CrossMark

Overutilization and underutilization of operating rooms - insights from behavioral health care operations management

Andreas Fügener[1] · Sebastian Schiffels[2] · Rainer Kolisch[2]

Abstract The planning of surgery durations is crucial for efficient usage of operating theaters. Both planning too long and too short durations for surgeries lead to undesirable consequences, e.g. idle time, overtime, or rescheduling of surgeries. We define these consequences as operating room inefficiency. The overall objective of planning surgery durations is to minimize expected operating room inefficiency, since surgery durations are stochastic. While most health care studies assume economically rational behavior of decision makers, experimental studies have shown that decision makers often do not act according to economic incentives. Based on insights from health care operations management, medical decision making, behavioral operations management, as well as empirical observations, we derive hypotheses that surgeons' behavior deviates from economically rational behavior. To investigate this, we undertake an experimental study where experienced surgeons are asked to plan surgeries with uncertain durations. We discover systematic deviations from optimal decision making and offer behavioral explanations for the observed biases. Our research provides new insights to tackle a major problem in hospitals, i.e. low operating room utilization going along with staff overtime.

1 Introduction

Numerous recent studies encourage researchers to take into account human behavior in operations management (e.g. [4, 24, 31]). Behavior plays a crucial role in health care operations management [8], since health care services are provided by people who may be influenced by cognitive biases, social preferences, and cultural norms [31]. Especially in the health care sector, public service motivation and professional norms may influence decision making in addition to economic incentives [3]. 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. An example is provided by [18], who applied an experimental study with nursing students and demonstrate that both user interface and appropriate statistical methods might affect the quality of appointment schedules. While behavioral experiments are a well-established

✉ Andreas Fügener
andreas.fuegener@wiwi.uni-augsburg.de

Sebastian Schiffels
sebastian.schiffels@tum.de

Rainer Kolisch
rainer.kolisch@tum.de

[1] Universitäres Zentrum für Gesundheitswissenschaften am Klinikum Augsburg (UNIKA-T), School of Business and Economics, Universität Augsburg, Universitätsstraße 16, 86159, Augsburg, Germany

[2] TUM School of Management, Technische Universität München, Arcisstr. 21, 80333 München, Germany

research methodology for studying human issues in many research areas including several business disciplines and medical research, studies combining findings from behavioral operations management with health care applications are still rare.

In this study, we approach the field of behavioral health care operations management by investigating surgeons' planning behavior in the operating room (OR), one of the most important resources in hospitals. [26] cite more than 100 studies on operating room management and [9] 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 arise in the operating theater [13] and more than 60 % of hospital admissions are for surgical operations [38]. Furthermore, processes in downstream units, such as intensive care units or general wards, are triggered by the OR department [21–23]. Although the high importance of optimizing the usage of this scarce resource is evident, there still is much room for improvement. [40] report poor utilization of ORs and [37] argue that 10-40 % of all scheduled elective surgeries are canceled or rescheduled at least once. Sometimes OR managers make bad decisions resulting in staff working overtime (e.g. [46]).

Low OR utilization, rescheduling of surgeries and staff overtime are inherent consequences of planning of stochastic surgery durations. However, good planning of surgery durations may help to reduce these negative effects. Obviously, centralized planning cannot account for the specific patient knowledge of the responsible surgeon. Therefore, it is common practice in most hospitals that each surgery duration is planned independently by the surgeon in charge of the patient. Using the planned durations as an input a central OR manager is typically responsible for scheduling and sequencing of surgeries. In the literature surgeons' behavior is mainly discussed in the context of medical decision making. There are a few studies indicating non-optimal behavior of surgeons considering operating room management. [49] conduct a literature review on non-technical skills of doctors in the OR and they conclude that non-technical skills such as planning know-how, resource management, and communication are often neglected, despite being vital for efficient OR management. [10] presents an example where doctors only considered fairness when planning the ORs but ignored negative consequences for other units. [1] state that OR management is often based on convenience and tradition rather than on optimization. A systematic underestimation of surgery durations is found by [17, 46] 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 are present in the OR 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, how the newsvendor model can be used in an OR setting is still an open issue.

Our research was motivated by a project with a medium sized hospital that asked for help with one of their major concerns, low operating room utilization and a high amount of overtime. The surgeons in charge of planning surgery durations received no guidelines or any information about the associated consequences of poor planning. Planned and realized durations were not monitored. Using real OR data from this hospital we demonstrate the complexity of planning of surgery durations. We verify that variability in surgery durations exists and we analyze the negative consequences of planning too long and too short surgery durations. Thus, we conclude that a newsvendor equivalent minimal cost model fits the problem of planning surgery durations. This is in line with several studies where trade-off decisions in OR management are modeled according to the newsvendor framework, even though the framework is rather stylized (e.g. [36, 44, 46] apply this framework for different tasks in operating room management). To test the behavioral effects of planning surgery durations we undertake an experimental study with senior surgeons. We chose doctors with experience in OR management since previous studies observed that, even if the direction of behavioral effects is the same, the magnitude of effects for students and experienced professionals as subjects may differ (e.g. [7]). To be able to differentiate effects resulting from different contexts and from different subjects, we also tested our design with business students as subjects, as done in most inventory studies, in our first follow-up study. Two further follow-up studies to test different distributions of surgery durations and the effect of planning multiple surgeries were applied. Based on the theory of professional norms [25] we are able to explain significant differences between biases from inventory management and OR planning in the newsvendor framework. As no consistent definition of the consequences of planning too long and too short surgery durations exists in the literature, we employ an experimental study with two different cost scenarios. Within our experimental study, we demonstrate significant non-optimal planning of surgery durations by experienced surgeons and we show that optimal decisions would lead to a potential cost reduction of about 3.3 % in our stylized setting. Even though external validity in laboratory experiments is somehow limited [29], this number provides an indication of the practical relevance of our problem.

The remainder of our paper is organized as follows: In the following section we present the problem framework of planning surgery durations and derive our hypotheses. In Section 3 the experimental setup is explained and the results
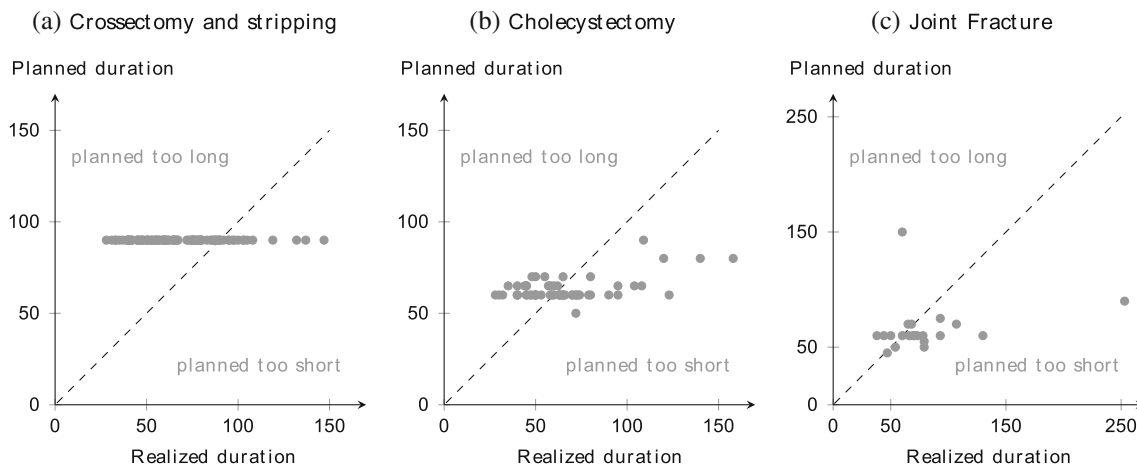
Fig. 1 Comparison of planned and realized durations of three different surgeries

are discussed. We draw conclusions and analyze managerial implications in the final Section 4.

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 to undesirable consequences (e.g. [34]). To obtain first insight into planning behavior in real life, we analyzed 6 months (12/2011 - 05/2012) of surgery data from a German university hospital. Of all surgery types, only 5.6 % were conducted more than ten times during the planning horizon of six months. In analogy to [14], the hospital has to rely on physician estimates for most surgeries. The durations of elective surgeries are usually planned a few days before the surgery. We compare the planned and the realized durations of three common types of 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 illustrate the planned and the realized durations of these three surgeries in Fig. 1. Consistently with the literature (e.g. [33]), the distributions of the surgery duration can be modeled as lognormal distributions. We state the corresponding p-values (K-S-test) and parameters in Table 1 along with some additional information.

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. In fact, crossectomy and stripping surgeries were always planned with 90 minutes. 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 different ways, which may be a consequence of different cost structures, motivations, or incentives. Third, some surgeries are systematically planned too long, while others are systematically planned too short.

Planning of surgery durations is a complex task due to two main characteristics of the problem. First, variability in surgery durations exists. Second, both planning too long and too short durations results in different negative

Table 1 Comparison of planned and realized durations of three different surgeries

	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 %
p-value lognormal distribution (K-S-test)	0.50	0.79	0.47
Corresponding parameters (σ/μ)	0.35/4.16	0.37/4.12	0.40/4.28

consequences. As a result, a trade-off decision minimizing these consequences has to be made.

2.1 Variability of surgery durations

There are two reasons for variability in 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 extends the duration. With diversity of situation we take into account a priori known factors, such as patient age or OR-team experience. Estimating the distribution of surgery durations is discussed widely in the literature. [45] and [33] use lognormal distributions to model surgery times, while [43] 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. [48] compare time estimates of forecasting modules 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. [14] analyze cases with rare historical data, where decision makers need to rely on surgeon estimates. They conclude that only in cases with substantial changes in surgery or anesthetic procedures updating these estimates may improve OR management. [19] demonstrate that, in addition to the surgeons' 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. Regression models allow to correct systematically biased surgery duration estimations by considering intercept (e.g. in cases where the estimate is usually 10 minutes too long) and slope (e.g. in cases where the estimate is usually 10 % too long). After the introduction of a new computerized planning system, [19] observe a significant shift in the estimation bias.

2.2 Consequences of planning too long or too short

The second important issue in 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 [44], we define this as underutilization. If the realized surgery duration is above the planned surgery duration multiple consequences might occur. Following surgeries may have to wait or even be rescheduled which involves a considerable organizational effort, reduces patient and staff satisfaction

as well as medical quality. Furthermore, the scheduled surgery or following surgeries might end after regular working hours, i.e. staff works overtime. Overtime caused by planning too short surgery durations is defined as overutilization. Planning surgeries too long or too short also affects patient waiting times. [28] differentiate between two aspects of patient waiting time: indirect waiting time, i.e. the time between the time the patient requests an appointment and the appointment at the hospital, and direct waiting time, i.e. the time between the appointment time and the time the patient is actually served. Indirect waiting time is supposed to be negatively correlated with OR utilization, i.e. the higher the utilization, the shorter is the expected patient indirect waiting time, as more patients can be treated in systems with high OR utilization. Direct waiting time, however, is supposed to have a positive correlation with OR utilization, i.e. the higher the utilization, the higher is the expected patient direct waiting time. High OR utilization in combination with variability of surgery durations leads to an increasing number of process disruptions and longer waiting times for scheduled patients. A similar trade-off between indirect waiting time, surgery start-time reliability (affecting direct waiting time), and hospital profits is discussed in [32].

To obtain insight into the consequences of inaccurate planning, we analyzed data from the hospital mentioned above. We performed regression analysis to determine the effects of planning too long and too short on OR under- and overutilization, respectively. In Fig. 2a) we relate for each OR and each day the number of minutes surgeries were planned too long with total realized operating time. Given fixed operating room time, less realized operating time leads to more idle time. We observe that the more minutes surgeries were planned too long, the more idle time occurred (0.348 Minutes of less realized operating time per minute planned too long, $p = 0.007$). We further compared the number of minutes planned too short with the minutes of overtime (between 4pm and 10pm). As presented in Fig. 2b), 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 or for being rescheduled) and patient satisfaction (for rescheduling their surgeries and for increased waiting times). [36] state that "the costs of OR idle time were perceived, on average, as approximately 60 % higher than the costs of schedule overrun," while [46] assume that the costs of OR overutilization

Fig. 2 Consequences of planning too long and too short (in minutes)

(a) Consequences of planning too long

Total operating time

(b) Consequences of planning too short

OR overtime
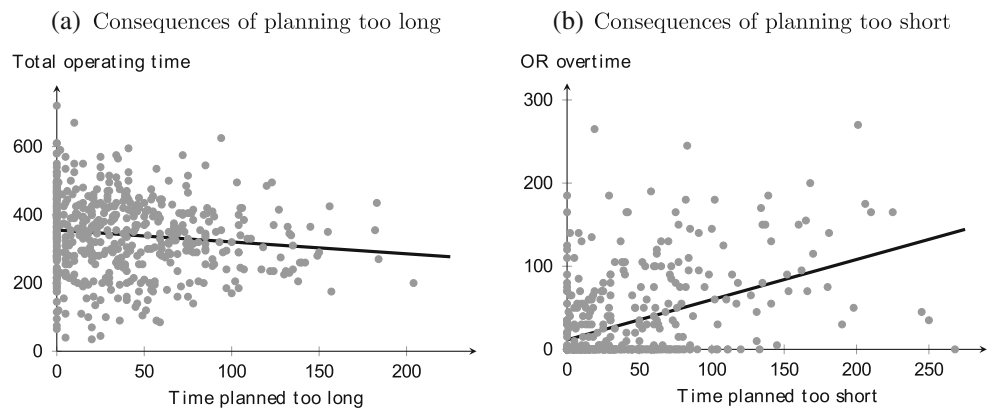
are twice as high as the costs of OR underutilization. Thus, different ratios of these costs exist in the literature, which might be caused by different assessments of under- and overutilization, different hospitals, and different decision makers (e.g. surgeons, anesthesia department managers, or nursing managers). As a consequence, we consider different cost ratios during this study.

2.3 Minimal cost model

To minimize the expected costs of under- and overutilization, we propose a minimal cost analysis model similar to [15, 44] define the sum of cost-weighted under- and overutilization as OR inefficiency. It has been shown that long term decisions, such as capacity dimensioning or staffing, have a major influence on OR over- and underutilization (e.g., [46]). As the data analysis in the previous section demonstrates, short term decisions like planning surgery durations also impact OR over- and underutilization (see Fig. 2). In this paper, we focus on the effects of planning surgery durations. Although surgeons often perform a series of surgeries, each surgery duration is usually planned individually. In the three hospitals that cooperated with us for this study, it is common practice that surgery durations are planned by the surgeon in charge of the patient without consideration of interdependencies between surgeries. Thus, we define a variation of the minimal cost analysis model concentrating on the costs for one surgery. For the sake of clarity we define costs that include all negative consequences of planning too long or too short as follows: c^u are the costs for each minute of underutilization, c^o for each minute of overutilization, and c for each minute of used OR capacity. Depending on the planned duration p and the realized duration D the OR inefficiency for one surgery is

$$C(p, D) = c^u \cdot \max\{p-D, 0\} + c^o \cdot \max\{D-p, 0\} + c \cdot D. \quad (1)$$

The minimal cost analysis model is mathematically equivalent to the well-known newsvendor problem, which is also

used for example by [36] to conduct a structural estimation of the costs for OR under- and overutilization. As in the newsvendor problem the planned duration p^* that minimizes the expected costs $E[C(p)]$ is

$$p^* = F^{-1}\left(\frac{c^o}{c^o + c^u}\right), \quad (2)$$

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" and $\frac{c^o}{c^o+c^u}$ as "critical ratio" in the following. In both the minimal cost analysis model and the newsvendor problem, individuals face a decision under uncertainty with known distribution and a trade-off between planning (ordering) too long (too many) or too short (too little) durations (products) has to be made. The optimal solution can be derived analytically. On the other hand, the two problems are obviously not the same. Important differences are the main task - planning time versus quantities; the different context - operating room planning versus inventory settings; and the different decision makers - surgeons with no management training versus inventory managers. Furthermore, the consequences of not reserving enough time and not ordering enough quantities vary as well: Overtime with additional (penalty) costs and other negative consequences in the OR case since, once started, surgeries have to be completed versus opportunity costs for lost sales in the inventory situation.

2.4 Hypotheses

The complexity of planning a surgery's duration is considerable, even if optimal durations can theoretically be derived with the newsvendor model. In practice surgeons, who are lacking training in capacity management, plan surgery durations. All studies using the minimal cost model have in common that a rational decision maker is assumed but they do not take into account that a human decision maker may

not act rationally in an economic sense. As several studies show that people do not behave optimally in the related inventory situation (e.g. [12, 35, 42]), and since some studies have observed biased surgeon behavior in general, we expect that surgeons do not plan optimally. Due to the similarities, 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. [42] are the first to describe this pattern and they also discuss possible explanations for the observed behavior. An explanation they found support for is that decision makers use the mean as an anchor and only insufficiently adjust towards the optimal solution. This bias was replicated in numerous follow-up studies (e.g. [7, 30]). Combining these findings with our empirical observations for surgeons planning behavior for different surgeries and the OR literature (e.g. [1]) we derive the first hypothesis of our experimental study:

H1: *Surgeons consistently plan too long (too short) in cases where the optimal duration p^* is below (above) the average duration μ of a surgery.*

Furthermore, we expect additional effects to those of classical newsvendor studies. 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 negative consequences, since operations have to be finished. Therefore, the planning of surgeries has similarities to a situation where penalties occur when ordering too little and demand has to be fulfilled. [41] analyze the impact of additional costs ("penalty costs") instead of opportunity costs in a trade-off situation and find strong support that people are more sensitive to penalty costs than to opportunity costs. In their experiments involving two situations with identical optimal solutions, decision makers made more of an effort to avoid underestimating when there are penalties associated than in a situation where opportunity costs occur. This behavior results in a bias upwards in the penalty setting. Considering this effect we derive our next hypothesis:

H2: *Surgeons avoid overutilization rather than underutilization and as a consequence, planned durations are biased upwards.*

Decision behavior is often sensitive to task and contextual factors. [30] compared a classical newsvendor situation (operations setting) with a context-free but mathematically equivalent neutral setting and discovered that the bias towards the mean demand was much stronger in the operations setting than in the neutral setting. Decision makers in a health care setting probably behave in a different way than those in the inventory management setting used in previous studies. Besides economic incentives, decisions may also be influenced by non-economic motivations [3, 39] define the concept of public service motivation, that is opposed to the concept of pure economic incentives. This motivation stems from rather altruistic goals, such as helping others or serving society. [25] state that the level of influence of incentives on individual decision making depends on professional norms. The health care industry is one with particularly strong professional norms [3], for example, doctors take the Hippocratic oath. In case of hospitals, the level of influence by economic incentives on professionals is low where patient health is at stake [3, 25] compare the effects of economic incentives, public service motivation, and professional norms on health professional's behavior. They conclude that economic incentives show little effect in situations where professional norms apply. In case of planning surgery durations, a professional norm directly linked to the Hippocratic oath is not to risk the optimal surgery process of all affected patients. In case of planning too short, the planned surgeries of following patients will be affected. Patients will have to wait longer, or surgeries may even have to be canceled. This can put pressure on the surgical team and may thus affect the quality of medical care in a negative way. In case of planning too long, OR time will be denied to other patients. Taking into account that professional norms influence decision makers, we would expect doctors to differ from inventory managers in two ways: First, they are expected to stronger neglect their economic incentive of minimizing OR inefficiency, the weighted costs of over- and underutilization. Instead, they may focus more on meeting the expected surgery duration, thus aiming for the case where following surgeries are not affected and, at the same time, no OR time is wasted. Second, the direct effect on patient health seems to be stronger in case of planning too short. Surgeons focusing on patient health will ensure that enough time is planned for the treatment to avoid stress due to unforeseen events or pressure from following surgeries. Thus, we would expect the doctors to show a stronger weighting of overutilization costs than of underutilization costs. Therefore, we expect the newsvendor behavior to be influenced by context-specific effects and derive our third hypothesis:

H3: *Surgeons confronted with planning surgery durations show a stronger shift to the mean and a stronger weighting of overutilization costs than decision makers in comparable inventory newsvendor studies.*

In most inventory studies, students took the role of decision makers. [7] observed that the direction of behavioral effects is the same for students and experienced managers and that only the magnitude of effects may differ. The study of [3] obtained consistent results for health care professionals in a variety of sectors containing individuals with different expertise and educational backgrounds. Thus, we expect that students as subjects would only partially adopt obvious professional norms, such as exposing patients to the least amount of risk as possible. Additionally testing H3 with students enables us to differentiate between task-driven and subject-driven effects.

We expect that the answers to these hypotheses will provide valuable insights into surgeons' behavior when planning surgery durations. Identifying and understanding behavioral biases in surgery planning could be an important step to improve the trade-off between overutilization and underutilization in hospitals. Increasing the OR utilization, or decreasing overtime and rescheduling, should not be driven by behavioral biases since both can affect the financial performance, patient satisfaction, and medical quality.

3 Experimental study

To investigate surgeons' behavior when planning surgeries we set up an experimental study. In Section 3.1 we describe the experimental setup and we discuss the results in Section 3.2.

3.1 Experimental setup

The empirical findings from Section 2 show that based on the variability of surgery durations and the negative consequences of deviations from the realized duration the planning of a surgery can be described as a decision under uncertainty minimizing the expected negative consequences for underutilization and overutilization. The variability of surgeries includes both uncertainty and diversity of situation. In the experiment we provide information about the stochastic distribution (i.e. uncertainty) of the surgery duration for a specific situation (i.e. surgery team, specific type of surgery and patient characteristics). To avoid different assessments of the situation we neither communicate details nor change these diversity aspects. In the main study, we apply a uniform distribution for simplicity. This also increases the comparability to the inventory literature, where the uniform distribution is used in most studies even though normal or lognormal distributions better fit real life distributions. [5] have shown that for an inventory setting the same behavioral effects are observed for different demand distributions. We test whether the findings remain identical in a robustness test with different distributions. For this robustness test, we consider a normal and a lognormal distribution, both based on real surgery durations. As discussed in the previous section, the literature leaves some ranges for the specific costs of underutilization and overutilization as these may differ between hospitals. To account for different situations on the one hand, and to be comparable to the 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. As our research question focuses on the behavior of surgeons when planning surgeries, we chose only doctors with relevant experience in scheduling surgeries as professional subjects in our main study. The main experiment was carried out with 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 minimal cost analysis model. We used a between subject design with 20 participants in each treatment. The experiments were conducted in hospitals. We set up 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. We provided information on the distribution of the surgery duration, the OR costs c per reserved minute (underutilization costs $c^u = c$) and the increased costs per minute overtime s (overutilization costs $c^o = s - c$). For simplification, each minute scheduled too long results in a minute of underutilization, and each minute scheduled too short results in a minute of overutilization. Further details are depicted in Table 2.

We also set up a number of follow-up studies: In the first follow-up study, we replicated the main experiment with 107 students as subjects. Thus, we are able to differentiate between effects caused by different settings and effects caused by different subjects. In the second follow-up study, the experiment using students as subjects was repeated with 96 subjects considering two distributions based on real surgery data, a symmetric normal distribution and an asymmetric lognormal distribution. Both distributions were designed to fit a real surgery distribution (Crossectomy and stripping, see Fig. 1a in Section 2) and were not rejected by a K-S-test (normal distribution: $p = 0.507$, lognormal distribution: $p = 0.504$). Thus, both estimations represent a

Table 2 Costs and optimal planning times for low and high quantile case

	Low quantile case	High quantile case
Distribution of surgery duration	U(100,200)	U(100,200)
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 planned time	125	175

realistic distribution for durations of crossectomy and stripping. In the third follow-up study, we take into account that in reality planning of surgery durations might be influenced by other surgeries. We analyze the influence of preceding and succeeding surgeries of other doctors on planning surgery durations. Additionally, we consider a setting where a doctor has to plan a series of consecutive surgeries. We tested these issues in an experimental study with 26 experienced doctors. The experiment consists of five situations: first, doctors have to plan the duration of a single surgery (repetition of main experiment). Second, they have to plan the duration of a surgery that is followed by a surgery of another doctor. Third, they have to plan the duration of a surgery that is preceded by a surgery planned with a short duration, thus having a high probability that the realized duration exceeds the planned duration. Forth, they have to plan the duration of a surgery that is preceded by a surgery planned with a long duration, thus having a low probability that the realized duration exceeds the planned duration. Fifth, they have to plan the duration for a series of three consecutive surgeries, where errors average out.

All experiments were programmed and conducted with the software z-Tree [20]. In the main experiment, either the low quantile case or the high quantile case was tested for each subject. After an initial screen, where the subjects had to enter a planned duration, feedback about the realized durations and the occurred costs was provided. The subjects performed 20 decision periods. The surgery duration 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. On average it took 25 minutes to undertake the experiment. Money was the only incentive used. Payments were based on total costs and ranged between 19 Euros and 39 Euros with a mean of 33 Euros for one experiment. Thus, the average payment matched the income of experienced doctors. The follow-up studies with students were based on the same setup with 15 rounds. For the first follow-up study, 53 subjects participated in the low quantile case and 54 subjects in the high quantile case. For the second follow-up study, 24 subjects participated in each of the four combinations of low quantile / high quantile case and normal / lognormal distribution

of surgery duration. The follow-up studies with students were performed in six groups (two for the first follow-up study, four for the second follow-up study). For each group, an iPod Shuffle was drawn in a lottery among one of the six best performing students as an incentive. The follow-up study with doctors consisted of the five situations, as explained before. Each situation was presented as low quantile case and as high quantile case, leading to ten tasks per doctor in total. A questionnaire followed the experiment. To avoid learning and demand-chasing effects, the realizations of the surgery durations were provided after the 10 tasks have been performed in this study. The experiment took on average 20 minutes, and the payments ranged between 16 and 33 Euros (mean: 27 Euros).

3.2 Results

As expected, we observed average planned durations for all subjects in the main experiment 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 Fig. 3. 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 represented by dotted lines.

In Fig. 3, it is apparent that the average planned durations per subject are closely 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: $p = 0.993$). 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 significantly above the optimal duration of 125 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 $p < 0.005$). Thus we can

Fig. 3 Average Planned
Durations
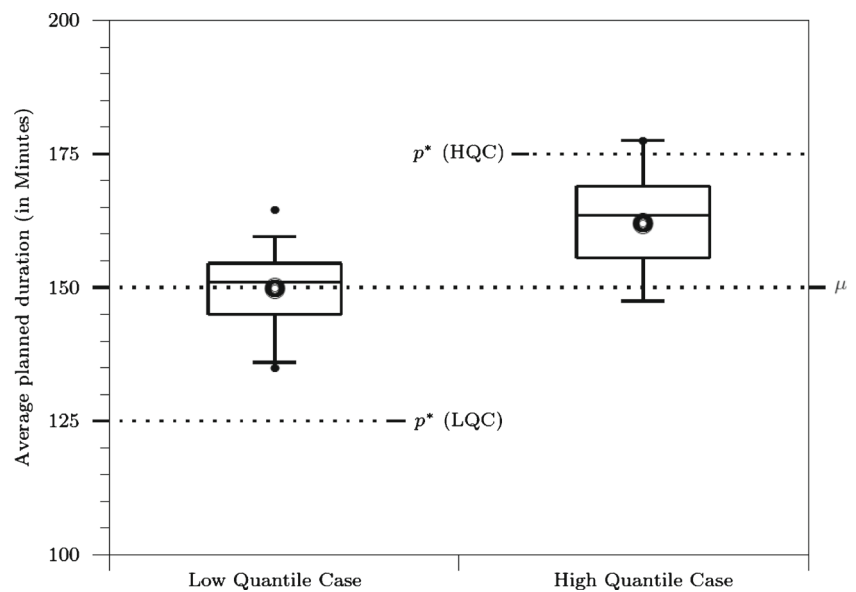
confirm Hypothesis 1: Surgeons consistently plan too long (too short) in cases where the optimal duration p^* is below (above) the average duration μ of a surgery. To measure the degree of non-optimality, we define in Eq. 3 the percentage of avoidable costs due to non-optimal planning ("costs of non-optimal planning"). For each decision of the 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.

$$I(p) = \frac{E[C(p)] - E[C(p^*)]}{E[C(p)]} \qquad (3)$$

We calculated the average costs of non-optimal planning according to Eq. 3 for all participants and all rounds. We found average costs of non-optimal planning of 3.3 % in the low quantile and 3.4 % in the high quantile case. Consistent with inventory management studies, we further gained some insight into learning behavior (trend in the planned durations towards the optimum) and adjustment behavior (the tendency to adjust period-to-period the planned duration in the direction of the previous realized duration). Using linear regression for both the low and the high quantile case, no significant learning could be observed. This comes as no surprise, as [6] show significant learning effects in the long run only. In line with [30], we found that subjects are much more likely to adjust their planned duration in the direction of the previous realized duration than away from it in both cases. To consider Hypothesis 2, we asked all subjects in the questionnaire after the experiment whether they had sought to avoid overutilization or underutilization. The results depicted in Table 3 show that in both cases the subjects tended to avoid overutilization.

We expect this to result in planned durations that are biased upwards. In the low quantile case, this should lead to planned quantities that are further away from the optimal duration, as the effect adds up to the bias towards the mean, than in the high quantile case, where both biases partially compensate for each other. 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, $p < 0.005$). We confirm Hypothesis 2: Surgeons avoid overutilization rather than underutilization. As a consequence, planned durations are biased upwards. To test Hypothesis 3, we compared our data with the corresponding data from [41]. There, subjects were 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. To increase comparability, we adapted their study design, their critical ratios, and their incentive scheme (to reflect different income levels, the payouts in our experiments with experienced surgeons were considerably larger). A visual comparison of both data sets is illustrated in Fig. 4 (to enhance visual comparability we shifted the data of [41] up by 100). [41] used a non-linear regression model to determine the relative shift to the mean α and the relative weighting of "penalty" costs β. The behavioral model which is based on the optimal planned surgery time as given in Eq. 2 is presented in Eq. 4, whereby the dependent and independent variables for the OR planning problem (the inventory management problem) are the following: p_t is the average planned duration (the average order quantity) in round t, μ is the expected duration (the expected demand), F^{-1} is the inverse function of the duration (the demand), c^o

Table 3 Motivation when planning surgery durations

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

are the costs for overutilization (the penalty costs for stock shortage), and c^u are the underutilization costs (the out-of-pocket costs for unused stock). A positive shift to the mean is described by $\alpha > 0$ and a higher weighting of overutilization costs (penalty costs) is described by $\beta > 1$. Restricting the data from [41] to the two settings corresponding to our study, i.e. the respective critical ratios of the penalty problem, we obtain values of 0.39 for α and 2.29 for β. Applying the same analysis on our data, we obtain an α value of 0.69 and a β value of 2.96. The values of α (one-tailed Welsh's t-test, $p < 0.005$) and β (one-tailed Welsh's t-test, $p = 0.021$) are both significantly higher in our case. Thus, we confirm Hypothesis 3: Surgeons confronted with planning surgery durations show a stronger shift to the mean and a stronger weighting of overutilization costs than decision makers in comparable inventory newsvendor studies.

$$p_t = \alpha \cdot \mu + (1 - \alpha) \cdot F^{-1} \left(\frac{\beta \cdot c^o}{\beta \cdot c^o + c^u} \right) + \epsilon_t \quad (4)$$

To demonstrate that our findings are not solely caused by different educational backgrounds, we repeated our study with business students as subjects, a comparable subject group to that used in [41], in our first follow-up study. In this setup, we obtain values of 0.59 for α and 2.82 for β. The values for α (one-tailed Welsh's t-test, $p < 0.005$) and β (one-tailed Welsh's t-test, $p = 0.059$) are significantly

above the values observed in [41]. Thus, we conclude that H3 did not rely on a subject effect only. However, the effect for α is slightly and significantly lower (one-tailed Welsh's t-test, $p = 0.005$), the one for β is slightly but not significantly lower (one-tailed Welsh's t-test, $p = 0.376$) than those for experienced doctors. We demonstrate that subjects put in a specific context partially adopt obvious professional norms in decision making independent of their educational background. However, our results indicate that the effects are less pronounced than with professionals.

In the second follow-up study, we tested whether the type of distribution influences the results. A symmetric normal distribution and an asymmetric lognormal distribution were applied to the low quantile and the high quantile case. [5] compared a newsvendor setting applying a normal and a uniform distribution and found comparable behavioral effects for both distributions. The mean bias α, however, seemed to be a bit more pronounced for the normal distribution. On average, they found the α-values to be around 14 % above the respective values of the uniform distribution. We obtain values of 0.74 for α and 2.45 for β for the normal distribution, and 0.81 for α and 2.77 for β for the lognormal distribution. We draw three main conclusions from these results: First, we can confirm our main findings for more realistic distributions. Second, consistent with [5], the normal and lognormal distributions lead to significantly higher

Fig. 4 Comparison to the study of Schiffels et al. [40]
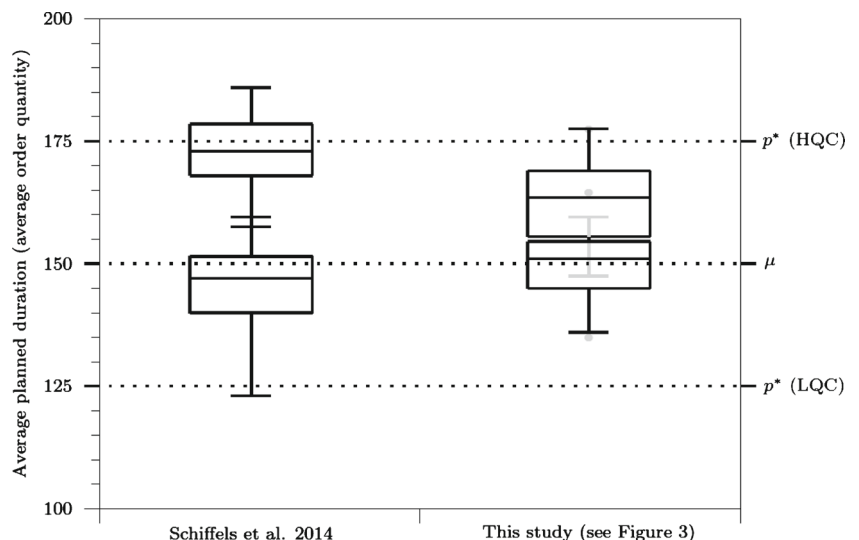
values for α compared to the uniform distribution (one-tailed Welsh's t-test, $p < 0.005$). This is intuitive as the optimum is much closer to the mean in the normal and lognormal distribution. Third, there seems to be no significant difference regarding the overestimation of overutilization as the values for β remain stable (two-tailed Welsh's t-test, normal versus uniform: $p = 0.511$, lognormal versus uniform: $p = 0.928$, normal versus lognormal: $p = 0.597$).

The third follow-up study focuses on the influence of other surgeries. We started with the first situation of planning the duration of a single surgery in isolation (repetition of our main experiment). In addition, we tested identical settings with additional but not cost relevant information: in the second situation, the surgery is followed by a another surgery planned and undertaken by another doctor. In the third situation, the surgery follows another surgery planned with a short duration, thus having a high probability that the realized duration exceeds the planned duration. In the forth situation, the surgery follows another surgery planned with a long duration, thus having a small probability that the realized duration exceeds the planned duration. As expected, the planned durations of the repetition matched the results of our main experiments (LQC: 149.0, Wilcoxon $p = 0.833$, HQC: 163.5, Wilcoxon $p = 0.527$). Knowing that another surgery will immediately follow did not lead to significant differences (LQC: 149.2, Wilcoxon $p = 0.971$, HQC: 167.7, Wilcoxon $p = 0.477$). Knowing that the surgery to be planned immediately follows another surgery did not lead to significantly different planned durations, neither if the preceding surgery was probably planned too short (LQC: 138.5, Wilcoxon $p = 0.118$, HQC: 166.2, Wilcoxon $p = 0.854$) nor if it was probably planned too long (LQC: 140.4, Wilcoxon $p = 0.272$, HQC: 166.5, Wilcoxon $p = 0.501$). The results of these stylized settings are consistent with a questionnaire we distributed among the 26 doctors after the experiment where we asked whether planning surgery durations in practice is affected by preceding or succeeding surgeries. A scale between 1 and 7 was employed (1 = plan much shorter than usual, 4 = plan the same as usual, 7 = plan much longer than usual). In the questionnaire, we further distinguished between situations with high and low OR utilization. In both utilization cases, doctors stated that planning surgery durations is not affected by following surgeries (low utilization: 3.9, Wilcoxon $p = 0.700$, high utilization: 3.9, Wilcoxon $p = 1.00$). If preceding surgeries exist, doctors plan slightly shorter durations (3.7, Wilcoxon $p = 0.054$) in cases of high OR utilization, while they plan slightly longer durations (4.3, Wilcoxon $p = 0.029$) in cases of low OR utilization. Furthermore, we tested planning a set of consecutive surgeries. Based on our interviews with surgeons, surgery durations are typically planned independently. However, situations exist where series of consecutive surgeries are planned. For references

on planning the duration of a series of cases we refer to [2, 47], and [16]. In our final experimental setting three surgeries immediately follow each other, and planning errors do average out. The resulting common duration is a convolution of three uniform distributions, includes values between 300 and 600 minutes, and has an expected value of 450 minutes. The optimal durations are 414 minutes for the LQC and 485 Minutes for the HQC. Percentiles and a graphical representation of the density function (that resembles a normal distribution) were provided to all subjects. The doctors planned on average 426.3 minutes in the LQC, and 495.0 minutes in the HQC. These values planning three consecutive surgeries are not significantly different to the planned durations of a single surgery multiplied by a factor of three (LQC: 447, Wilcoxon $p = 0.470$, HQC: 490, Wilcoxon $p = 0.811$). Thus, doctors seem to neglect the pooling effects among surgeries.

4 Conclusion

Many studies have shown that human behavior has a great impact on operations management decisions. Although the operating theater is the most expensive resource in hospitals, and its efficient usage is crucial, the behavior of health care decision makers in hospitals is generally ignored in research. It was challenging and took a tremendous effort to run the experiments with 66 experienced surgeons. However, we believe that this effort was necessary to gain acceptance with health care professionals and to close the gap between inventory management problems and the health care context. We are able to replicate basic biases known from previous newsvendor experiments in our study. Furthermore, we also demonstrate that different motivation schemes affected by professional norms in the health care sector lead to significantly different results. To be consistent with the inventory literature, we chose a uniform distribution of surgery durations in our main experimental setup. As surgery durations tend to follow a lognormal distribution [45] we validate the effects in an additional study. Consistent with [5], we observe the same behavioral biases for different, both symmetric and asymmetric, distributions. The findings of our main study are robust regarding subjects, distributions of duration, and hold for tasks of planning multiple surgeries as well. 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 and thus provides opportunities for future research. The newsvendor approach, and especially the simplified experimental framework, are stylized models of reality to investigate the

behavioral effects considering the trade-off between planning too long and too short surgery durations. In practice, there are many other factors that influence planning surgery durations, such as capacity limitations (e.g. the only slot available is shorter than the desired planned time), scheduling restrictions (e.g. durations are only planned in 15 minute intervals), and interpersonal effects (e.g. the doctor with the succeeding surgery is particularly unhappy in case of delays). Further research on these factors might lead to a better understanding of planning behavior. Furthermore, we assume a cost minimization model to define the negative effects of overutilization and underutilization. However, different hospitals might employ different incentive schemes. Future research could provide insight whether those lead to different planning behavior, and which schemes are suitable to minimize OR inefficiency. Additional impact on waiting and admission times arises if sequencing and appointment scheduling decisions are discussed. [11] and [27] discuss effects of sequencing decisions and appointment scheduling on patient waiting times. Thus, analyzing these issues from a behavioral point of view is an interesting opportunity for future research. In our experiment, we did not find any significant learning behavior. For long run experiments we would expect small learning effects as described by [6]. The strongest learning effects can be expected if the time intervals between planning surgeries in the same situation 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 we gave full disclosure of the distribution in our experiment, we do not investigate how surgeons account for diversity of situation and concentrate on uncertainty of the duration. Important aspects that are out of scope of our paper are, e.g., staff assignment or patient characteristics. An interesting empirical research project would be to analyze surgeons' behavior considering their assessment of different information on diversity aspects.

As planning of surgery durations is a task of high economic impact for all hospitals, and as we have shown significant and systematic non-optimal behavior of experienced surgeons, important managerial implications may be derived. From our findings one can infer 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 with training. The research described in this paper helped the hospital that triggered this study in several ways: First, the associated consequences of planning too long and too short durations were analyzed and evaluated. Profitability issues the hospital was facing could be partially traced back to low OR utilization. Second, a target critical ratio was defined by the hospital management to decrease planned surgery durations in most departments in order to increase OR utilization. Third, the hospital management, OR management and the surgeons in charge were informed about behavioral biases when planning surgery duration. Based on this, guidelines for planning surgery durations were defined. In consistency with the findings of our experimental studies, seven out of ten medical specialties sharing the same OR resources systematically planned between 5 % and 25 % too long, while three specialties systematically planned around 5 % too short. Each specialty was provided with feedback whether they should rather plan more or less time than previously to meet the target critical ratio according to the new guidelines. Besides the recommendations the project helped the hospital management to gain a better understanding of the complexity as well as the behavioral biases their employees encounter. It also provided the surgeons with a better understanding of managerial targets. The experimental study with experienced surgeons was of great importance for the project in order to gain acceptance in the hospital. It seemed to be infeasible to convince surgeons to adapt their behavior based on inventory studies.

As the health care sector is the largest industry in industrialized countries in terms of number of employees, and human decision making plays an important role, more research should be conducted in this field. We hope to encourage future research since we are convinced that many biases in the field of behavioral health care operations management are still to be discovered, and managing these biases could greatly impact the health care sector.

Appendix

Instructions of the main experiment: The instructions are translated from German and shortened since the original instructions also contain screenshots. 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 result-ing amount which is converted by a factor of ECU ... = €1. Depending on your performance, the payoff will be between €5 and €5.

References

1. Abouleish AE, Dexter F, Epstein RH, Lubarsky DA, Whitten CW, Prough DS (2003) Labor costs incurred by anesthesiology groups because of operating rooms not being allocated and cases not being scheduled to maximize operating room efficiency. Anesth Analg 96(4):1109–1113
2. Alvarez R, Bowry R, Carter M (2010) Prediction of the time to complete a series of surgical cases to avoid cardiac operating room overutilization. Can J Anesth/J Can d'anesthésie 57(11):973–979
3. Andersen LB (2009) What determines the behaviour and performance of health professionals? public service motivation, professional norms and/or economic incentives. Int Rev Adm Sci 75(1):79–97
4. Bendoly E, Croson R, Goncalves P, Schultz K (2010) Bodies of knowledge for research in behavioral operations. Prod Oper Manag 19(4):434–452
5. Benzion U, Cohen Y, Peled R, Shavit T (2008) Decision-making and the newsvendor problem: an experimental study. J Oper Res Soc 59(9):1281–1287
6. Bolton GE, Katok E (2008) Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. Manuf Serv Oper Manag 10(3):519–538
7. Bolton GE, Ockenfels A, Thonemann U (2012) Managers and students as newsvendors. Manag Sci 58(12):2225–2233
8. Brailsford S, Schmidt B (2003) Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. Eur J Oper Res 150(1):19–31
9. Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. Eur J Oper Res 201(3):921–932
10. Carter M (2002) Diagnosis: Mismanagement of resources. OR/MS Today 29(2):26–32. http://www.orms-today.org/orms-4-02/frmismanagement.html
11. Cayirli T, Veral E, Rosen H (2006) Designing appointment scheduling systems for ambulatory care services. Health Care Manag Sci 9(1):47–58. doi:10.1007/s10729-006-6279-5
12. De Véricourt F, Jain K, Bearden JN, Filipowicz A (2013) Sex, risk and the newsvendor. J Oper Manag 31(1):86–92
13. Denton B, Viapiano J, Vogl A (2007) Optimization of surgery sequencing and scheduling decisions under uncertainty. Health Care Manag Sci 10(1):13–24
14. Dexter EU, Dexter F, Masursky D, Kasprowicz KA (2010) Prospective trial of thoracic and spine surgeons' updating of their estimated case durations at the start of cases. Anesth Analg 110(4):1164–1168
15. Dexter F, Traub RD (2002) How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesth Analg 94(4):933–942
16. Dexter F, Wachtel RE (2014) Ophthalmologic surgery is unique in operating room management. Anesth Analg 119(6):1243–1245
17. Dexter F, Macario A, Ledolter J (2007a) Identification of systematic underestimation (bias) of case durations during case scheduling would not markedly reduce overutilized operating room time. J Clin Anesth 19(3):198–203

18. Dexter F, Xiao Y, Dow AJ, Strader MM, Ho D, Wachtel RE (2007b) Coordination of appointments for anesthesia care outside of operating rooms using an enterprise-wide scheduling system. Anesth Analg 105(6):1701–1710

19. Eijkemans MJC, van Houdenhoven M, Nguyen T, Boersma E, Steyerberg EW, Kazemier G (2010) Predicting the unpredictable a new prediction model for operating room times using individual characteristics and the surgeons estimate. Anesthesiology 112(1):41–49

20. Fischbacher U (2007) z-tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10(2):171–178

21. Fügener A (2015) An integrated strategic and tactical master surgery scheduling approach with stochastic resource demand. J Bus Logist (Forthcoming)

22. Fügener A, Hans EW, Kolisch R, Kortbeek N, Vanberkel PT (2014) Master surgery scheduling with consideration of multiple downstream units. Eur J Oper Res 239(1):227–236

23. Fügener A, Edenharter G, Kiefer P, Mayr U, Schiele J, Steiner F, Kolisch R, Blobner M (2015) Improving icu and Ward Utilization by Adapting Master Surgery Schedules: A Case Study. Anesthesia & Analgesia Case Reports (Forthcoming)

24. Gino F, Pisano G (2008) Toward a theory of behavioral operations. Manuf Serv Oper Manag 10(4):676–691

25. Goodrick E, Salancik GR (1996) Organizational Discretion in Responding to Institutional Practices: Hospitals and Cesarean Births. Administrative Science Quarterly, pp 1–28

26. Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: A survey. Health Care Manag Sci 14(1):89–114

27. Gul S, Denton BT, Fowler JW, Huschka T (2011) Bi-criteria scheduling of surgical services for an outpatient procedure center. Prod Oper Manag 20(3):406–417

28. Gupta D, Denton B (2008) Appointment scheduling in health care: Challenges and opportunities. IIE Trans 40(9):800–819

29. Katok E (2011) Laboratory Experiments in Operations Management. TutORials in Operations Research, pp 15–35

30. Kremer M, Minner S, VanWassenhove LN (2010) Do random errors explain newsvendor behavior Manuf Serv Oper Manag 12(4):673–681

31. Loch CH, Wu Y (2007) Behavioral Operations Management. Now Publishers Inc

32. Lovejoy WS, Li Y (2002) Hospital operating room capacity expansion. Manag Sci 48(11):1369–1387

33. May JH, Strum DP, Vargas LG (2000) Fitting the lognormal distribution to surgical procedure times. Decis Sci 31(1):129–148

34. May JH, Spangler WE, Strum DP, Vargas LG (2011) The surgical scheduling problem: Current research and future opportunities. Prod Oper Manag 20(3):392–405

35. Moritz BB, Hill AV, Donohue KL (2013) Individual differences in the newsvendor problem: Behavior and cognitive reflection. J Oper Manag 31:72–85

36. Olivares M, Terwiesch C, Cassorla L (2008) Structural estimation of the newsvendor model: An application to reserving operating room time. Manag Sci 1(1):41–55

37. Pandit JJ, Carey A (2006) Estimating the duration of common elective operations: Implications for operating list management. Anaesthesia 61(8):768–776

38. Peltokorpi A (2011) How do strategic decisions and operative practices affect operating room productivity Health Care Manag Sci 14(4):370–382

39. Perry JL, Wise LR (1990) The Motivational Bases of Public Service. Public administration review, pp 367–373

40. Rhodes M, Barker P (2007) Operating room utilization. Surg Endosc 21(12):2339–2340

41. Schiffels S, Fügener A, Brunner JO, Kolisch R (2014) On the assessment of costs in a newsvendor environment: Insights from an experimental study. Omega 43:1–8

42. Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. Manag Sci 46(3):404–420

43. Silber JH, Rosenbaum PR, Zhang XM, Even-Shoshan O (2007) Estimating anesthesia and surgical procedure times from medicare anesthesia claims. Anesthesiology 106(2):346–355

44. Strum DP, Vargas LG, May JH, Bashein G (1997) Surgical suite utilization and capacity planning: A minimal cost analysis model. J Med Syst 21(5):309–322

45. Strum DP, May JH, Vargas LG (2000) Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. Anesthesiology 92(4):1160–1167

46. Wachtel RE, Dexter F (2010) Review of behavioral operations experimental studies of newsvendor problems for operating room management. Anesth Analg 110(6):1698–1710

47. Wang J, Yang K (2014) Using type iv pearson distribution to calculate the probabilities of underrun and overrun of lists of multiple cases. Eur J Anaesthesiol (EJA) 31(7):363–370

48. Wright IH, Kooperberg C, Bonar B, 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

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