aggregated)	0.11	1	0
ACL	0.06	1	0
Meniscus	0.04	1	0
Patellae	0.05	1	0
Hand (aggregated)	0.29	2	0
CTS	0.08	2	0
Plastic (aggregated)	0.39	1	0
Carsinoma	0.02	1	0
BCC	0.09	1	0
Malignant melanoma	0.15	1	0
Cancer mammae	0.10	1	0
SCC	0.03	1	0
Hip (primary)	0.37	1	0
Hip (revision)	0.09	1	0
Knee (primary)	0.21	1	0
Knee (revision)	0.04	1	0
Reconstructive (aggregated)	0.32	2	0
Back (aggregated)	0.18	2	0
Tumour (aggregated)	0.14	2	0

Table B.20
Base case: slots.

Location	# of slots each day	Time available per slot [min]
Outpatient clinic	1	240
Operating theatre	1	480

two follow-up consultations in the OC. However, these are not counted in the total OC queue before the week following surgery. Therefore, there are relatively more OC consultations left to be performed in the `Spec` policy, compared with the other policies, than indicated by comparing the "Queue OCs". Note that the `Spec` policy performs well in I_4 where the bed capacity is much increased. This indicates that coordination is not crucial when downstream capacities are high.

7. Managerial insights

In this section we list insights based on the findings of this paper, relevant for surgical departments where patients require both OC consultations and surgery.

- A hospital department is often measured by how fast it can provide initial consultations for its patients. Increasing the throughput of initial consultations without coordinating with downstream activities can harm the system efficiency and increase the total throughput time of patients. In our case, implementing the `Act` instead of the `Spec` policy decreases the average time in the system by 0.38 weeks (averaging over all resource capacity cases) even though the `Spec` has access to more capacity.
- Coordination between resources is particularly important when downstream resources are scarce. In our case, I_0 is most limited in terms of bed- and OR capacity, and this is the case where the throughput time of patients differs the most (1.84 weeks).
- Including information about what activity types to perform within each time slot is valuable when coordination is an issue. In our example, scheduling the surgery categories in addition to the specialty and the number of surgeons made it easier to achieve a high bed utilisation and to provide a high throughput of patients.
- Scheduling patients according to a pre-defined pattern settled at a tactical level can suffer from inflexibility causing unused capacities. Establishing a patient calling list or other flexible mechanisms can increase efficiency. In the case studied here, implementing the `Comb` instead of the `Act` policy decreases the average throughput time of patients by 1.71 weeks (averaging over all resource capacity settings).

8. Conclusion

In this paper we argue that surgical departments should consider the OC and the operating theatre simultaneously to ensure efficient handling of patients. Furthermore, we demonstrate how optimisation can be used as a tool to develop efficient master schedules for both facilities. Through simulation, we conclude that scheduling activity types, in addition to specialties, in the master schedules allow for a simple and efficient scheduling of individual patients. The value of scheduling activity types in the master schedules is to a large extent caused by the scheduling of patients for surgery, which is a complex task due to the need for coordination between the surgical activities and the bed capacity. Furthermore, a successful implementation of a patient calling list, which enables patients to be summoned for a consultation on short notice (the coming week), is beneficial for increasing patient throughput. In our problem formulation, the resource capacities are levelled based on a demand for initial consultations, and a derived demand for surgery and downstream OR consultations. However, there can be situations where there are considerable queues also for surgery and downstream OC consultations at the beginning of a planning horizon. If the queue for initial consultations is long, and we schedule the capacities to decrease this queue, the corresponding capacities scheduled for surgery and downstream OC consultations will be sufficient to handle more than expected demand for treatment, and these queues will decrease as well. However, if the queues for initial consultations are

Table C.21
The queue of patients for the OC when starting the simulation. Except from the queue of initial consultations that are not yet scheduled, the remaining consultations are scheduled within the four first weeks.

Subspecialty	Scheduled for I cons	Scheduled for Alt. treatment cons	Scheduled FU cons	I cons not yet scheduled
Arthroscopy	52	0	12	23
Hand	58	30	102	42
Plastic	88	18	84	37
Arthroplasty	58	0	42	42
Reconstructive	56	6	42	44
Back	32	0	12	43
Tumour	30	18	42	20

short, while the others are long, these queues should be handled as separate demands, independent of the derived downstream demand.

Algorithm 1: The algorithm for scheduling patients to the OC rooms according to the solutions of the optimisation model. Referred to as the Act scheduling policy.

input: Queue_referred, Queue_treatment, Queue_follow_up, OC_availability
for $d \leftarrow 1$ **to** 5 **do**
    for $l \leftarrow 1$ **to** $Number\_of\_OC$ **do**
        for $j \leftarrow 1$ **to** $Specialties$ **do**
            $N_1 = min\{x_{j1ld}, length(Queue\_referred)\}$;
            $N_2 = min\{x_{j2ld}, length(Queue\_treatment)\}$;
            $N_3 = min\{x_{j3ld}, length(Queue\_follow\_up)\}$;
            for $i \leftarrow 1$ **to** N_1 **do**
                Subtract the consultation duration from OC_availability;
                Remove the first patient from Queue_referred;
            end
            for $i \leftarrow 1$ **to** N_2 **do**
                Subtract the consultation duration from OC_availability;
                Remove the first patient from Queue_treatment;
            end
            for $i \leftarrow 1$ **to** N_3 **do**
                Subtract the consultation duration from OC_availability;
                Remove the first patient from Queue_follow_up;
            end
        end
    end
end

There are some limitations in our study. First, we apply a deterministic optimisation model to solve a problem that has several stochastic parameters. Applying a stochastic model could allow us to capture more of the inherent uncertainties when generating the master schedules, and introduce flexible mechanisms to handle the uncertainties. Furthermore, due to our relatively simple simulation model, we cannot give the complete picture of how the schedules and the scheduling policies will perform in a real-life setting. A recommendation for future research is to consider both uncertain patient arrivals and activity demands in the optimisation model. Capturing these uncertainties and proposing mechanisms for handling them will be of great value to hospitals that face variations in demand.

Table C.22
The queue of patients for surgery when starting the simulation. The surgeries are scheduled within the four first weeks.

Surgery category	# of patients in queue
Arthroscopy (aggregated)	6
ACL	4
Meniscus	4
Patellae	4
Hand (aggregated)	16
CTS	6
Plastic (aggregated)	34
Carsinoma	4
BCC	6
Malignant melanoma	12
Cancer mammae	12
SCC	4
Hip (primary)	22
Hip (revision)	6
Knee (primary)	12
Knee (revision)	4
Reconstructive (aggregated)	18
Back (aggregated)	6
Tumour (aggregated)	6

CRediT authorship contribution statement

Thomas Reiten Bovim: Contact with the orthopaedic department, Collecting and preparing data , Definition of the problem, Development and building of optimisation model, Development and building of simulation model, Responsible for the computational study, Writing – original draft. **Anita Abdullahu:** Contact with the orthopaedic department, Collecting and preparing data, Definition of the problem, Development and building of optimisation model. **Henrik Andersson:** Development of optimisation model, Development of simulation model, Discussing what analysis to include, Writing – original draft. **Anders N. Gullhav:** Contact with the orthopaedic department, Collecting and preparing data, Definition of the problem, Development of optimisation model, Development of simulation model, Discussing what analysis to include, Writing – original draft.

Appendix A. Algorithms for describing the DES model

See Algorithms 1 and 2.

Appendix B. Data for the base case

See Tables B.13–B.20.

Appendix C. Starting conditions for the DES model

See Tables C.21 and C.22.

Algorithm 2: The algorithm for utilising idle OC room capacity after having scheduled patients according to the solutions of the optimisation model. Referred to as the $\mathcal{Spec}$ scheduling policy.

```
input : Queue_referred, Queue_treatment,
        Queue_follow_up, OC_availability
for d ← 1 to 5 do
    for l ← 1 to Number_of_OC do
        for j ← 1 to Specialties do
            check=1;
            while length(List_of_candidates) > 0 or check = 1
             do
                check = 0;
                List_of_candidates=[];
                if consultation duration of Queue_referred[0]
                 fits into remaining OC_availability then
                    Append patient to List_of_candidates;
                end
                if consultation duration of Queue_treatment[0]
                 fits into remaining OC_availability then
                    Append patient to List_of_candidates;
                end
                if consultation duration of Queue_follow_up[0]
                 fits into remaining OC_availability then
                    Append patient to List_of_candidates;
                end
                Choose the patient from List_of_candidates
                 that has waited the longest;
                Subtract the consultation duration of the
                 chosen patient from OC_availability;
                Remove the chosen patient from
                 List_of_candidates;
                Remove the chosen patient from the
                 corresponding queue;
            end
        end
    end
end
```

References

[1] J.V. Essen, E. Hans, J. Hurink, A. Oversberg, Minimizing the waiting time for emergency surgery, Oper. Res. Health Care 1 (2) (2012) 34–44.

[2] N. Freeman, M. Zhao, S. Melouk, An iterative approach for case mix planning under uncertainty, Omega 76 (2018) 160–173.

[3] M.B. Vik, H. Finnestrand, R.L. Flood, Systemic problem structuring in a complex hospital environment using viable system diagnosis – keeping the blood flowing, Syst. Pract. Action Res. 35 (2022) 203–226.

[4] P.J.H. Hulshof, N. Kortbeek, R.J. Boucherie, E.W. Hans, P.J.M. Bakker, Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS, Health Syst. 1 (2) (2012) 129–175.

[5] B. Cardoen, E. Demeulemeester, J. Belin, Operating room planning and scheduling: A literature review, European J. Oper. Res. 201 (3) (2010) 921–932.

[6] X. Li, N. Rafaliya, M.F. Baki, B.A. Chaouch, Scheduling elective surgeries: the tradeoff among bed capacity, waiting patients and operating room utilization using goal programming, Health Care Manage. Sci. 20 (1) (2017) 33–54.

[7] A. Moosavi, S. Ebrahimnejad, Robust operating room planning considering upstream and downstream units: A new two-stage heuristic algorithm, Comput. Ind. Eng. 143 (2020) 106387.

[8] I. Adan, J. Bekkers, N. Dellaert, J. Jeunet, J. Vissers, Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources, European J. Oper. Res. 213 (1) (2011) 290–308.

[9] A.T. Schneider, J.T. van Essen, M. Carlier, E.W. Hans, Scheduling surgery groups considering multiple downstream resources, European J. Oper. Res. 282 (2) (2020) 741–752.

[10] A. Fügener, E.W. Hans, R. Kolisch, N. Kortbeek, P.T. Vanberkel, Master surgery scheduling with consideration of multiple downstream units, European J. Oper. Res. 239 (1) (2014) 227–236.

[11] 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.

[12] J. Marynissen, E. Demeulemeester, Literature review on multi-appointment scheduling problems in hospitals, European J. Oper. Res. (ISSN: 0377-2217) 272 (2) (2019) 407–419.

[13] I.A. Bikker, N. Kortbeek, R.M. van Os, R.J. Boucherie, Reducing access times for radiation treatment by aligning the doctor's schemes, Oper. Res. Health Care 7 (2015) 111–121.

[14] T.B.T. Nguyen, A.I. Sivakumar, S.C. Graves, A network flow approach for tactical resource planning in outpatient clinics, Health Care Manage. Sci. 18 (2) (2015) 124–136.

[15] 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 Manage. Sci. (ISSN: 1572-9389) 17 (4) (2014) 379–392.

[16] P.J. Hulshof, R.J. Boucherie, E.W. Hans, J.L. Hurink, Tactical resource allocation and elective patient admission planning in care processes, Health Care Manage. Sci. 16 (2013) 152–166.

[17] A.G. Leeftink, I.A. Bikker, I.M.H. Vliegen, R.J. Boucherie, Multi-disciplinary planning in health care: a review, Health Syst. 9 (2) (2020) 95–118.

[18] B. Liang, A. Turkcan, M.E. Ceyhan, K. Stuart, Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic, Int. J. Prod. Res. 53 (24) (2015) 7177–7190.

[19] A. Braaksma, N. Kortbeek, G. Post, F. Nollet, Integral multidisciplinary rehabilitation treatment planning, Oper. Res. Health Care (ISSN: 2211-6923) 3 (3) (2014) 145–159.

[20] T. Monks, C.S.M. Currie, B.S. Onggo, S. Robinson, M. Kunc, S.J.E. Taylor, Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines, J. Simul. 13 (1) (2019) 55–67.