Operating room planning and scheduling:

A literature review

Brecht Cardoen†*, Erik Demeulemeester‡, Jeroen Beliën‡ †

†Katholieke Universiteit Leuven, Faculty of Business and Economics, Department of Decision Sciences and Information Management, Naamsestraat 69, B-3000 Leuven, Belgium, brecht.cardoen@econ.kuleuven.be, erik.demeulemeester@econ.kuleuven.be, jeroen.beliën@econ.kuleuven.be

‡Hogeschool-Universiteit Brussel, Campus Economische Hogeschool, Centrum voor Modellering en Simulatie, Stormstraat 2, B-1000 Brussel, Belgium, jeroen.beliën@ehsal.be

Abstract

This paper provides a review of recent research on operating room planning and scheduling. We evaluate the literature on multiple fields that are related to either the problem setting (e.g. performance measures or patient classes) or the technical features (e.g. solution technique or uncertainty incorporation). Since papers are pooled and evaluated in various ways, a diversified and detailed overview is obtained that facilitates the identification of manuscripts related to the reader’s specific interests. Throughout the literature review, we summarize the significant trends in research on operating room planning and scheduling and we identify areas that need to be addressed in the future.

Keywords: health care, operating room, scheduling, planning, literature review

1 Introduction

The managerial aspect of providing health services to patients in hospitals is becoming increasingly important. Hospitals want to reduce costs and improve their financial assets, on the one hand...
hand, while they want to maximize the level of patient satisfaction, on the other hand. One unit that is of particular interest is the operating theater. Since this facility is the hospital’s largest cost and revenue center [65, 80], it has a major impact on the performance of the hospital as a whole. Managing the operating theater, however, is hard due to the conflicting priorities and the preferences of its stakeholders [57], but also due to the scarcity of costly resources. Moreover, health managers have to anticipate the increasing demand for surgical services caused by the aging population [48]. These factors clearly stress the need for efficiency and necessitate the development of adequate planning and scheduling procedures.

In the past 60 years, a large body of literature on the management of operating theaters has evolved. Magerlein and Martin [81] review the literature on surgical demand scheduling and distinguish between advance scheduling and allocation scheduling. Advance scheduling is the process of fixing a surgery date for a patient, whereas allocation scheduling determines the operating room and the starting time of the procedure on the specific day of surgery. Blake and Carter [11] elaborate on this taxonomy in their literature review and add the domain of external resource scheduling, which they define as the process of identifying and reserving all resources external to the surgical suite necessary to ensure appropriate care for a patient before and after an instance of surgery. They furthermore divide each domain in a strategic, administrative and operational level, although these boundaries may be vague and interrelated. Przasnyski [100] structures the literature on operating room scheduling based on general areas of concern, such as cost containment or scheduling of specific resources. Other reviews, in which operating room management is covered as a part of global health care services, can be found in [16, 99, 106, 121].

The aim of this literature review paper is threefold. First, we want to provide an updated
overview on operating room planning and scheduling that captures the recent developments in this rapidly evolving area. In order to maintain a homogeneous set of contributions, we do not enlarge the scope to operating room management and hence exclude from this review topics such as business process reengineering, the impact of introducing new medical technologies, the estimation of surgery durations or facility design. In other words, we restrict the focus to operating room capacity planning and surgery scheduling (timetabling). Second, we want to structure the obtained information in such a way that research contributions can easily be linked to each other and compared on multiple facets, which should facilitate the detection of contributions that are within a specific researcher’s area of interest. In Section 2, we describe how the structure of this review paper contributes to this goal. Third, pooling literature in a detailed manner enables the identification of issues that are currently (not) well covered and examined.

We searched the databases Pubmed, Web of Science, Current Contents Connect and Inspec on relevant manuscripts on operating room planning and scheduling. Furthermore, references that were cited in the manuscripts were reviewed for additional publications, which eventually led to a set of 246 manuscripts. As can be seen from Table 1, this set largely consists of articles published in scientific journals. Note that almost half of the contributions appeared in or after 2000, which clearly illustrates the increasing interest of researchers in this domain. Since the total number of manuscripts is large and our main interest is directed towards the recent advances proposed by the scientific community, we restrict the set of manuscripts to those published in or after 2000. We furthermore limit the contributions that are incorporated in this review to those that are written in English in order to augment the paper’s accessibility. A detailed bibliography of the entire set of manuscripts, however, is provided on http://www.econ.kuleuven.be/public/NDBAA92.
Table 1: Number of manuscripts in the original set, categorized according to publication type and publication year

<table>
<thead>
<tr>
<th></th>
<th>1950-1999</th>
<th>2000-Present</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal</td>
<td>106</td>
<td>81</td>
<td>187</td>
</tr>
<tr>
<td>Proceedings</td>
<td>15</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>Working paper</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Ph.D. dissertation</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>132</td>
<td>114</td>
<td>246</td>
</tr>
</tbody>
</table>

2 Organization of the paper

When a researcher is interested in finding papers on, for instance, operating room utilization, a taxonomy based on solution technique does not seem very helpful. Equivalently, a taxonomy based on performance measures is not accurate when the reader wants to identify papers that deal with stochastic optimization. Therefore, we propose a literature review that is structured using descriptive fields. Each field analyzes the manuscripts from a different perspective, which may be either problem or technically oriented. In particular, we distinguish between 7 fields:

- **Patient characteristics (Section 3)**: reviewing the literature according to the elective (in-patient or outpatient) or non-elective (urgency or emergency) status of the patient.

- **Performance measures (Section 4)**: discussion of the performance criteria such as waiting time, patient deferral, utilization, makespan, financial value, preferences or throughput.

- **Decision level (Section 5)**: indicating what type of decision has to be made (date, time, room or capacity) and whether this decision is situated on the discipline, the surgeon or the patient level.
• **Type of analysis (Section 6):** distinguishing between an optimization problem, a decision problem, a scenario analysis, a data envelopment analysis or a complexity analysis.

• **Solution technique (Section 7):** overview of the solution procedures retrieved from the manuscript set, such as mathematical programming methods, constructive and improvement heuristics, simulation or analytical approaches.

• **Uncertainty (Section 8):** indicating whether researchers incorporate arrival or duration uncertainty (stochastic approach) or not (deterministic approach).

• **Applicability of research (Section 9):** information on the testing (data) of research and its implementation in practice.

Each section consists of a brief discussion of the specific field based on a selection of appropriate manuscripts and clarifies the terminology when needed. Furthermore, a detailed table is included in which all relevant manuscripts are listed and categorized. Pooling these tables over the several fields should enable the reader to reconstruct the content of specific papers. They furthermore act as a reference tool to obtain the subset of papers that correspond to a certain characteristic.

Although the introduction of descriptive fields may be seen as a first attempt to classify and categorize the literature on operating room planning and scheduling, this is not a formal objective of this review paper. In order to obtain a transparent and simple classification scheme, the number of fields should be reduced and information should be aggregated, which would probably lead to a loss of information. However, this should be a topic for future research as successful classifications based on fields are already provided in the domain of, for instance, machine scheduling [15] or project scheduling [24].
3 Patient characteristics

Two major patient classes are considered in the literature on operating room planning and scheduling, namely elective and non-elective patients. The former class represents patients for whom the surgery can be well planned in advance, whereas the latter class groups patients for whom a surgery is unexpected and hence needs to be performed urgently.

As shown in Table 2, the literature on elective patient planning and scheduling is rather vast compared to the non-elective counterpart. Although many researchers do not indicate what type of elective patients they are considering, some distinguish between inpatients and outpatients. Inpatients refer to hospitalized patients who have to stay overnight, whereas outpatients typically enter and leave the hospital on the same day. Adan and Vissers [1], for instance, consider both inpatients and outpatients in their research. They formulate a mixed integer programming model to identify the cyclic number and mix of patients that have to be admitted to the hospital in order to obtain the target utilization of several resources such as the operating theater or the intensive care unit (ICU). In their case, outpatients are treated as inpatients with a length of stay of one day who do not necessarily need specialized resources such as the ICU.

When considering non-elective patients, a distinction can be made between urgent and emergent surgery based on the responsiveness to the patient’s arrival (i.e. the waiting time until the start of the surgery). Whereas the surgery of emergent patients (emergencies) has to be performed as soon as possible, urgent patients (urgencies) refer to non-elective patients that are sufficiently stable so that their surgery can possibly be postponed for a short period. Wullink et al. [120], for instance, examined whether it is preferred to reserve a dedicated operating room or to reserve some capacity in all elective operating rooms in order to improve the responsiveness to emergencies. Using discrete-event simulation they found that the responsiveness, the amount
Table 2: Patient characteristics

<table>
<thead>
<tr>
<th>elective</th>
<th>[1, 3, 7, 6, 9, 18, 20, 38, 39, 40, 54, 72, 87, 88, 90, 98, 108, 110, 113, 115, 122]</th>
</tr>
</thead>
<tbody>
<tr>
<td>inpatient</td>
<td>[1, 3, 8, 18, 20, 21, 22, 27, 31, 32, 37, 38, 39, 40, 53, 66, 72, 87, 88, 90, 98, 108, 110, 115, 122]</td>
</tr>
<tr>
<td>outpatient</td>
<td>[1, 3, 8, 18, 20, 21, 22, 27, 31, 32, 37, 38, 39, 40, 53, 66, 72, 87, 88, 90, 98, 108, 110, 115, 122]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>non-elective</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>urgent</td>
<td>[10, 17, 49, 82, 88, 98]</td>
</tr>
<tr>
<td>emergent</td>
<td>[20, 64, 75, 76, 77, 87, 97, 98, 111, 115, 120, 122]</td>
</tr>
<tr>
<td>not specified</td>
<td>[71, 72, 112]</td>
</tr>
</tbody>
</table>

of overtime and the overall operating room utilization significantly improved when the reserved capacity was spread over multiple operating rooms. Bowers and Mould [17] group orthopaedic urgencies into trauma sessions and use Monte-Carlo simulation to determine which session length balances the amount of session overruns with an acceptable utilization rate. They furthermore provide both a discrete-event simulation model and an analytical approximation to explore the effects of including elective patients in the trauma session.

4 Performance measures

As depicted in Table 3, various criteria are proposed to evaluate the performance of the planning and scheduling methods. One common evaluation measure relates to the waiting time of patients or surgeons. Denton et al. [26], for instance, examine how case sequencing affects patient waiting time, operating room idling time (i.e. surgeon waiting time) and operating room overtime. They formulate a two-stage stochastic mixed integer program (MIP) and propose a set of effective solution heuristics that are furthermore easy to implement. Note that patient
Waiting time may also be interpreted as the stay on a surgery waiting list. As illustrated by VanBerkel and Blake [115], this issue is closely related to throughput analysis. In their study they use discrete-event simulation to examine how a change in throughput triggers a decrease in waiting time. In particular, they affect throughput by changing the capacity of beds in the wards and by changing the amount of available operating room time. Note that their operating theater of interest is spread over multiple sites, which is rare in the literature (see Section 5).

The utilization of resources is a second performance measure, next to waiting time, that is well addressed in the literature. Especially the utilization rate of the operating room has been the subject of recent research. Dexter et al. [29, 32, 33, 39, 44, 45, 46] evaluate procedures based on the OR efficiency, which is a measure that incorporates both the underutilization and the overutilization of the operating room. As shown in Table 3, we relate underutilization to undertime and overutilization to overtime, although they do not necessarily represent the same concept. Utilization actually refers to the workload of a resource, whereas undertime or over-time includes some timing aspect. It is hence possible to have an underutilized operating room complex, although overtime may occur in some operating rooms. Consider, for instance, two operating rooms with a daily capacity of 4 hours. When we assume that operating room 1 (room 2) has a surgical workload of 2 (5) hours, only 7 out of 8 operating room hours are used. Although this operating theater is underutilized, one hour of overtime in operating room 2 is incurred. We prefer, though, to group the terms since it is unclear in many manuscripts which view is applied. Since the operating room schedule affects other facilities in the hospital, researchers also focused, to a lesser extent, on the utilization of resources other than the operating room. In Section 3 we already introduced the example by Adan and Vissers [1] in which the deviation between the target utilization of resources such as the ICU staff, ICU beds or regular ward beds
Table 3: Performance criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Patient</th>
<th>Surgeon</th>
<th>Throughput</th>
<th>Utilization overutilization / overtime</th>
<th>Utilization underutilization / undertime</th>
</tr>
</thead>
<tbody>
<tr>
<td>throughput</td>
<td>[3, 5, 20, 53, 64, 103, 104, 109, 115]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>utilization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>underutilization / undertime operating room</td>
<td>[1, 29, 32, 33, 39, 44, 45, 46, 47, 50, 51, 52, 62, 67, 68, 69, 77, 78, 86, 89, 91, 94, 108, 110, 118, 122]</td>
<td>[1, 118]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICU</td>
<td>[1, 118]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>overutilization / overtime operating room</td>
<td>[1, 17, 20, 23, 25, 26, 28, 29, 32, 33, 39, 44, 45, 46, 47, 50, 51, 52, 59, 60, 63, 67, 68, 69, 75, 76, 77, 78, 82, 84, 86, 89, 91, 94, 95, 98, 102, 104, 108, 109, 110, 111, 112, 118, 120]</td>
<td>[1, 20, 118]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICU</td>
<td>[1, 91, 118]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACU</td>
<td>[21, 22, 27]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>general</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>operating room</td>
<td>[3, 5, 10, 17, 18, 20, 32, 40, 43, 49, 53, 63, 64, 78, 89, 90, 97, 104, 109, 110, 111, 112, 113, 120]</td>
<td>[18, 20, 49, 64]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ward</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>holding area</td>
<td>[8, 21, 22, 66, 82, 83]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>patient volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>makespan</td>
<td>[51, 52, 66, 82, 93, 98]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>patient deferral / refusal</td>
<td>[2, 17, 20, 49, 54, 64, 70, 95, 96, 97, 104, 109]</td>
<td>[2, 17, 20, 49, 54, 64, 70, 95, 96, 97, 104, 109]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>financial</td>
<td>[12, 19, 30, 31, 32, 35, 36, 37, 38, 40, 41, 60, 72, 79, 87, 90]</td>
<td>[12, 19, 30, 31, 32, 35, 36, 37, 38, 40, 41, 60, 72, 79, 87, 90]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>[4, 10, 13, 14, 23, 34, 46, 53, 58, 62, 75, 76, 77, 79, 82, 90, 92, 95, 98, 101, 102, 109, 110, 114, 117]</td>
<td>[4, 10, 13, 14, 23, 34, 46, 53, 58, 62, 75, 76, 77, 79, 82, 90, 92, 95, 98, 101, 102, 109, 110, 114, 117]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
is minimized. Vissers et al. [118] furthermore provide a case study in which they illustrate this approach for a department of cardiothoracic surgery.

The literature provides a diverse set of manuscripts that incorporate a leveling objective. Marcon and Dexter [83], for instance, use discrete-event simulation to examine how standard sequencing rules, such as *longest case first* or *shortest case first*, may assist in reducing the peak number of patients in both the holding area and the post anesthesia care unit (PACU). A similar analysis of such sequencing rules is provided in [82]. In this paper, however, the authors restrict the focus to the PACU and study, amongst other, its makespan and the peak number of patients. The makespan represents in this case the completion time of the last patient’s recovery. In both studies, operating rooms are sequenced independently which resulted in a reduced complexity. It should be clear, though, that this can be done simultaneously as well (e.g. [22, 66]).

The quality of a planning or scheduling procedure may also be evaluated by the number of deferred, refused or canceled patients. Kim and Horowitz [70], for instance, study how to include quotas in the surgery scheduling process in order to streamline the admittance to the ICU. In particular, they try to reduce the number of canceled elective surgeries that result from ICU bed shortages without significantly worsening the waiting times of other patients who are seeking admission to the ICU.

Financial criteria make up an alternative performance perspective. Dexter et al. [30, 32, 35, 36, 37, 40] examine how adequate planning and scheduling contributes to an increased contribution margin, which they define as revenue minus variable costs. It should be noted that research efforts are not limited to the identification of the best practice. Dexter et al. [31], for instance, formulate a linear programming model in which the variable costs are maximized in order to determine the worst case scenario.
One other evaluation approach is to determine how well operating room planning or scheduling procedures fulfill the preferences of its stakeholders. Cardoen et al. [21, 22], for instance, solve a case sequencing problem in which they try, amongst other, to schedule surgeries of children and prioritized patients as early as possible on the surgery day. At the same time, they want patients with a substantial travel distance to the ambulatory surgery center to be scheduled after a certain hour.

The final category of Table 3 depicts manuscripts that describe other performance measures than those that were addressed in the previous paragraphs. This category groups criteria related to, for instance, the use of additional capacity of specific resources [58, 95, 102, 114, 117], delays in PACU admissions [34, 82] or operating room target allocation [13, 14, 23, 101].

5 Decision level

A variety of planning and scheduling decisions with a resulting impact on the performance of the operating theater are studied in the literature. In Table 4, we provide a matrix that indicates what type of decisions are examined in the manuscripts, such as the assignment of a date (e.g. on Monday, in January), a time indication (e.g. at 11 a.m.), an operating room (e.g. operating room 2, operating room of type A) or the allocation of capacity (e.g. three hours of operating room time). The manuscripts are furthermore categorized according to the decision level they address, i.e. to whom the particular decisions apply.

The discipline level unites contributions in which decisions are taken for a medical specialty or department as a whole. Blake et al. [13] and Blake and Donald [14], for instance, report on an integer programming model and an improvement heuristic to construct a cyclic timetable that minimizes the underallocation of a specialties’ operating room time with respect to its
Table 4: Type and level of decisions

<table>
<thead>
<tr>
<th>discipline level</th>
<th>surgeon level</th>
<th>patient level</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>date</td>
<td>[6, 13, 14, 19, 23, 101, 103, 109, 122]</td>
<td>[7, 8, 9, 20, 69, 96]</td>
<td>[1, 20, 23, 40, 41, 42, 49, 50, 51, 52, 59, 60, 62, 63, 67, 68, 69, 70, 71, 75, 76, 77, 89, 91, 94, 95, 96, 98, 102, 103, 104, 108, 109, 113, 114, 117, 118]</td>
</tr>
<tr>
<td>time</td>
<td>[6, 64]</td>
<td>[7, 8, 9, 20]</td>
<td>[5, 20, 21, 22, 23, 25, 26, 27, 28, 34, 41, 51, 52, 60, 64, 66, 67, 68, 71, 78, 82, 83, 85, 93, 98, 102, 104, 109, 111, 116]</td>
</tr>
<tr>
<td>room</td>
<td>[13, 14, 23, 101, 104, 109, 122]</td>
<td>[8, 9, 69, 96]</td>
<td>[18, 21, 22, 23, 28, 39, 44, 45, 50, 51, 52, 59, 62, 63, 67, 68, 69, 71, 75, 77, 82, 84, 85, 91, 95, 96, 98, 102, 103, 104, 109, 113, 114, 116, 120]</td>
</tr>
<tr>
<td>capacity</td>
<td>[10, 17, 19, 23, 41, 47, 60, 64, 90, 103, 104, 109, 122]</td>
<td>[12, 20, 30, 31, 35, 36, 37, 38, 43, 69, 72, 96]</td>
<td>[1, 2, 17, 20, 23, 34, 54, 62, 64, 69, 87, 95, 97, 103, 113, 114, 118, 120]</td>
</tr>
<tr>
<td>other</td>
<td>[103, 112]</td>
<td>[42, 89, 96]</td>
<td>[42, 49, 89, 96, 103, 110]</td>
</tr>
</tbody>
</table>

The model determines for each specialty what operating room types are assigned to what days of the week, i.e. a decision concerning date and room. At the surgeon level, Beliën et al. [8] introduce a software tool in which decisions for specific surgeons, instead of disciplines, are considered. For each surgeon, the planner has to decide on what day and in which room surgeries have to be performed. Since operating rooms may be divided in a morning and an afternoon session, the block assignments also incorporate a time indication. The impact of the cyclic timetable decisions on the use of various resources, such as nurses, artroscopic...
towers or lasers, is visualized and guides the planner in improving the constructed surgery schedule. Since the amount of operating room time for each surgeon in the planning horizon is predetermined, no capacity decisions have to be made. Next to the discipline and surgeon level, Table 4 also specifies a patient level. On this level, decisions are made for individual patients or patient types. Although patient types may represent the distinction between, for instance, elective or non-elective patients, they frequently refer to surgical procedure types. This view is incorporated, for instance, by van Oostrum et al. [114]. Starting from a list of recurring procedure types, i.e. types that are frequently performed and hence have to be scheduled in each planning cycle, they decide what mix of procedures will be performed on what day and in which operating room. They aim at the minimization of the number of operating rooms in use, on the one hand, and the leveling of the hospital bed requirements, on the other hand. A two-phase decomposition approach is formulated that is heuristically solved by column generation and mixed integer programming. Although all the above papers claim to construct a cyclic master surgery schedule, it should be clear that the granularity of the outcome differs according to the decision level or perspective chosen by the authors. A similar reasoning applies to case mix planning since the available amount of operating room time (capacity) may be divided according to disciplines, surgeons or patient types.

Although most manuscripts take only one decision level into account, this does not necessarily have to be the case. Testi et al. [109], for instance, report on a hierarchical three-phase approach to determine operating theater schedules. In the first phase, which they refer to as session planning, they determine the number of sessions to be scheduled weekly for each discipline. Since they distribute the available operating room time over the set of disciplines, this problem can be regarded as a case mix planning problem. Phase 2 formulates a master surgery scheduling
problem in which they assign an operating room and a day in the planning cycle to the sessions of each discipline. Both phases are solved by integer programming and are situated on the discipline level. Phase 3, on the contrary, is formulated in terms of individual patients. A discrete-event simulation model is presented to evaluate decisions concerning date, room and time assignments. When patients are scheduled consecutively in an operating room, i.e. without incorporation of idle time, the planned surgery start times (time decision) are determined by sequencing the patients.

We added both a row and a column (other) to Table 4 to provide entries for manuscripts that study the operating room planning and scheduling problems in a way that is not well captured by the main matrix. Manuscripts that are categorized in this column or row, for instance, examine the decision on surgeon-patient combinations [42, 89, 96] or decide in which hospital or site capacity has to be preserved [49, 115].

In the introduction (see Section 1) we already mentioned that operating room planning and scheduling decisions affect facilities throughout the entire hospital. Therefore it seems to be useful to incorporate facilities, such as the ICU or PACU, in the decision process and try to improve the global performance. If not, improving the operating room schedule may worsen the practice and efficiency of those related facilities. In Table 5, we classify the manuscripts according to whether they study the operating theater in isolation or integrate it with other facilities. It is somehow surprising to see that almost half of the contributions limit their scope to an isolated operating theater. One of the major reasons to simplify the research scope probably stems from the increased complexity, both in formulation and in computation, of the decision process caused by the integration. Note that this integration should not be limited to facilities that are situated within one hospital, as studies on multi-facility or multi-site operating room planning
Table 5: Integration of the operating room planning and scheduling process

| isolated operating room | \[2, 4, 10, 13, 14, 17, 23, 25, 26, 28, 33, 36, 38, 39, 40, 41, 43, 44, 45, 46, 47, 50, 51, 52, 58, 59, 60, 62, 63, 71, 72, 75, 76, 77, 78, 84, 85, 86, 89, 92, 94, 96, 97, 101, 102, 104, 110, 111, 112, 113, 120\] |
| integrated operating room | \[1, 3, 5, 7, 6, 8, 9, 12, 18, 19, 20, 21, 22, 27, 30, 31, 34, 35, 37, 42, 49, 52, 53, 54, 60, 62, 64, 66, 67, 68, 82, 83, 87, 88, 91, 93, 95, 98, 103, 108, 109, 114, 115, 116, 117, 118, 122\] |

and scheduling are currently emerging [49, 103, 115].

6 Type of analysis

A substantial part of the literature on operating room planning and scheduling consists of contributions in which a problem is stated and consecutively optimized. As indicated in Table 6, these combinatorial optimization approaches are either exact, i.e. eventually leading to a solution for which optimality can be proven, or heuristic in nature. We furthermore distinguish between single and multiple objective approaches based on the number of performance criteria that need to be optimized. Although it is often stated that heuristic approaches are indispensable to solve practical or real-sized problems efficiently, a lot of powerful exact approaches seem to be suggested in the literature, even when multiple criteria are considered.

Since the computational effort to solve optimization problems does not only depend on the objective function, but also on the type of constraints that are incorporated in the analysis, we list in Table 7 what type of constraints are addressed in the literature. We limit the scope to the occurrence of hard constraints, i.e. constraints or limitations that are never allowed to be violated, as soft constraints are often incorporated as part of the objective function (see Section 4). A first category of hard constraints are those related to the use of resources. As
these resources are costly and limited in capacity, they are often binding and hence have a substantial impact on the set of feasible solutions. Note that hospitals may even impose a limit on the allowed amount of operating room overtime or undertime. Second, we identify precedence constraints or constraints related to time lags. Due to contamination risks, for instance, it is obliged to schedule infected patients at the end of the surgery day or to insert idle time between surgeries which allows for an extended cleaning of the operating room [21, 22, 85]. A third category consists of constraints related to certain release or due dates, whereas a fourth and last category represents the demand-related constraints. Pham and Klinkert [98], for instance, incorporate all but demand-related constraints in their surgical case scheduling problem. They model their optimization problem as a multi-mode blocking job shop problem and develop
Table 7: Type of hard constraints retrieved from operating room optimization approaches

<table>
<thead>
<tr>
<th>resource constraints</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td>holding area</td>
<td>[27, 87, 98]</td>
</tr>
<tr>
<td>ward</td>
<td>[1, 12, 19, 30, 31, 35, 37, 62, 87, 95, 98, 103, 108, 114, 118]</td>
</tr>
<tr>
<td>ICU</td>
<td>[1, 30, 31, 35, 37, 62, 67, 68, 87, 91, 98, 103, 114, 118]</td>
</tr>
<tr>
<td>PACU</td>
<td>[21, 22, 27, 52, 68, 87, 93, 98]</td>
</tr>
<tr>
<td>equipment</td>
<td>[21, 22, 23, 59, 67, 68, 77, 95, 98, 103, 108]</td>
</tr>
<tr>
<td>surgical staff</td>
<td>[1, 7, 12, 21, 22, 27, 51, 58, 59, 63, 67, 68, 69, 71, 89, 91, 94, 96, 98, 102, 103, 108, 109, 113]</td>
</tr>
<tr>
<td>budget</td>
<td>[12, 31]</td>
</tr>
<tr>
<td>regular operating room time</td>
<td>[1, 2, 7, 6, 9, 12, 13, 14, 23, 30, 31, 35, 36, 37, 51, 52, 62, 67, 68, 69, 72, 75, 76, 77, 85, 86, 89, 91, 94, 95, 96, 98, 101, 102, 103, 109, 113, 118, 122]</td>
</tr>
<tr>
<td>operating room overtime / undertime</td>
<td>[23, 51, 52, 59, 62, 67, 68, 77, 91, 98, 102, 109, 114]</td>
</tr>
<tr>
<td>other</td>
<td>[25, 93, 102, 116, 117]</td>
</tr>
<tr>
<td>precedence constraints/time lags</td>
<td>[21, 22, 67, 85, 89, 96, 98, 111]</td>
</tr>
<tr>
<td>release / due date constraints</td>
<td>[23, 50, 51, 52, 59, 67, 68, 71, 75, 76, 77, 94, 98, 102]</td>
</tr>
<tr>
<td>demand constraints</td>
<td>[1, 2, 7, 6, 9, 12, 13, 14, 19, 23, 30, 31, 35, 36, 37, 58, 62, 69, 72, 87, 94, 96, 101, 103, 109, 114, 118, 122]</td>
</tr>
</tbody>
</table>

A MIP formulation to minimize performance criteria such as the resulting makespan or the incurred operating room overtime. Each job or surgery is described as a predetermined sequence of activities and a maximum allowed waiting time between the processing of two consecutive activities is specified (precedence and time lag). Precedence relations or priorities may further be imposed to surgeries in order to resolve conflicts on shared resources. Furthermore, they allow to incorporate urgency deadlines for certain activities (due date) or lower bounds on the execution time (release date). An example of demand-related constraints is, for instance, provided by Santibanez et al. [103], who study the impact of simultaneously changing the master surgery schedule of multiple hospitals on throughput or the peak use of post-surgical resources. In their MIP formulation, they restrict the amount of operating room blocks (i.e. demand for operating
room time) that is assigned to the surgical specialties within each hospital between a lower and upper bound. Equivalently, they state lower and upper throughput bounds for procedure types (i.e. demand for surgery).

As indicated in Table 6, not every analysis related to the planning and scheduling of the operating room is formulated as a traditional optimization problem. Velasquez and Melo [117], for instance, exploit the structure of their scheduling problem in which they assign one specific surgery to a specific day in the planning horizon so that penalties related to the use of additional resources or time window violations are avoided, and divide the set of solutions into equivalence classes. Such equivalence classes group solutions with the same objective value. Optimizing the problem hence boils down to solving a decision problem: ’Can we obtain a feasible solution in the best equivalence class, yes or no?’ When no solution exists, a next (inferior) class is examined until a feasible solution is obtained.

Although data envelopment analysis (DEA) may also be considered to be an optimization approach, we introduce a separate entry for this type of analysis in Table 6. In this methodology, linear programming is used to determine the weights of both inputs and outputs that optimize a decision making unit’s efficiency score. Comparison of a unit with the scores of other units may suggest areas that need to be improved. Basson and Butler [4], for instance, apply DEA to operating room activity. They analyze how rankings of sites based on their operating room efficiency scores differ when the types of inputs (e.g. staffing pattern) and outputs (e.g. number of cases performed per equipped operating room) that are taken into account vary.

Instead of limiting the focus to the optimization of one specific problem setting, researchers may also focus on the impact that results from changes to the operating room setting under study. We refer to this type of analysis as scenario analysis since multiple scenarios or settings...
are compared to each other with respect to the performance criteria. As indicated in Table 6, the literature provides a large set of contributions in which scenario analyses are addressed. Niu et al. [88], for instance, describe a simulation model in which scenarios are tested with adapted resource capacities. In particular, they examine how the length of stay of patients varies according to changes in the number of operating rooms, chairs in the holding unit, beds in the PACU or transporters.

Finally, researchers may also analyze the computational complexity of their combinatorial problem or its corresponding solution approach. Lamiri et al. [76], for instance, prove using the 3-partition problem that their stochastic optimization problem is strongly NP-hard and hence very difficult to solve. We refer the interested reader to Garey and Johnson [55] for an introduction to problem complexity and technical details on this type of analysis. A primer on calculating the computational complexity of algorithmic solution procedures is, for instance, provided in [107].

7 Solution technique

The literature on operating room planning and scheduling exhibits a wide range of solution methodologies that are retrieved from the domains of operations management and operations research. We refer to Gass and Harris [56] or Winston and Goldberg [119] for a brief introduction to the various solution techniques that are listed in Table 8.

Mathematical programming methods tend to be well applied in the literature on operating room planning and scheduling. Mulholland et al. [87], for instance, report on the application of linear programming to determine the mix of patients that optimizes the financial outcome of both physicians and the hospital, taking into account the resulting consumption of multiple
resources such as the ICU, PACU, ward or holding unit. In contrast to linear programming models, quadratic programming models feature a nonlinear objective function. Beliën and Demeulemeester [6] try to minimize the expected total bed shortage, which is not linearly dependent on the decision variables, by adapting the master surgery schedule. They provide heuristic solution methods based on, for instance, simulated annealing and quadratic or mixed integer programming to solve both the original problem and an approximate problem setting in which the objective function is linearized. When dealing with multiple objectives, goal programming
may serve as a flexible optimization technique. For each objective, a target value or goal is specified. The objective is to minimize the penalized deviations from the targets. Ozkarahan [91], for instance, formulates a goal programming approach in which surgeries, if they are scheduled, are assigned to operating rooms and in which, amongst other, intensive care capabilities or operating room and surgeon preferences are addressed. Mathematical formulations of operating room planning and scheduling problems with a realistic size often result in a huge set of decision variables. Instead of specifying and adding this entire set of variables in advance, column generation generates and adds variables when needed and hence optimizes the problem with only a subset of the variables. Lamiri et al. [77], for instance, describe a column generation approach that assigns patients to surgery days and operating rooms in such a way that patient related costs and operating room utilization costs are minimized. They propose a dynamic programming algorithm to solve the pricing problem, i.e. the subproblem in which promising variables are generated. As column generation cannot force the decision variables to be integer, the authors use the fractional output as input for various constructive and improvement heuristics. However, column generation can also be intertwined with an enumerative branch-and-bound framework in order to obtain integer solutions. This methodology is referred to as branch-and-price and is applied, for instance, by Fei et al. [50]. They assign surgical cases, who may be characterized by a surgery deadline, to specific days and operating rooms so that the total unexploited or overtime operating cost is minimized. Similarly to [7, 22, 52, 77], they solve the appropriate pricing problem through dynamic programming. Other mathematical programming approaches that are retrieved in the literature on operating room planning and scheduling are based on, for instance, lagrangian relaxation [36, 93].

The literature on operating room planning and scheduling also provides, next to mathe-
matical programming methods, a substantial amount of simulation approaches. As shown in Table 8, we distinguish between discrete-event and Monte-Carlo simulation. While discrete-event simulation represents a system as it evolves over discrete or countable points in time (dynamic), Monte-Carlo simulation represents a system at a particular point in time (static) [119]. Lebowitz [78], for instance, applies Monte-Carlo simulation to evaluate and quantify the impact of sequencing procedures on waiting time and operating room utilization criteria. A discrete-event simulation model is designed by Sciomachen et al. [104] in order to evaluate the utilization of operating rooms or medical disciplines, patient throughput and the number of overruns or patient deferrals. In particular, they examined the impact of changing, amongst other, the master surgery schedule and the case sequencing rules on the listed performance criteria. Note that their study largely corresponds with the third phase that is examined in [109].

Dedicated heuristic procedures broadly fall into two main categories, namely constructive and improvement heuristics. Whereas constructive heuristics generally build solutions to planning and scheduling problems from scratch, improvement heuristics perform operations on an existing schedule to transform a solution in an improved one. Guinet and Chaabane [59], for instance, present a primal-dual constructive heuristic that assigns patients to surgery days and operating rooms. Their algorithm, which is an extension of the Hungarian method, minimizes both operating room overtime costs and patient hospitalization costs, i.e. costs resulting from the waiting time between the hospitalization date and the intervention date. Hans et al. [63] propose various priority-based constructive heuristics to maximize the capacity utilization of the operating theater and minimize the risk of overtime by introducing an amount of planned slack time. However, they also elaborate on improvement heuristics such as a random exchange method, which only accepts changes or swaps that yield an improved solution, or a simulated
annealing approach, which accepts worse solutions with a low probability in order to leave local optima. Next to simulated annealing, the literature provides contributions that apply other kinds of meta-heuristics. Hsu et al. [66], for instance, solve a case sequencing problem by tabu search to minimize both the required number of PACU nurses and the completion time of the PACU’s last patient. Roland et al. [102], on the other hand, report on the construction of a genetic algorithm that heuristically minimizes the costs related to operating room openings and overtime. In particular, their scheduling problem, which is closely related to the well-known resource-constrained project scheduling problem, questions what date, operating room and starting time indication should be assigned to the set of surgeries. They validate the performance of the genetic algorithm through a comparison with a MIP approach.

Finally, Table 8 also reports on solution techniques that are rather rarely applied to the domain of operating room planning and scheduling. Lovejoy and Li [79], for instance, analytically examine whether it is preferred to increase capacity by extending the working hours in the current operating rooms or by building new operating rooms. They evaluate both scenarios with respect to the waiting time to get on the schedule, the start-time reliability of procedures and hospital profits. Cardoen et al. [21] describe, next to a MIP model, a dedicated branch-and-bound procedure to solve a multi-objective case sequencing problem that is also addressed in [22]. In contrast to the MIP approaches, their dedicated branching and bounding procedures are not based on LP relaxations.

8 Uncertainty

One of the major problems associated with the development of accurate operating room schedules or capacity planning strategies is the uncertainty inherent to surgical services. Whereas deter-
ministic planning and scheduling approaches ignore such uncertainty or variability, stochastic approaches try to incorporate it. In Table 9, we list the relevant manuscripts based on their uncertainty incorporation. Two types of uncertainty that seem to be well addressed in the stochastic literature are arrival uncertainty and duration uncertainty. The former points, for instance, at the unpredictable arrival of emergency patients or at the lateness of surgeons at the beginning of the surgery session, whereas the latter represents deviations between the actual and the planned durations of activities related to the surgical process. Harper [64], for instance, presents a detailed hospital capacity simulation model that enables system evaluations by means of scenario analyses. The participation of multiple hospitals in the development phase resulted in a generic framework that allows to incorporate uncertainty or trends in the arrival profiles of patient groups as well as duration variability (e.g. length of stay or surgery durations). Persson and Persson [97] describe a discrete-event simulation model to study how resource allocation policies at the department of orthopaedics affect the waiting time and utilization of emergency resources, taking into account both patient arrival uncertainty and surgery duration variability. Note that simulation is often preferred as the solution technique to study complex stochastic set-

| deterministic | [1, 2, 4, 7, 8, 12, 13, 14, 19, 21, 22, 23, 31, 37, 50, 51, 52, 58, 59, 62, 66, 67, 68, 69, 71, 72, 83, 85, 87, 89, 90, 91, 93, 95, 96, 98, 101, 102, 103, 108, 109, 111, 116, 117, 118, 122] |
| stochastic | [6, 9, 10, 17, 18, 20, 36, 40, 41, 43, 49, 60, 64, 70, 75, 76, 77, 79, 86, 88, 92, 95, 97, 104, 109, 111, 115, 120, 122] |
| arrival | [3, 5, 6, 9, 10, 17, 20, 25, 26, 27, 28, 41, 43, 44, 49, 54, 60, 62, 63, 64, 70, 75, 76, 78, 79, 82, 83, 84, 86, 88, 92, 94, 97, 104, 109, 110, 111, 112, 113, 114, 115, 120, 122] |
| duration | [20, 35, 37] |
tings, as it features an extensive modeling flexibility. Next to arrival and duration uncertainty, other types of uncertainty may be addressed. Dexter and Ledolter [35], for instance, examine to what extent uncertainty in the estimated contribution margin of surgeons (characterized by e.g. standard deviations) may lead to inferior allocations of operating room capacity when the goal is to maximize a hospital’s expected financial return.

Remarkably, only few manuscripts explicitly refer to resource uncertainty (see [20] for an example), while this topic currently is a hot topic in, for instance, project management or project scheduling [74]. It should be noted, though, that resource uncertainty often coincides with arrival uncertainty. For example, the arrival of emergencies may result in a claim of both the surgeon who is needed to perform the emergent surgery and a specific operating room. These claims actually result in resource breakdowns as the elective program cannot be continued and hence has to be delayed.

9 Applicability of research

Many researchers provide, next to the development of a model or a formulation, a thorough testing phase in which they illustrate the applicability of their research. Whether this applicability points at computational efficiency or at showing to what extent objectives may be realized, a substantial amount of data is desired. From Table 10, we notice that most of this data stems from reality. This evolution is noteworthy and results from the improved hospital information systems from which data can be easily extracted. Unfortunately, a single testing of procedures or tools based on real data does not imply that they finally get implemented in practice. Although Lagergren [73] indicates that this lack of implementation in the health services seems to have improved considerably, this literature review only lists few contributions for which its implemen-
Table 10: Applicability of research

<table>
<thead>
<tr>
<th>Category</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>no testing</td>
<td>[29, 33, 42, 58, 60]</td>
</tr>
<tr>
<td>data for testing</td>
<td>[7, 6, 25, 39, 41, 45, 47, 49, 50, 52, 59, 67, 68, 71, 75, 76, 77, 78, 82, 84, 89, 93, 96, 98]</td>
</tr>
<tr>
<td>theoretic data</td>
<td>[1, 2, 3, 4, 5, 8, 9, 17, 18, 19, 20, 21, 22, 26, 27, 28, 30, 31, 35, 36, 37, 38, 39, 40, 43, 44, 45, 47, 51, 53, 54, 62, 63, 64, 66, 69, 70, 72, 79, 83, 85, 86, 87, 89, 90, 91, 92, 94, 95, 96, 97, 98, 101, 103, 104, 108, 109, 111, 112, 113, 114, 115, 116, 117, 118, 120, 122]</td>
</tr>
<tr>
<td>based on real data</td>
<td>[10, 12, 13, 14, 53, 64, 109]</td>
</tr>
</tbody>
</table>

The implementation and use in practice is confirmed. Examples of research approaches that are implemented in practice are already discussed in, for instance, Section 5 [13, 14, 109] or Section 8 [64]. Note, though, that details on the implementation phase are hardly provided. A result of this poor implementation is that a substantial gap may exist between theory and practice. However, only limited research is performed to quantify this gap and to indicate what expertise is currently in use in hospitals. Using a survey, Sieber and Leibundgut [105] recently noticed that the current state of operating room management in Switzerland is far from excellent. It is somehow contradictory to see that in a domain as practical as operating room planning and scheduling, so little research seems to be effectively applied. However, increasing the implementation rate does not only depend on the efforts of the scientific community. Possibly, practitioners lack some kind of awareness of the power of operations management techniques. Therefore, educational applications should be developed to introduce planning and scheduling concepts to the managers of the future. Hans and Nieberg [61], for instance, recently report on an educational tool that specifically focuses on the management of the operating room.
10 Conclusion

In this paper, we reviewed manuscripts on operating room planning and scheduling that have recently appeared. We analyzed the contributions on various levels, which we referred to as fields. Within each field, we highlighted the most important trends and we illustrated important concepts through the citation of key references. Since each discussion is accompanied by a detailed table, which provides even more information than is addressed in the text, readers may easily identify manuscripts that have specific features in common. They furthermore allow to track specific contributions over the different fields and visually indicate what area of research is well addressed or should be subject to future research.

In short, we noticed that most of the research that appeared in or after 2000 is directed to the planning and scheduling of elective patients. The study of issues related to the waiting time of various stakeholders and the utilization of resources seems to be well addressed. Most of the researchers analyze and/or solve their problem, which is frequently formulated at the patient level, by means of mathematical programming methods or simulation. This results in a steady amount of both optimization approaches and scenario analyses. Although the operating theater can be linked with an upstream and downstream process, such integration only occurs for about half of the contributions. Operating room planning and scheduling problems are furthermore studied both in a deterministic setting and a stochastic setting. Although the incorporation of uncertainty is more realistic, a lot of researchers prefer the deterministic approach due to computational complexity. Unfortunately, bridging the gap with reality and implementing the advances in practice seems to be very difficult and should be further addressed in the future.
Acknowledgements

We acknowledge the support given to this project by the Fonds voor Wetenschappelijk Onderzoek - Vlaanderen (FWO - Vlaanderen) as Aspirant (Brecht Cardoen) and as Post-doctoral researcher (Jeroen Beliën).

References


[28] F. Dexter. A strategy to decide whether to move the last case of the day in an operating room to another empty operating room to decrease overtime labor costs. Anesthesia and Analgesia, 91:925–928, 2000.


F. Dexter and A. Macario. Changing allocations of operating room time from a system based on historical utilization to one where the aim is to schedule as many surgical cases as possible. *Anesthesia and Analgesia*, 94:1272–1279, 2002.


F. Dexter, R.D. Traub, and A. Macario. How to release allocated operating room time
to increase efficiency: Predicting which surgical service will have the most underutilized operating room time. *Anesthesia and Analgesia*, 96:507–512, 2003.


[47] R.H. Epstein and F. Dexter. Uncertainty in knowing the operating rooms in which cases were performed has little effect on operating room allocations or efficiency. *Anesthesia and Analgesia*, 95:1726–1730, 2002.


[86] C. McIntosh, F. Dexter, and R.H. Epstein. The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room pro-


[93] V. Perdomo, V. Augusto, and X. Xie. Operating theatre scheduling using lagrangian


[117] R. Velasquez and M.T. Melo. Tactical operating theatre scheduling: Efficient appoint-


