TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Technische Dienstleistungen und Operations Management

Operating Room Planning - Incorporating Stochastic and Behavioral Effects

Dipl.-Kfm. Andreas Fügener

Vollständiger Abdruck der von der Fakultät TUM School of Management der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Stefan Minner
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Rainer Kolisch
	2.	Associate Professor dr.ir. Erwin Hans University of Twente / Niederlande

Die Dissertation wurde am 08.01.2013 bei der Technischen Universität München eingereicht und durch die Fakultät TUM School of Management am 15.04.2013 angenommen.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Health care operations management

Health care services are not only one of the largest, but also one of the fastest growing sectors in all industrialized countries. This is due to demographic changes, rising standards of living, and technological advances [70]. As cost and competitive pressure increases it comes as no surprise that the health care sector is a growing part of the operations research/management science community [39]. Hans et al. [39] propose a framework for clustering health care planning problems (see Table 1.1). The first dimension is the managerial area. They differentiate between medical planning (planning of medical services), resource capacity planning (planning of renewable resources), materials planning (planning of consumable resources/materials), and financial planning (planning of costs and revenues). All chapters of this thesis focus on the field of resource capacity planning. The second dimension is the decomposition into hierarchical levels. The strategic level considers problems with long time horizons and great impact. An example of a strategic planning problem is the creation of new capacities, e.g. the construction of new patient wards. Operational planning usually involves short term decisions. It can further be distinguished into offline and online operational planning. Offline planning problems are those that can be solved in advance, while monitoring and reactive processes are defined as online planning problems. Examples of offline and online operational planning are appointment scheduling and rescheduling in case of emergencies, respectively. The tactical level is

	Medical planning	Resource capacity planning	Materials planning	Financial planning
Strategic	Research, development of medical protocols	Case mix and capacity planning	Supply chain and warehouse design	Investment plans, insurance strategy
Tactical	Treatment and protocol selection	Block planning, admission planning	Supplier selection, tendering	Budget and cost allocation
Operational (offline)	Diagnosis and planning of individual treatment	Appointment scheduling	Purchasing, determining order sizes	DRG billing, cash flow analysis
Operational (online)	Triage, diagnosing emergencies and complications	Monitoring, emergency coordination	Inventory replenishing, rush ordering	Billing complications and changes

Table 1.1: Sample framework of health care planning and control [39]

between the strategic and the operational level. Tactical problems might be those where mid-term allocation of resources defined on the strategic level is made. Tactical decisions often set the frame for operational planning. This thesis addresses a tactical and an offline operational planning problem.

1.2 Operating room planning and management

Planning and management of the operating theater has gained increasing attention in the operations management literature. Guerriero and Guido [38] cite more than 100 studies on operating room (OR) management and Cardoen et al. [18] write "in the last 60 years, a large body of literature on the management of operating theaters has evolved". This comes as no surprise as around 40% of hospital expenses [22] arise in the operating theater, and more than 60% of hospital admissions are for surgical operations [69]. Operating room planning exists on all hierachichal levels. Examples are construction of new operating theaters (strategic level), the creation of master surgery schedules (MSSs) (tactical level), scheduling of surgeries (offline operational level), and rescheduling of surgeries in case of emergencies (online operational level). Vanberkel et al. [89] argue that "surgery care does not operate in isolation". Departments like the post-anesthesia care unit (PACU), the intensive care unit (ICU), the general patient wards, rehab, and the emergency department are affected by the OR. Another important factor that influences OR performance is human behavior. A tactical approach considering effects on downstream units and an offline operational approach analyzing behavioral effects are discussed in this thesis.

1.3 Structure of the thesis

Besides the introduction (Chapter 1) and the conclusions (Chapter 6), this thesis contains four chapters. Chapters 2 and 3 discuss the stochastic tactical operating room management problem of creating MSSs considering downstream units. Chapter 2 discusses the theoretical foundations and Chapter 3 presents a case study in a German university hospital. Chapter 4 presents the effects of penalty aversion on the behavior in the well-known newsvendor problem. While it does not discuss an operating room management issue, it provides the theoretical foundation for Chapter 5. Here, surgeons' behavior when planning surgery durations is discussed. The thesis ends with Chapter 6 providing conclusions and discussing possible directions for future research. In what follows we give a more detailed summary of the four core chapters of this thesis.

1.3.1 Master surgery scheduling with consideration of multiple downstream units (Chapter 2)

Chapter 2 considers a tactical MSS problem in which block OR time is assigned to different surgical specialties. While most MSS approaches in the literature examine only impact of designing the MSS on the operating theater and operating staff, the scope is expanded to include downstream resources, such as the ICU and the general wards required for recovery by the patients once they leave the OR. In this chapter a stochastic analytical approach, which calculates the exact demand distribution for the downstream resources of a given MSS, is proposed. Measures to define downstream costs resulting from the MSS as well as exact and heuristic algorithms to minimize these costs are discussed. This chapter is based on Fügener et al. [34].

1.3.2 Improving ICU and ward utilization by adapting master surgery schedules: A case study (Chapter 3)

The close relationship between an MSS and the bed demand in the downstream units ICU and ward is outlined in Chapter 3. Using historical data retrieved from the clinical information system of a German university hospital and a simple patient flow model, an algorithm for predicting bed demand based on MSSs is applied. Scenario simulations are performed with three different MSSs. The impact on the required number of beds in the downstream units is analyzed. Chapter 3 presents potential improvements of the current MSS and evaluates two alternative MSSs: one lowering the weekend ICU utilization by 20%, the other one reducing the maximum number of ward bed requests by 7%. The application of the algorithm provides insight into the impact of MSS designs on the bed demand in downstream units. It enables the development of MSSs that avoid both peaks in bed requests and high weekend occupancy levels in the ICU and the ward. This chapter is based on Fügener et al. [33].

1.3.3 On the assessment of costs in a newsvendor environment: Insights from an experimental study (Chapter 4)

Chapter 4 addresses the question how the assessment of costs influences decisions in a newsvendor setting. It is expected that different cost types lead to different behavior. This investigation considers a newsvendor problem with opportunity costs and a newsvendor problem with penalty costs. In addition, three cases with different margins for each of the two problems are differentiated. In an experimental study the average order quantities in the newsvendor problem with penalty costs exceeded the average order quantities in the newsvendor problem with opportunity costs and a mean anchor effect, as familiar from a number of previous studies, exists. A different weighting of costs can be seen as the main driver for the different order quantities. Thus, a biased perception of different cost types exists and decision makers are more sensitive to penalty costs than to opportunity costs. Based on the observations in this chapter, both situations where the cost weighting and the mean anchor effect compensate for each other and thus lead to "good" decisions as well as situations where the two effects compound and therefore lead to "bad" decisions can be identified. As penalty costs are present in many newsvendor situations, e.g. the planning of surgery durations, the insights allow for the application of the findings from behavioral studies of the newsvendor problem to a broader context. This chapter is based on Schiffels et al. [74].

1.3.4 Underutilization and overutilization of operating rooms: Insights from behavioral health care operations management (Chapter 5)

The planning of surgery durations discussed in Chapter 5 is crucial for efficient usage of operating theaters. Planning too long or too short slots for surgeries leads to operating room inefficiency, e.g. idle time, overtime, or rescheduling of surgeries. Since surgery durations are stochastic, the overall objective of planning surgery durations is to minimize the expected costs of operating room inefficiency. While most health care studies assume rational behavior of decision makers, experimental studies have shown that decision makers often act irrational. Based on insights from health care operations management, medical decision making, behavioral operations management, as well as empirical observations, hypotheses that surgeons' behavior deviates from rational behavior in many ways are derived. To investigate this, an experimental study where experienced surgeons were asked to plan surgeries with uncertain durations was set up. Systematic deviations from optimal decision making are discovered and behavioral explanations for the observed biases are given. This research provides new insights to tackle a major problem in hospitals, i.e. low operating room utilization going along with staff overtime. This chapter is based on Fügener et al. [35].

Chapter 2

Master surgery scheduling with consideration of multiple downstream units

2.1 Introduction

Due to an aging society and technological advances, the demand for health care services is rising in industrialized countries [42]. At the same time, cost cuts and human resource shortages lead to increasing pressure on hospital resources. Therefore, the importance of optimizing the usage of scarce resources in hospitals is self-evident. The most expensive resource in most hospitals is the operating room (OR) [38]. ORs are clearly connected with other "downstream" resources, for example, the post-anesthesia care unit (PACU), the intensive care unit (ICU), and the general patient wards, referred to hereafter as "ward." When planning the operating rooms and the downstream units, decision makers face a trade-off between the high complexity of a holistic view and the danger of suboptimal solutions resulting from focusing on isolated units [89].

Many hospitals use a so-called block-booking system when planning surgeries. In this system a medical specialty, e.g. urology, is assigned to blocks denoting a specific amount of time, e.g. a day, in one OR. These blocks can be combined into cyclical master surgery schedules (MSSs), where every block is repeated after a fixed cycle, e.g. every two weeks. In planning and scheduling problems can be categorized according to levels of a decision hierarchy [39]: The strategic, tactical, offline-operational — i.e. planning in advance — and the online-operational — i.e. reacting/monitoring — level. In block-booking systems, decisions are made on all hierarchical levels. At the strategic level the number of blocks assigned to the specialties during a MSS cycle is determined. At the tactical level, OR-days are allocated to specialties in an MSS, such that the strategic allocation is met. At the operational level, patients are scheduled (offline) and rescheduled in case of emergencies or unexpected changes (online). An overview of OR planning may be found in Hans and Vanberkel [40].

In this chapter, we discuss the tactical MSS problem, concentrating on the effect the MSS has on downstream care units. Surgeries performed in each block of the MSS create a flow of patients through the ICU to the ward, or directly from the OR to the ward, before they leave the hospital. We exclude the PACU in our tactical problem and denote the ICU and the ward as downstream units. We define a model to calculate the distributions of recovering patients in the downstream units expected from the MSS. Based on this, we propose an approach for planning the MSS with the objective to minimize downstream costs by leveling bed demand and reducing weekend bed requests. Anderson et al. [4] show that a high level of utilization in hospital wards leads to a higher discharge rate of patients which might reduce the quality of care. On days with high patient inflow to the ICU the danger of readmissions [5] and the probability of rejected ICU requests [60] strongly increases. Therefore, downstream units should also be considered in surgery planning for medical reasons.

The remainder of this chapter is organized as follows: In Section 2.2 we provide a brief overview of the relevant literature. Section 2.3 presents an algorithm for calculating the distribution of recovering patients in the downstream units — ICU and ward. Section 2.4 offers a generic model to determine optimal MSSs and a discussion of relevant objective functions to determine downstream costs. In Section 2.5 we present a branch-and-bound algorithm and different heuristics to minimize these costs. We test the algorithms in Section 2.6 in an experimental investigation using data obtained from a Dutch hospital. Finally, we discuss managerial implications, limitations, and potential extensions of our study.

2.2 Literature review

Operating rooms are among the most expensive resources in hospitals and a large number of studies about OR scheduling exists [18]. For recent literature reviews on OR scheduling, see Cardoen et al. [18] and Guerriero and Guido [38]. Articles about health care models that include both the OR and downstream units are reviewed in Vanberkel et al. [89]. In this section, we will focus our review on articles that combine OR scheduling with the effect on downstream units, such as ICUs or wards.

Adan and Vissers [3] present a deterministic integer programming approach to schedule patients based on fixed capacities in the OR, the ICU, and the ward. The ICU and ward capacities are the number of beds available for each specialty, while the OR capacity is the total available operating time per day. Additionally, the capacity of the nursing staff is considered. Based on this, a daily admission profile for different specialties that minimizes the deviation from resource utilization targets is obtained. Santibanez et al. [73] examine various trade-offs made during tactical OR planning. They also apply a deterministic mixed-integer program and compare different objectives, e.g. maximizing throughput of patients or leveling the bed requests of downstream units like ICUs or wards. They differentiate between beds and nursing levels as well as between ORs and surgeons. A more detailed (i.e. closer to operational) model to construct an MSS where patient types are assigned to blocks is formulated by [88]. They seek to minimize the required OR capacity and to level hospital ward bed requirements. To incorporate the uncertainty of OR durations, they introduce probabilistic constraints. They solve the model in two steps. First, OR capacities are optimized without consideration of hospital-beds using so-called Operating Room Day Schedules (ORDS), i.e., lists of surgery types that are assigned to one OR day. Then, the ORDSs are assigned to OR days in order to level hospital-bed demand. Therefore, leveling hospital-bed demand is only possible using the precomputed set of ORDS. These three studies model multiple downstream units with deterministic approaches, while our study employs a stochastic approach.

Models for creating MSSs with leveled bed occupancy in downstream units are presented in Beliën and Demeulemeester [8]. Contrary to the articles presented above, both the number of patients and the length of stay in the hospital are assumed to be stochastic. A multinomial distribution is used to model the length of stay. The authors aim to minimize the expected bed shortage and employ a mixed-integer programming and simulated annealing approach. The approach of Beliën and Demeulemeester [8] differs from our approach in only allowing one downstream resource (ward), while we model the patient flow including the ICU and ward and thus consider multiple downstream units.

Min and Yih [61] propose an operational scheduling of elective surgeries that considers both uncertainty and downstream capacity constraints. They formulate a stochastic surgery scheduling problem minimizing the sum of costs directly related to patients and expected overtime costs. The downstream capacities are modeled as constraints. In contrast to their approach, that considers the operational surgery planning level, we focus on the tactical level.

Our study is based on the approach of Vanberkel et al. [91] where binomial distributions and discrete convolutions are used to calculate the exact distribution of recovering patients in the ward resulting from a given MSS. Vanberkel et al. [91] present algorithms to determine the distributions of ward occupancy, patient admissions, patient discharges, and the number of patients on each day of their recovery period. The major contribution of that paper is an exact algorithm to evaluate the effect of an MSS on ward occupancy. A case study where the algorithm is implemented in a Dutch hospital is presented in Vanberkel et al. [90]. The authors use the algorithm to test several MSSs and to choose one with a more favorable ward occupancy pattern. However, the study contains some limitations, which we address in the study at hand. First, the algorithm only includes the ward as a single downstream unit. As the ICU is an important bottleneck in hospitals [53], we incorporate ICU bed requests in the model as a valuable extension. Second, Vanberkel et al. [91] contains no quantitative approach to evaluate the costs resulting from an MSS. As different downstream costs exist, e.g. costs for providing fixed capacities or costs for weekend staffing, we develop a model to assign costs to specific MSSs. Third, we improve on the selection process for new MSSs by introducing several exact and heuristic algorithms to minimize downstream costs.

To the best of our knowledge, the current study presents the first exact stochastic approach to calculate patient occupancy distributions for the ICU and ward following an MSS. In addition, we present exact and heuristic algorithms to minimize costs resulting from patients in downstream units.

2.3 Recovering patients in downstream units

In this section, we describe a model that calculates the exact distribution of post-operative inpatients in the ICU and the ward resulting from a given MSS cycle. We do not further distinguish between different ICUs or wards in this study. However, the presented model can easily be extended to include several ICUs and wards as well as stays in the PACU. We now present the general underlying assumptions regarding the process, the data requirements, and the detailed model.

After an operation several patient paths exist. In most cases, patients are admitted to the ward. In more acute cases, patients are sent to the ICU to recover. Alternatively, patients might be discharged without being sent to the ward or die. Patients in the ward will be transferred to the ICU if their condition has become unstable. Most patients leave the system only after recovering in the ward, but they might also leave the hospital directly from the ICU (e.g. in case of death or if transferred to another hospital). The patient paths are outlined in Figure 2.1.

In studying data from a large University Hospital in Munich, Germany, we found that more than 98% of inpatients follow one of three paths. The vast majority (92%) follow the path OR → ward → discharge. About 5% follow OR → ICU → ward → discharge. Just above 1% follow the path OR → ICU → discharge, i.e. the previous path with a zero day stay in the

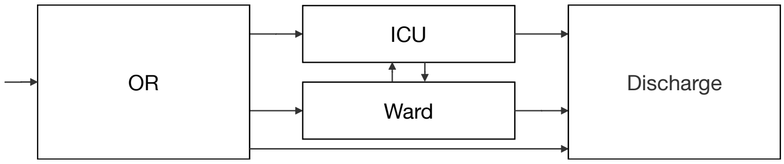

Figure 2.1: Patient paths

ward. It is very rare for patients to return to the ICU after being transferred from the ICU to the ward (just above 1%). Based on this data we have simplified the modeled patient pathway as depicted in Figure 2.2.

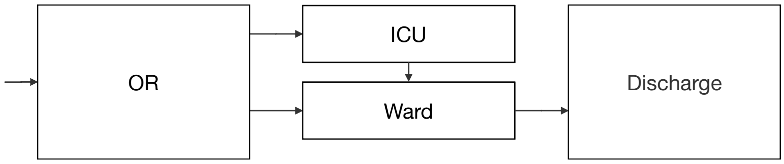

Figure 2.2: Simplified patient paths

The number of patients sent to the ICU or the ward after one surgery block is modeled by a discrete empirical distribution. This distribution may also include emergency patients who were operated on during this block. A stay in the ICU is denoted by "I," a stay in the ward of patients who directly came from the operating room by "WO," and a stay in the ward of patients who were transferred from the ICU by "WI." The lengths of stay (in days) in the ICU or the ward, after being transferred from the OR or from the ICU, are also modeled by discrete empirical distributions. Such distributions are easily obtained from historical records.

The main sets and indices used in the following model are shown in Table 2.1. The index n is used to determine days after surgery, where 1 denotes the day of surgery. Days after a transfer to the ward from the ICU will be denoted by u. Required historical or estimated data for every specialty $j \in \mathcal{J}$ are as follows:

- $a_j(p)$ represents the probability that $p \in \{0, \ldots, P_j\}$ patients are operated on during a surgery block of specialty j.
- b_j represents the probability that a patient of specialty j is admitted

Description	Index $\in$ Set
Surgery specialties	$j \in \mathcal{J}$
Operating rooms (ORs)	$i \in \mathcal{I}$
Patients	$p \in \{0, \ldots, P_j\}$
Days in the ICU after surgery	$n \in \{1, \ldots, N_j^I\}$
Days in the ward after surgery	$n \in \{1, \ldots, N_j^{WO}\}$
Days in the ward after ICU	$u \in \{0, \ldots, N_j^{WI}\}$
Days in the MSS cycle	$\ell \in \mathcal{L}$
Weekdays in the MSS cycle	$q \in \mathcal{Q}$
Weekend days in the MSS cycle	$\ell \in \mathcal{L} \setminus \mathcal{Q}$

Table 2.1: Sets and Indices

to the ICU immediately after surgery. $1 - b_j$ is the probability that the patient is admitted to the ward.

- $c_j^I(n)$ represents the probability that a patient from surgery of specialty j stays exactly $n \in \{1, \ldots, N_j^I\}$ days in the ICU after surgery.
- $c_j^{WO}(n)$ represents the probability that a patient from surgery of specialty j stays exactly $n \in \{1, \ldots, N_j^{WO}\}$ days in the ward after surgery.
- $c_j^{WI}(u)$ represents the probability that a patient from surgery of specialty j stays exactly $u \in \{0, \ldots, N_j^{WI}\}$ days in the ward after being released from the ICU. A stay of zero days implies a direct release from the ICU.

The model works in three steps (see Figure 2.3). First, we calculate the distributions of recovering patients from a single surgery block in the ICU and the ward, respectively. This step is carried out for each surgical specialty. In the next step we calculate the distributions for a single cyclical block. It is important to note that we assume the MSS to be cyclical. Therefore, each block will be repeated for each new MSS cycle. In the third step we combine all blocks from a cyclical MSS. The first two steps do not depend on the specific MSS, we only need information about the definition of surgery blocks (e.g. length of a block) and the length of the MSS cycle. Therefore, these steps can be calculated during preprocessing. Due to the structure of the problem, the third step has to be calculated for each MSS we want to evaluate.

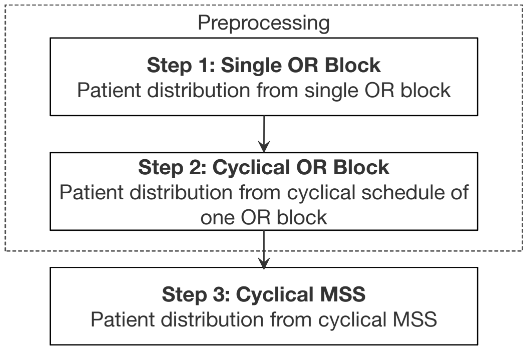

Figure 2.3: Process steps

2.3.1 Calculation of the distributions of patients resulting from a single OR block (Step 1)

In the following, we present the algorithm to derive the distributions of patients resulting from a single OR block. First, using the probability of an ICU admission and the empirical length of stay distributions, we analyze the pathway of a single patient through the hospital (see Figure 2.4). After surgery, a patient can be admitted either to the ICU or to the ward. On each day n, a patient in the ICU may either stay or be transferred to the ward. A patient in the ward may either stay or be released from the hospital. We assume that the probability for a patient to be discharged from the ward after being transferred from the ICU only depends on the length of stay in the ward.

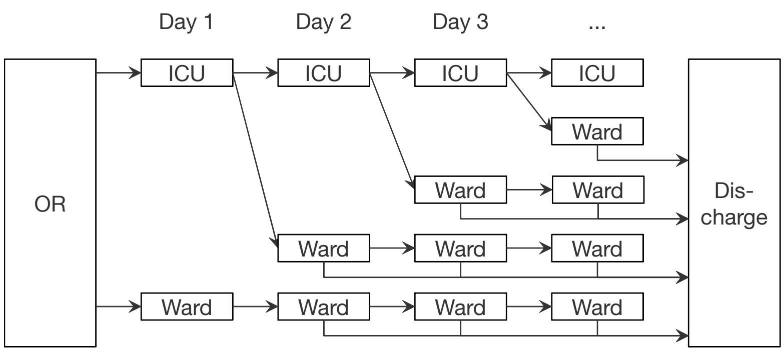

Figure 2.4: Patient paths modeled as a Markov chain

Equation (1) calculates the probability $d^I_{j,n}$ for a patient of specialty j in the ICU on day n (i.e. who has not been transferred by then) to be transferred to the ward that day. Analogously, Equation (2) calculates the probability $d^{WO}_{j,n}$ that a patient who is in the ward n days after surgery is discharged on that day; and Equation (3) calculates the probability $d^{WI}_{j,u}$ for a discharge u days after the transfer from the ICU to the ward. Patients who leave the hospital after staying in the ICU are modeled to have a stay of zero days in the ward. The calculations follow the logic of Vanberkel et al. [91].

$$d^I_{j,n} = \frac{c^I_j(n)}{\sum_{k=n}^{N^I_j} c^I_j(k)} \qquad j \in \mathcal{J}, n \in \{1, \dots, N^I_j\} \tag{1}$$

$$d^{WO}_{j,n} = \frac{c^{WO}_j(n)}{\sum_{k=n}^{N^{WO}_j} c^{WO}_j(k)} \qquad j \in \mathcal{J}, n \in \{1, \dots, N^{WO}_j\} \tag{2}$$

$$d^{WI}_{j,u} = \frac{c^{WI}_j(u)}{\sum_{k=u}^{N^{WI}_j} c^{WI}_j(k)} \qquad j \in \mathcal{J}, u \in \{0, \dots, N^{WI}_j\} \tag{3}$$

We denote the latest possible day with a positive probability of a patient staying in the ICU and in the ward as N^I_j and $N^W_j = \max(N^{WO}_j, N^I_j + N^{WI}_j)$, respectively. Now, we calculate in Equation (4) for all specialties $j \in \mathcal{J}$ and each day $n \in \{1, \dots, N^I_j\}$ the probabilities $e^I_{j,n}$ that a patient of specialty j who had surgery on day 1 is in the ICU. Accordingly, the same is done for patients staying in the ward ($e^W_{j,n}$ for $n \in \{1, \dots, N^W_j\}$). For the probability that a patient of specialty j is in the ICU on day n we get:

$$e^I_{j,n} = \begin{cases} b_j, & n = 1 \\ (1 - d^I_{j,n-1}) e^I_{j,n-1}, & n \in \{2, \dots, N^I_j\} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

On day 1, this probability equals b_j, i.e. the probability that the patient is directly transferred to the ICU after surgery. For the following days, the probability decreases as patients might be transferred to the ward. In order to calculate $e^W_{j,n}$, we differentiate between patients who were directly transferred to the ward after leaving the OR and those who were transferred via the ICU. The probability that the patient came directly from the OR and is in the ward on day n is denoted by $e^{WO}_{j,n}$, whereas the probability that the patient is in the ward on day n after staying m days in the ICU is $e^{WI}_{j,m,n}$.

$$e^{WO}_{j,n} = \begin{cases} 1 - b_j, & n = 1 \\ (1 - d^{WO}_{j,n-1}) e^{WO}_{j,n-1}, & n \in \{2, \dots, N^{WO}_j\} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

$$e^{WI}_{j,m,n} = \begin{cases} (1 - d^{WI}_{j,0}) d^{I}_{j,m} e^{I}_{j,m}, & m \in \{1 \dots N^{I}_j\}, n = m + 1 \\ (1 - d^{WI}_{j,n-m-1}) e^{WI}_{j,m,n-1}, & m \in \{1 \dots N^{I}_j\}, \\ & n \in \{m + 2, \dots, m + N^{WI}_j\} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The calculation of $e^{WO}_{j,n}$ in Equation (5) is analogous to $e^{I}_{j,n}$. To calculate $e^{WI}_{j,m,n}$ in Equation (6), the different transfer times from the ICU are taken into account. After staying m days in the ICU ($n = m+1$), $e^{WI}_{j,m,n}$ equals the product of (a) the probability $(1 - d^{WI}_{j,0})$ that the patient did not leave the hospital immediately, (b) the probability $d^{I}_{j,m}$ that he was transferred to the ward after m days in the ICU, and (c) the probability $e^{I}_{j,m}$ that the patient was in the ICU on day m. Therefore, the probability $e^{W}_{j,n}$ that a patient is in the ward on day n is calculated in Equation (7) by adding the probability $e^{WO}_{j,n}$ that he came directly from the OR and the probabilities $e^{WI}_{j,m,n}$ that he stayed m days in the ICU before for all possible number of days $m < n$.

$$e^{W}_{j,n} = \begin{cases} e^{WO}_{j,1}, & n = 1 \\ e^{WO}_{j,n} + \sum_{m=1}^{n-1} e^{WI}_{j,m,n}, & n \in \{2, \dots, N^{W}_j\} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Now, we calculate for each day n the probability distribution for the number of patients in the ICU, $f^{I}_{j,n}(p)$, in Equation (8) and in the ward, $f^{W}_{j,n}(p)$, in Equation (9). The probability that out of k patients who had surgery, p patients are in the ICU or ward on day n can be determined using a binomial distribution [91]. Next, we have to sum these probabilities weighted by $a_j(k)$ for all possible k (number of patients that had surgery) that could lead to p patients on day n.

$$f^I_{j,n}(p) = \sum_{k=p}^{P_j} \binom{k}{p} (e^I_{j,n})^p (1 - e^I_{j,n})^{k-p} a_j(k) \tag{8}$$

$j \in \mathcal{J}, n \in \{1, \ldots, N^I_j\}$.

$$f^W_{j,n}(p) = \sum_{k=p}^{P_j} \binom{k}{p} (e^W_{j,n})^p (1 - e^W_{j,n})^{k-p} a_j(k) \tag{9}$$

$j \in \mathcal{J}, n \in \{1, \ldots, N^W_j\}$.

2.3.2 Calculation of the distributions of patients resulting from a cyclical OR block (Step 2)

As the MSS schedule is cyclical, each block will be repeated in every cycle. For example, in a weekly cycle a urological block on Monday will take place on every Monday. As the maximum recovery time of patients usually exceeds the cycle time, patients having their surgery in different cycles might be recovering at the same time. The number of overlapping cycles depends on the cycle length $L = |\mathcal{L}|$ and the maximum length of stay N^I_j for patients in the ICU and N^W_j for patients in the ward. To obtain the distributions of patients on the days of one cycle, we perform discrete convolutions — see (10) and (11) — of the patient distributions of all overlapping cycles for the ICU and the ward, respectively. We use the symbol $*$ for discrete convolution. $F^I_{j,\ell}$ ($F^W_{j,\ell}$) represents the distribution — on the ℓth day of a cycle — of the number of recovering patients of specialty j in the ICU (ward) which result from a cyclical surgery block on day 1 of all previous cycles including the current cycle.

$$F^I_{j,\ell} = f^I_{j,\ell} * f^I_{j,\ell+L} * \ldots * f^I_{j,\ell+\lfloor (N^I_j - \ell)/L \rfloor L} \qquad j \in \mathcal{J}, \ell \in \mathcal{L} \tag{10}$$

$$F^W_{j,\ell} = f^W_{j,\ell} * f^W_{j,\ell+L} * \ldots * f^W_{j,\ell+\lfloor (N^W_j - \ell)/L \rfloor L} \qquad j \in \mathcal{J}, \ell \in \mathcal{L} \tag{11}$$

2.3.3 Calculation of the distributions of patients resulting from a cyclical MSS (Step 3)

To calculate a cyclical MSS, we obtain the patient distributions coming from each block (i, q), where i denotes the operating room and q the day of the cycle. We assume that surgery blocks are only provided on weekdays. For a given MSS, x is set and each $x_{i,q,j}$ has a value of 1 if specialty j is assigned to block (i, q) and a value of 0 otherwise. $\overline{F}^{I}_{i,q,l}$ in (12) ($\overline{F}^{W}_{i,q,l}$ in (13)) is the distribution of the number of recovering patients in the ICU (ward) on day ℓ of the MSS cycle coming from surgery in OR i on day q of the MSS cycle.

$$\overline{F}^{I}_{i,q,\ell} = \begin{cases} \sum_{j\in\mathcal{J}} F^{I}_{j,\ell-q+1} x_{i,q,j}, & \ell \geq q \\ \sum_{j\in\mathcal{J}} F^{I}_{j,\ell-q+1+L} x_{i,q,j}, & \text{otherwise.} \end{cases} \tag{12}$$
$$i \in \mathcal{I},\ q \in \mathcal{Q},\ \ell \in \mathcal{L},$$
$$\overline{F}^{W}_{i,q,\ell} = \begin{cases} \sum_{j\in\mathcal{J}} F^{W}_{j,\ell-q+1} x_{i,q,j}, & \ell \geq q \\ \sum_{j\in\mathcal{J}} F^{W}_{j,\ell-q+1+L} x_{i,q,j}, & \text{otherwise.} \end{cases} \tag{13}$$
$$i \in \mathcal{I},\ q \in \mathcal{Q},\ \ell \in \mathcal{L}$$

Now we have to convolve the distributions of all blocks to obtain the patient distribution resulting from the MSS. F^{I}_{ℓ} in (14) (F^{W}_{ℓ} in (15)) denotes the distribution of recovering patients in the ICU (ward) on day ℓ of the MSS cycle. $\sup\{\mathcal{I}\}$ denotes the last operating room, $\sup\{\mathcal{Q}\}$ the last weekday with an active surgery slot.

$$F^{I}_{\ell} = \overline{F}^{I}_{1,1,\ell} * \overline{F}^{I}_{1,2,\ell} * \ldots * \overline{F}^{I}_{\sup\{\mathcal{I}\},\sup\{\mathcal{Q}\},\ell} \qquad \ell \in \mathcal{L} \tag{14}$$
$$F^{W}_{\ell} = \overline{F}^{W}_{1,1,\ell} * \overline{F}^{W}_{1,2,\ell} * \ldots * \overline{F}^{W}_{\sup\{\mathcal{I}\},\sup\{\mathcal{Q}\},\ell} \qquad \ell \in \mathcal{L} \tag{15}$$

The steps presented in this section calculate for a given MSS the distribution of patients for every day ℓ in the MSS cycle for the ICU and the ward. Note that $x_{i,q,j}$ is assumed to be set for now but will become a variable when we are searching for a good MSS. In the next sections we present methods to minimize downstream costs of an MSS using these distributions.

2.4 Generic model and discussion of objectives

In Section 2.4.1 we present a generic model that minimizes downstream costs using a general assignment problem. We then discuss different downstream cost functions for this model in Section 2.4.2.

2.4.1 Generic model

We define a generic assignment problem — i.e. not considering hospital specifics — that minimizes the downstream costs $c(x)$. $c(x)$ is a function of the distribution of patients in the downstream units calculated in steps 1 to 3 in the previous section resulting from the MSS x, i.e. the assignment of all blocks (i, q) to a specialty j. The formulation of the generic model contains constraints (17)-(20) to ensure that no more than one specialty is assigned to any surgery block, that the sum of required blocks d_j for every specialty j does not exceed the available number of blocks, and that the maximum number of blocks of specialty j on day q is s_{qj}, e.g. the number of available surgeons.

$$\text{Min } c(x) \tag{16}$$

s.t.

$$\sum_{j\in\mathcal{J}} x_{i,q,j} \leq 1 \qquad i \in \mathcal{I}, q \in \mathcal{Q} \tag{17}$$

$$\sum_{i\in\mathcal{I}} \sum_{q\in\mathcal{Q}} x_{i,q,j} \geq d_j \qquad j \in \mathcal{J} \tag{18}$$

$$\sum_{i\in\mathcal{I}} x_{i,q,j} \leq s_{qj} \qquad q \in \mathcal{Q}, j \in \mathcal{J} \tag{19}$$

$$x_{i,q,j} \in \{0, 1\} \qquad i \in \mathcal{I}\ q \in \mathcal{Q}, j \in \mathcal{J}. \tag{20}$$

Equations (17) and (18) are the assignment problem constraints. Equation (17) ensures that at most one specialty is assigned to each block, while (18) ensures that the number of blocks assigned to each specialty equals the number of required blocks. The maximum number of blocks of each specialty per day is modeled in (19). Equation (16) represents a generic objective function.

This model can easily be adjusted to deal with specific constraints, e.g. some specialties have to operate in specific ORs.

2.4.2 Discussion of downstream cost functions

We define four cost components that drive downstream costs: fixed costs, overcapacity costs, staffing costs, and additional weekend staffing costs. Our main goal is to incorporate cost estimates based on patient distributions in the generic model to obtain optimized MSSs.

Fixed costs. We consider the costs for creating and maintaining fixed capacities. We define $c^{I,f}$ and $c^{W,f}$ as the costs for creating and maintaining the capacity for one patient in the ICU and the ward per cycle, respectively. An example of fixed costs is the costs associated with an ICU bed. The model determines the required capacity of these resources to ensure certain service levels α^I and α^W. We denote $Q^I_\ell(\alpha^I)$ as the α^I-quantile of the distribution F^I_l of the number of patients in the ICU on day ℓ. $Q^W_\ell(\alpha^W)$ is the α^W-quantile for the distribution F^W_l of patients in the ward. For example, $Q(.99)$ denotes the capacity that will not be exceeded with a probability of 99%. The number of beds we need to provide in the ICU and in the ward are therefore $cap^I(\alpha^I) = \max_{\ell \in \mathcal{L}}(Q^I_\ell(\alpha^I))$ and $cap^W(\alpha^W) = \max_{\ell \in \mathcal{L}}(Q^W_\ell(\alpha^W))$, respectively. We obtain total fixed costs of

$$cost^f = c^{I,f} cap^I(\alpha^I) + c^{W,f} cap^W(\alpha^W). \tag{21}$$

Overcapacity costs. Overcapacity costs are costs that incur due to requiring capacity beyond cap^I and cap^W. This situation occurs, for example, when patients must be transferred to ICUs or wards in other hospitals as capacity limits, depending on the service levels α^I and α^W, are reached. We assign costs of $c^{I,o}$ and $c^{W,o}$ for each patient above existing capacities per day. The expected number of these patients per day is $exc^I(\alpha^I) = \sum_{\ell=1}^{Q} \sum_{p=cap^I(\alpha^I)+1}^{UB^I} pF^I_\ell(p)$ in the ICU and $exc^W(\alpha^W) = \sum_{\ell=1}^{Q} \sum_{p=cap^W(\alpha^W)+1}^{UB^W} pF^W_\ell(p)$ in the ward. UB^I and UB^W is an upper bound of the number of patients that request a bed in the ICU and the ward, respectively. A simple upper bound is the product of the number of overlapping cycles and the maximum number of patients per cycle. We obtain overcapacity costs of

$$cost^o = c^{I,o} exc^I(\alpha^I) + c^{W,o} exc^W(\alpha^W). \tag{22}$$

Both the fixed costs and the overcapacity costs depend on the service levels α^I and α^W. The higher the service level is, the higher the fixed costs and the lower the overcapacity costs are. Setting the appropriate service levels α^I and α^W should be done on a strategic level and is not treated in this study. A discussion of the trade-off between idle capacities (i.e. fixed costs) and waiting time (i.e. overcapacity costs) can be found in Patrick and Puterman [68].

Staffing costs. Staffing costs are dependent on the number of patients, i.e. occupied beds. The staffing decision for every bed is made in advance. Therefore, we assume a service level of β^I and β^W for staffing beds. For example, a hospital might staff the .75 quantile of demand to be understaffed no more than 25 % of the time. The consideration of a constant nurse-to-patient ratio would lead to a step-function and an integer number of solutions. However, this would significantly increase the complexity of the model. For simplicity, we assume the costs for staffing one bed per day are constant with $c^{I,s}$ for the ICU and $c^{W,s}$ for the ward. The total number of beds to be staffed during one cycle in the ICU and in the ward are therefore $sta^I(\beta^I) = \sum_{\ell \in \mathcal{L}} Q^I_\ell(\beta^I)$ and $sta^W(\beta^W) = \sum_{\ell \in \mathcal{L}} Q^W_\ell(\beta^W)$, respectively. The staffing costs, $cost_s$, with standard wages for all days are

$$cost^s = c^{I,s} sta^I(\beta^I) + c^{W,s} sta^W(\beta^W). \tag{23}$$

Weekend staffing costs. Usually, there are additional costs for staffing beds on weekends. The additional costs for one bed per day are $c^{I,we}$ and $c^{W,we}$. The total number of beds to be staffed on the weekends of one cycle in the ICU and the ward are therefore $sta^{I,we}(\beta^I) = \sum_{\ell \in \mathcal{L} \setminus \mathcal{Q}} Q^I_\ell(\beta^I)$ and $sta^{W,we}(\beta^W) = \sum_{\ell \in \mathcal{L} \setminus \mathcal{Q}} Q^W_\ell(\beta^W)$, respectively. The additional costs on weekends are

$$cost^{we} = c^{I,we} sta^{I,we}(\beta^I) + c^{W,we} sta^{W,we}(\beta^W). \tag{24}$$

Many combinations of downstream costs are possible. In our case study, we employ downstream costs of $c(x) = cost^f + cost^{we}$. The resulting objective function is

$$\text{Min } c^{I,f} cap^I + c^{W,f} cap^W + c^{I,we} sta^{I,we} + c^{W,we} sta^{W,we}. \quad (25)$$

2.5 Solution algorithms

The generic model presented in the previous section is a classical assignment problem. Although the assignment problem is well-known to be NP-hard, there are efficient procedures, such as branch-and-bound, to solve even large instances to optimality. However, the calculation of the objective function value is, due to the convolution of distributions as carried out in Section 2.3, quite extensive. Hence, in addition to an optimal branch-and-bound procedure, we discuss two different heuristic approaches to solve the master surgery scheduling problem:

1. Exact objective function and heuristic solution method
2. Approximated objective function and exact solution method

Relating to 1, we apply an incremental improvement heuristic, a 2-Opt heuristic, and simulated annealing. Relating to 2, we consider two approximated objective functions: the first uses expected values only, while the second employs a combination of expected values and variances. The last two approaches show some similarities to Beliën and Demeulemeester [8], who minimize expected shortage of ward beds by linearization of their problem.

2.5.1 Straightforward branch-and-bound

The straightforward branch-and-bound (SBB) algorithm is based on complete enumeration but avoids redundant symmetrical solutions. These could be caused by having different combinations of the same specialties on the same day in different ORs. The algorithm fills up block after block of the

MSS using a depth-first search. It assigns all blocks, i.e. combinations of days q and operating rooms i, to specialties j starting with the specialty with the lowest index. After each block of a day is assigned to a specialty, the next day is started. To avoid redundant solutions, remaining blocks on the same day will only be filled with specialties with the same or a higher index. An example for the solutions is presented in Figure 2.5. Here, we show for 5 blocks (1 day with 1 OR, 2 days with 2 ORs) and 3 specialties (specialty 1 and 2 with 2 blocks each, specialty 3 with one block) all 11 possible non-redundant solutions (compared to a total of $\frac{5!}{2!2!} = 30$ solutions to assign these 3 specialties to 5 blocks). As an example, the solution 3 shows an assignment of the two required blocks of specialty 1 to OR 1, days 1 and 3. The two required blocks of specialty 2 are assigned to operating rooms 1 and 2 on day 2, and the block of specialty 3 is assigned to operating room 2 on day 3.

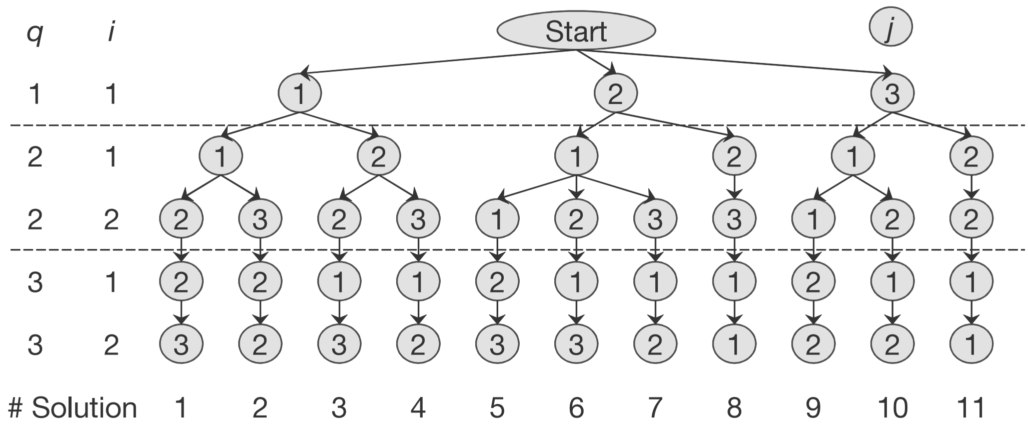

Figure 2.5: All non-redundant solutions for an example with 5 blocks and 3 specialties

After assigning a specialty to a block the algorithm updates the distributions of patients in the ICU and the ward. A lower bound of the objective function, leaving the remaining blocks empty, is computed and compared with an upper bound, which is the best known feasible solution. A good first upper bound may be obtained by simulated annealing, as detailed in the following section. A partial solution is fathomed as soon as its lower bound is not strictly smaller than the upper bound. If a new feasible solution is obtained after assigning the last block to a specialty is better than the current upper bound, the upper bound is updated. We present the example of Figure 2.5 with upper and lower bounds, fathoming of non-optimal solutions and the optimal solution (dark with white numbers) in Figure 2.6. While this method is exact, it may only be applied to small problem instances.

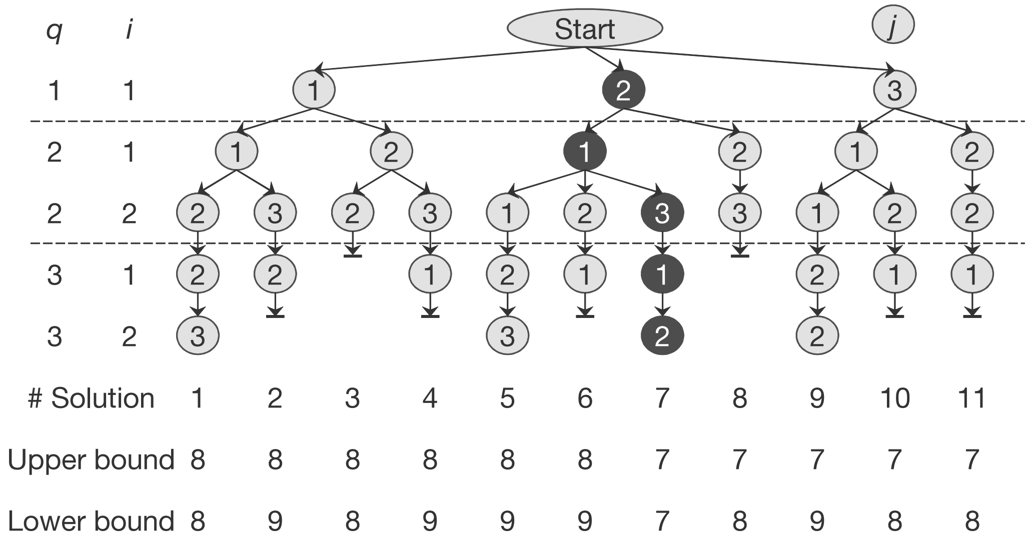

Figure 2.6: Optimal solution for an example with 5 blocks and 3 specialties

2.5.2 Exact objective function and heuristic solution method

Incremental improvement heuristic. The incremental improvement heuristic (IIH) is motivated by the way MSSs are altered in practice. Usually, an MSS already exists and the hospital is not willing to allow many changes, since the MSS affects many upstream and downstream departments (e.g. the outpatient clinic). The proposed heuristic will seek the best option if only one swap of two blocks is allowed. Therefore, it will realize the swap with the maximum incremental improvement. This method may be used to show improvements for a defined maximum number of swaps.

2-Opt heuristic. We repeat the incremental improvement heuristic until no further improvement of the objective function is observed. In this case it is equivalent to the 2-Opt (2OH) approach known from the traveling salesman literature [52].

Simulated annealing. IIH and 2OH presented above are quite likely to get stuck in a local optimum. Its results are sensitive to the starting point. To overcome this weakness we propose a simulated annealing approach (SA) with the same neighborhood as IIH and 2OH. Contrary to IIH, the SA might find solutions that are completely different from the starting solution. The SA will accept every move, i.e. swap of two blocks, that improves the objective function. A swap causing an increase in the objective function will be accepted with a probability which decreases over time. We implement the SA using a geometric cooling schedule with $t_k = cf \cdot t_{k-1}$, where cf denotes the cooling factor and t_k the temperature level in round k. The lower it is, the faster the cool down occurs and thus the faster the SA terminates. More details of the algorithm are given in Section 2.6.

2.5.3 Approximated objective function and exact solution method

Heuristic with approximated objective function based on expected values. We approximate the quantiles Q^I_ℓ and Q^W_ℓ used in the exact objective function by their expected values, $E(F^I_{j,l})$ for the ICU and $E(F^W_{j,l})$ for the ward, multiplied with parameters $a^{I,E}$ and $a^{W,E}$ for fixed capacities and $b^{I,E}$ and $b^{W,E}$ for weekend staffing, respectively. We estimate the parameters by the average quotient of the respective quantiles $Q(\cdot)$ and the expected values $E(\cdot)$ of the starting solution. Table 2.2 states the approximated quantiles. We denote the heuristic using expected values as EV.

As defined in the previous section, the objective function is

$$\text{Min } c^{I,f} cap^I + c^{W,f} cap^W + c^{I,we} sta^{I,we} + c^{W,we} sta^{W,we}, \tag{26}$$

where cap^I and cap^W denote the capacity levels, $sta^{I,we}$ and $sta^{W,we}$ the cumulated beds to be staffed on weekends during one MSS cycle for the ICU and the ward, respectively. For the approximation of the objective function,

Quantile	Exact Model	EV Heuristic
Patients in the ICU relevant for fixed capacities	$Q^I_\ell(\alpha^I)$	$a^{I,E}E(F^I_\ell)$
Patients in the ward relevant for fixed capacities	$Q^W_\ell(\alpha^W)$	$a^{W,E}E(F^W_\ell)$
Patients in the ICU relevant for weekend staffing	$Q^I_\ell(\beta^I)$	$b^{I,E}E(F^I_\ell)$
Patients in the ward relevant for weekend staffing	$Q^W_\ell(\beta^W)$	$b^{W,E}E(F^W_\ell)$

Table 2.2: Approximations for EV

the constraints (27) to (29) need to be added to the generic model presented in Section 2.4.1. We only show the constraints for the ICU, the ones for the ward are formulated analogously.

$$E(F^I_\ell) = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{q=1}^{\ell} E(F^I_{j,\ell-q+1})x_{i,q,j} + \tag{27}$$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}}\sum_{q=\ell+1}^{L} E(F^I_{j,\ell-q+1+L})x_{i,q,j} \qquad \ell\in\mathcal{L}$$

$$a^{I,E}E(F^I_\ell) \le cap^I \qquad \ell\in\mathcal{L} \tag{28}$$

$$\sum_{\ell\in\mathcal{L}\setminus\mathcal{Q}} b^{I,E}E(F^I_\ell) = sta^{I,we} \tag{29}$$

In (27), the values for the expected number of patients in the ward are determined for each day. In (28) the required capacity for the ward is calculated. Finally, in (29) the number of patients per weekend day relevant for staffing is determined.

Heuristic with approximated objective based on expected values and variances. The EV neglects the distribution of patients as it only considers the expected values. With the heuristic with approximated objective based on expected values and variances (EVV), we assume the distributions of patients to be normally distributed and approximate the quantiles using the expected value and the approximated standard deviation. To avoid a square root function, the standard deviations $SD(F_l^I)$ and $SD(F_l^W)$ are approximated by a linear function of the variance ($V(F_{j,l}^I)$ for the ICU and $V(F_{j,l}^W)$ for the ward). We employ one linear factor for the ICU, sr^I in Constraint (31), and one for the ward, sr^W. We use the factors that minimize the squared errors for the variances of the starting solutions. The z-values for the quantiles are $z^{I,cap}$ and $z^{W,cap}$ for the capacity levels and $z^{I,sta}$ and $z^{W,sta}$ for the weekend staffing levels. Table 2.3 states the approximated quantiles.

Quantile	Exact Model	EVV Heuristic
Patients in the ICU relevant for fixed capacities	$Q_\ell^I(\alpha^I)$	$E(F_\ell^I) + z^{I,cap} SD(F_\ell^I)$
Patients in the ward relevant for fixed capacities	$Q_\ell^W(\alpha^W)$	$E(F_\ell^W) + z^{W,cap} SD(F_\ell^W)$
Patients in the ICU relevant for weekend staffing	$Q_\ell^I(\beta^I)$	$E(F_\ell^I) + z^{I,sta} SD(F_\ell^I)$
Patients in the ward relevant for weekend staffing	$Q_\ell^W(\beta^W)$	$E(F_\ell^W) + z^{W,sta} SD(F_\ell^W)$

Table 2.3: Approximations for EVV

The objective function (25), the assignment problem constraints (17)-(20), and the constraints to determine the expected values (27) stay unchanged. Constraints to determine the variances (30) and the approximated standard deviations (31) need to be added. The constraints determining the capacities (32) and the beds to be staffed at weekends (33) had to be changed. The approximation of the square root function to determine the standard deviation in (31) can be carried out in many ways. The most simple one is to use a linear function. This method works well if there is no big difference in the possible variances. For larger differences in the variances, a piecewise linear function as described in van Essen et al. [87] may also be applied. Again, we only present the constraints for the ICU (30)-(33), the constraints for the ward are formulated analogously.

$$V(F_\ell^I) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{q=1}^{\ell} V(F_{j,\ell-q+1}^I) x_{i,q,j} + \tag{30}$$

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{q=\ell+1}^{L} V(F^I_{j,\ell-q+1+L}) x_{i,q,j} \qquad \ell \in \mathcal{L}$$

$$sr^I V(F^I_\ell) = SD(F^I_\ell) \qquad \ell \in \mathcal{L} \qquad (31)$$

$$E(F^I_\ell) + z^{I,cap} SD(F^I_\ell) \leq cap^I \qquad \ell \in \mathcal{L} \qquad (32)$$

$$\sum_{\ell \in \mathcal{L} \setminus \mathcal{Q}} E(F^I_\ell) + z^{I,sta} SD(F^I_\ell) = sta^{I,we} I \qquad (33)$$

2.6 Case study

We tested all algorithms with data used in Vanberkel et al. [91]. Additional data for the ICU was created based on expert interviews with Dutch hospital managers. The MSS cycle has a length of two weeks and there are seven specialties. There is no limit to the number of blocks of any specialty on any given day other than the number of ORs. We distinguish three examples:

- A small MSS with only one OR and a second OR on Wednesdays. The number of OR blocks is therefore 12.
- A medium MSS with three ORs (30 blocks).
- A large MSS with nine ORs (90 blocks).

The percentage of blocks of all specialties is approximately equal. Although in reality some hospitals can have more than 100 ORs, we assume that in most cases no more than 9 ORs will be grouped together in a common MSS. As discussed in the previous sections, the downstream costs to be minimized are the fixed costs $cost^f$ and the additional weekend staffing costs $cost^{we}$. The values of the cost parameters are the outcome of interviews with OR managers and shown in Table 2.4.

To obtain comparable starting solutions for all cases, we employ a simple construction heuristic. We begin with an empty MSS and start at the first day with filling the slots with the first specialty until all required slots have been assigned. Then we continue with the remaining specialties. As we cannot compute the optimal solution for the medium and the large case,

Description	Notation	Value
Service level for fixed capacities	α^I, α^W	0.99
Fixed costs ICU bed/two weeks	$c^{I,f}$	€ 5,000
Fixed costs ward bed/two weeks	$c^{W,f}$	€ 500
Service level for staffing	β^I, β^W	0.75
Additional costs staffing ICU bed at weekends/day	$c^{I,we}$	€ 700
Additional costs staffing ward bed at weekends/day	$c^{W,we}$	€ 120

Table 2.4: Input parameters

we use the starting solution as a reference point. We compare the simple branch-and-bound (SBB), the incremental improvement heuristic (IIH), the 2-Opt heuristic (2OH), simulated annealing (SA), the heuristic with approximated objective function based on expected values (EV), and the heuristic with approximated objective based on expected values and variances (EVV). For the IIH we continue swapping blocks until a maximum of one third of all blocks are swapped. For the simulated annealing (SA), a cooling factor of $cf = 0.9$ is chosen as proposed by Aarts et al. [1]. We configured the SA with 1,500 iterations for each temperature level, an initial temperature of $t_0 = 9,000$, and a stopping criterion of 6 consecutive temperature levels without a new best solution. For each case we compare computation time, total cost, relative improvement of the starting solution, and the percentage of changed blocks of the tested algorithms.

Algorithm	Computation time	Total costs	Improve-ment	Changed blocks (%)
Starting Solution		72,500		
SBB	13,319s	64,500	11.0%	67%
IIH	5s	68,600	5.4%	33%
2OH	7s	68,600	5.4%	33%
SA	260s	65,000	10.3%	92%
EV	1s	69,480	4.2%	83%
EVV	1s	69,620	4.0%	67%

Table 2.5: Results small case

A summary of the results of the 12 block example can be found in Table 2.5. In this example the IIH has swapped four blocks, which results in an improvement of 5.4% compared to the starting solution. The 2OH reaches the same solution. Therefore, after swapping four blocks no swap of two blocks improves the solution. The SA takes longer, changes more blocks but reaches a much better solution than the 2OH (an additional 5% in cost savings is gained). The two heuristics using an approximated objective function are very fast (around 1 second), but achieve improvements of about 4% only. The optimal SBB shows that a maximum improvement of 11% is possible.

Algorithm	Computation time	Total costs	Improvement	Changed blocks (%)
Starting Solution		153,120		
SBB	N/A	N/A	N/A	N/A
IIH	145s	137,860	10.0%	33%
2OH	481s	136,140	11.1%	60%
SA	685s	136,500	10.9%	80%
EV	6s	136,000	11.2%	73%
EVV	6s	136,000	11.2%	87%

Table 2.6: Results medium case

Table 2.6 provides the results for the example with 30 blocks where all heuristics achieve improvements of at least 10%. EV and EVV approximating the objective functions work considerably better than in the small example. Not only are they much faster than the other heuristics, but they also achieve the best results of more than 11% improvement. However, they change between 73 and 87% of the blocks of the initial MSS.

The results of the large example with 90 blocks (see Table 2.7) show greatly increasing computing times. The heuristics EV and EVV, which solve the approximative objective exactly, show the highest computational times. However, at the same time they achieve the highest improvements of all tested methods. IIH and 2OH suffer from the large problem size as well, since for each swap all possibilities have to be calculated. In fact, in this example for each iteration more than 4,000 possible swaps have to be evaluated. The only heuristic that requires less than one hour is the SA with

Algorithm	Computation time	Total costs	Improvement	Changed blocks (%)
Starting Solution		397,680		
SBB	N/A	N/A	N/A	N/A
IIH	11,902s	368,620	7.3%	28%
2OH	31,101s	365,260	8.2%	39%
SA	1,832s	361,300	9.1%	76%
EV	107,826s	358,480	9.9%	82%
EVV	181,954s	358,560	9.8%	81%

Table 2.7: Results large case

an average improvement of 9.1%, a higher achievement than those of the IIH and the 2OH.

Regarding these results we draw three conclusions. First, there is only a small potential in improving MSSs with a small number of blocks. Second, while IIH and 2OH work fairly well for small problems, they are outperformed by the SA, the EV and the EVV heuristics in the large case. Nevertheless, IIH and 2OH could still be of interest for the case where the majority of the blocks of the initial MSS should be kept. Third, most algorithms discussed in this chapter require long computation times for the large example. Only the SA requires computational time of less than one hour. EV and EVV could be run using a time limit. Even if the optimal solution of the approximated objective function might not been found within the time limit, good results can be achieved after a relatively short time period. However, as we are discussing a tactical problem, relatively high computational times of up to a few hours can be tolerated in practical settings.

2.7 Conclusion

In this chapter we presented an algorithm for calculating the exact distributions of patients both in the ICU and the ward resulting from a given cyclical MSS. We further discussed measures as fixed capacities and staffing levels to estimate the downstream costs of an MSS and proposed algorithms to find an MSS with minimized costs. We considered multiple heuristics. Two simple heuristics that swap MSS blocks, a simulated annealing algorithm that finds good solutions in a reasonable period of time, and a simple branch-and-bound, that only works in realistic time for small problems. We further tested optimal solution methods with approximated objective functions. They showed excellent results for medium and large instances, but required long computation times for large instances.

For large instances there is further room for research on heuristics. For example, one could investigate on a combination of the approximated objective function, such as EV or EVV, with a non-optimal solution method to gain satisfactory results within reasonable time. Furthermore, there are many possible extensions. Upstream units like the outpatient clinic can be incorporated as surgeons work there too, so scheduling in both departments could be coordinated. Effects on the post-anesthesia care unit could also be incorporated. Operations on weekends (for emergency patients only) as well as pre-operative stays in ICUs and wards or patients with no surgery could be included. Moreover, for practice, relevant constraints such as differently equipped ORs, minimum time between blocks of the same specialty, etc. may be considered. The algorithms and heuristics proposed in this chapter can be adapted to these extensions.

Summarizing our findings, we conclude there is significant potential in cost savings and quality improvements in considering downstream units when designing tactical operating room schedules. Accounting for weekend staffing and leveling bed requests may further contribute to employee satisfaction and decrease negative medical effects.

Chapter 3

Improving ICU and ward utilization by adapting master surgery schedules: A case study

3.1 Introduction

In the last years the pressure on hospitals to work in a cost efficient way has constantly increased [42]. Medical progress, an aging population and financial restrictions are forcing hospitals to review each individual process in the treatment of patients. By analyzing hospital expenses it becomes obvious that the bulk of costs is generated by surgical interventions [22]. Consequently many studies deal with organizational improvements concerning the cost-intensive operating room (OR) departments [38, 18, 41]. While the management of the OR is vital for hospital performance, it should not be considered in isolation [89]. Among other things, the degree of OR utilization, which is an important goal in many hospitals [86], has a considerable effect on downstream units [23]. In particular, most admissions of patients to the intensive care unit (ICU) and to the wards result from the OR workload. Approaches solely aiming at maximizing the degree of OR utilization can lead to a high variation in the occupancy of both the ICU and the regular wards [7, 60]. Furthermore, Anderson et al. [4] showed an increased hospital discharge rate when the utilization of the OR is high. Baker et al. [5] revealed that the readmission rate within 72h after discharging from the ICU, following days of high patient influx, is significantly increased. The variability of scheduled caseload, in particular, should be reduced to avoid high stress on the ICU [60]. In addition, periods of high patient demand are always associated with triage decisions that impact patients' outcome and organizational efforts in the post-operative units [79]. Consequently, a smoother influx of postoperative patients to the downstream units should be taken into consideration when creating an OR schedule. The most common scheduling technique is the use of Master Surgery Schedules (MSSs) [88, 90]. In an MSS surgical specialties, e.g. urology, are assigned to time slots, e.g. days, in specific ORs. An MSS can be considered as an essential part of the tactical OR planning and can improve resource utilization and patient flow within the hospital [88]. Vanberkel et al. [90] describe a successful implementation of an improved MSS in a comprehensive cancer center, where the effects of the MSS on the regular wards were the main focus.

Expanding the work of Vanberkel et al. [91], our study considers the capacity utilization of the highly cost-intensive ICU as well. It is based on the algorithm presented in Fügener et al. [34] (see Chapter 2), but extending it to account for emergency surgeries on weekends and applying it with real life data from a German University hospital. This work aims specifically at analyzing the effects of employing different MSS scenarios on both the ICU and ward from a holistic perspective. Our model determines for different MSSs the probabilities for bed requests in both units.

3.2 Methods

Patient flow information which serve as a data base for our model are taken from the "Klinikum München rechts der Isar" (MRI), a tertiary care university hospital containing 1,100 beds serving all surgical disciplines (except heart surgery). The hospital runs 36 ORs on a daily basis. The ORs are located in several decentralized operating room suites. Due to interdisciplinary ORs, the so called OR-suite "Zentral-OP-2 (ZOP2)" as a functional autonomously operating sub-unit allowed an optimized MSS. Consequently, we focused solely on this unit with eight ORs. In this OR-suite the department of neurosurgery (NCS) runs 3 ORs, urology (URS) 3 ORs and sport orthopedics (SP) 2 ORs. One OR block of capacity represents a full day of operating time and the MSS (see Table 1) is repeated on a weekly basis. All ORs are open from Monday to Friday — during the weekend and on holidays only emergency operations are performed. The emergency surgeries on weekends can be represented within an MSS by introducing a fourth imaginary surgical specialty “WE” that comprises all cases during the weekend. In this study, we pool the ICU and ward bed capacities of these specialties to a single ICU and ward, respectively.

The historical data utilized for the following analysis has been extracted from the MRI hospital information system and exported to two Excel-files. One file contained all OR movements, whereas the other provided insight into the patients' ICU and ward movements afterwards. Based on the patient ID, the records were merged. Data processing was performed using MySQL 5, a relative database management system and Python 2.7, an object oriented programming language. For our analysis, we considered data from the first half of 2010. During this period, 2,480 patients were surgically treated in ZOP2, resulting in 2,690 surgeries. Regarding these figures, it is obvious that — in contrast to the main patient flow pattern — some patients have been operated on more than once. Whenever a patient had to be operated more than once, we generated a separate record for each surgery of this patient. Of course, this method involves losing information about pre-operative stays and their influence on bed occupancy, but these influences are neglected for the purpose of our study.

The mathematical concept with all equations and more rigorous explanations is provided in more detail Chapter 2. Therefore, we only briefly review the basic principles and present short explanations. Our model considers historical data to calculate the ICU and ward occupancy resulting from a given MSS. Analyzing the historical data, the patient flow follows the subsequent patterns as shown in Figure 3.1, where the percentages given on the left correspond to the percentage of patients who follow the flow. A closer look at the patient flow in the MRI shows that after surgery most patients will either be sent directly to the ward or first to the ICU before being transferred to the ward. Fewer patients leave the hospital directly after their stay in the ICU,

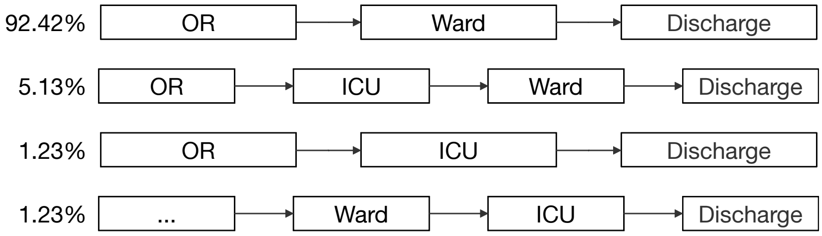

Figure 3.1: Visualization of possible patient flows

or are transferred to the ICU after first staying in the ward. The first three types of patient paths are represented by the model in Figure 3.2, whereas any patient flow from the ward to the ICU are not considered. As the figures

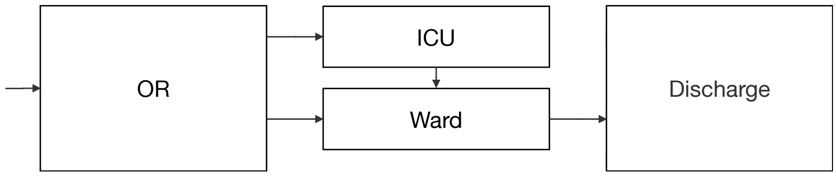

Figure 3.2: Possible transitions of patient movings after a surgery [34] (see Chapter 2)

demonstrate, such patient flows occur less often. They usually indicate a complication of a patient's health condition. Taking into account the small number of cases, we rearranged the chronological order of the ICU and ward stays, i.e. all ICU stays are aggregated right after the surgery, independently of their real chronological occurrence. For each specialty we use the following as input:

- The probability distribution of the number of surgeries per OR slot (Figure 3.3)
- The probability for patients to be sent to the ICU directly after surgery (Figure 3.4)
- The probability distributions of the number of days a patient has to recover in the ICU, in the ward after surgery, and in the ward after having already occupied a bed in the ICU (Figures 3.5-3.7)

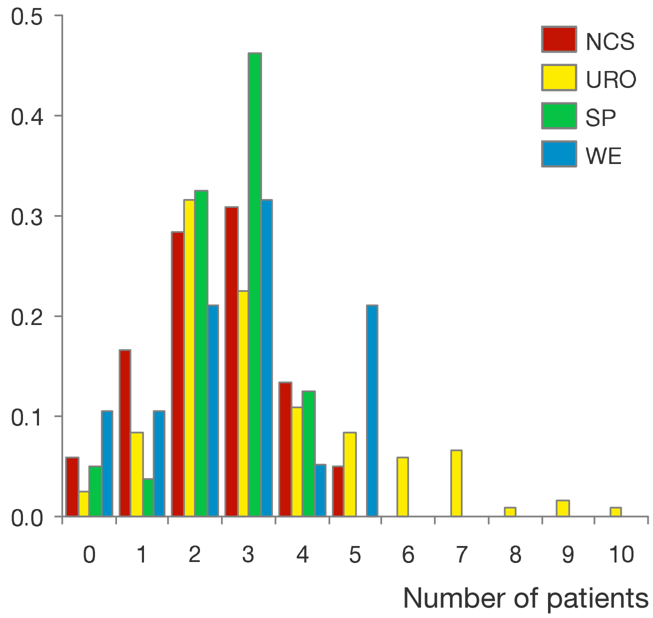

Figure 3.3: Probability distribution of number of patients per OR block

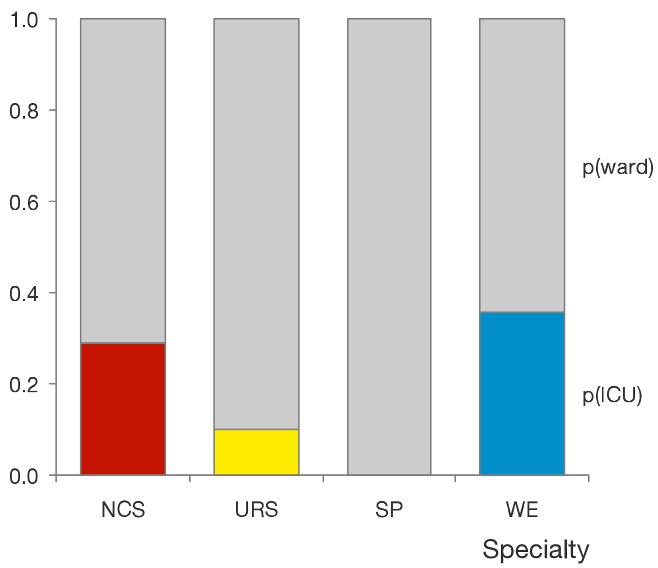

Figure 3.4: Probability to be sent to ICU after surgery

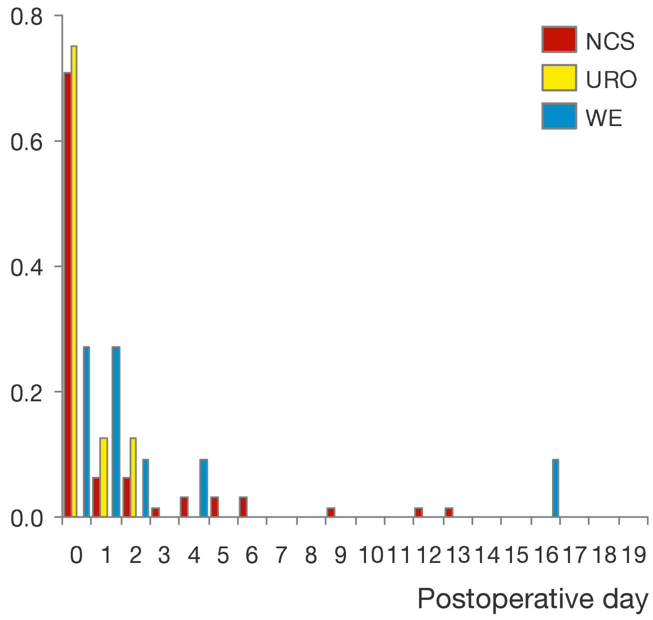

Figure 3.5: Probability distribution of length of stay in the ICU (days)

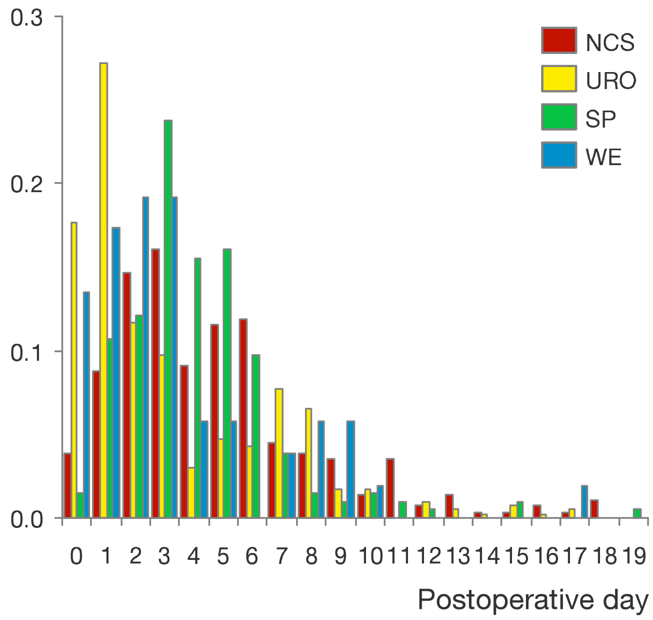

Figure 3.6: Probability distribution of length of stay in the ward (days)

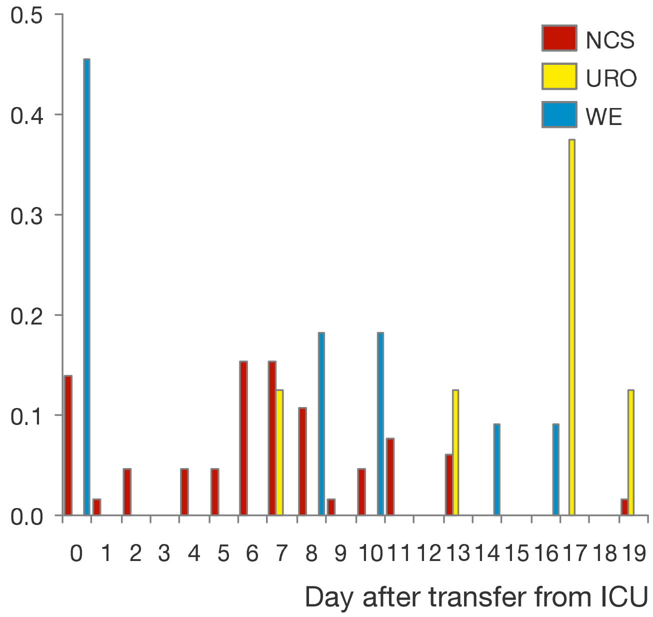

Figure 3.7: Probability distribution of length of stay in the ward after being transferred from the ICU (days)

The model to determine the occupancy levels in the ICU and the ward works in three steps. In step 1, the probabilities, that p patients of specialty j are still in the ICU or ward n days after surgery are calculated for a single surgery block of specialty j. For this, we used a so called Markov chain to model the patient flow. The model is illustrated in Figure 3.8, each arrow represents a possible transition of a patient. We use these transition probabilities and the binomial distribution to calculate numerical values for the probabilities that p patients of specialty j are in the ICU or ward n days after the surgery. In step 2, we consider that each OR block is repeated weekly. Therefore, on each day of a week patients coming from OR blocks from different weeks can be recovering in the ICU or ward at the same time. We incorporate this by convolving the distributions from step 1. This can be done as the sum of independent random variables is computed by convolving their probability distributions. After step 2, we obtain the probabilities that p patients are in the ICU or ward the ℓth day of a week for a weekly OR block of specialty j assigned to the first day of the week. In step 3, we convolve the distributions for all blocks of the final MSS. As a result, we obtain the

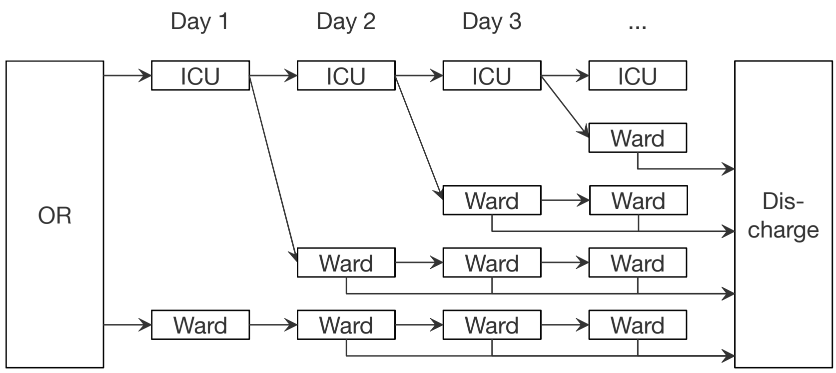

Figure 3.8: Modeling the patient flow by a Markov chain [34] (see Chapter 2)

probabilities that p patients are in the ICU or ward the ℓth day of a week resulting from the MSS. These probabilities may be directly used to calculate exact quantiles for ICU and ward bed occupancy.

3.3 Results

We tested three MSSs using our algorithm. MSS A represents the status quo at MRI, where each specialty is assigned to specific ORs for all days of the week (see Figure 3.9). MSS B, presented in Figure 3.10, is designed to avoid weekend bed requests and therefore assigns all ORs on the first days to NCS, where patients had the longest average stays, and on the last days to SP, where patients had short average stays and no ICU patients. MSS C, presented in Figure 3.11, is a mix between MSS A and MSS B to avoid peaks in ICU and ward utilization as well as high weekend occupancy levels. Here, NCS is assigned to three ORs on all days of the week as in MSS A. The remaining ORs are assigned to URS in the beginning and to SP towards the end of the week.

Figures 3.12-3.14 illustrate the quantiles of the three MSSs. The red, green and blue lines represent the 25%, 50% and 75% quantile of occupied beds on the given day of the week - i.e. the real capacity utilization should not exceed these limits in more than 75%, 50% and 25% of all occurring cases, respectively. At MRI, the 75% quantile was considered relevant for staffing

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
OR1	SP	SP	SP	SP	SP		
OR2	SP	SP	SP	SP	SP		
OR3	URS	URS	URS	URS	URS		
OR4	URS	URS	URS	URS	URS	WE	WE
OR5	URS	URS	URS	URS	URS		
OR6	NCS	NCS	NCS	NCS	NCS		
OR7	NCS	NCS	NCS	NCS	NCS		
OR8	NCS	NCS	NCS	NCS	NCS		

Figure 3.9: Master Surgery Schedule A

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
OR1	NCS	NCS	URS	URS	SP		
OR2	NCS	NCS	URS	URS	SP		
OR3	NCS	NCS	URS	URS	SP		
OR4	NCS	NCS	URS	URS	SP	WE	WE
OR5	NCS	NCS	URS	URS	SP		
OR6	NCS	NCS	URS	URS	SP		
OR7	NCS	NCS	URS	SP	SP		
OR8	NCS	URS	URS	SP	SP		

Figure 3.10: Master Surgery Schedule B

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
OR1	NCS	NCS	NCS	NCS	NCS		
OR2	NCS	NCS	NCS	NCS	NCS		
OR3	NCS	NCS	NCS	NCS	NCS		
OR4	URS	URS	URS	SP	SP	WE	WE
OR5	URS	URS	URS	SP	SP		
OR6	URS	URS	URS	SP	SP		
OR7	URS	URS	URS	SP	SP		
OR8	URS	URS	URS	SP	SP		

Figure 3.11: Master Surgery Schedule C

decisions. Hence, we only discuss this occupancy level in detail. The quantiles of MSS A are presented in Figure 3.12. In the ICU there is a constant occupation of 6 beds from Monday until Friday. On Saturday the occupation diminishes by 1 bed to 5 beds. In the ward there is an increasing demand for beds during the weekdays until Friday, where peak occupation reaches 123 beds. During the weekend, demand decreases to 107 beds on Saturday and 91 beds on Sunday.

Figure 3.12 shows the results after performing the scenario analysis for MSS B, where the ICU peak occupation of 8 beds occurs on Monday and Tuesday and decreases until Thursday to 4 beds. Peak occupation of ward beds occurs on Thursday, where 125 beds are requested in the 75% quantile. This number decreases to 103 beds on Saturday and 90 beds on Sunday.

Introducing MSS C reveals a constant number of 6 occupied ICU beds which decreases to 5 on Thursday (Figure 3.14). The maximum number of occupied ward beds is 115 on Wednesday and Friday. On Saturday and Sunday the number of bed requests is 105 and 94, respectively.

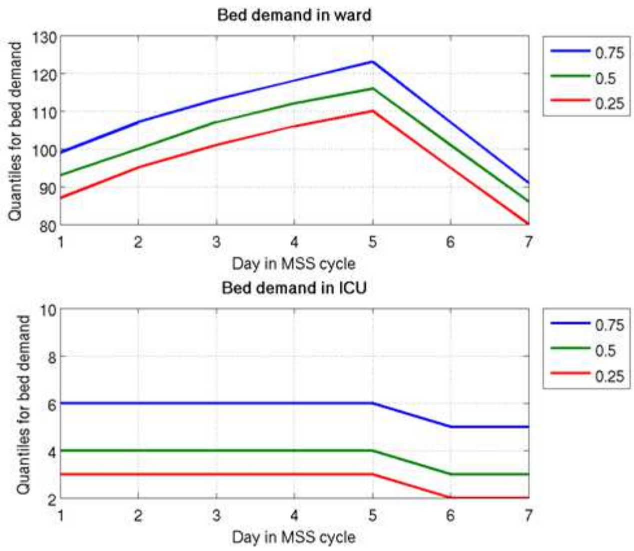

Figure 3.12: Quantiles of bed demand for Master Surgery Schedule A

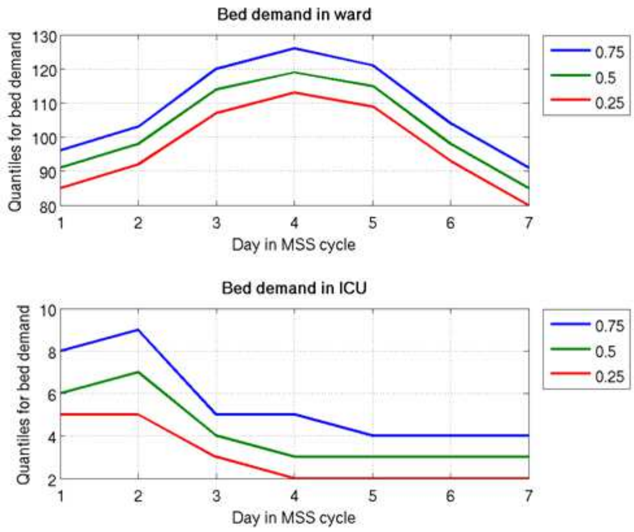

Figure 3.13: Quantiles of bed demand for Master Surgery Schedule B

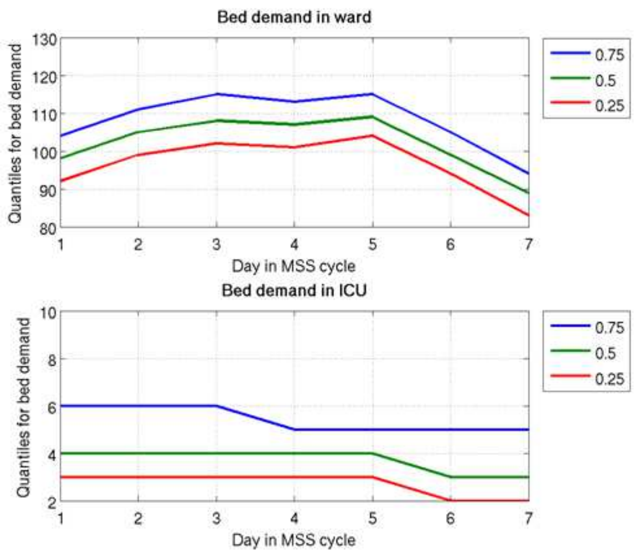

Figure 3.14: Quantiles of bed demand for Master Surgery Schedule C

3.4 Discussion

MRI focuses on two goals when considering ICU and ward bed requests. First, occupancies on Saturdays and Sundays should be lower than during the week since working on the weekends is not very attractive for most employees and overtime has to be paid. Second, bed requests during the week should be leveled as fixed capacities for peak demand have to be provided on all days, as it is easier to staff at relatively constant levels, and as constant inflow of patients in the ICU is important for medical reasons. Table 3.1 summarizes the average bed requests per day, the peak bed requests, the relative additional bed request of the peak demand compared to the minimum bed demand during the week, the average bed request on Saturdays and Sundays, and the relative reduction of the average bed requests on the weekend compared to the weekdays for the ICU and the ward.

MSS A shows a peak on Friday, where ward bed requests are about 24% higher than on Mondays. In the cost-intensive ICU the degree of capacity utilization throughout the week remains constant. This is due to that the OR blocks of NCS and URS, both operating on patients with a high probability to be sent to the ICU, are spread evenly across the week. The main reason for the high fluctuation in ward bed demand is that the department of neurosurgery operates on many patients throughout the week and many of those patients have long postoperative stays in the hospital. This leads to a growing number of patients during the week. Weekend bed requests are on average 5 beds for the ICU and 99 beds for the ward.

MSS B is designed to move forward the peak demand in order to reduce weekend bed requests. Shifting all blocks of NCS towards the beginning of the week results in a high ICU bed demand on Mondays and Tuesdays with a peak demand of 8 beds. The peak demand in the ward of 125 beds is also higher than in MSS A. The average weekend bed requests were reduced to 4.0 and 96.5 beds for the ICU and ward, respectively. However, MSS B will not be implemented for two reasons: First, the immense variation in ICU bed requests during the week leads to difficult staffing situations. Second, it is unrealistic to increase the number of ORs run by NCS from 3 to 8 per day on Mondays and Tuesdays. In this case, a lot of problems would rise such as

	MSS A	MSS B	MSS C
ICU			
Average bed requests/day	5.7	5.3	5.4
Peak bed requests	6.0	8.0	6.0
Relative peak vs. minimum (Mon-Fri)	0%	100%	20%
Average weekend bed requests (Sat-Sun)	5.0	4.0	5.0
Relative weekend (Sat-Sun) vs. week (Mo-Fri)	-20%	-45%	-12%
Ward			
Average bed requests/day	108.3	108.3	108.1
Peak bed requests	123.0	125.0	115.0
Relative peak vs. minimum (Mon-Fri)	24%	30%	11%
Average weekend bed requests (Sat-Sun)	99.0	96.5	99.5
Relative weekend (Sat-Sun) vs. week (Mo-Fri)	-13%	-17%	-12%

Table 3.1: Peak and weekend bed requests for ICU and ward (75% quantile)

staffing of the operating rooms, supplying them with appropriate equipment, and the organization of the logistics around the OR suite.

To ensure a constant patient flow from NCS to the ICU, we assigned 3 ORs per day in MSS C, similar to MSS A. URS has a high percentage of patients with a very short postoperative stay in the hospital so we scheduled the URS blocks from Mondays until Wednesdays and the SP blocks on Thursdays and Fridays. This schedule and MSS A show a similar constant ICU bed demand pattern. However, the peak of ward bed requests of 115 beds is 7% and 8% below the peaks of MSS A and B, respectively. The average weekend ward bed requests are only 1% and 3% above those of MSS A and B.

Although MSS C provides a good fit regarding the goals of reducing of weekend occupancy and leveling of bed requests during the week, some difficulties of implementing MSS C remain. URS has to run 5 ORs on three days instead of 3 daily ORs, and SP has to run 5 ORs on two days instead of 2 daily ORs. This implies dramatic changes in staffing and logistics, as well as necessary rescheduling in the outpatient clinic. In order to implement modifications of MSSs, arguments have to be presented to the stakeholders in an

OR suite. The results demonstrate that it is possible to improve bed demand in postoperative units by changing the OR schedule. As a first step, minor modifications concerning the order of the operation within one department could be realized. Surgeries with a long post-operative stay could be generally carried out on special days. Modeling different surgery types within medical specialties could offer a valuable extension of our approach. In our study we did not take into consideration any financial aspects of different MSSs. Bed occupancy levels as well as changes in MSSs could be connected with costs. Assigning costs to MSSs could be another matter for further research. First steps in this direction have already been performed in Fügener et al. [34] (see chapter 2). Due to its design, the suggested model can be extended to deal with more complex organizational structures with several separated ICUs and wards so that our assumption regarding an aggregated ICU and ward could be dropped. Another interesting extension would be to adapt the model to include pre-operative stays as these also account for bed occupancy.

With the approach discussed in this chapter, the influence of a given MSS on the ICU and ward can be investigated and modeled mathematically. Extending Fügener et al. [34] (see chapter 2), we account for emergency surgeries taking place at the weekends by introducing another hypothetical specialty "WE" in order to reflect their impact as well. We have started with implementing the current MSS and tested two additional scenarios adjusting the MSS. As the performed modifications were rather straightforward, one could easily predict the effects on bed demand and in fact the results of the simulations match the expected outcome. Therefore, our model should be considered an important tool for planning downstream units, e.g. to determine staffing levels or to derive capacity extensions. The model we presented in this chapter may be used in other hospitals as well. As discussed above, further extensions to better fit specific hospitals are easily implementable. We believe that considering both ICU and ward occupancy levels when designing MSSs can greatly improve both medical and financial hospital performance.

Chapter 4

On the assessment of costs in a newsvendor environment: Insights from an experimental study

4.1 Introduction

In the newsvendor problem a decision maker has to decide on the number of ordered products under stochastic demand. Once the uncertainty is resolved, the costs incurred from the mismatch between the decision and the realization become apparent. The decision maker observes that his decision was too "high" or too "low". The newsvendor model provides a theoretically grounded approach to determine the optimal order quantity, i.e. the order quantity that minimizes the expected mismatch costs.[1] However, experimental studies show that decision makers systematically deviate from the optimal order quantity. In their seminal paper Schweitzer and Cachon [76] observe a pattern of behavior where subjects order too few high margin products and too many low margin products. According to the anchoring and adjustment heuristic [85] this too low/too high pattern can be explained

[1] A minimization of the expected mismatch costs is equivalent to a maximization of the expected profit, see Silver et al. [78] or Khouja [49].

by that individuals anchor on the mean demand and insufficiently adjust toward the optimal order quantity. A number of follow-up studies have confirmed the too low/too high pattern, e.g., in experimental newsvendor studies considering doubled payoffs and reduced order frequency [15], the effect of learning [12], different demand distributions [9], participants with different educational backgrounds [13], different frames [51, 75], multilocation inventory systems [43], different payment schemes [21] or cross-cultural differences between Western and Eastern countries [30]. Order decisions in the newsvendor problem tend to be biased towards the anchor of mean demand, which we call "mean anchor effect". For a recent review considering experimental studies of the newsvendor problem we refer to Kremer and Minner [50].

Although many studies discuss behavioral aspects in the newsvendor problem, there is hardly any research on the assessment of the different cost types. Since costs are one of the essential influencing variables in the newsvendor problem, the assessment of costs may have a strong effect on human decision making. Depending on the field of application, costs like out-of-pocket costs, opportunity costs, or penalty costs can be relevant for the decision about the order quantities. A detailed definition of the cost types in the context of our study will be given in Section 4.2. Previous studies have shown that these cost types may have a diverse influence on behavior in several situations. The indirect character of opportunity costs is a reason why they are often neglected in decision making. Northcraft and Neale [65] state that opportunity costs are abstract possibilities which can lead to a biased assessment of the cost/benefit picture of a decision maker. This biased opportunity cost perception is documented in numerous papers. The results from an experimental study by Becker et al. [6] suggest that decision makers consider opportunity costs as less important than out-of-pocket costs and even ignore them in some cases. A study by Friedman and Neumann [32] leads to consistent results. They conclude that decision makers underweight opportunity costs when only partial information is available. While Becker et al. [6] as well as Friedman and Neumann [32] investigate a setting with a certain environment, Hoskin [44] considers the assessment of opportunity costs in an uncertain environment. Seventeen years before the seminal paper of Schweitzer and Cachon [76], the experimental study of Hoskin [44] had already addressed human behavior in the newsvendor problem. The results show that decision makers deviate from the order quantities that optimize expected profits. However, the study has a number of technical shortcomings which do not allow deriving consistent and reliable results.[2] Since previous research has shown that decision makers underweight or even neglect foregone payoffs, Ho et al. [43] hypothesize that the psychological aversion to leftovers is greater than the disutility of stockouts. They develop and experimentally test a newsvendor framework where they add psychological costs of overordering and underordering. A main weakness of their additive approach is that an underweighting of foregone losses is modeled as additional (positive) psychological costs, which seems counterintuitive. Furthermore, the case that decision makers neglect foregone payoffs is even incompatible. The question why in many situations decision makers underweight opportunity costs compared to out-of-pocket costs is addressed by Thaler [83]. He argues that the endowment effect supports the different weighting of these costs. While opportunity costs are often underweighted other cost types tend to be overweighted by decision makers. McCaffery and Baron [59] refer to Richard Thaler's real-world observation: "when a gas station charged a 'penalty' for using credit cards ($2.00 versus $1.90, say), people paid cash; when a gas station across the street gave a 'bonus' for using cash ($1.90 versus $2.00), people used credit cards". McCaffery and Baron [59] state that, due to penalty aversion, individuals will rather avoid penalties than obtain bonuses. The tendency of people to avoid penalties is documented in several experimental studies and holds true in diverse economical contexts. For example, tax rules [58] or contracts [55] are less likely to be accepted when they are presented as penalties rather than as bonuses. The consequences of penalty aversion are decisions where penalty costs are higher weighted than out-of-pocket costs – another example that the different assessment of cost types can lead to a different behavior.

Involving different types of costs, a wide range of business decisions require that a decision is made before the occurrence of a random event. The underlying trade-off, concerning the costs of the mismatch between the decision and the realization, is captured by the newsvendor model. However, experimental studies of the newsvendor problem typically consider out-of-pocket costs (overage costs) and opportunity costs (underage costs) as mis-

[2] The number of participants per experimental setting as well as the number of periods were too small, participants had to estimate the demand distribution based on the past data on demand, and some participants received changed information already after few periods. Furthermore, several product types had to be ordered and the margins of the products were chosen unfavorably.

match costs. To investigate how the assessment of costs influences a decision maker, we consider two newsvendor situations involving different types of cost. Motivated by the literature, we expect an underweighting of opportunity costs and an overweighting of penalties. Therefore, we consider a situation where penalty costs (respectively additional reorder costs) instead of opportunity costs occur in the underage case. An example is a newsvendor situation involving a second order for an additional premium as considered by Cachon and Terwiesch [17] where "the second order opportunity eliminates lost sales (...) [but] therefore, the penalty for ordering too little in the first order is that one may be required to purchase additional units in the second order at a higher cost." We refer to this kind of newsvendor problem involving out-of-pocket costs and penalty costs as the "penalty cost problem" whereas the classical newsvendor problem as considered by Schweitzer and Cachon [76] is referred to as the "opportunity cost problem". Since only the type of costs is different the balancing problem remains mathematically identical and the decision maker is still facing the same underlying tradeoff concerning ordering too little and ordering too much [17]. Gavirneni and Isen [36] show that most people are able to compute the overage and underage costs accurately, but fail to determine the optimal inventory level. Therefore, a different behavior in the penalty cost and the opportunity cost problem implies that the assessment of costs changes for different cost types. Consequently, to investigate our research question we set up an experimental study where we differentiate between these two problems. Since previous research has shown that people anchor on the mean demand, we further distinguish between three cases with different margins for each of the two problems.

The main contribution of this chapter is twofold. Firstly, to the best of our knowledge, our work is the first which systematically investigates how the assessment of costs influences a decision maker in a newsvendor situation. We propose a behavioral approach, including a higher weighting of penalty costs than of opportunity costs and order decisions which are biased towards the mean. Our model explains large portions of the observed behavior in our experimental study. A different weighting of costs can be seen as the main driver for higher order quantities in the penalty cost problem compared to the opportunity cost problem. Based on our findings we identify situations in the newsvendor problem which are particularly unfavorable for the performance of a decision maker. Furthermore, our insights allow detecting situations where the behavioral effects partially compensate for each other and therefore lead to a better performance of decision makers.

Secondly, our experimental study gives important insights into how people behave in newsvendor situations which are affected by penalty costs. For many business decisions, the underage case of the underlying newsvendor trade-off is influenced by penalty or reorder costs and not by opportunity costs. Typical areas where expensive re-orders, contractual penalties or second production runs occur instead of lost profits are procurement problems if too little was ordered (e.g. Cachon and Terwiesch [17]), inventory problems if too little was stored (e.g. Eppen [29]), or production problems if too little was produced (e.g. Donohue [26]).[3,4] In order to apply the findings from behavioral studies of the newsvendor problem to a broad field of business situations it is important to check the validity and to identify limitations. The results of our study clarify that the behavior in a newsvendor situation which is affected by penalty costs is significantly different from the behavior in a situation which is influenced by opportunity costs.

The chapter is organized as follows: Section 4.1 provides an introduction and a literature review before we define our hypotheses in Section 4.2. The experimental setup and design is described in Section 4.3, and we discuss the results in Section 4.4. Finally, in Section 4.5 we draw conclusions and discuss managerial implications.

4.2 Definitions and Research Hypotheses

To investigate the influence of different cost types on decision making, we differentiate between two from a mathematical point of view identical newsvendor problems with the only difference that the type of costs in the underage case is different. We consider one situation where penalty costs occur and

[3] Identical to the opportunity cost problem, the overage case of the penalty cost problem involves out-of-pocket costs like production costs, holding costs, and purchasing costs.

[4] In a broader context the penalty cost problem is also inherent in stochastic project management settings, such as the determination of feeding buffers (e.g. Trietsch [84]) or due dates (e.g. Zhu et al. [95]) assuming costs for starting activities earlier and tardiness penalties. Furthermore, a typical application in health care management is the reservation of operating room capacity under uncertainty considering costs for operating room time and overtime costs (e.g. Olivares et al. [66], Wachtel and Dexter [92]).

one situation where opportunity costs occur. We expect a different behavior in the penalty cost and the opportunity cost problem.

The opportunity cost problem is equal to the classical newsvendor problem as described, e.g., in the seminal paper of Schweitzer and Cachon [76] and in most follow-up newsvendor studies. A vendor orders goods for the next period where he faces an uncertain demand d. The cumulated demand distribution $F(D)$ is known. Purchasing costs per item are c and the selling price is p. Consistent with the newsvendor literature, we define the purchasing costs as out-of-pocket costs. If the demand exceeds the order quantity q the foregone opportunity to make more profit by selling more products leads to lost sales and thus to lost profits which are also referred to as opportunity costs. The opportunity costs per item which cannot be delivered, termed as "underage costs", is $c_u = p - c$. If demand is less than the order quantity, assuming a salvage value of 0, the costs for each unit ordered too much, called "overage costs", are $c_o = c$.

Analog to the opportunity cost problem, in the penalty cost problem a vendor orders q units for the next period where he faces an uncertain demand d with a known cumulated demand distribution $F(D)$. For each unit he orders before demand takes place, he has purchase costs of c (out-of-pocket costs). If the demand exceeds the order quantity, he has to reorder units for higher reorder costs of $s > c$ to satisfy the excess demand.[5,6] The costs for each unit ordered too little ("underage costs") are the additional "penalty" costs of the reorder, i.e. $c_u = s - c$. If demand is less than the order quantity, the costs for each item ordered in excess of the realized demand ("overage costs") are equal to the purchase costs, i.e. $c_o = c$.[7]

In both newsvendor situations the expected costs of overestimating and underestimating demand have to be minimized. The only difference between both situations is the different type of costs in the underage case (see Table

[5]In contrast to Cachon and Terwiesch [17] we consider a reorder obligation instead of a reorder possibility. It is obvious that the second order should equal the unfulfilled demand. In order to avoid additional behavioral biases, we prefer to maintain a situation involving only one decision.

[6]An obligation to reorder may be interpreted as a commitment for a service level of 100%.

[7]Considering a reorder possibility instead of a reorder obligation does not change the overage and underage costs and therefore the optimal order quantity, given that the selling price is above the costs (see Eeckhoudt et al. [27]).

4.1). The underage costs correspond to penalty costs in the penalty cost problem, while they correspond to opportunity costs in the opportunity cost problem. Since only the type of costs is different, we can determine the

	Penalty cost problem	Opportunity cost problem
Underage costs	Penalty costs	Opportunity costs
Overage costs	Out-of-pocket costs	Out-of-pocket costs

Table 4.1: Summary of cost types

"optimal order quantity" q^* for both problems with the classical newsvendor formula

$$q^* = F^{-1}\left(\frac{c_u}{c_u + c_o}\right) \tag{34}$$

with a problem specific definition of the underage costs as given above. By simple algebraic reformulation we obtain

$$c_o \cdot F\left(q^*\right) = c_u \cdot \left(1 - F\left(q^*\right)\right) \tag{35}$$

which shows the trade-off a decision maker faces: The optimal order quantity q^* can be derived from balancing the probability of being over and under stocked weighted with the overage and underage costs, respectively. In order to depict a biased assessment of costs we include the underage cost weight $\beta > 0$ which specifies how much the underage costs, relative to the overage costs, influence a decision. Since the overage costs are equal to the purchase costs in both problems, we scale these out-of-pocket costs with a weight of 1. An underage cost weight of $\beta > 1$ indicates that a decision maker has a stronger weighting of underage costs relative to overage costs. An underage cost weight of $\beta < 1$ indicates that the decision maker weights the underage costs lower than the overage costs. We denote the consequences of the cost weight on the order quantity as the "assessment of costs effect" (ACE). To integrate the biased assessment of the different costs in the balancing problem we extend Equation (3) by the overage cost weight 1 and the underage cost weight β and obtain

$$1 \cdot c_o \cdot F\left(q^{ACE}\right) = \beta \cdot c_u \cdot \left(1 - F\left(q^{ACE}\right)\right) \tag{36}$$

where q^{ACE} denotes the adapted optimal order quantity.[8] Reformulation of (36) leads to

$$q^{ACE} = F^{-1}\left(\frac{\beta \cdot c_u}{\beta \cdot c_u + c_o}\right). \tag{37}$$

We assume that the weight of the underage costs β depends on the type of costs only and not on absolute values. In the penalty cost problem the underage costs correspond to penalty costs that occur because an expensive reorder has to be placed. Since individuals are trying to avoid penalties, we derive our first hypothesis with β_{pen} as underage cost weight in the penalty cost problem:

> *H1: In the penalty cost newsvendor problem people have a higher weighting of penalty costs compared to out-of-pocket costs, that is $\beta_{pen} > 1$.*

In the opportunity cost problem the underage costs have the character of opportunity costs. As decision makers tend to underweight opportunity costs, we derive our second hypothesis with β_{opp} as underage cost weight in the opportunity cost problem:

> *H2: In the opportunity cost newsvendor problem people have a lower weighting of opportunity costs compared to out-of-pocket costs, that is $\beta_{opp} < 1$.*

Our central research question is whether decision makers behave differently in the penalty cost and in the opportunity cost problem. On the one hand, we expect opportunity costs to be lower weighted than out-of-pocket

[8] The incorporation of an underweight factor in the newsvendor model is similar to Chen et al. [21]. They show that the payment timing affects ordering behavior, and they can explain this behavior by the effect that decision-makers underweight order-time payments.

costs and on the other hand we expect penalty costs to be higher weighted than out-of-pocket costs. This leads to our third hypothesis:

> *H3: The weighting of opportunity costs in the opportunity cost newsvendor problem is lower than the weighting of penalty costs in the penalty cost newsvendor problem, that is* $\beta_{opp} < \beta_{pen}$.

As many newsvendor studies have shown that human behavior depends on the margin, we consider several cases. In the opportunity cost problem, the margin is defined as $\frac{p-c}{p}$, while in the penalty cost problem the margin is defined as $\frac{s-c}{s}$. Therefore, the margins are equal to the critical ratios. We differentiate between a "high margin case" where the critical ratio exceeds 0.5, a "medium margin case" where the critical ratio equals 0.5, and a "low margin case" where the critical ratio is less than 0.5. Assuming symmetric demand distributions, this leads to optimal order quantities above, equal to, and below the mean demand. As we consider a medium margin case we can discuss a situation where deviations from the optimal order quantity may not solely be explained by the mean anchor effect. Benzion et al. [9] proposed the following formula to consider the mean anchor effect (MAE) where the order quantity q^{MAE} is determined by a linear combination of the mean demand μ and the optimal order quantity q^* with mean anchor weight α:

$$q^{MAE} = \alpha \cdot \mu + (1 - \alpha) \cdot q^*. \tag{38}$$

For $0 < \alpha < 1$ the resulting order quantity is consistent with the mean anchor effect. We assume the mean anchor effect to be symmetric, so the strength of the shift towards the mean neither depends on the order quantity being above or below the mean nor its distance from the mean. We further assume the mean anchor weight to be the same for the opportunity cost and the penalty cost problem. As the mean anchor effect has been documented in many previous studies, it can be assumed to have a significant effect on the order decision. This leads to our fourth hypothesis:

> *H4: The mean anchor effect exists, that is* $0 < \alpha < 1$.

For an integrated model of human behavior, a combined consideration of both the assessment of costs effect and the mean anchor effect is needed. To model the human order decision we combine both effects in a straightforward way. We denote the resulting effect as "combined effect" (CE). The logic of the combined effect is as follows: The assessment of costs effect leads to the adapted optimal order quantity considering the cost weights of the decision maker. This adapted optimal order quantity is biased by the mean anchor effect towards the mean, resulting in the order quantity q^{CE}. The formula for the combined effect is then

$$q^{CE} = \alpha \cdot \mu + (1 - \alpha) \cdot F^{-1} \left(\frac{\beta \cdot c_u}{\beta \cdot c_u + c_o} \right) \tag{39}$$

where, depending on the problem, β stands for β_{opp} in the opportunity cost problem and β_{pen} in the penalty cost problem, respectively.

We employ a 2×3 design where we combine two problems (opportunity cost problem, penalty cost problem) with three margin cases (high margin, medium margin, low margin) and thus obtain six different combinations. To compare the opportunity cost and the penalty cost problem, we set the selling price p equal to the reorder costs s. As we consider the same purchase costs c in both problems, the critical ratios are equal. By assuming an identical demand distribution, we achieve the same optimal order quantities. This enables a clear comparison of human behavior in the opportunity cost problem and the penalty cost problem, as the identical optimal order quantity can be used as a reference point. To achieve the different margin cases, we vary the costs only.

Based on our hypotheses we consider the consequences of the combined effect, including the assessment of costs effect and the mean anchor effect on the order decisions. The higher weighting of penalty costs leads to an increase in the order quantity in the penalty cost problem while in the opportunity cost problem the order quantity is reduced by a lower weighting of opportunity costs. Furthermore, the mean anchor effect leads to a shift towards the mean demand. The comparison given in Table 4.2 clarifies that based on our hypotheses, the order quantities of the penalty cost problem should exceed the ones of the corresponding opportunity cost problem in all margin cases.[9] We note that the assessment of costs effect and the mean

[9] Extreme examples could lead to a situation where the assessment of costs effect leads to quantities above the mean in the low margin case or below the mean in the high margin case. In these cases the mean anchor effect will change direction as depicted in Table 4.2.

	Penalty cost problem			Opportunity cost problem	
	ACE	MAE		ACE	MAE
High margin case	$\Uparrow$	$\Downarrow$	>	$\Downarrow$	$\Downarrow$
Medium margin case	$\Uparrow$	$\Downarrow$	>	$\Downarrow$	$\Uparrow$
Low margin case	$\Uparrow$	$\Uparrow$	>	$\Downarrow$	$\Uparrow$

Table 4.2: Expected consequences for the order quantities

anchor effect lead in the same direction in both the high margin case of the opportunity cost problem and the low margin case of the penalty cost problem, respectively. We therefore expect results in these situations that are especially far away from the optimal order quantity. On the other hand in the remaining situations the two effects work in opposite directions, therefore they should partially compensate for each other and thus the deviations from the optimal order quantity should be not as big.

4.3 Experimental Setup

To test our hypotheses we set up a laboratory study using a 2×3 between-subjects design where we distinguish between six combinations of problem and case, as given in Table 4.2. In all six experiments we examine a discrete uniform demand distribution with the boundaries 0 and 100. The realization of the demand was randomly drawn in advance and is used for all six experiments.

Furthermore, we consider the same critical ratio for the opportunity cost and the penalty cost problem in each case. The parameters are set to $s = 12$ in the penalty cost problem and to $p = 12$ in the opportunity cost problem. The costs are set to $c = 3$ in the high margin case, to $c = 6$ in the medium margin case, and to $c = 9$ in the low margin case. The obtained optimal order quantities of $q^* = 75$, $q^* = 50$, and $q^* = 25$, are above, equal to, and below the mean demand of $\mu = 50$.

All experiments were conducted at the "Munich Experimental Labora-

The overall order quantities will still be greater in the penalty cost problem than in the corresponding opportunity cost problem.

tory for Economic and Social Sciences" (MELESSA). For every experiment 25 separated PC terminals were ready to use. Participants were recruited from the subject pool of the MELESSA with the help of a recruitment-software. All participants were students without profound knowledge of the newsvendor problem and they came from different fields of study. Each student participated in one experimental study only and altogether 148 students participated in the six different experiments. We ran four experiments with 25 participants and two experiments with 24 participants. Despite an overbooking of 3, only 24 students participated in the high margin case of the opportunity cost problem and in the high margin case of the penalty cost problem. The experiment was programmed and conducted with the software z-Tree [31]. Before the experiments the instructions were read aloud (see Appendix). Every period started with a decision screen where the participants had to make their order decision. After every decision they received information about the realization of the demand, their order quantity, and the resulting profit or costs of this period on the information screen. The profits or costs were displayed in "experimental currency units" (ECU). In all six experiments the purchase decision was repeated for thirty periods. The duration of one experiment was about 45 minutes. By control questions we ensured that all subjects understood their job within the experiment. After completing the session the accumulated earnings were paid privately and in cash. In all six experiments we chose the factor and the fixed amount such that an income of €14 could be obtained if the optimal order quantity was placed in each period. The performance oriented compensation was explained in the instructions and therefore known in advance. Across all six experiments the subjects earned on average €10.47 including a show-up fee of €4. The standard deviation was €1.72.

4.4 Results

4.4.1 General Results

As in previous studies we observe average order quantities per period and over all periods which are significantly higher in the high margin case than in the medium margin case (one-tailed Wilcoxon, $p < 0.005$), and significantly lower in the low margin case than in the medium margin case (one-tailed

Wilcoxon, $p < 0.005$). This holds true for the penalty cost problem (PCP) and for the opportunity cost problem (OCP). For each case the average order quantities of the subjects are shown in the Figures 4.1, 4.2, and 4.3.

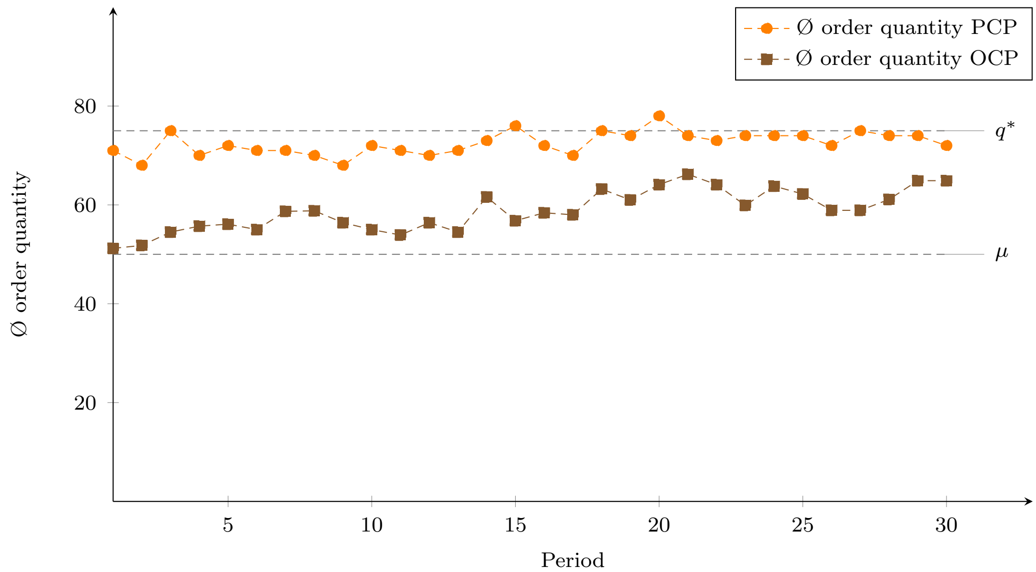

Figure 4.1: Average order quantities in the high margin case

For both the penalty cost and the opportunity cost problem, Figures 4.1 to 4.3 illustrate that the average order quantities differ from the mean demand as well as from the optimal order quantities. As provided in Table 4.3, only for the medium margin case of the opportunity cost problem is the difference not significant. Furthermore, the average order quantities in the penalty cost problem significantly exceed the average order quantities in the opportunity cost problem for each margin case (see Table 4.3). Our results show that the order quantities are especially far away from the optimal order quantity in both the high margin case of the opportunity cost problem and the low margin case of the penalty cost problem. This is in line with our expectations outlined in Section 4.2.[10] To investigate learning effects we conducted a regression analysis on the average order quantities for each of the six experiments, where we define learning as a trend towards the optimal order quantity. As illustrated in Table 4.4, we observe significant learning

[10]Considering the high and the low margin case, we observe a too low/too high pattern for the penalty cost problem as well as for the opportunity cost problem.

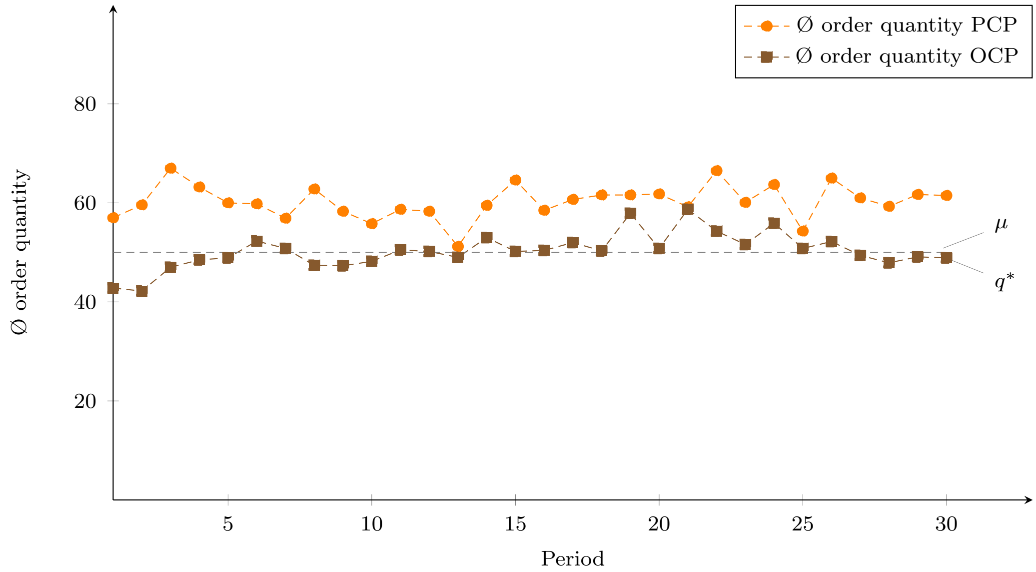

Figure 4.2: Average order quantities in the medium margin case

in the high margin case, but there is no significant learning in the medium margin case and the low margin case. Over all six experiments no consistent

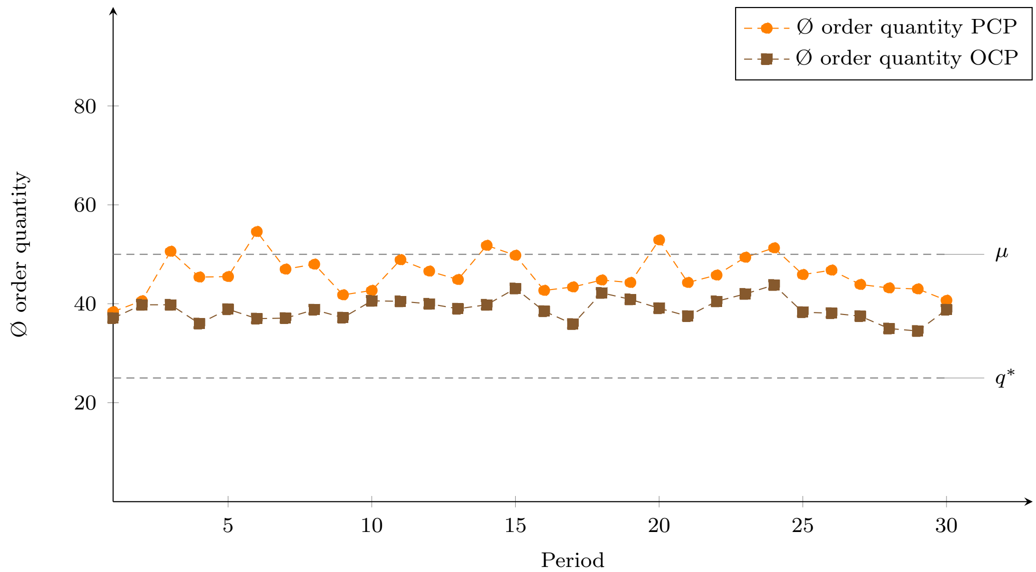

Figure 4.3: Average order quantities in the low margin case

	High margin case		Medium margin case		Low margin case	
	PCP	OCP	PCP	OCP	PCP	OCP
Opt. order quantity q^*	75.0		50.0		25.0	
Mean average order quantity	72.6	58.9	60.3	50.3	46.0	38.9
Difference from μ	$p < 0.005$	$p < 0.005$	$p < 0.005$	$p = 0.829$	$p < 0.005$	$p < 0.005$
Difference from q^*	$p < 0.005$	$p < 0.005$			$p < 0.005$	$p < 0.005$
PCP is significantly higher than OCP	$p < 0.005$		$p < 0.005$		$p < 0.005$	

Table 4.3: One-tailed Wilcoxon test for average order quantities

learning or trend pattern can be approved.

	Penalty cost problem	Opportunity cost problem
High margin case	0.141 (p <0.005)	0.367 (p <0.005)
Medium margin case	0.061 (p =0.425)	0.192 (p =0.009)
Low margin case	-0.023 (p =0.783)	0.000 (p =0.996)

Table 4.4: Trend values of the regression analysis on the average order quantities

We further analyzed how the average order quantities over the 30 periods of the decision makers are distributed. Figure 4.4 provides a box plot diagram for each of the six experiments with the lower quartile, the median, and the upper quartile. The end of the "whiskers" show the lowest and the highest datum within the 1.5 interquartile range. Outliers are illustrated as well. The box plots show that the average order quantities of the individuals (shown on the y-axis) are significantly higher in the penalty cost problem than in the opportunity cost problem for all three margin cases (one-tailed Mann-Whitney U, $p < 0.005$ for all three cases). The box plots also clarify the

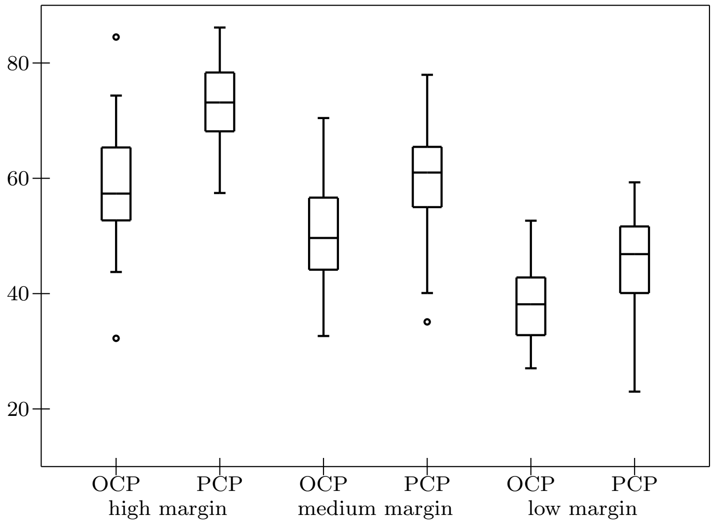

Figure 4.4: Box plot diagram of the average order quantities of the participants

systematic difference of the average order quantities in the three margin cases for both problems. Our results provide a good prediction of the behavior of an “average” decision maker since the box plots illustrate that most of the observed average orders of individuals are distributed closely around the median.

4.4.2 Testing of the hypotheses

Our general results confirm that humans do not behave optimally in the newsvendor setting. However, their decisions are not random. For both problems the level of the average order quantities in the different margin cases correspond to the level of the optimal order quantities. Furthermore, systematic differences between the penalty cost and opportunity cost problems exist. Consequently, we investigate whether a combined effect, including a higher weighting of penalty costs than of opportunity costs and order decisions which are biased towards the mean, is consistent with the observed behavior. Therefore, we adapt Formula (39) describing the combined effect to test our hypotheses and to evaluate our explanatory approach. Based on the six experiments, we estimate the three relevant parameters for Formula (40) wherein q_t is the average order quantity in period t. For the estimation we integrated an error term ϵ_t.

$$q_t = \alpha \cdot \mu + (1-\alpha) \cdot F^{-1}\left(\frac{\beta \cdot c_u}{\beta \cdot c_u + c_o}\right) + \epsilon_t \tag{40}$$

For the six experiments with 30 periods each, we estimate one common mean anchor weight α according to our assumptions in Section 4.2. For the three experiments in the penalty cost problem, we estimate a common penalty cost weight β_{pen} and for the three experiments in the opportunity cost problem, we estimate a common opportunity cost weight β_{opp}. The parameters are estimated using a least square estimation (R^2=0.88) where the test statistic follows asymptotically the standard normal distribution. We obtain $\beta_{pen} = 2.42$, and we can show that $\beta_{pen} > 1$ is significant (one-tailed z-test, $p < 0.005$). Consequently, the first hypothesis could be verified: Penalty costs are higher weighted than out-of-pocket costs. As $\beta_{opp} = 0.95$, the opportunity cost weight is lower than one. However, the difference is quite small. Since $\beta_{opp} < 1$ (one-tailed z-test, $p = 0.073$), we find support for the second hypothesis: Opportunity costs are lower weighted than out-of-pocket costs. Furthermore, we can verify the third hypothesis as β_{opp} is significantly lower than β_{pen} (one-tailed Welch's t-test, $p < 0.005$): Penalty costs are higher weighted than opportunity costs. For the mean anchor weight we obtain $\alpha = 0.49$.[11] The fourth hypothesis could also be confirmed since we can show that $0 < \alpha < 1$ is significant (two-tailed z-test, $p < 0.005$) and thus a mean anchor effect exists.

Our results clarify that the mean anchor effect can be seen as the strongest driver for the non-optimal order quantities since the mean demand is weighted by almost 50%. However, the mean anchor effect cannot explain the large differences of the order quantities between the opportunity cost problem and the penalty cost problem. Hence, our results show that the different weighting of costs can be seen as the main driver for higher average order quantities in the penalty cost problem compared to the opportunity cost problem.

Our approach leads to an mean absolute error of 3.2 (standard deviation

[11] This is in line with the results from previous studies in Western countries. Considering a high margin case and a low margin case of the opportunity cost problem, e.g. Bostian et al. [15] obtain a mean anchor weight of 0.47 and the data from Bolton and Katok [12] correspond to a mean anchor weight of 0.54.

of 2.4) concerning the order quantity.[12] This is very low compared to an mean absolute error of 11.1 when using the optimal order quantity as an estimator, of 10.0 when using the mean demand as an estimator, and of 6.2 when using the mean anchor effect (see Formula 38) as an estimator. The comparison demonstrates the high explanatory quality of our approach.[13]

4.5 Conclusion

Our research highlights differences in human decision making in situations involving different types of costs. Motivated by the literature, we expect that opportunity costs are underweighted compared to out-of-pocket costs while penalty costs are overweighted. In order to investigate the research question we set up an experimental study in a newsvendor setting which provides a simple yet realistic environment to investigate the assessment of penalty costs and opportunity costs in two mathematical identical situations. To the best of our knowledge we are the first to compare the assessment of these two cost types in an operations management setting. We observe that individuals order significantly more in a newsvendor setting with penalty costs than in a newsvendor setting with opportunity costs. We propose a behavioral approach which incorporates decision biases in the newsvendor model to explain the observed behavior. Besides the assessment of costs effect we also include the mean anchor effect. Based on our approach we tested our hypotheses and could confirm the mean anchor effect as well as a different weighting of different cost types. We found that penalty costs are higher weighted than opportunity costs. Our approach is valuable to predict actual ordering behavior and, furthermore, it allows us to quantify the extent of the psychological biases. Based on our findings we conclude that the performance of a newsvendor depends clearly on the underlying situation. Situations where the assessment of costs effect and the mean anchor effect lead

[12] The remaining error can be partially explained by demand chasing, see Schweitzer and Cachon [76]. Since the underlying demand vector is identical in all experiments, demand chasing systematically influences the average order quantities per round.

[13] Our approach reduces the error by 68% compared to the mean demand as an estimator and by 71% compared to the optimal order quantity as an estimator. The isolated mean anchor effect reduces the error by 38% compared to the mean demand and by 44% compared to the optimal order quantity. This clarifies that the mean anchor effect explains only a part of the behavioral deviations.

in the same direction result in particularly bad performance while in situations where the two effects partially compensate each other result in a better performance of human decision makers. Since in many business decisions the underlying newsvendor trade-off is influenced by penalty costs instead of opportunity costs, our study gives important insights in order to apply behavioral findings of newsvendor studies in a broader field. Typical examples where contractual penalties occur are inventory problems or purchase situations. It can be misleading to relate the insights from the opportunity cost newsvendor problem to situations incurring penalty costs instead of lost sales. This already happens, e.g., in research concerning the bullwhip effect [64] or operating room planning [92]. The main finding of our research is that a biased perception of opportunity costs as well as a biased perception of penalty costs can explain the observed behavior. We show that decision makers are more sensitive to penalty costs than to opportunity costs. Consequently, we conclude that people have a different assessment of different cost types.

Our work has several limitations. We assume the mean anchor effect to be symmetric even if the effect is stronger in the low margin context than in the high margin context in many studies. A first promising research considering the asymmetry in ordering behavior is done by Moritz [62]. He finds support that cognitive dissonance explains a portion of this behavior. As there are considerable differences in the asymmetry and since the asymmetry is even reverse in several studies (e.g., Ho et al. [43], Rudi and Drake [72], Lurie and Swaminathan [56]) a further investigation is needed. If an asymmetry of the mean anchor effect could be validated and measured in terms of different mean anchor weights for different margin cases, it could be easily included in our approach. We agree with Bostian et al. [15] that the exploration of the asymmetry is one of the most promising directions for further research. Another interesting research area is cross-cultural differences between Western and Eastern countries. Even though, the mean anchor effect is a predominant cross-cultural effect, Feng et al. [30] have shown that the strength of the effect is stronger in Eastern cultures. The investigation of differences and similarities in the assessment of costs between Eastern and Western cultures is a promising area for further research. Furthermore, the explanatory power of our approach could be increased by the integration of additional behavioral factors. Rudi and Drake [72] state that besides the "level behavior" the "adjustment behavior" can be seen as the main driver of the mismatch costs. Therefore, consideration of demand chasing could be worthwhile. Another example is the integration of a learning factor which would be especially useful for long-term consideration, e.g. 100 periods as investigated by Bolton and Katok [12]. These extensions would lead to a more complex approach but they could also enable an even more realistic description and prediction of the behavior.

Based on our approach and our findings we conclude several managerial implications: From an internal company point of view our insights could be used in a control process to detect situations which lead to systematic deviations from the optimal order quantity that are particularly unfavorable. Identifying these situations may allow corrective actions. Another internal aspect related to the planning process is that one could create situations where the deviations of the order decisions are relatively small and therefore the decision maker performs better. From a supplier perspective it may be possible to create situations where the behavior of the decision maker leads to deviations which are favorable for the supplier. Another important aspect for a supplier is to identify situations where the customer systematically orders too little. Consulting the customer may help to improve the situation for both, e.g. by a modified contract.[14]

[14]The performance of different contracting mechanisms in a two-echelon supply chain in which the retailer faces the opportunity cost newsvendor problem is investigated by Katok and Wu [48]. Based on our results further research concerning the performance of contracting mechanisms in the penalty cost newsvendor problem as well as further research concerning the use of different cost types to increase contract performance in general would be interesting.

Chapter 5

Underutilization and overutilization of operating rooms: Insights from behavioral health care operations management

5.1 Introduction

Recently, Gino and Pisano [37] encouraged researchers to take into account human behavior in operations management. Health care operations management has a particularly strong behavioral influence [16], since health care services are provided by people who may be influenced by cognitive biases, social preferences, and cultural norms [54]. Even though people issues are vital for the processes in health care, very little research investigates the effects of human behavior on process performance in this industry. A promising opportunity to come up with more realistic health care operations management theories and to develop models which take into account human behavior is provided by experimental research. While behavioral experiments are a well-established research methodology for studying human issues in many disciplines including several business disciplines and medical research, combining findings from behavioral operations management with health care applications is a virtually untouched area. In line with Gino and Pisano [37] who argue that “human beings are critical to the functioning of the vast majority of operating systems, influencing both the way these systems work and how they perform,” we recommend a classification of behavioral health care operations management focusing on the human beings involved in hospitals. These can be divided into hospital staff (service providers) and patients (service receivers). Patients can be further distinguished into inpatients and outpatients, while hospital staff can be further differentiated into hospital management, surgeons and nurses (see Figure 5.1). Most studies discussing patients’ behavior relate to patient satisfaction, e.g. perceived waiting time for outpatients [45] or general patient satisfaction [11] for inpatients. The coverage of staff behavior in hospitals is diverse. Hospital management’s behavior is well covered by general behavioral operations management studies. Nurses’ behavior is often discussed in the context of job satisfaction, see Chang et al. [20], Irvine and Evans [46], and Jamal and Baba [47]. While surgeons make both medical and management decisions, in the literature their behavior is mainly discussed in the context of medical decision making. Bornstein et al. [14] and Bland and Altman [10] observe biased surgeons’ decision making. Moskowitz et al. [63] state that doctors might show non-optimal behavior “dealing with uncertainty, risks, and trade-offs in critical decisions” and claim that the anchoring and adjustment heuristic exists in medical decision making. To further classify the field of behavioral health care operations management, we consider the four main hierarchical levels in health care decision making presented by Hans et al. [39]: The strategic (e.g. case mix planning), the tactical (e.g. master surgery scheduling), the offline operational (e.g. surgery scheduling), and the online operational (e.g. rescheduling of surgeries) levels. In Figure 5.1 we assign the people making these decisions in hospitals to the hierarchical decision levels. Typically, the decisions on the strategic level are made by management, the tactical decisions are either made by management or surgeons, while most operational decisions are made by surgeons or nurses.

In this study, we approach the field of behavioral health care operations management by investigating surgeons’ behavior in the operating room (OR), one of the most researched resources in hospitals. Guerriero and Guido [38] cite more than 100 studies on operating room management and Cardoen et al. [18] write “in the last 60 years, a large body of literature on the management of operating theaters has evolved”. This comes as no surprise as

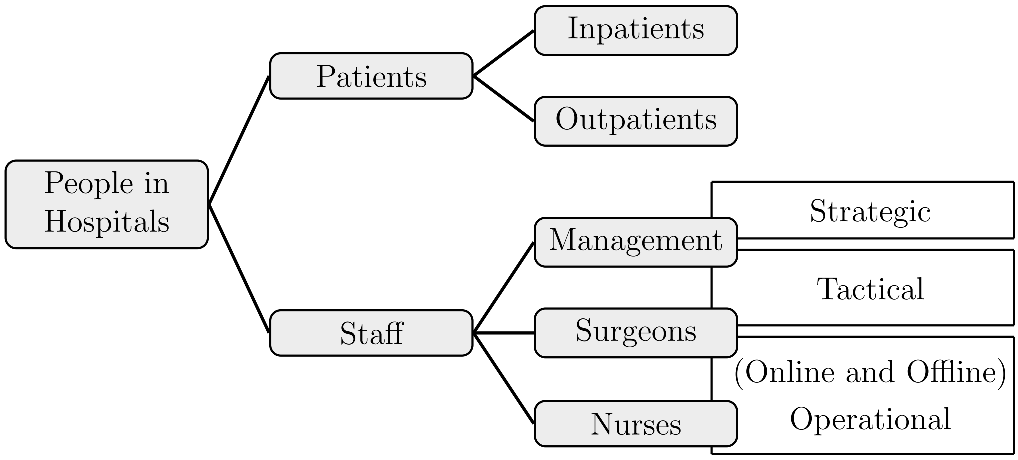

Figure 5.1: Classification for Behavioral Health Care Operations Management

around 40% of hospital expenses arise in the operating theater [22] and more than 60% of hospital admissions are for surgical operations [69]. Although the high importance of optimizing the usage of this scarce resource is evident, there still seems to be much room for improvement. Rhodes and Barker [71] report poor utilization of ORs and Pandit and Carey [67] argue that 10-40% of all scheduled elective surgeries are cancelled or rescheduled at least once. A major consequence of poor decisions made by OR managers is staff working overtime as mentioned by Wachtel and Dexter [92].

Low OR utilization, rescheduling of surgeries and staff overtime are consequences of poor planning of surgery durations. Obviously, centralized planning cannot account for the specific patient knowledge of the responsible surgeon. Therefore, it is common practice in most hospitals that surgery durations are planned independently by the surgeon in charge of the patient. There are a few studies indicating non-optimal behavior of surgeons considering operating room management. Yule et al. [94] conduct a literature review on non-technical skills of doctors in the OR and they conclude that non-technical skills such as planning skills, resource management, and communication are often neglected, despite being vital for efficient OR management. Carter [19] show one example where doctors only considered fairness when planning the ORs but neglected negative consequences on other units and Abouleish et al. [2] state that OR management is often based on convenience and tradition rather than on efficiency optimization. A systematic underestimation of surgery durations is found by Dexter et al. [24]. Wachtel and Dexter [92] discuss that the newsvendor model could be used to determine the time period, where staff is required in the OR. They also provide a literature review on behavioral newsvendor studies as they suspect biases known from the newsvendor model to hold in the staffing problem as well. However, they do not account for differences between the operating room staffing problem and the inventory newsvendor problem. Furthermore, they do not carry out an experimental investigation. Hence, the details on how to use the newsvendor model in an OR setting are still an open issue.

In our study, we undertake an experimental study with senior surgeons to test the behavioral effects of planning surgery durations. The main contribution of our research is to provide better understanding of surgeons' behavior when planning surgery durations. We can show systematic deviations from optimal planning. Awareness of these deviations helps in managing decision situations, such as communicating transfer prices for OR utilization or in developing debiasing methods to improve the planning results. Furthermore, we show the significance of human behavior in health care operations management. With this, we hope to encourage researchers to engage in the promising field of behavioral health care operations management. The remainder of our chapter is organized as follows: In the next section we present the problem of planning surgery durations. We explain similarities and differences of the surgery duration planning problem compared to the newsvendor problem familiar from inventory management, and discuss behavioral biases as known from newsvendor studies. In Section 5.3 we present our research question and derive our hypotheses. In Section 5.4 the experimental setup is explained. After discussing the results in Section 5.5 and providing an assessment of the experiment in Section 5.6, we draw conclusions and analyze managerial implications in the final section 5.7.

5.2 Planning of Surgery Durations

Planning of surgery durations is a challenging task for surgeons since every patient is different, surgery durations are uncertain, and bad planning leads

	Crossectomy	Cholecystectomy	Joint fracture
Surgeries planned too long	83%	52%	29%
Surgeries planned too short	14%	47%	67%
Mean (st. dev) planned durations	90.0 (0.0)	63.3 (6.5)	66.0 (21.5)
Mean (st. dev) realized durations	68.1 (24.0)	65.9 (26.8)	79.9 (45.2)
Average plan deviation	+ 31%	- 4%	- 17%

Table 5.1: Comparison of planned and realized durations of three different surgeries

to undesirable consequences. To obtain first insight into planning behavior in real life, we have analyzed 6 months (12/2011 - 05/2012) of surgery data from a German university hospital. We compare the planned and realized durations of three exemplary operations from different specialties: Varicose veins crossectomy and stripping, cholecystectomy, i.e. the surgical removal of the gallbladder, and a specific joint fracture surgery. We compare the planned and the realized durations of these three surgeries in Figures 5.2-5.4 and present some additional information in Table 5.1. Crossectomy and stripping was systematically planned too long (one tailed, Wilcoxon $p < 0.005$), cholecystectomy surgeries were on average planned close to the expected duration (two tailed, Wilcoxon $p = 0.978$), and joint fracture surgeries were significantly planned too short (one tailed, Wilcoxon $p = 0.030$). All surgeries have in common that the planned durations showed less variation than the realized ones. We derive three main findings from these data. First, it is obviously not possible to always plan the exact surgery duration, as surgery times are stochastic. Second, different specialties seem to plan their surgeries in a different way, which may be a consequence of different cost structures. Third, some surgeries are systematically planned too long, while others are systematically planned too short. In general, surgeons have a consistent behavior, since their decisions are not random.

Planning of surgery durations is a complex task due to two characteristics of the problem.

- First, variability in surgery durations exists.
- Second, as both planning too long and too short durations results in different negative consequences, a trade-off decision has to be made.

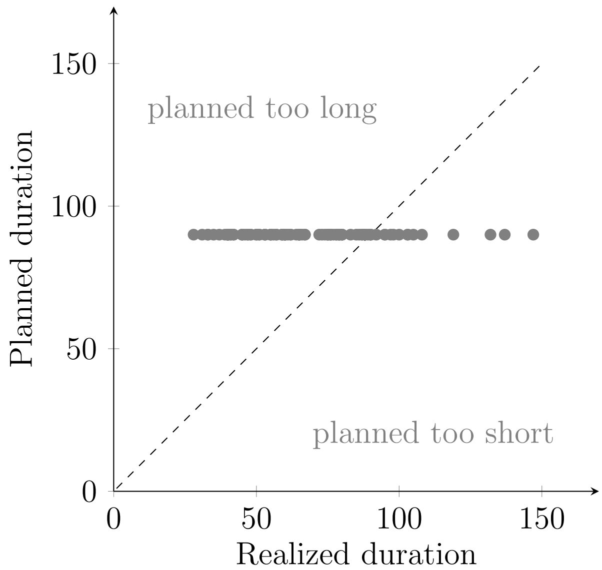

Figure 5.2: Comparison of planned and realized durations — Crossectomy

There are two reasons for variability of surgery durations: Uncertainty and "diversity of situation". Uncertainty in surgery durations is caused by many factors that cannot be pre-determined. A typical example is unexpected bleeding that increases the duration. With "diversity of situation" we take into account a priori known factors, such as patient age or OR-team experience. There is a lot of literature on estimating the distribution of surgery durations. Strum et al. [81], May et al. [57] and Stepaniak et al. [80] use lognormal distribution to model surgery times, while Silber et al. [77] estimate surgical and anesthesia procedure times using data obtained from the US Medicare system. All these studies show that there is significant uncertainty in surgery durations. Furthermore, there are several empirical studies showing that surgeons' estimates do not meet the realized durations. Wright et al. [93] compared time estimates of software scheduling systems to those made by surgeons. Even though the software systems could not outperform the surgeons, modeling could help the surgeons to improve their time estimates. Eijkemans et al. [28] demonstrated that, in addition to the surgeons'

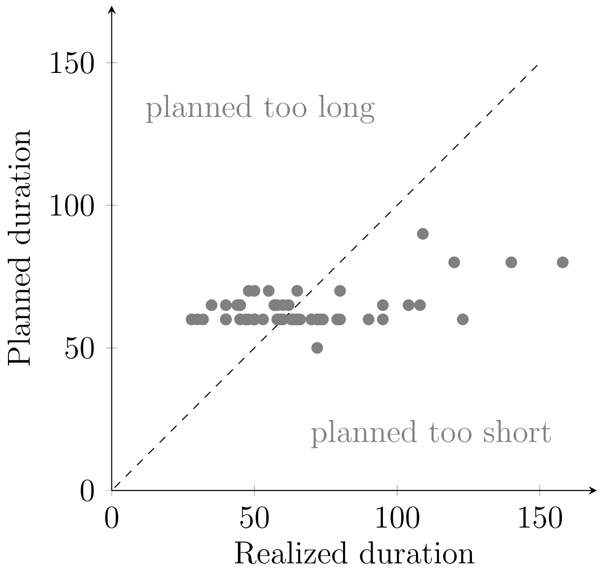

Figure 5.3: Comparison of planned and realized durations — Cholecystectomy

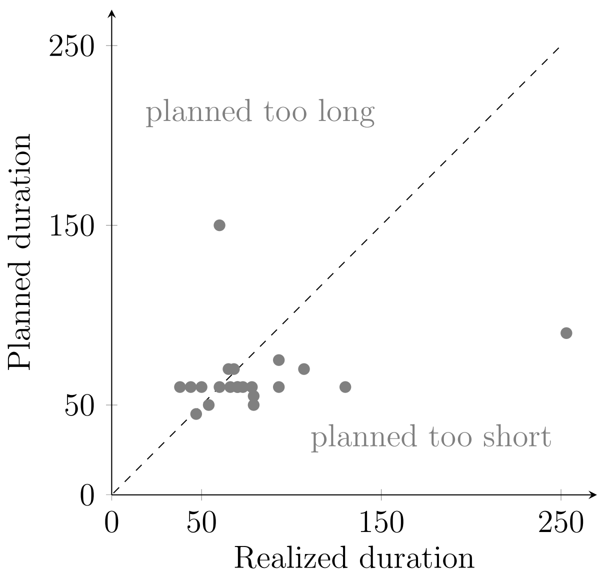

Figure 5.4: Comparison of planned and realized durations — Joint Fracture

estimates, diversity of situation factors such as surgery and team characteristics and, to a lesser extent, patient characteristics like age and body mass index proved to be relevant for surgery times. These results demonstrate that surgeons' estimates do not include all diversity factors.

The second driver of the complexity of planning surgery durations is that both planning with too long and too short time estimates for surgeries leads to undesirable consequences. If the realized surgery duration falls below the planned duration, OR idle time will be the consequence. In line with Strum et al. [82], we define this as underutilization. If the realized surgery duration is above the planned surgery duration multiple consequences might occur. The scheduled surgery or following surgeries might end after regular working hours, i.e. staff works overtime. Following surgeries may have to be rescheduled which involves a considerable organizational effort and reduces patient and staff satisfaction. Overtime or time of rescheduled surgeries caused by planning too short surgery durations is defined as overutilization. Since some compensation of under- and overutilization may occur, the expected under- and overutilization of an OR is less than the time surgeries were planned too long and too short, respectively. To obtain a rough idea about the consequences of inaccurate planning, we used data from the hospital mentioned above. We performed regression analyses to determine the effects on planning too long and too short on OR under- and overutilization, respectively. In Figure 5.5 we compare for each OR and each day the number of minutes surgeries were planned too long with total operating time performed. We observe that the more minutes surgeries were planned too long, the less was the total operating time (0.348 Minutes of operating time per minute planned too long, $p = 0.007$) and thus the more idle time occurred. We further compared the number of minutes planned too short with the minutes of overtime (between 4pm and 10pm). As presented in Figure 5.6, the more minutes surgeries were planned too short, the more overtime occurred (0.483 minutes of OR overtime time per minute of planned too short, $p < 0.005$). Both underutilization and overutilization of ORs are associated with additional costs. Typically, costs for underutilization are created by idle OR and staff capacities, while costs for overutilization represent the additional overtime payments and costs for reorganizing the schedule. These costs can also include further negative effects on employee satisfaction (for working unplanned overtime) and patient satisfaction (for rescheduling their surgeries and for increased waiting times). Olivares et al. [66] state that "the costs of OR idle time were perceived, on average, as approximately 60% higher than the costs of schedule overrun," while Wachtel and Dexter [92] assume that the costs of OR overutilization are twice as high as the costs of OR underutilization. Thus, there is no clear ratio of these costs in the literature, which might be caused by different assessments of under- and overutilization in different hospitals. To minimize the expected costs of under- and overutilization, Strum et al. [82] propose a minimal cost analysis model. The sum of cost-weighted under- and overutilization is defined as OR inefficiency by Dexter and Traub [25]. The minimal cost analysis model is mathematically equivalent to the well-known newsvendor problem, which is also used by Olivares et al. [66] to conduct a structural estimation of the costs for under- and overutilization. All studies using the minimal cost model have in common that a rational decision maker is assumed. They do not take into account that a human decision maker may not act rationally.

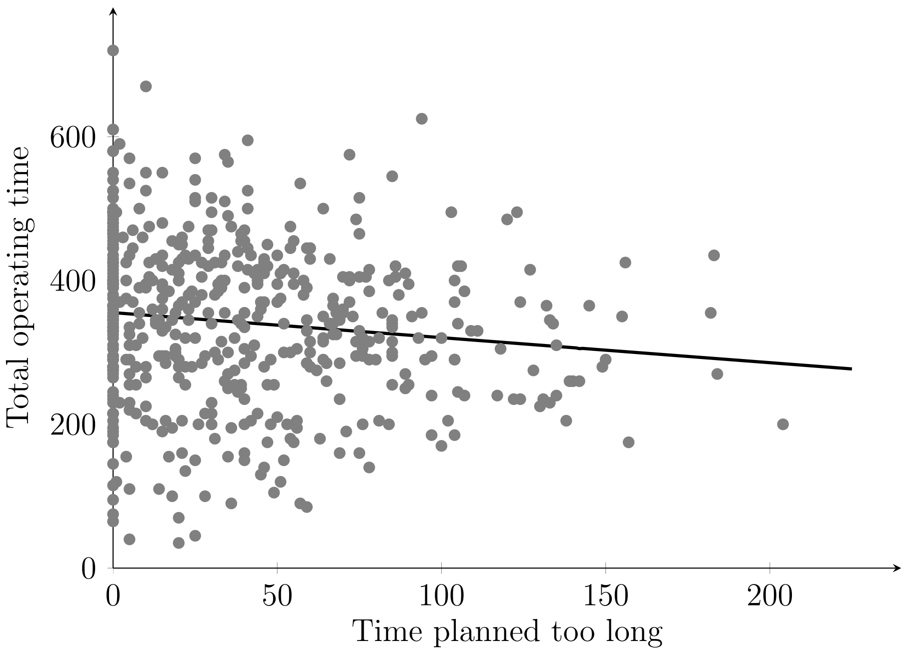

Figure 5.5: Consequences of planning too long and too short (in minutes)

As decision biases in the newsvendor problem are well-researched in the

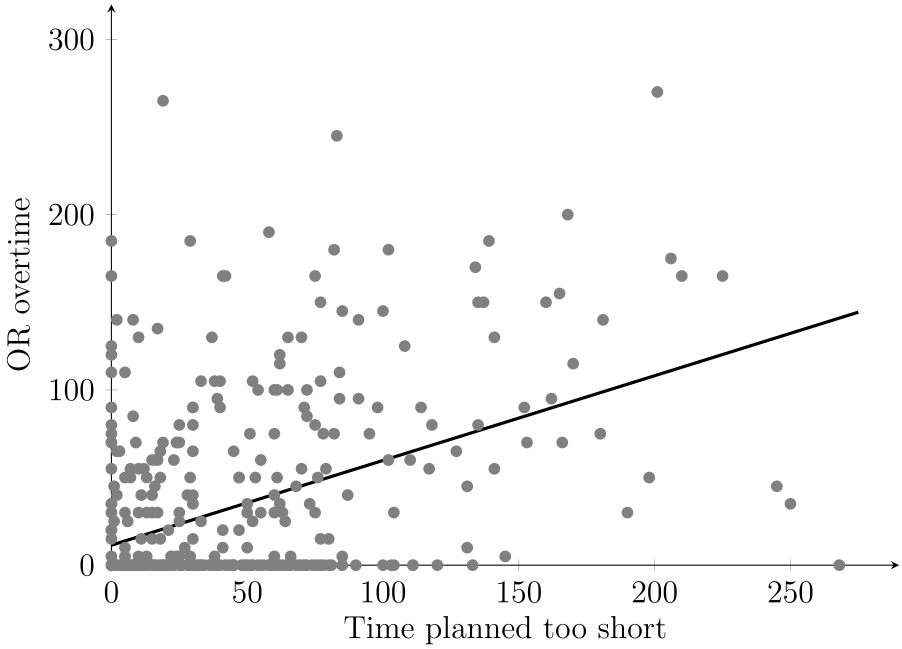

Figure 5.6: Consequences of planning too long and too short (in minutes)

field of inventory management, we compare the surgery planning problem to the inventory management situation. The structure of the problem seems to be the same. In both problems, the decision situation can be characterized as follows.

- Decision under uncertainty with known distribution.
- Trade-off between planning (ordering) too long (too many) or too short (too little) durations (products).
- Optimal solution (i.e. cost minimum or profit maximum) can be derived analytically.

On the other hand, the two problems are obviously different. Some important differences include the following:

- Planning time versus quantity.

- Consequences when not reserving enough time (not ordering enough quantities): Overtime with additional (penalty) costs in the OR case versus opportunity costs for lost sales in the inventory situation.
- Different decision makers: Surgeons (with no management training) versus inventory managers.

Due to the similarities, however, we expect that some behavioral effects in the inventory problem can be found in the OR planning problem as well. One bias that is consistently found in all newsvendor studies is the mean anchor effect, where orders are too high when the optimum lies below mean demand and too low when the optimum lies above the mean demand. Schweitzer and Cachon [76] are the first to describe this pattern and they also discuss possible explanations for the observed behavior. They find support that decision makers use the mean as an anchor and only insufficiently adjust towards the optimal solution. They also find some support that decision makers anchor at the previous order quantity and adjusts towards the previous demand which is called “chasing demand heuristic”. Based on their seminal paper a number of follow-up studies analyze further behavioral effects. Benzion et al. [9] encounter similar biases for different demand distributions and Bolton and Katok [12] analyze learning effects detecting small but significant long term learning effects when a decision is repeated over 100 times. In contrast to the classical newsvendor problem, where ordering too little results in lost profits, the consequences of planning too short in our context differ. Too short planning of surgeries is associated with additional costs, since operations have to be finished. Therefore, the planning of surgeries is similar to a newsvendor situation where penalties occur when ordering too little and demand has to be fulfilled. Schiffels et al. [74] analyze the impact of “penalty” costs instead of opportunity costs and they find that order quantities are consistently higher in contexts with penalties than in those with opportunity costs for ordering too little. Kremer et al. [51] compared a classical newsvendor situation (operations frame) with a context-free but mathematically equivalent situation (neutral frame). They discovered that the bias towards the mean demand was much stronger in the operations frame than in the neutral frame. Furthermore, Bolton et al. [13] observed that even if the direction of behavioral effects is the same the magnitude of effects may differ for students and experienced managers. Therefore, context as well as professional background matters when considering behavioral biases. Even though newsvendor biases are likely to be relevant for our OR planning problem, biases in the OR context might be different from those in the inventory context.

We conclude that the complexity of planning a surgery's duration is considerable, even if optimal planning can theoretically be obtained with the newsvendor model. In everyday life surgeons lacking training in capacity management plan the surgery durations. As several studies show that inventory managers do not behave optimally in the related inventory situation, and since some studies observe biased surgeon behavior in general, we expect that surgeons do not plan optimally. To the best of our knowledge, we are the first to employ an experimental study to analyze surgeons' behavior.

5.3 Research Question and Hypotheses

To investigate surgeons' behavior when planning surgery durations we set up an experimental study. The experimental design pursued three main goals: Simplicity, realism, and comparability to literature. Simplicity of an experiment guarantees that it is understood by participants, i.e. that errors by misinterpretation are avoided and that behavioral effects can be isolated. Realism is vital to apply findings to real life situations. As noted previously, many studies have been conducted on inventory management. In order to insure comparability of our study to already known findings from the inventory management literature, we designed our study similar to those.

The experimental task that we consider is the planning of one surgery duration. We define the costs c^u for each minute of underutilization, c^o for each minute of overutilization, and c for each minute of used OR capacity. (Note: In classical newsvendor studies, costs for ordering too much (underutilization) are defined as overage costs, and costs for ordering too little (overutilization) are defined as underage costs.) Using these costs, the costs for one surgery depend on the planned duration p and the realized duration D as follows:

$$C(p, D) = c^u \cdot \max\{p - D, 0\} + c^o \cdot \max\{D - p, 0\} + c \cdot D. \tag{41}$$

Analogous to the newsvendor problem, the planned duration p^* that minimizes the expected costs $E[C(p)]$ is:

$$p^* = F^{-1}\left(\frac{c^o}{c^o + c^u}\right), \tag{42}$$

where F^{-1} denotes the inverse of the cumulative distribution function of the realized duration D. For the sake of brevity, we denote p^* as "optimal duration" in the following. $\frac{c^o}{c^o+c^u}$ is defined in the literature as the critical ratio. As stated before, the specific costs for underutilization and overutilization may differ across hospitals. To account for different situations on the one hand, and to be comparable to the inventory literature on the other hand, we differentiate between two exemplary cases. The "low quantile case" with relatively high underutilization costs c^u indicates a hospital where idle capacities are of greater concern, while the "high quantile case" with relatively high overutilization costs c^o indicates a hospital where overtime is of greater concern. Furthermore, we provide information about the distribution of the surgery duration. To avoid different assessments of potential causes for diversity of situation, we exclude details about patient or OR-team characteristics. Thus, we can refer the observed behavior to the underlying decision problem. As we present a distribution for the duration, every subject has the same information about the uncertainty of the surgery duration. Since our goal is to investigate how doctors behave in real life situations, the subjects of our experiment are doctors with relevant experience, i.e. doctors who are responsible for scheduling their surgery durations.

Based on our literature review, e.g. [2, 92, 76] and our empirical observations in Section 5.2 we derive the first two hypotheses of our experimental study:

H1: The surgeons behavior is not random but is consistent.

H2: Doctors do not behave optimally.

To measure the degree of non-optimality, we define the relative cost increase (see Equation 43) as the percentage of avoidable costs due to non-optimal planning. For each decision of our subjects in our experimental study we calculate the difference between the expected costs of the planned duration and of the optimal decision and divide this difference by the expected costs of the planned duration.

$$I(p) = \frac{E[C(p)] - E[C(p^*)]}{E[C(p)]} \tag{43}$$

The most striking bias known from the inventory newsvendor literature discussed above is the mean anchor effect, or mean bias [76]. As this bias can be found in all studies, we formulate the next two hypotheses:

H3: Doctors plan longer durations than optimal in cases where the optimal duration p^* is below the average duration μ of a surgery.

H4: Doctors plan shorter durations than optimal in cases where the optimal duration p^* is above the average duration μ of a surgery.

As explained in the previous section, planning the length of a surgery corresponds to the penalty cost newsvendor problem, since operations have to be finished. Besides a mean anchor effect Schiffels et al. [74] observe order quantities that are consistently higher in a penalty cost problem than in a situations with lost profits for ordering too little. Taking these findings into account, we formulate the fifth hypothesis:

H5: Planned durations are more strongly biased away from the optimal duration in cases where the optimal duration p^* is below the average duration μ than in cases where the optimal duration p^* is above the average duration μ.

As the experiment from our study involves asking surgeons about planning surgery durations, we expect context and background driven effects to be different than in studies, where undergraduate students performed inventory experiments [51, 13]. Thus, we derive our sixth hypothesis:

H6: Surgeons' behavior is different compared to decision makers' behavior in inventory newsvendor studies.

As demand chasing is a heuristic that often appears in newsvendor experiments, we expect doctors to react to the realized duration of the previous surgery. If the realized duration of the previous surgery (D_{t-1}) was longer than the planned duration (p_{t-1}), we expect doctors to plan a longer duration (p_t) for the following surgery, if it took shorter, we expect them to plan a shorter duration. To detect demand chasing, we count each decision as moving towards the previous duration if $A_t = \frac{p_t - p_{t-1}}{D_{t-1} - p_{t-1}}$ is positive, if A_t is negative we count the decision as moving away from the previous duration [51]. If $A_t = 0$, we count the decision as repeat plan. We formulate our seventh hypothesis:

H7: Demand chasing affects surgeons' behavior.

We expect that the answers to these hypotheses provide valuable insight in surgeons' behavior when planning surgery durations. To test the hypotheses, we set up an experimental study.

5.4 Experimental Setup

As our research question focuses on the behavior of surgeons when planning surgeries, we chose only doctors with relevant experience in scheduling surgeries as subjects. The experiment was carried out by 40 doctors from three German university hospitals, 20 in the low quantile case and 20 in the high quantile case. They were all senior physicians or chief physicians with an average age of 43 years. None of them had previous knowledge of the newsvendor model. We used a between subject design with 20 participants in each treatment. Since it was not possible to bring 40 senior or chief physicians to our computer laboratory, all experiments were conducted in hospitals. We conducted the experiments in a separated office room with a computer and we ensured that the physicians had no time pressure and that they were not disturbed or interrupted during the experiment. At the beginning of the experiment we provided the instructions (see Appendix). The subjects were asked to schedule the duration of one surgery at a time. The following information was provided:

- A uniform distribution between 100 and 200 minutes for surgery duration.
- OR costs c per reserved minute (underutilization costs $c^u = c$).

	Low Quantile Case	High Quantile Case
Costs for planned time c	90	30
Costs for overtime s	120	120
Underutilization costs $c^u = c$	90	30
Overutilization costs $c^o = s - c$	30(=120-90)	90(=120-30)
Critical ratio	0.25	0.75
Optimal duration	125	175

Table 5.2: Costs and optimal planning times for low and high quantile case

- Increased costs per minute overtime s (overutilization costs $c^o = s - c$).

The uniform distribution was chosen for simplicity and for comparability to the inventory literature. In most studies, the uniform distribution is used even though normal or lognormal distributions better fit real life distributions. Benzion et al. [9] have shown that for an inventory setting the same behavioral effects are observed for different demand distributions. For simplification, we assume that each minute scheduled too long results in a minute of underutilization, and each minute scheduled too short results in a minute of overutilization. Details are depicted in Table 5.2.

The experiment was implemented in z-Tree [31]. Either the low quantile case or the high quantile case was tested for each subject. After an initial screen, where the subjects had to place a planned duration between 100 and 200 minutes, feedback about the realized demands and the occurred costs was provided. The subjects performed 20 decision periods. The demand for each round was randomly drawn in advance and the same for all subjects. After planning the 20 surgery durations the subjects answered a questionnaire. The average duration of each experiment was 25 minutes. Money was the only incentive used. Payments were based on total costs and ranged between 19 Euro and 39 Euro with a mean of 33.2 Euro. Thus, the average payment matched the income of experienced doctors.

5.5 Results

As expected, we observed average planned durations of all subjects that are significantly higher in the high quantile case (HQC) (162.2) than in the low quantile case (LQC) (149.5) (one tailed, Wilcoxon $p < 0.005$). In neither case did doctors plan the optimal duration. The box plots of the average planned durations per subject are presented in Figure 5.7. The box indicates the quartiles, and the ends of the "whiskers" show the lowest and the highest datum within the 1.5 interquartile range. Outliers are illustrated by dots beyond the whiskers. The average planned duration of all subjects is marked with a bold circle for both cases. The average duration of 150 minutes and the optimal durations for both the low quantile case (125 minutes) and the high quantile case (175 minutes) are illustrated by dotted lines. In Figure

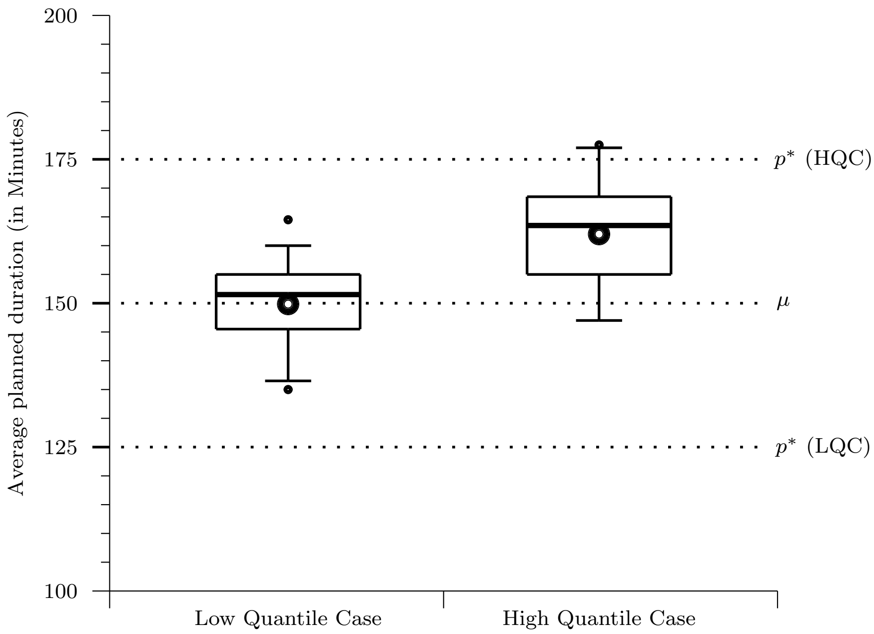

Figure 5.7: Average Planned Durations

5.7, it is apparent that the average planned durations per subject are distributed around the average planned duration of all subjects in both cases. In fact, they are approximately normally distributed (K-S test for normal distribution, low quantile case: $p = 0.714$, high quantile case: $p = 0.993$). Therefore, we can confirm our first hypothesis: The surgeons behavior is not random but is consistent. As stated previously, the planned durations differ from the optimal duration in both cases. On average, in the low quantile case the planned durations are highly above the optimal duration of 125

	Low Quantile Case	High Quantile Case
Avoid overutilization costs (c^o)	65%	65%
Indifferent	15%	10%
Avoid underutilization costs (c^u)	20%	25%

Table 5.3: Motivation when planning surgery durations

minutes (Wilcoxon $p < 0.005$) and close to the the mean duration of 150 minutes (Wilcoxon $p = 0.493$). In the high quantile case the planned durations are below the optimal duration of 175 minutes (Wilcoxon $p < 0.005$) and above the mean duration (Wilcoxon $p < 0.005$). We calculated the average relative cost increase (43) for all participants and all rounds. We found average relative cost increase of 3.3% in the low quantile and 3.4% in the high quantile case. The relative cost increase is higher in the high quantile case, as the total costs are lower in this case. This analysis shows that in our two scenarios about $\frac{1}{30}$ of OR costs (i.e. the part depending on staff and fixed capacities) may be saved by planning optimally. This confirms our second hypothesis: Doctors do not behave optimally. Our third and fourth hypotheses can be confirmed as well: Doctors plan longer durations than optimal in cases where the optimal duration is below the average duration of a surgery, and shorter durations than optimal in cases where the optimal duration is above the average duration. However, these biases are asymmetric, as the planned durations in the low quantile case are further away from the optimal duration than in the high quantile case. In the low quantile case, the planned duration is on average 24.8 minutes above the optimal duration, while in the high quantile case, the planned duration is on average 12.8 minutes below the optimal duration. Therefore, the bias away from the optimal duration is significantly stronger in the low quantile case (one-tailed Mann-Whitney U, $p < 0.005$), which confirms our fifth hypothesis: Planned durations are more strongly biased away from the optimal duration in cases where the optimal duration is below the average duration than in cases where the optimal duration is above the average duration. To gain more insights, we asked all subjects after the experiment whether they had sought to avoid overutilization or underutilization. The results depicted in Table 5.3 show that consistent in both cases the subjects tended to avoid overutilization. To test our sixth hypothesis, we compared our data with the corresponding data from the study of Schiffels et al. [74] (see Chapter 4). There, subjects where asked to order newspapers in a penalty cost based scenario with critical ratios of 0.25 and 0.75 and a uniform demand distribution between 0 and 100. In order to compare the results we shifted the data from [74] (see Chapter 4) up by 100. Therefore, differences in the order quantities/planned durations can be referred back to the different contexts. In the low quantile case, the values were significantly lower in Schiffels et al. [74] (see Chapter 4) (average: 146.0) compared to our study (average: 149.5) (one-tailed Mann-Whitney U, $p < 0.005$). In the high quantile case, the values were significantly higher in Schiffels et al. [74] (see Chapter 4) (average: 172.6) compared to our study (average: 162.2) (one-tailed Mann-Whitney U, $p < 0.005$). In both cases planned durations were more strongly biased towards the mean in our study. We conclude that different contexts, in our example reserving OR time versus ordering newspapers, do influence the behavior and confirm our sixth hypothesis: Surgeons' behavior is different compared to decision makers' behavior in inventory newsvendor studies. Consistent with inventory management studies, we further gained some insight into learning behavior. We assume learning if a significant trend in the planned durations towards the optimum exists. We determine the trend using linear regression. For the low quantile case it is evident that no learning occurred. Over the 20 decision periods, the trend of 0.165 is not significant ($p = 0.603$). As the mean planned duration of 149.5 minutes is already above the optimal solution of 125 minutes, the trend even moves away from the optimum (see Figure 5.8). In the high quantile case, the trend of 0.189 leads in the direction of the optimal duration (see Figure 5.9), but is not significant either ($p = 0.595$). This comes as no surprise, as Bolton and Katok [12] show significant learning effects in the long run only. To test our final hypothesis, we analyze whether demand chasing might explain the surgeons' behavior. We show the directions of planned durations for the low quantile case and the high quantile case in Table 5.4. Consistent with Kremer et al. [51], for all subjects, we compared the probabilities to adjust their planned duration in the direction of the previous realized duration and away from it. We found that subjects are more likely to adjust their planned duration in the direction of the previous realized duration than away from it, both in the low quantile case ($p < 0.005$) and in the high quantile case ($p < 0.005$). Therefore, we can confirm our seventh hypothesis: Demand chasing occurs.

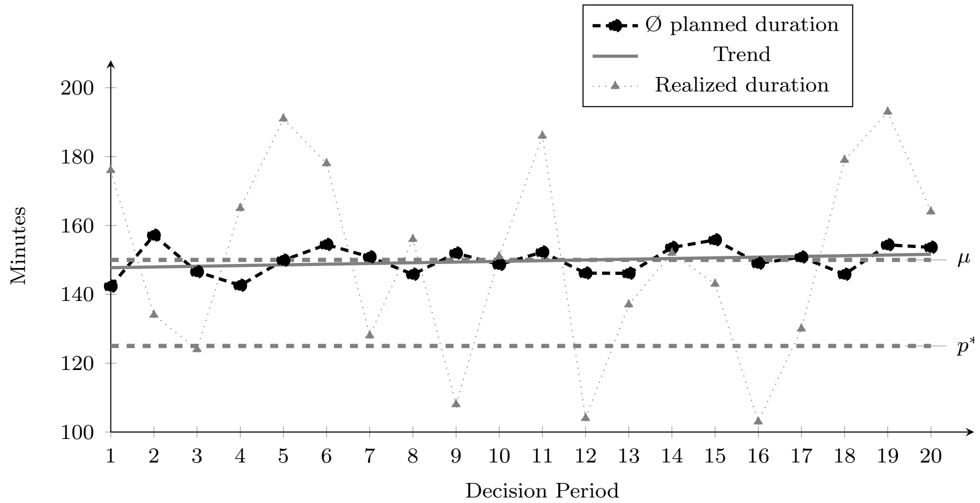

Figure 5.8: Average Planned Duration, Realized Duration and Trend per Period (LQC)

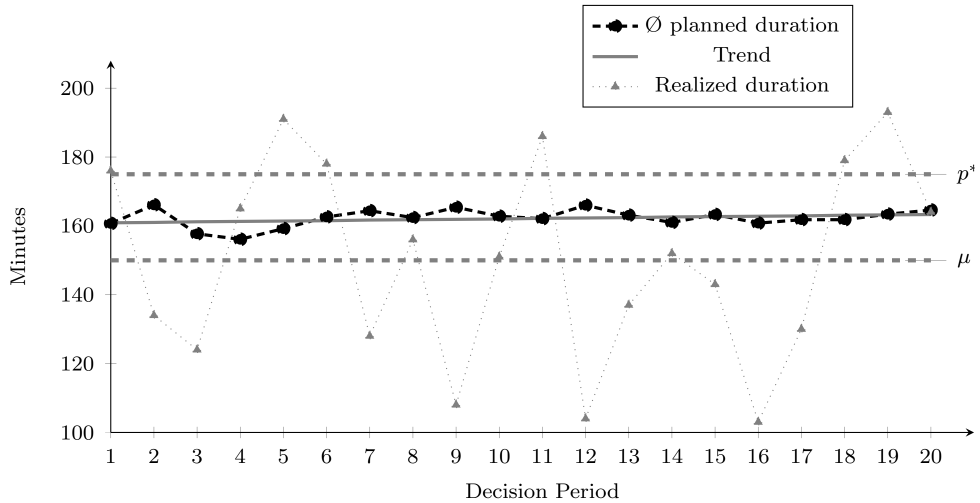

Figure 5.9: Average Planned Duration, Realized Duration and Trend per Period (HQC)

	Low Quantile Case	High Quantile Case
Repeat plan	30%	49%
Towards previous duration	45%	37%
Away from previous duration	25%	14%

Table 5.4: Directions for planned duration

5.6 Assessment of Experiment

We conducted an ex-post assessment of our experiment to validate whether the newsvendor model fits the reality in planning surgery durations and whether the experimental set up suited our goals simplicity, realism, and comparability to literature. The newsvendor problem is a cost-minimization model based on a trade-off decision under uncertainty. We asked our subjects whether cost pressure influences their decisions and whether there is uncertainty when planning surgery durations. They reported that cost pressure has a great impact on their decisions (average: 4.6 out of 7, 1 being no impact, 7 being huge impact). They reported as well that they do encounter uncertainty when planning surgeries (average: 4.4 out of 7, 1 being no uncertainty, 7 being huge uncertainty). We conclude that the newsvendor framework fits well for modeling planning surgery durations. Furthermore, we investigated whether our experiment was simple, realistic, and comparable as demanded in Section 5.3. We obviously chose a simple experimental setting, and all subjects stated they understood their task. On the other hand, it is obvious that our experiment simplifies the real life situation, so we analyzed whether the task was still realistic enough for the results to be of any value. In the questionnaire, we asked the subjects whether the experiment reflected the real life situation. The response, an average of 3.8 out of 7 (1 being not realistic, 7 being extremely realistic), showed the experiment was considered realistic, but did not reflect all relevant information. To gain insight into required information, we asked what was missing. The top three responses (reflecting about 90% of replies) were information on the surgery team, the type of surgery and the patient characteristics. These are all measures of diversity of situation as discussed in Section 5.2. Considering diversity of situation is vital for precise estimations of surgery duration distributions, especially the three characteristics surgery team, type of surgery and patient characteristics are of significant influence [28]. As we gave full disclosure of the distribution in our experiment, we do not account for diversity of situation and concentrate on uncertainty of the duration. Therefore, the "missing" information does not affect the results. By offering no further information about diversity measures like surgery team, surgery type, and patient characteristics, we avoided different interpretations of the duration distribution. We conclude our experiments to be realistic and the results to be applicable to real life. As our experiment was configured analogously to most inventory studies, e.g., we chose a uniform demand distribution and compared two margin cases with critical ratios of 0.25 and 0.75 (see e.g. Schweitzer and Cachon [76], Bolton and Katok [12], Schiffels et al. [74] (see Chapter 4)), the findings were compared to those of inventory studies and explanations could be adapted taking into account context sensitivities

5.7 Conclusion

Many studies have shown that human behavior has great effects on operations management decisions. Although the operating theater is the most expensive resource in hospitals, and especially the operational planning level is crucial for its efficient usage, the behavior of health care decision makers in hospitals is widely ignored in research. In our study, we demonstrated significant non-optimal planning of surgery durations by experienced surgeons and we showed that relative cost improvements of about 3.3% could be achieved in our setting. In a university hospital with an annual budget of 150 million Euro and annual OR costs for staff and fixed capacities of about 45 million Euro, the savings potential is thus about 1.5 million Euro a year. We verified that biases known from inventory newsvendor studies exist as well, but that the different context and the penalty cost situation changed the magnitude of the biases. Planned durations showed greater biases in situations where idle capacities were expensive compared to situations where overtime was expensive. Our study demonstrates that even in a simplified environment the planning behavior of surgeons is not efficient, systematic biases can be observed, and avoidable costs accrue.

Our work has several limitations. To be consistent with the inventory literature, we chose a uniform distribution of surgery durations in our experimental setup. Typically, surgery durations tend to follow a lognormal distribution [81]. Benzion et al. [9] show that in an experimental setting, different distributions yield the same behavioral biases. However, a possible next step could be to set up experiments with more realistic duration distributions. We further ignore variability of durations due to diversity-factors such as different patient and OR-team characteristics. By this we were able to concentrate on the impact of uncertainty on trade-off decisions. An interesting research project would be to analyze surgeons' behavior when diversity-information is additionally available. In our experiment, we did not find any significant learning behavior. For long run experiments we would expect small learning effects as described by Bolton and Katok [12]. The strongest learning effects can be expected if the time intervals between the same surgery situations are not too long. The same situation, i.e. the same combination of surgery type, OR team, and patient characteristics like age and body mass index does not appear that often during short time intervals. Therefore, the investigation of cross-learning effects, i.e. learning over a sequence of different surgeries, is a promising field for further research.

As planning of surgery durations is a task of high economic impact for all hospitals, and as we are able to show significant and systematic non-optimal behavior of experienced surgeons, important managerial implications may be derived. We showed that in hospitals where idle capacities are more expensive than overtime, surgeons planned too long, while they planned too short when overtime costs exceed costs for idle capacities. Hospital management could react to these findings and create incentives for planning optimal surgery durations, develop debiasing methods to obtain better planning results, or improve the planning skills of surgeons by training. In addition to these managerial implications, we hope to encourage future research in this field. We discovered some systematic errors in planning surgery durations. We were able to demonstrate that biases in other fields like inventory management are relevant in health care settings as well. Although the health care sector is in terms of number of employees the largest industry in industrialized countries, and human decision making plays an important role, little research has been conducted in the field of behavioral health care operations management. We are convinced that many biases in this field are still to be discovered, and managing these biases could greatly impact the health care sector. Therefore, we strongly encourage researchers, both from the field of behavioral operations management and from health care operations management, to conduct further research in this promising area.

Chapter 6

Conclusion

6.1 Summary and conclusions

In the thesis at hand, we investigated the field of operating room management from two directions. First, we discussed the influences a tactical MSS has on the downstream units ICU and the regular wards in Chapters 2 and 3. Second, behavioral aspects of operational operating room planning are presented in Chapters 4 and 5. In Chapter 2, we developed an analytical model to detect the exact distributions of post-operative patients in the downstream units ICU and ward resulting from an MSS. We further defined possible cost functions based on these distributions and developed algorithms to create cost-minimizing MSSs. These algorithms were tested and compared in a numerical case study with three scenarios. The costs in the downstream units could be reduced in a range from 4 to 12%. The data was based on real life data of a Dutch hospital, however, the scenarios were artificial. In Chapter 3, a real life case study with data from a German university hospital was carried out. The model was adjusted to meet the requirements of the hospital. For instance, to consider emergency surgeries on weekends, an artificial specialty WE was introduced. In the case study three potential MSSs were compared and discussed. As a result, peak occupancies in the ward could be reduced by about 7% through adapting the MSS. Thus the model proved to be valuable when designing or adapting MSSs.

Chapter 4 provides a theoretical foundation for behavior in penalty-based newsvendor situations. We showed that people ordered more goods in situations where ordering too little was penalized by costs, than in situations where ordering too little resulted in opportunity costs. This is of high relevance, as all previous newsvendor studies assumed an opportunity cost based problem, and many situations with penalties exist. One example of a penalty cost based problem is the planning of surgery durations. If surgeries are planned too short, penalties for staff overtime and potential rescheduling of following surgeries occur. In Chapter 5, we discussed the problem of planning surgery durations and behavioral biases of surgeons in detail. We showed that surgeons behave in non-optimal ways and reveiled systematic biases similar to those in newsvendor situations. Surgeons reserved more time than optimal in situations, where the optimal planned duration was below the expected duration, and less time than optimal in situations, where the optimal planned duration was above the expected duration. However, some context related differences occurred. For instance, the shift towards the expected duration was stronger than in inventory newsvendor studies.

6.2 Final Remarks and future research directions

Each approach presented in this thesis created new ideas for future research. Incorporating stochasticity and downstream units in operating room planning could be further expanded. A possible future research project might involve applying similar approaches for strategic problems. They could be used, e.g., to obtain more information on the effect of OR capacity extension projects on downstream units. This could help reducing bottleneck situations in these units.

Another promising future research area is the further investigation of behavioral effects in operating room planning and other areas of health care operations management. As human behavior greatly influences decision making in health care, further insights are valuable. Possible research projects might be to investigate behavioral effects in admission planning, staffing, and surgery scheduling. Behavioral investigation of hospital management or nursing staff could also help shed light on health care decision making.

Appendix A

Instructions for Experiments

A.1 Instructions experiment Chapter 4

Instructions: The instructions are translated from German and shortened since the original instructions also contain examples and screen shots. Furthermore, they do not contain the price and cost values and the payment figures since they are different for the three margin cases. Differences in the penalty cost and the opportunity cost problem are set in italics. The instructions consist of four parts though Part 1 is identical for both problems.

1. General information: You are about to participate in an experiment in decision making. You will receive a fixed payment of €4 for your appearance. Furthermore, in the course of the experiment you can earn a considerable amount of money depending on your decisions. In the experiment, all monetary amounts are specified in Experimental Currency Units (ECU). They are converted according a fixed exchange rate into Euro (see payment determination) at the end of the experiment. The experiment is followed by a short questionnaire and, afterwards, you will be paid in cash. All your decisions and answers will be treated confidentially. Please read the following instructions carefully. If you have any questions, please raise your hand. An instructor will come to your place and answer your questions. During the experiment you have to switch off your cell phone and communication with other participants is prohibited. If you fail to comply with these rules, we

will exclude you from the experiment and you will receive no payment.

Opportunity cost problem (Part 2-4):

2. Experimental task: Your job is to determine the order quantity of a product before you know the demand. You know that the demand is equally probable for any value between 0 and 100. If your order quantity exceeds the demand, the remaining products are worthless. *If the demand exceeds your order quantity, the unsatisfied demand expires.* For each product you order, you pay a price of ECU ... to the wholesaler (the costs per product unsold correspond to ECU ...). *For each product sold you will receive a price of ECU ... from your customers (the opportunity costs for each product ordered too little corresponds to ECU ...).*

- You cannot sell more products than are demanded.
- *You cannot sell more products than you have ordered.*

3. Experimental procedure: The experiment consists of 30 rounds and the demand in each round is independent of past demand. Every round consists of two screens. The first screen summarizes the information already given in the instructions. Furthermore, you have to enter the number of products you want to order in a red box and press the button “OK”. Please take sufficient time to make your decisions. Once all participants have confirmed their entry, the second screen appears. On the second screen, your order quantity is given again and you receive information about the realized demand. Furthermore, the resulting *gains/losses* are listed. When all participants have pressed “OK”, the next round starts.

4. Payment determination: You receive a fixed payment of €4 for your appearance. Furthermore, you can earn additional money dependent on your performance in the course of the experiment. At the end of the experiment, the *gains/losses* in ECU incurred in all rounds are added together. Your payoff is the resulting amount which is converted by a factor of ECU ... = €1 plus the €4 you receive for your appearance. In the event that you have generated a total loss, you still receive your show up fee.

Penalty cost problem (Part 2-4):

2. Experimental task: Your job is to determine the order quantity of a product before you know the demand. You know that the demand is equally probable for any value between 0 and 100. *The demand of the customers has to be satisfied.* If your order quantity exceeds the demand, the remaining products are worthless. *If the demand exceeds your order quantity, you have to reorder products instantly at a higher price.* For each product you order, you pay a price of ECU ... to the wholesaler (the costs per product unsold correspond to ECU ...). *For each product ordered too little, you have to pay a price of ECU ... to the wholesaler (the additional costs for each product ordered too little correspond to ECU ...).*

- You cannot sell more products than are demanded.
- *You have to reorder products if demand exceeds the order quantity.*

3. Experimental procedure: The experiment consists of 30 rounds and the demand in each round is independent of past demand. Every round consists of two screens. The first screen summarizes the information already given in the instructions. Furthermore, you have to enter the number of products you want to order in a red box and press the button "OK". Please take sufficient time to make your decisions. Once all participants have confirmed their entry, the second screen appears. On the second screen, your order quantity is given again and you receive information about the realized demand. Furthermore, the resulting *costs* are listed. When all participants have pressed "OK", the next round starts.

4. Payment determination: You receive a fixed payment of €4 for your appearance. Furthermore, you can earn additional money dependent on your performance in the course of the experiment. At the end of the experiment, the *costs* in ECU incurred in all rounds are added together. *These costs will be deducted from a fixed budget of ECU ..., which is available to fulfill the task.* Your payoff is the resulting amount which is converted by a factor of ECU ... = €1 plus the €4 you receive for your appearance. In the event that you have generated a total loss, you still receive your show up fee.

A.2 Instructions experiment Chapter 5

Instructions: The instructions are translated from German and shortened since the original instructions also contain screen shots. Furthermore, they do not contain the cost values and the payment figures since they are different for the instruction of the low quantile case and the high quantile case. The instructions consist of four parts.

1. General information: You are about to participate in an experiment in decision making. In the course of the experiment you can earn a considerable amount of money depending on your decisions. In the experiment, all monetary amounts are specified in Experimental Currency Units (ECU). They are converted according a fixed exchange rate into Euro (see payment determination) at the end of the experiment. The experiment is followed by a short questionnaire and, afterwards, you will be paid in cash. All your decisions and answers will be treated confidentially. Please read the following instructions carefully. If you have any questions, please ask.

2. Experimental task: Consider the following simplified decision situation about planning of surgery durations. Your job is to reserve time for a surgery in the operating room. You don't know how long the surgery will take but you know that the duration of that surgery is equally probable for any value between 100 and 200. Every reserved minute of the operating room is associated with costs. If your reserved time exceeds the duration, the remaining time can not be used otherwise. If the duration exceeds your reserved time, the additional time needed is associated with higher costs. The surgery can not be interrupted. For each minute you reserve the operating room, the costs are ECU ... (the costs per minute reserved too much correspond to ECU ...). For each minute the operating room is needed beyond the reserved time, the costs are ECU ... (the additional costs for each minute reserved too little correspond to ECU ...).

- The cost per minute reserved time even occur if the duration is shorter than the reserved time.
- The operation must be carried out until the end.

3. Experimental procedure: The experiment consists of 20 rounds and the surgery duration in each round is independent of past surgery durations. Every round consists of two screens. The first screen summarizes the information already given in the instructions. Furthermore, you have to enter the minutes you want to reserve the operating room (between 100 and 200 minutes) in the red box and press the button "OK". Please take sufficient time to make your decisions. Afterwards, the second screen appears. On the second screen, your reserved time is given again and you receive information about the realized duration. Furthermore, the resulting costs are listed. After pressing "OK", the next round starts. You have to plan 20 independent surgeries.

4. Payment determination: You can earn money dependent on your performance in the course of the experiment. At the end of the experiment, the costs in ECU incurred in all rounds are added together. These costs will be deducted from a fixed budget of ECU ..., which is available to fulfill the task. Your payoff is the resulting amount which is converted by a factor of ECU ... = €1. Depending on your performance, the payoff will be between €5 and €55.

Appendix B

Abbreviations, Notations, and Symbols

B.1 General Abbreviations

2OH	2-Opt heuristic
ACE	Assessment of costs effect
CE	Combined effect
ECU	Experimental currency units
EV	Heuristic with approximated objective function based on expected values
EVV	Heuristic with approximated objective based on expected values and variances
HQC	High quantile case
ICU	Intensive care unit
IIH	Incremental improvement heuristic
K-S	Kolmogorow-Smirnow
LQC	Low quantile case
MAE	Mean anchor effect

MELESSA	Munich Experimental Laboratory for Economic and Social Sciences
MRI	Klinikum München rechts der Isar
MSS	Master surgery schedule
NCS	Department of neurosurgery
OCP	Opportunity cost problem
OR	Operating room
ORDS	Operating room day schedules
PACU	Post-anesthesia care unit
PCP	Penalty cost problem
SA	Simulated annealing
SBB	Straightforward branch-and-bound
SP	Department of sport orthopedics
URS	Department of urology
WE	Imaginary specialty for cases on weekends
ZOP2	Zentral-OP-2

B.2 Notations and Symbols Chapter 2

Sets and indices

$j \in \mathcal{J}$	Surgery specialties
$i \in \mathcal{I}$	Operating rooms
$p \in \{0, \ldots, P_j\}$	Patients
$n \in \{1, \ldots, N_j^I\}$	Days in the ICU after surgery
$m \in \{1, \ldots, N_j^I\}$	Days in the ICU after surgery
$n \in \{1, \ldots, N_j^{WO}\}$	Days in the ward after surgery
$u \in \{0, \ldots, N_j^{WI}\}$	Days in the ward after ICU

$\ell \in \mathcal{L}$	Days in the MSS cycle
$q \in \mathcal{Q}$	Weekdays in the MSS cycle
$\ell \in \mathcal{L} \setminus \mathcal{Q}$	Weekend days in the MSS cycle

Parameters

$a^{I,E}$	Estimated quotient of required fixed capacities and expected number of patients in the ICU
$a^{W,E}$	Estimated quotient of required fixed capacities and expected number of patients in the ward
$a_j(p)$	Probability that $p \in \{0, \dots, P_j\}$ patients are operated on during a surgery block of specialty j
α^I	Service level for fixed capacities in the ICU
α^W	Service level for fixed capacities in the ward
$b^{I,E}$	Estimated quotient of patients to be staffed and expected number of patients in the ICU
$b^{W,E}$	Estimated quotient of patients to be staffed and expected number of patients in the ward
b_j	Probability that a patient of specialty j is admitted to the ICU immediately after surgery
β^I	Service level for staffing in the ICU
β^W	Service level for staffing in the ward
$c^{I,f}$	Costs for creating and maintaining the capacity for one patient in the ICU per cycle
$c^{W,f}$	Costs for creating and maintaining the capacity for one patient in the ward per cycle
$c^{I,o}$	Costs for each patient above existing capacities in the ICU per day
$c^{W,o}$	Costs for each patient above existing capacities in the ward per day
$c^{I,s}$	Costs for staffing one patient in the ICU per day

$c^{W,s}$	Costs for staffing one patient in the ward per day
$c^{I,we}$	Extra costs for staffing one patient in the ICU per day on weekends
$c^{W,we}$	Extra costs for staffing one patient in the ward per day on weekends
$c^{I}_{j}(n)$	Probability that a patient of specialty j stays $n \in \{1, \ldots, N^{I}_{j}\}$ days in the ICU after surgery
$c^{WO}_{j}(n)$	Probability that a patient of specialty j stays $n \in \{1, \ldots, N^{WO}_{j}\}$ days in the ward after surgery
$c^{WI}_{j}(u)$	Probability that a patient of specialty j stays $u \in \{0, \ldots, N^{WI}_{j}\}$ days in the ward after being released from ICU
cf	Cooling factor for SA
d_j	Sum of required blocks for specialty j
$d^{I}_{j,n}$	Probability for a patient of specialty j in the ICU to be transferred to the ward on day n
$d^{WO}_{j,n}$	Probability for a patient of specialty j in the ward n days after surgery to be released that day
$d^{WI}_{j,u}$	Probability for a patient of specialty j in the ward u days after transfer from ICU to be released that day
$e^{I}_{j,n}$	Probability that patient of specialty j who had surgery on day 1 is in the ICU on day n
$e^{WO}_{j,n}$	Probability that patient of specialty j who had surgery on day 1 and was transferred from the OR is in the ward on day n
$e^{WI}_{j,m,n}$	Probability that patient of specialty j who had surgery on day 1 and stayed in the ICU before is in the ward on day n
$e^{W}_{j,n}$	Probability that patient of specialty j who had surgery

	on day 1 is in the ward on day n
$f_{j,n}^{I}$	Probability distribution for the number of patients in the ICU on day n from one block of specialty j
$f_{j,n}^{W}$	Probability distribution for the number of patients in the ward on day n from one block of specialty j
$F_{j,\ell}^{I}$	Probability distribution for the number of patients in the ICU on day ℓ from one cyclical block of specialty j
$F_{j,\ell}^{W}$	Probability distribution for the number of patients in the ward on day ℓ from one cyclical block of specialty j
s_j	Maximum number of blocks for specialty j on day q
sr^I	Estimated quotient of variance and standard deviation of number of patients in the ICU
sr^W	Estimated quotient of variance and standard deviation of number of patients in the ward
t_k	Temperature level in round k
$z^{I,cap}$	z-Value for standard deviation of distribution of required fixed capacities in the ICU
$z^{W,cap}$	z-Value for standard deviation of distribution of required fixed capacities in the ward
$z^{I,sta}$	z-Value for standard deviation of distribution of patients to be staffed in the ICU
$z^{W,sta}$	z-Value for standard deviation of distribution of patients to be staffed in the ward

Decision variables

$x_{i,q,j}$	1, If specialty j is assigned to block (i, q), 0 otherwise

Functions

$c(x)$	Downstream costs
cap^I	Number of beds needed in the ICU

cap^W	Number of beds needed in the ward
$cost^f$	Fixed costs in downstream units
$cost^o$	Overcapacity costs in downstream units
$cost^s$	Staffing costs in downstream units
$cost^{we}$	Extra staffing costs on weekends in downstream units
$E(F_\ell^I)$	Expected number of patients in the ICU on day ℓ
$E(F_\ell^W)$	Expected number of patients in the ward on day ℓ
exc^I	Expected number of patient days above capacities in the ICU during one cycle
exc^W	Expected number of patient days above capacities in the ward during one cycle
$\overline{F}^I_{i,q,l}$	Probability distribution for the number of patients in the ICU on day ℓ from the cyclical block (i, q)
$\overline{F}^W_{i,q,l}$	Probability distribution for the number of patients in the ward on day ℓ from the cyclical block (i, q)
F_l^I	Probability distribution for the number of patients in the ICU on day ℓ from the MSS
F_l^W	Probability distribution for the number of patients in the ICU on day ℓ from the MSS
$Q_\ell^I(\alpha^I)$	α^I-quantile of the distribution F_l^I on day ℓ.
$Q_\ell^W(\alpha^W)$	α^W-quantile of the distribution F_l^W
$SD(F_\ell^I)$	Standard deviation of number of patients in the ICU on day ℓ
$SD(F_\ell^W)$	Standard deviation of number of patients in the ward on day ℓ
sta^I	Number of patients to be staffed in the ICU
sta^W	Number of patients to be staffed in the ward
$V(F_\ell^I)$	Variance of number of patients in the ICU on day ℓ
$V(F_\ell^W)$	Variance of number of patients in the ward on day ℓ

B.3 Notations and Symbols Chapter 3

Indices

j	Surgery specialties
p	Patients
n	Days after surgery
ℓ	Days in the MSS cycle

B.4 Notations and Symbols Chapter 4

Indices

t	Period

Parameters

α	Mean anchor weight
β	Underage cost weight
β_{opp}	Underage cost weight in the opportunity cost problem
β_{pen}	Underage cost weight in the penalty cost problem
c	Purchasing costs per item
c_o	Overage costs
c_u	Underage costs
d	Uncertain demand

ϵ_t	Error term in period t
$F(D)$	Cumulated demand function
μ	Mean demand
p	Selling price per item
s	Higher reorder costs

Decision variables

q	Order quantity
q^*	Optimal order quantity
q^{ACE}	Adapted optimal order quantity
q^{CE}	Order quantity after CE
q^{MAE}	Order quantity after MAE
q_t	Order quantity in period t

B.5 Notations and Symbols Chapter 5

Indices

t	Period

Parameters

c	Costs per minute of used OR capacity
c^o	Costs per minute of overutilization
c^u	Costs per minute of underutilization
D	Realized duration
D_t	Realized duration in period t
F^{-1}	Inverse cumulative distribution function of D
μ	Mean realized duration

p	Planned duration
s	Increased costs per minute of overutilization

Decision variables

p	Planned duration
p^*	Optimal planned duration
p_t	Planned duration in period t

Functions

A_t	Planning adjustment score
$C(p, D)$	Total costs
$E[C(p)]$	Expected costs with planned duration p
$E[C(p^*)]$	Expected costs with optimal planned duration p^*
$I(p)$	Relative cost increase of planned duration p

Bibliography

[1] Aarts, E., Korst, J., and Michiels, W. (2005). Simulated annealing. In Burke, E. K. and Kendall, G., editors, *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, pages 187–211. Springer, New York.

[2] Abouleish, A. E., Dexter, F., Epstein, R. H., Lubarsky, D. A., Whitten, C. W., and Prough, D. S. (2003). Labor costs incurred by anesthesiology groups because of operating rooms not being allocated and cases not being scheduled to maximize operating room efficiency. *Anesthesia & Analgesia*, 96(4):1109–1113.

[3] Adan, I. J. B. F. and Vissers, J. M. H. (2002). Patient mix optimisation in hospital admission planning: A case study. *International Journal of Operations and Production Management*, 22(4):445–461.

[4] Anderson, D., Price, C., Golden, B., Jank, W., and Wasil, E. (2011). Examining the discharge practices of surgeons at a large medical center. *Health Care Management Science*, 14(4):338–347.

[5] Baker, D. R., Pronovost, P. J., Morlock, L. L., Geocadin, R. G., and Holzmueller, C. G. (2009). Patient flow variability and unplanned readmissions to an intensive care unit. *Critical Care Medicine*, 37(11):2882–2887.

[6] Becker, S. W., Ronen, J., and Sorter, G. H. (1974). Opportunity costs – an experimental approach. *Journal of Accounting Research*, 12(2):317–329.

[7] Bekker, R. and Koeleman, P. M. (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Management Science*, 14(3):237–249.

[8] Beliën, J. and Demeulemeester, E. (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, 176(2):1185–1204.

[9] Benzion, U., Cohen, Y., Peled, R., and Shavit, T. (2008). Decision-making and the newsvendor problem: an experimental study. *Journal of the Operational Research Society*, 59(9):1281–1287.

[10] Bland, J. M. and Altman, D. G. (1994). Statistics notes: Some examples of regression towards the mean. *British Medical Journal*, 309:780.

[11] Bleich, S. N., Özaltin, E., and Murray, C. J. (2009). How does satisfaction with the health-care system relate to patient experience? *Bulletin of the World Health Organization*, 87(4):271–278.

[12] Bolton, G. E. and Katok, E. (2008). Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, 10(3):519–538.

[13] Bolton, G. E., Ockenfels, A., and Thonemann, U. (2012). Managers and students as newsvendors. *Management Science*, forthcoming.

[14] Bornstein, B. H., , and Emler, A. C. (2001). Rationality in medical decision making: a review of the literature on doctors´ decision-making biases. *Journal of Evaluation in Clinical Practice*, 7(2):97–107.

[15] Bostian, A. A., Holt, C. A., and Smith, A. M. (2008). Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*, 10(4):590–608.

[16] Brailsford, S. and Schmidt, B. (2003). Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. *European Journal of Operational Research*, 150(1):19–31.

[17] Cachon, G. and Terwiesch, C. (2009). *Matching Supply with Demand: An Introduction to Operations Management*. McGraw-Hill, Boston.

[18] Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.

[19] Carter, M. (2002). Diagnosis: Mismanagement of resources. *OR/MS Today*, 29(2):26–32.

[20] Chang, E. M., Hancock, K. M., Johnson, A., Daly, J., and Jackson, D. (2005). Role stress in nurses: Review of related factors and strategies for moving forward. *Nursing & Health Sciences*, 7(1):57–65.

[21] Chen, L., Kok, A. G., and Tong, J. (2010). The effect of payment timing on inventory decisions in a newsvendor experiment. Working Paper, Fuqua School of Business, Duke University.

[22] Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24.

[23] Dexter, F., Epstein, R. H., Marcon, E., and de Matta, R. (2005). Strategies to reduce delays in admission into a postanesthesia care unit from operating rooms. *Journal of PeriAnesthesia Nursing*, 20(2):92–102.

[24] Dexter, F., Macario, A., and Ledolter, J. (2007). Identification of systematic underestimation (bias) of case durations during case scheduling would not markedly reduce overutilized operating room time. *Journal of Clinical Anesthesia*, 19(3):198–203.

[25] Dexter, F. and Traub, R. D. (2002). How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesthesia & Analgesia*, 94(4):933–942.

[26] Donohue, K. L. (2000). Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science*, 46(11):1397–1411.

[27] Eeckhoudt, L., Gollier, C., and Schlesinger, H. (1995). The risk-averse (and prudent) newsboy. *Management Science*, 41(5):786–794.

[28] Eijkemans, M. J. C., van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G. (2010). Predicting the unpredictable a new prediction model for operating room times using individual characteristics and the surgeon´s estimate. *Anesthesiology*, 112(1):41–49.

[29] Eppen, G. D. (1979). Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, 25(5):498–501.

[30] Feng, T., Keller, L. R., and Zheng, X. (2011). Decision making in the newsvendor problem: A cross-national laboratory study. *Omega*, 39(1):41–50.

[31] Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

[32] Friedman, L. A. and Neumann, B. R. (1980). The effects of opportunity costs on project investment decisions: A replication and extension. *Journal of Accounting Research*, 18(2):407–419.

[33] Fügener, A., Edenharter, G., Kiefer, P., Mayr, U., Schiele, J., Steiner, F., Kolisch, R., and Blobner, M. (2012a). Improving ICU and ward utilization by adapting master surgery schedules: A case study. Working Paper, TUM School of Management, Technische Universität München.

[34] Fügener, A., Hans, E. W., Kolisch, R., Kortbeek, N., and Vanberkel, P. T. (2012b). Master surgery scheduling with consideration of multiple downstream units. Working Paper, TUM School of Management, Technische Universität München.

[35] Fügener, A., Schiffels, S., and Kolisch, R. (2012c). Underutilization and overutilization of operating rooms - insights from behavioral health care operations management. Working Paper, TUM School of Management, Technische Universität München.

[36] Gavirneni, S. and Isen, A. M. (2010). Anatomy of a newsvendor decision: Observations from a verbal protocol analysis. *Production and Operations Management*, 19(4):453–462.

[37] Gino, F. and Pisano, G. (2008). Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, 10(4):676–691.

[38] Guerriero, F. and Guido, R. (2011). Operational research in the management of the operating theatre: A survey. *Health Care Management Science*, 14(1):89–114.

[39] Hans, E. W., van Houdenhoven, M., and Hulshof, P. J. H. (2011). A framework for health care planning and control. In Hall, R., editor, *Handbook of Health Care Systems Scheduling*, volume 168 of *Springer International Series in Operations Research & Management Science*, pages 303–320. Springer, New York.

[40] Hans, E. W. and Vanberkel, P. T. (2011). Operating theatre planning and scheduling. In Hall, R., editor, *Handbook of Health Care Systems Scheduling*, volume 168 of *Springer International Series in Operations Research & Management Science*, pages 105–130. Springer, New York.

[41] Harper, P. R. (2002). A framework for operational modelling of hospital resources. *Health Care Management Science*, 5(3):165–173.

[42] Hay, J. W. (2003). Hospital cost drivers: An evaluation of 1998-2001 state-level data. *American Journal Managed Care*, 9(1):13–24.

[43] Ho, T.-H., Lim, N., and Cui, T. H. (2010). Reference dependence in multilocation newsvendor models: A structural analysis. *Management Science*, 56(11):1891–1910.

[44] Hoskin, R. E. (1983). Opportunity costs and behavior. *Journal of Accounting Research*, 21(1):78–95.

[45] Huang, X.-M. (1994). Patient attitude towards waiting in an outpatient clinic and its applications. *Health Services Management Research*, 7(1):2–8.

[46] Irvine, D. M. and Evans, M. G. (1995). Job satisfaction and turnover among nurses: Integrating research findings across studies. *Nursing Research*, 44(4):246–253.

[47] Jamal, M. and Baba, V. V. (1992). Shiftwork and department-type related to job stress, work attitudes and behavioral intentions: A study of nurses. *Journal of Organizational Behavior*, 13(5):449–464.

[48] Katok, E. and Wu, D. Y. (2009). Contracting in supply chains: A laboratory investigation. *Management Science*, 55(12):1953–1968.

[49] Khouja, M. (1999). The single-period (news-vendor) problem: Literature review and suggestions for further research. *Omega*, 27(5):537–553.

[50] Kremer, M. and Minner, S. (2008). The human element in inventory decision making under uncertainty – a review of experimental evidence in the newsvendor model. *ZfB - Special Issue*, 4:83–97.

[51] Kremer, M., Minner, S., and VanWassenhove, L. N. (2010). Do random errors explain newsvendor behavior? *Manufacturing & Service Operations Management*, 12(4):673–681.

[52] Lin, S. and Kernighan, B. (1973). An effective heuristic algorithm for the travelling salesman problem. *Operations Research*, 21(2):489–516.

[53] Litvak, N., van Rijsbergen, M., Boucherie, R. J., and Houdenhoven, M. V. (2008). Managing the overflow of intensive care patients. *European Journal of Operational Research*, 185(3):998–1010.

[54] Loch, C. H. and Wu, Y. (2005). *Behavioral Operations Management*, volume 1. now Publishers Inc.

[55] Luft, J. (1994). Bonus and penalty incentives – contract choice by employees. *Journal of Accounting and Economics*, 18(2):181–206.

[56] Lurie, N. H. and Swaminathan, J. M. (2009). Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human Decision Processes*, 108:315 – 329.

[57] May, J. H., Strum, D. P., and Vargas, L. G. (2000). Fitting the lognormal distribution to surgical procedure times. *Decision Sciences*, 31(1):129–148.

[58] McCaffery, E. J. and Baron, J. (2004). Framing and taxation: Evaluation of tax policies involving household composition. *Journal of Economic Psychology*, 25(6):679–705.

[59] McCaffery, E. J. and Baron, J. (2005). The political psychology of redistribution. *U.C.L.A. Law Review*, 52(6):1745–1792.

[60] McManus, M. L., Long, M. C., Cooper, A., Mandell, J., Berwick, D. M., Pagano, M., and Litvak, E. (2003). Variability in surgical caseload and access to intensive care services. *Anesthesiology*, 98(6):1491–1496.

[61] Min, D. and Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206(3):642–652.

[62] Moritz, B. B. (2010). *Cognition and Heterogeneity in Supply Chain Planning: A Study of Inventory Decision Making*. PhD thesis, University of Minnesota.

[63] Moskowitz, A. J., Kuipers, B. J., and Kassirer, J. P. (1988). Dealing with uncertainty, risks, and tradeoffs in clinical decisions. *Annals of Internal Medicine*, 108(3):435–449.

[64] Niranjan, T. T., Metri, B. A., and Aggarwal, V. (2009). Behavioral causes of bullwhip effect: Breaking the mould. *International Journal of Services and Operations Management*, 5(3):350–374.

[65] Northcraft, G. B. and Neale, M. A. (1986). Opportunity costs and the framing of resource allocation decisions. *Organizational Behavior and human decision processes*, 37:348–356.

[66] Olivares, M., Terwiesch, C., and Cassorla, L. (2008). Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science*, 1(1):41–55.

[67] Pandit, J. J. and Carey, A. (2006). Estimating the duration of common elective operations: implications for operating list management. *Anaesthesia*, 61(8):768–776.

[68] Patrick, J. and Puterman, M. L. (2008). Reducing wait times through operations research: Optimizing the use of surge capacity. *Healthcare Policy*, 3(3):75–88.

[69] Peltokorpi, A. (2011). How do strategic decisions and operative practices affect operating room productivity? *Health Care Management Science*, 14(4):370–382.

[70] Pham, D.-N. and Klinkert, A. (2008). Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185:1011–1025.

[71] Rhodes, M. and Barker, P. (2007). Operating room utilization. *Surgical Endoscopy*, 21(12):2339–2340.

[72] Rudi, N. and Drake, D. (2010). Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. Working Paper, INSEAD, France.

[73] Santibanez, P., Beliën, J., and Atkins, D. (2007). Surgical block scheduling in a system of hospitals: An application to resource and wait list management in a british columbia health authority. *Health Care Management Science*, 10(3):269–282.

[74] Schiffels, S., Fügener, A., Brunner, J. O., and Kolisch, R. (2012). On the assessment of costs in a newsvendor environment: Insights from an experimental study. Working Paper, TUM School of Management, Technische Universität München.

[75] Schultz, K. L., McClain, J. O., Robinson, L. W., and Thomas, L. J. (2007). The use of framing in inventory decisions. Working Paper, Cornell University.

[76] Schweitzer, M. E. and Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, 46(3):404–420.

[77] Silber, J. H., Rosenbaum, P. R., Zhang, X. M., and Even-Shoshan, O. (2007). Estimating anesthesia and surgical procedure times from medicare anesthesia claims. *Anesthesiology*, 106(2):346–355.

[78] Silver, E. A., Pyke, D. F., and Peterson, R. P. (1998). *Inventory Management and Production Planning and Scheduling.* John Wiley & Sons, New York, 3 edition.

[79] Sinuff, T., Kahnamoui, K., Cook, D. J., Luce, J. M., and Levy, M. M. (2004). Rationing in critical care task f. rationing critical care beds: a systematic review. *Critical Care Medicine*, 32(7):1588–1597.

[80] Stepaniak, P. S., Heij, C., Mannaerts, G. H. H., de Quelerij, M., and de Vries, G. (2009). Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: A multicenter study. *Anesthesia & Analgesia*, 109(4):1232–1245.

[81] Strum, D. P., May, J. H., and Vargas, L. G. (2000). Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology*, 92(4):1160–1167.

[82] Strum, D. P., Vargas, L. G., May, J. H., and Bashein, G. (1997). Surgical suite utilization and capacity planning: A minimal cost analysis model. *Journal of Medical Systems*, 21(5):309–322.

[83] Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1(1):39–60.

[84] Trietsch, D. (2006). Optimal feeding buffers for projects or batch supply chains by an exact generalization of the newsvendor model. *International Journal of Production Research*, 44(4):627–637.

[85] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series*, 185(4157):1124–1131.

[86] Tyler, D. C., Pasquariello, C. A., and Chen, C.-H. (2003). Determining optimum operating room utilization. *Anesthesia & Analgesia*, 96:1114–1121.

[87] van Essen, J. T., Bosch, J., and van der Veen, E. (2011). Approximating concave functions with piecewise linear functions. *Working Paper, Twente University, Netherlands.*

[88] van Oostrum, J. M., van Houdenhoven, M., Hurink, J. L., Hans, E. W., Wullink, G., and Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, 30(2):355–374.

[89] Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., and Litvak, N. (2010). A survey of health care models that encompass multiple departments. *International Journal of Health Management and Information*, 1(1):37–69.

[90] Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., van Lent, W. A. M., and van Harten, W. H. (2011a). Accounting for inpatient wards when developing master surgical schedules. *Anesthesia & Analgesia*, 112(6):1472–1479.

[91] Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., van Lent, W. A. M., and van Harten, W. H. (2011b). An exact approach for relating recovering surgical patient workload to the master surgery schedule. *Journal of the Operational Research Society*, 62(10):1851–1860.

[92] Wachtel, R. E. and Dexter, F. (2010). Review of behavioral operations experimental studies of newsvendor problems for operating room management. *Anesthesia & Analgesia*, 110(6):1698–1710.

[93] Wright, I. H., Kooperberg, C., Bonar, B., and Bashein, G. (1996). Statistical modeling to predict elective surgery time: Comparison with a computer scheduling system and surgeon-provided estimates. *Anesthesiology*, 85(6):1235–1245.

[94] Yule, S., Flin, R., Paterson-Brown, S., and Maran, N. (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*, 139(2):140–149.

[95] Zhu, G., Bard, J. F., and Yu, G. (2007). A two-stage stochastic programming approach for project planning with uncertain activity durations. *Journal of Scheduling*, 10(3):167–180.