Three Essays on

Optimization of the Intensive Care Unit (ICU) Management Decisions

Inaugural Dissertation to Obtain the Academic Degree of a Doctor
in Business Administration at the University of Mannheim

submitted by

Jie Bai

Dean:	*Joachim Lutz*
Referent:	*Prof. Dr. Raik Stolletz*
Correferent:	*Prof. Dr. Jens Brunner*

Day of oral examination: 11.02.2022

To my parents,

to my kids Xitong and Yitong,

and to Haoyang

Acknowledgements

It is really an unexpectedly long journey to finish this dissertation. Something happened, something changed, and difficulties came out from time to time. But all in all, I feel fortunate to have spent these years in my life to be doctorate student. The most important thing in this journey along with me, is the supports from my colleagues and the family.

First and foremost, I am really grateful to my supervisor, Prof. Jens Brunner. When I joined the chair, I was excited about the research topic and the ideas, but on the other side, I had no experience of doing research. It was Jens who guided me through with his valuable suggestions and patiently supports. Because of raising two kids, I spent a lot of time and effort on the family. Jens never pushed or criticized me because of this. I appreciate his understanding, empathy, and encouragement. Without these, I might give up and would never finish this dissertation. Special thanks also to Prof. Stolletz for all the help and support in the last years.

I want to thank Prof. Andreas Fügener and Prof. Jochen Gönsch. Through the collaboration with them, I learned a lot about being rigorous and organized in research. I also want to thank my colleague Dr. Melanie Erhard, Dr. Sebastian McRae, Dr. Sebastian Kohl, Dr. Anna Muzalyova, Ms. Antje Hilbrig, and other colleagues in the chair. I was lucky to know them during my PhD. Thank you all for your help and support in research, teaching, administrative work, and even daily life.

Furthermore, I would like to thank my parents for their unconditional support and trust since I was a child, enabling me to achieve all my goals. To my kids, Xitong and Yitong, you are always my source of happiness and motivation. Thank you for all your unlimited love for your imperfect mother. Lastly, I don’t know how to describe my thanks to my husband Haoyang. Through his love, support, and understanding, I was kept on track.

Obersulm, September 2021

Jie Bai

Summary

The intensive care unit (ICU) is one of the most crucial resources in the hospital. Due to its function of catering to patients with severe disease and in need of constant and close monitoring, the ICU requires highly trained doctors and nurses with an above-average staff-to-patient ratio and specialized equipment. Improper ICU management causes many negative effects in the ICU itself and in other connected departments along the patient care path. This dissertation presents three papers on optimizing the ICU management decisions. The first paper provides the first structured and comprehensive review of ICU problems in OR/MS. The relevant papers covering the topics "importance of the ICU for hospital patient flow" and "the ICU management problems" are discussed based on a new framework. Furthermore, the modeling methods and solution approaches are classified and discussed in detail. Based on the analysis of existing papers, the future research topics are also addressed along three streams. The second paper proposes a discrete-time MDP model to find admission and early discharge policies that minimize these negative consequences. By minimizing the medical consequences, the approach demonstrated significantly outperforms a myopic policy as applied by most hospitals in practice. An efficiency frontier covering medical and monetary perspectives is developed and thereby is contributed to the ongoing discussion on the trade-off between medical quality and monetary costs. The third paper compares eleven different management policies based on different KPIs by a simulation study. In comparison to the baseline case running on a FCFS rule, it is shown that any management policy is superior regardless of the evaluation criteria. The 30 most frequently occurring patient paths, based on the practical dataset of more than 75,000 patient records from a German teaching hospital, are simulated.

Content

List of Figures

List of Tables

1. Introduction

The intensive care unit (ICU) is one of the most crucial resources in the hospital. Due to its function of catering to patients with severe disease and in need of constant and close monitoring, the ICU requires highly trained doctors and nurses with an above-average staff-to-patient ratio, as well as specialized equipment [1]. The high fixed cost of ICU strongly suggests a tight control of the capacity. As reported by the Federal Statistical Office of Germany, the number of hospitals with ICU beds has been continuously dropping in Germany, from 1286 in 2007 to 1160 in the year 2017. In total, 2,131,216 cases were treated by 28,031 ICU beds in 2017.

Meanwhile, the demand in the ICU is growing, especially in the aging time. The Federal Statistical Office's demography report states that the number of people aged 80 or over in Germany will continue to increase steadily and reach its peak of about 10 million in 2050, which will constitute between 12% and 13% of the population. What's worse, the stochastic and unpredictable features of the demand add challenges to manage the capacity optimally. Therefore, simply focusing on high utilization levels may lead to undesired consequences, such as over-beds, patient diverting, and early discharge. Further on, the patient's higher mortality rate and readmission rate are resulted in [2]. Moreover, ICU plays the bottleneck role in a hospital-wide patient flow. Once the patient-in-need admission is restricted by capacity, correlated up-and downstream departments in the hospital will be influenced, even overcrowded.

This dissertation focuses on the management decisions in the ICU and addresses different levels of management topics in three chapters. Starting from a "strategic" level study, a framework of the related literature in ICU management is developed. According to the research gaps, two research problems are proposed and discussed in the following two chapters. The first one tries to find optimal admission and discharge decisions in the ICU, while the second one discusses how the ICU decisions influence the hospital patient flow, namely the up- and downstream departments.

The ICU management topics have been developed for years. Many researchers in OR/MS area have contributed to the topic over the last decades. The research topics are various, such as staff scheduling, capacity planning, patient LOS and mortality prediction, etc. The applied methodologies cover almost all the standard techniques in OR/MS, such as simulation, integer programming, and dynamic programming. Which topics are more attractive to practice? To different topics, which methodology fits better? What are the research gaps? A holistic overview of ICU management problems in OR/MS was still missing.

The first article of this dissertation (Chapter 2)[1] is the first comprehensive and structured review of the application of OR/MS techniques in ICU management problems. We start our review by illustrating the critical role that the ICU plays for hospital patient flow based on 18 recently published papers. These papers involve the ICU in combination with other upstream and downstream departments, but do not show a clear focus on the ICU. Then we turn to 52 papers discussing the ICU management problem. To analyze these papers, we introduce a two-dimensional framework (time horizon of decisions and research topics) to discuss the existing literature on ICU management problems. Relevant topics are *ICU patient flow optimizatio*n *and control* (*patient admission*, *treatment/LOS*, *discharge*, and *readmission*), *bed capacity management*, and *staff scheduling*. We further review the literature according to the definition of uncertainties, modeling methods, and solution approaches, which have not been covered by any other review. Based on the evaluation of current papers, we highlight research gaps and future trends following our classification logic. Thus, we analyze gaps regarding research topics, modeling methods, and solution approaches.

From the first article, we find one of the research gaps is admission decision-making with dynamic programming approaches. Specifically, the following questions remain to be answered: during the capacity shortages, should higher priority be assigned to the new arrival patients or existing inpatients? Should several beds in the ICU be reserved for more critical patients, or rather prefer a high utilization level? Along with this consideration, how could we cluster the patients into different groups? None of these decisions are easy to make. Many ICU managers employ a myopic strategy (only considering direct and immediate effects) of patient control. When there are free capacities in the ICU, the "first-come, first-serve" policy is always applied. When a new patient arrives at a fully occupied ICU, different myopic policies (such as keeping the new patient waiting in the other department, early discharge of an existing patient) might be implemented in different situations to minimize the direct negative consequences. These decisions are typically made based on the judgment of the ICU

[1] Bai, J., Fügener, A., Schoenfelder, J., & Brunner, J. O. (2018). Operations research in intensive care unit management: a literature review. *Health care management science*, *21*(1), 1-24.

physicians. However, we should never forget that each decision leads to immediate consequences and delayed effects, affecting future decisions.

To find optimal management policies considering both the direct and future effects, we model the admission and discharge control of the ICU as a discrete-time Markov decision process (MDP) in our second article (Chapter 3)[2]. After modeling based on realistic assumptions, we evaluate the policies resulting from the MDP in two case studies capturing different management objectives – a medical and a monetary perspective – based on real-world data from a large German teaching hospital. The results show that the optimal policy from our MDP model can considerably reduce the negative effects from a medical perspective – the mortality due to capacity shortages may be reduced by 21% in our case study compared to myopic policies. We discuss the impact of different combinations of cost parameters on solutions and on the robustness of our model in case of over- or underestimation of cost parameters. Our approach provides a novel contribution in two directions: First, it enables an analytical demonstration of the trade-offs between medical and monetary goals when designing ICU admission and discharge policies. The impact of different goals is large, and deciding on the percentage of resources to be spent on intensive care is of great societal importance. Second, our model provides optimal holistic policies combining admission and discharge decisions in an ICU based on realistic assumptions. Those policies may lead to direct implications for ICU management, such as reserving a certain number of beds for internal emergencies or diverting ambulances if a certain threshold of critical patients is currently in the ICU.

As we discussed, the ICU plays the bottleneck role in a hospital-wide patient flow. The ICU management policy influence not only ICU itself, but also the correlated up-and downstream departments in the hospital. Currently, there are papers discussed different ICU management policies in different settings [3]. Early discharging current patients in the ICU to the downstream departments in order to create space for the new patients [2, 4], denying the admission requests from the upstream departments [5], and rescheduling the operations [6, 7] are the most frequently discussed policies to manage the ICU capacity. They are supposed to work well based on their own key performance indicators (KPIs). However, what are the performances of these policies in the same scenario? Which ones are better when considering the same set of KPIs? How do the ICU management policies influence the upstream and downstream departments? Before our paper, we haven't found one to answer these questions.

[2] Bai, J., Fügener, A., Gönsch, J., Brunner, J. O., & Blobner, M. (2021). Managing admission and discharge processes in intensive care units. *Health care management science*, 1-20.

Therefore, the patient flow centered by the ICU is simulated to evaluate the performances of eleven different management policies in Chapter 4[3]. Nine KPIs are compared, covering different aspects, as well as different departments. The simulation model is based on a dataset with 75,934 patient records in 2015 from one of the largest teaching hospitals in south Germany. The dataset covers 1,215 beds in general wards, 45 emergency beds, 20 IMC beds, and 30 ICU beds. The 30 most commonly occurring patient paths from the dataset are integrated in the simulation model. The results clearly indicate that the introduction of control policies positively impacts patient status, length of stay (LOS), and cost when compared to the baseline case (first-come-first-served policy, FCFS). The parameters of our model can be flexibly adjusted to the parameters from different hospitals. Therefore, it can work as a managerial reference to practice. The managers can choose proper policies according to their goals (specific KPIs).

Furthermore, it would be an exciting research direction to find a stylized way to model the problem with leading-edge techniques. A machine-learning-supported decision-making framework is proposed in the last chapter (Chapter 5), which includes the prediction by supervised learning and decision-making by reinforcement learning. The other potential research trends are discussed as well.

[3] Bai, Jie, Jens O. Brunner, and Steven Gerstmeyr. "Simulation and evaluation of icu management policies." 2020 Winter Simulation Conference (WSC). IEEE, 2020.

2. Operations Research in Intensive Care Unit Management: A Literature Review

Co-authors:

Andreas Fügener

Faculty of Management, Economics and Social Sciences, University of Cologne, Germany

Jan Schönfelder, Jens O Brunner

Faculty of Business and Economics, University of Augsburg

Published in:

Health Care Management Science, 21(1) (2018), pages 1-24, DOI: 10.1007/s10729-016-9375-1

Abstract:

Insufficient ICU capacity causes many negative effects not only in the ICU itself, but also in other connected departments along the patient care path. Operations research/management science (OR/MS) plays an important role in identifying ways to manage ICU capacities efficiently and in ensuring desired levels of service quality. Thus, numerous papers on the topic exist. The goal of this paper is to provide the first structured literature review on how OR/MS may support ICU management. We start our review by illustrating the important role the ICU plays in the hospital patient flow. Then we focus on the ICU management problem (single department management problem) and classify the literature from multiple angles, including decision horizons, problem settings, and modeling and solution techniques. Based on the classification logic, research gaps and opportunities are highlighted, e.g., combining bed capacity planning and personnel scheduling, modeling uncertainty with non-homogenous distribution functions, and exploring more efficient solution approaches.

2.1. Introduction and Motivation

The intensive care unit (ICU) is a crucial and expensive resource in the hospital. Halpern & Pastores [8] report that in 2005 the ICU costs represented around 15% of total hospital costs, and 4.1% of the U.S. health expenditures. Treating a patient in the ICU is much more expensive than in a regular ward – in the literature, ICU cost rates of four [9] to six [10] times the costs of a regular ward have been reported. In the U.S., more than 5.7 million patients are admitted to ICUs annually, up to 60% of emergency department admissions are transferred to an ICU, and approximately 20% of the acute care admissions result in an ICU admission [11]. The ICU is also a special department dedicated to the most critically ill patients. Therefore, poor decisions in the ICU may threaten patient lives, e.g., when a patient has to be deferred due to capacity shortages. Increasing medical requirements, expensive resources, and high uncertainty and variability often limit the capacity in many ICUs. The resulting overcrowding has many negative effects such as increased morbidity and mortality, overworked staff, decreased revenue, and even congestion of the patient flow in the entire hospital. The rising importance of multi-resistant infectious diseases is leading to additional challenges in the management of ICUs - in these cases, private rooms are recommended to reduce infection rates [12]. Thus, hospital managers have a great interest in understanding how to provide high quality care with limited capacity.

Several factors make ICU management an urgent and difficult problem. Due to the high degree of specialization, ICU resources are very expensive. To ensure that the patients with most severe and life-threatening diseases and injuries can be treated adequately, the ICU is staffed with highly trained nurses. Additionally, specialized expensive equipment is required to provide constant and close monitoring. From an economic point of view, a high utilization of the ICU is desirable. However, contrary to many other services, medical treatment of ICU patients cannot be delayed in most cases. Therefore, focusing ICU capacity management on reaching a high utilization level could lead to the following negative consequences: Scheduled surgeries might be postponed or canceled [13]. Over-beds might need to be used, leading to over-loaded staff and decreasing quality of care [14]. Finally, the patient might be discharged early, leading to an increased likelihood of readmission, longer second lengths of stay (LOS), and higher mortality rates [15]. Iapichino et al. [16] conclude that, in general, higher occupancy of ICUs leads to higher mortality rates. Hence, the tradeoffs between economical utilization and quality of care should be well balanced.

Compared to automated manufacturing plants, an ICU is a much less predictable planning system. There exist many sources of uncertainty, increasing the difficulty of ICU management. For example, patient arrival patterns are relatively hard to predict, because patients may be directly admitted to the ICU, arrive from a scheduled surgery, or transfer from the emergency department with or without a stop in the operating room. The length of stay is another source of uncertainty due to the variability of conditions when patients arrive [17]. Additionally, there is uncertainty on the supply side. The delivery time of relevant medical resources needed in the ICU, for example, is stochastic, and absenteeism of staff causes uncertainty in staff scheduling.

OR/MS addresses these difficult issues, and many researchers have contributed to the topic over the last decades. There are several reviews on healthcare operations that involve parts of the ICU planning problem. Hulshof et al. [18] present a taxonomy to review capacity planning decisions in different hierarchical levels and several departments. Lakshmi & Appa Iyer [19] review the applications of queueing theory in healthcare problems. The case mix planning problem, or service mix problem, is reviewed in Hof et al. [20]. Tierney & Conroy [21] provide a review on how to measure, report, and interpret ICU occupancy levels from a medical perspective. Further, they investigate optimal ICU occupancy levels. Additionally, a few other reviews focus on the medical perspective [22–24]. None of these reviews focus on ICU management, a holistic overview on ICU management problems in OR/MS is still missing.

Our contribution to the literature is manifold. We present the first comprehensive and structured review of the application of OR/MS techniques in ICU management problems. We start our review by illustrating the important role that the ICU plays for hospital patient flow based on 18 recently published papers. These papers involve the ICU in combination with other upstream and downstream departments, but do not show a clear focus on the ICU. Then we turn to 52 papers discussing the ICU management problem. To analyze these papers, we introduce a two dimensional framework (time horizon of decisions and research topics) to discuss the existing literature on ICU management problems. Relevant topics are *ICU patient flow optimization and control* (*patient admission*, *treatment/LOS*, *discharge*, and *readmission*), *bed capacity management*, and *staff scheduling*. We further review the literature according to the definition of uncertainties, modeling methods, and solution approaches, which have not been covered by any other review. Based on the evaluation of current papers, we highlight research gaps and future trends following our classification logic. Thus, we analyze gaps regarding research topics, modeling methods and solution approaches. Examples are sizing and pooling all corresponding ICU capacities in a region, combining bed capacity planning and personnel scheduling, applying non-homogenous distributions to model uncertainties, and exploring more efficient solution approaches.

The remainder of the paper is organized as follows. Section 2.2 provides an overview of hospital patient flows and discusses the role the ICU plays in holistic hospital settings. Section 2.3 proposes a framework to structure the papers according to ICU management and methodology applied. Section 2.4 discusses future research trends based on the evaluation and identification of gaps in the literature. Finally, Section 2.5 presents the summary and conclusion.

2.2. ICU in the Hospital Patient Flow

Patient pathways connect the ICU to other units inside and outside of the hospital. Decisions made in the ICU also influence upstream departments, such as the emergency department (ED) and the operating theatre (OT), and downstream departments, such as the intermediate care unit (IMC) and the general wards. In this section, we first describe the patient flow in the hospital and then discuss how other departments are influenced by decisions in the ICU.

2.2.1. The Hospital Patient Flow

A stylized illustration of a generic hospital patient flow model involving the ICU is shown in Figure 2.1. Patients may arrive at the ED of a hospital. These patients can be divided into three groups. Patients of the first group are not seriously ill and will be discharged directly from the ED. The patients of the second group are more critically ill and are sent to a ward or the ICU. The patients of the third group are sent to the OT department for surgery. Another arrival stream of patients is via elective, i.e. planned, surgeries. Inpatients are admitted to a ward prior to surgery. Outpatients, in contrast, are sent directly to the OT and typically leave the hospital after surgery without staying overnight at a ward. After surgery, patients are transferred to a ward or the ICU, with or without a stop in the post-anesthesia care unit (PACU). When ICU patients are in good enough condition, they are sent to a ward or the IMC unit. When the health conditions of IMC and ward patients deteriorate, they are sent back to the ICU. Patients are usually discharged from the wards. In rare cases patients are directly admitted at the ICU (e.g., transferred from other hospitals) or directly discharged from the ICU (e.g., transferred to other hospitals or died within the ICU).

As Figure 2.1 shows, patient paths connect the ICU with many other departments, and their performances are interdependent. Kramer et al. [25] demonstrate that at both the patient and the unit level the hospital and ICU LOS are strongly correlated. Blocking of the ICU may lead to increased waiting times in upstream units, e.g. in operating rooms, leading to a resulting overall reduction in patient throughput [26]. Vice versa, the ICU service quality can be influenced by the other departments. For instance, if the OT schedule does not consider scarce ICU capacities, it may schedule additional surgeries leading to decreasing ICU service levels [27]. Thus, holistic optimization of the hospital flow should lead to better results compared to managing hospital units in isolation. However, integrating hospital units leads to additional complexity, responsibility conflicts, and interface issues. Thus, despite of the theoretical dominance, there are no approaches in the literature that deal with a truly holistic optimization. However, some papers do integrate the ICU in the (partial) patient flow.

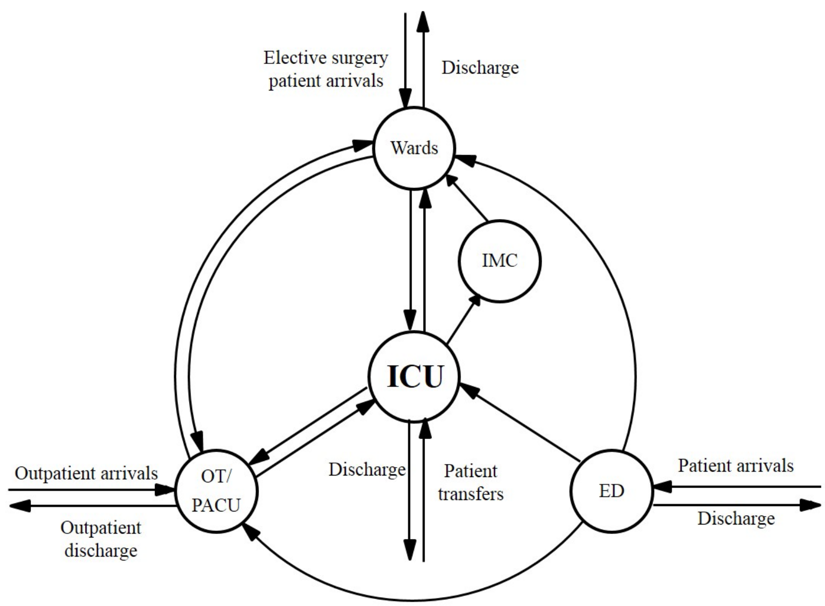


Figure 2.1 Hospital patient flow with the ICU at center

2.2.2. The Literature Integrating the ICU in the Patient Flow

As mentioned, ICU capacity planning is significantly important due to the interaction with different departments. The ICU capacity also influences the admission and discharge processes and the scheduling of connected departments. Additionally, the capacity limitation does not only block the ICU but also the upstream and downstream departments. Many publications discuss hospital patient flow optimization problems in multi-departments including the ICU. We employed a structured search method to detect the papers until December 2015. First, we searched several relevant databases, such as JSTOR, PubMed, ScienceDirect, Web of Science, and Wiley Online Library, and Google scholar using relevant key words, such as “Intensive care”, “ICU”, “Critical care” in combination with “Operations research”, “Operations management”, “Management science”, “Hospital patient flow”, “Multi-department”, “Operations room/theatre”, and “Ward”. In the next step, a backward and forward search of all references was performed. As this is not the core part of our review, we considered the papers published in the previous 15 years (from 2000 to 2015) in OR/MS journals or conference proceedings. In total, 18 recently published articles were detected (Table 2.1).

Three aspects are investigated in Table 2.1: the interacting departments in the hospital patient flow, the role an ICU plays in the respective model, and the solution methodologies. These publications study the interactions between several departments, namely the upstream departments ED and OT, and the downstream department wards, PACU and IMC. We find that 95% of the papers include more than two other departments, the OT department is relevant in 90% of the papers, and the general wards are taken into account in 56% of the papers. In some papers, several other departments such as the obstetrics units [13, 28–30] and telemetry [28–30] are referred to. The ICU plays different roles in the multi-departments optimization. In two thirds of the articles, the ICU is a part of the objective, e.g. if both ICU and OT utilization are optimized [31–36]. In the other cases, it acts as a constraint, e.g., if ICU capacity restricts an OT planning problem [6, 7, 37–46]. The solution methodologies are classified into two groups. One third are purely stochastic methods [31, 33–35, 41, 44], while two thirds are mainly deterministic methods [6, 7, 32, 36–40, 42, 43, 45, 46].

There are several papers that study how the capacity limitation in ICUs influences different stages of the patient flow. Adan & Vissers [39] apply case mix planning to study hospital admission control based on resource limitations, which include the number of ICU beds, operating theatre capacity, and nurse capacity. Similarly, Barz & Rajaram [44] study a hospital patient admission problem with multiple resource constraints. They formulate the elective patients' admission process with a Markov decision process (MDP) to maximize expected net contribution of overbooking costs. Allon et al. [41] focus on the emergency patient admissions influenced by the size of both the ICU and the ED. During overcrowding periods of the EDs incoming ambulances are diverted to neighboring hospitals, which is known as "ambulance diversion". They studied this phenomenon and conclude that when the spare capacity of the inpatient department (the ICU and wards) and the size of the ED increase, the ambulance diversion rate (i.e., the share of diverted ambulances) decreases. Besides the patient admission process, the ICU capacity can also influence the discharge and readmission process. Anderson et al. [43] find statistically significant evidence that surgeons adjust their discharge practices to accommodate the surgical schedule and number of available ICU beds. In addition, Anderson et al. [42] demonstrate that patients who are discharged from highly utilized post-operative units, such as the ICU, are more likely to be readmitted within 72 hours. Gartner & Kolisch [46] focus on hospital wide patient flow scheduling. They classify the majority of elective patients according to their diagnosis-related group (DRG) and clinical path way. Then, they optimize patient flow to maximize the contribution margin considering resource constraints including the ICU. Villa et al. [36] study the restructuring of patient logistics problems to make hospitals more patient-centered. Based on the analysis of three hospital redesign projects, they provide evidence that an organizational model based on a clinical framework (including the intensity of care and patients' length of stay in different departments, e.g., the ICU), can be beneficial both in efficacy and efficiency.

As an upstream department of the ICU, the OT is one of the most important sources of ICU patients. Specifically, most elective surgery patients and part of the emergency patients are transferred from the OT to the ICU. The capacity limitation of OTs and ICUs can cause cancelations of elective patient surgeries or emergency patient diversions. Therefore, Bowers [31] explores the interdependencies of resource availabilities and the daily demand for both ICU and operating theatres. He shows that the capacity for both the ICU and the OT department should be well balanced. McHardy et al. [35] develop a general model encompassing ICUs and operating theatres and analyze how different parameters (arrival rate distributions, lengths of stay distributions, number of beds in the ICU, number of OTs, etc.) influence the system. Fügener et al. [32] develop a cyclic master surgical schedule to reduce demand peaks in ICUs and general wards. Additionally, the ICU capacity also has a great impact on surgery scheduling. High bed occupancy levels often result in stressed-out staff, frequent surgery cancelations, and long surgical wait times [6]. Therefore, Chow et al. [6] and Price et al. [7] propose scheduling surgeries to reduce boarding in the post-surgical beds (PACU and ICU). Several researchers develop cyclic master surgical scheduling approaches that consider bed limitations in the ICU [37, 38, 40, 45].

Moreover, downstream bed shortages can also cause blocking in the ICU, keeping patients from moving forward [29]. Balancing ICU capacities with other inpatient bed departments is another research direction. Marmor et al. [33, 34] predict minimum bed requirements in ICUs and PACUs to achieve high patient service levels. Additionally, they explore the effects of process improvement, such as smoothing surgery schedules and transferring long-stay patients from the ICU, on patient service levels.

We conclude that the ICU plays an important part in the hospital patient flow. It has significant interactions with upstream and downstream departments, which should be more thoroughly investigated in the future. All publications discussed in this section focus on the hospital patient flow partially involving ICUs but do not focus on the ICU itself. Two thirds of the publications treat the ICU as an additional resource only. Nearly all papers focus on operating room planning problems, and all but one paper include at least two additional resources. In the following section, we will review in detail publications focusing on ICU management problems.

Publications	Other hospital departments[1]						ICU roles[2]		Stochastic methods[3]			Deterministic methods[3]	
	ED	OT	Ward	PCU	IMC	Other	Ct	Obj	Queueing	MDP	PA	MP	SA
Adan & Vissers [39]		×	×				×					×	
Adan et al. [40]	×	×			×		×					×	
Allon et al. [41]	×		×				×		×				
Anderson et al. [43]		×	×				×						×
Anderson et al. [42]				×	×		×						×
Barz & Rajaram [44]	×	×	×				×			×			
Bowers [31]		×						×			×		
Chow et al. [6]		×	×	×			×					×	
Fügener et al. [32]		×	×					×				×	
Fügener [45]		×	×				×					×	
Gartner & Kolisch [46]		×	×				×					×	
Marmor et al. [33]		×		×				×			×		
Marmor et al. [34]		×		×				×			×		
McHardy et al. [35]		×						×	×				
Price et al. [7]		×		×			×					×	
Van Houdenhoven et al. [37]		×					×					×	
Van Oostrum et al. [38]		×	×				×					×	
Villa et al. [36]	×	×	×	×		×		×					×
Total	**4**	**16**	**10**	**6**	**2**	**1**	**12**	**6**	**2**	**1**	**3**	**9**	**3**

1 Emergency department (ED); Operating theatre (OT); Inpatient wards (Ward); post-anesthesia care unit (PCU); Intermediate care (IMC)

2 Constraints (Ct); Objectives (Obj)

3 Markov decision process (MDP); Process analysis (PA); Mathematical programming (MP); Statistical analysis (SA)

Table 2.1: Hospital patient flow optimization problems including ICU.

2.3. State of the Art of ICU Management

In this section, we concentrate on papers that present OR/MS methods to solve planning problems directly arising in the ICU. Papers published in English between January 1980 and December 2015 are considered. We employed the same search logic as discussed in Section 2.2 (keyword search, backward and forward search). Additional search terms include "Scheduling", "Admission", "Discharge", "Capacity", and "Patient flow". This search resulted in more than 200 research papers. We included all papers published in OR/MS journals. As most papers from medical journals do not focus on OR/MS methods, we only included papers with "queueing", operations research", "simulation", "patient flow", "capacity" and "staffing". As a result, we detected 52 relevant papers in total, consisting of 31 papers from OR/MS journals, 16 from medical journals, and 5 conference proceedings. Figure 2.2 presents the number of publications by journal type over time. It is evident that ICU management have been increasingly attracting attention in the OR/MS literature.

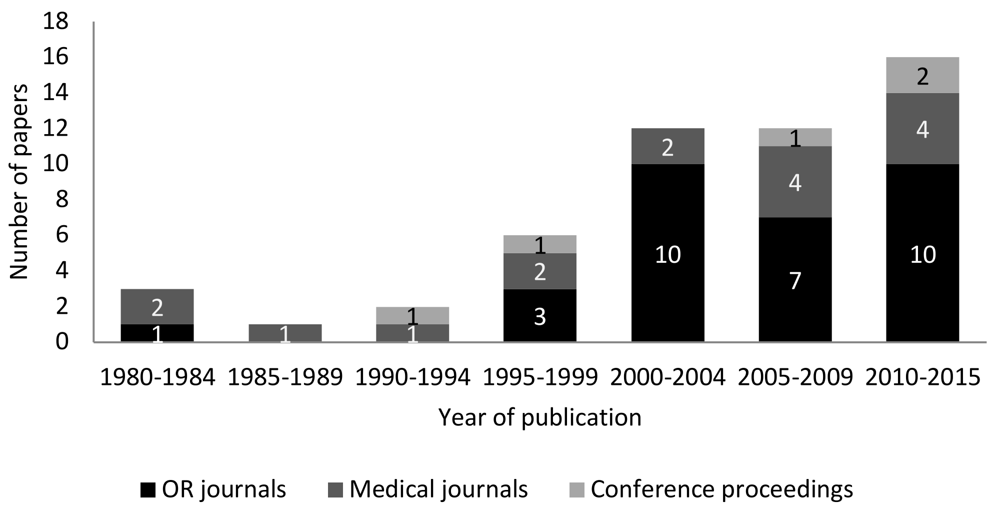


Figure 2.2: Number of papers per journal type and publication year.

Different types of ICUs are discussed in the papers we reviewed (see Figure 2.3, 2 reviews are excluded). Some papers focus on specific ICUs in the hospital, for example the cardiac ICU [15, 47–49], the coronary ICU [50], the neonatal/pediatric ICU [28–30, 51–56], the surgical ICU [57, 58], or a multidisciplinary ICU [59–67]. Some papers study several different ICUs and draw robust and generalized conclusions [68–71]. About 40% of the papers abstractly model and optimize the ICU management problem without specifying the type of ICU. Detailed information is available in the Appendix (Table A2).

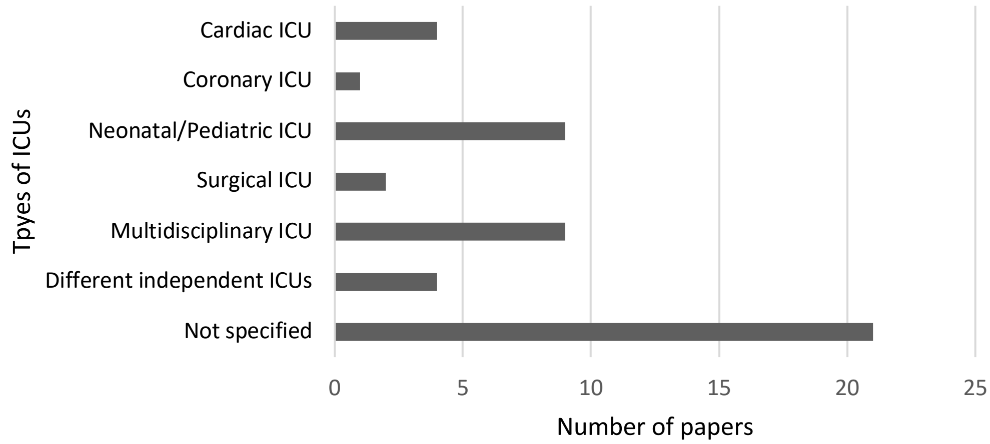


Figure 2.3: Number of papers per ICU type.

Due to different legislation and organizational practices, ICU management problems might differ between geographical regions. Therefore, we summarize the number of papers per country (author affiliation) in Figure 2.4. The papers are from 9 different countries, and 52% are from researchers situated in the U.S. We discuss the effects of type of ICU and geography on patient pathways and modeling methods in Section 2.3.4.

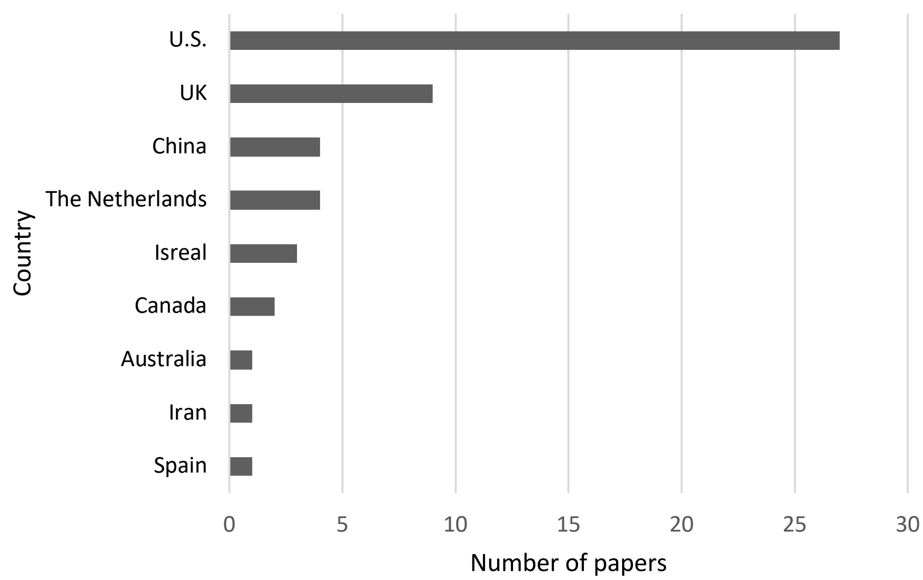


Figure 2.4: Number of papers per country.

Although the body of literature on ICU management problems is not extensive, many aspects of ICU management have been covered. To review the literature structurally, we develop a framework based on two dimensions – time horizon and type of planning problem (see Table 2.2). The time horizon of management decisions in the ICU is differentiated into *long-term, medium-term, and short-term* decisions. Long-term decisions are typically made when designing the ICU and will influence future activities. Medium-term decisions are based on long-term decisions and mostly focus on the optimization of existent, recurrent problems in the ICU. Short-term decisions impact daily activities in the ICU.

Some topics have been extensively discussed in the literature, such as capacity planning, occupancy rate control, and admission and discharge policies. Other topics, such as logistics and layout planning, have been rarely touched. Thus, to detect attractive parts for future research, we differentiate between three main types of management problems in the second classification dimension - *patient flow optimization and control, bed capacity management, and personnel planning*. Patient flow optimization and control involves strategies and policies for *patient admission, treatment/LOS, discharge, and readmission*, as well as the discussion and analysis of patient arrival processes and length of stay. Most papers on readmission discuss the negative influence of increasing early demand-driven discharges on readmission rates. Therefore, we classify readmissions into the group of discharge problems. Bed capacity management discusses questions such as the efficient use of bed resources. Personnel planning focuses on scheduling of nurses and physicians.

Integrating the decision levels and research topics leads to the creation of a two-dimensional framework (see Table 2.2). The vertical axis represents the three decision levels, while the horizontal axis illustrates the three main streams of the research topics in this area. We provide examples for each field in the framework.

	Patient flow optimization and control	**Bed capacity management**	**Personnel planning**
Long-term	Admission / discharge strategy Service mix	Regional coverage Capacity dimensioning (bed) Care unit portioning	Capacity dimensioning (staff)
Medium-term	Admission / discharge policy Discharge policy Bed reservation Triage and forecasting	Bed allocation Temporary bed capacity change	Cyclic shift scheduling
Short-term	Admission scheduling Elective admission rescheduling Discharge planning Acute admission / discharge handling	Patient to bed assignment Bed relocation in case of emergencies	Staff to shift assignment Staff rescheduling Nurse to patient assignment

Table 2.2: Exemplary research topics per research topic and time horizon.

Table 2.3 shows the number of papers in each category in the framework. Several papers discuss more than one aspect of the patient flow optimization and control problem, therefore they may be classified multiple times. One paper is classified as *admission* and *discharge & readmission*, two as *admission* and *treatment/LOS*, one as *treatment/LOS* and *discharge & readmission*, one as *admission*, *treatment/LOS, and discharge & readmission.* There are 19 papers in the *long-term*, 32 in the *medium-term*, and 2 in the *short-term* category. One paper [14] is classified as both *long-term* and *medium-term*. Most of the patient flow optimization and control papers fall into the medium-term planning category, while two thirds of the bed capacity management papers discuss long-term decisions, and the staff scheduling papers are distributed relatively equally. Only few papers exist on the short-term planning level. Detailed information on the papers in each category is available in the Appendix (Table A1). It is obvious that some areas have not yet been covered in the literature.

	Patient flow optimization and control			Bed capacity management	Personnel planning
	Admission	Treatment/LOS	Discharge & Readmission		
Long-term	-	-	-	17	2
Medium-term	15	10	8	4	1
Short-term	-	-	1	-	1

Table 2.3: Number of papers per research topic and time horizon.

Furthermore, we illustrate trends in research per topics in Figure 2.5. Apparently, patient flow optimization is drawing more and more attention, as up to 75% of the papers published since 2010 discuss this topic. In comparison, staff scheduling in ICUs has not been studied as extensively yet. The number of bed capacity management has stayed comparatively stable.

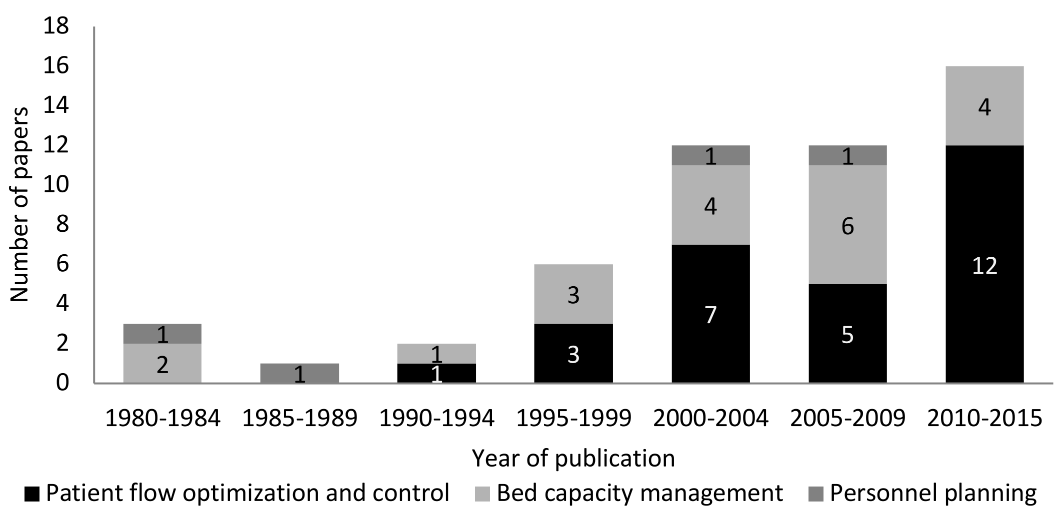


Figure 2.5: Number of papers per research topic and publication year.

Furthermore, there are additional methodological aspects that are attractive to researchers. First, one of the most important features of ICU planning is uncertainty. Therefore, a key issue of ICU management is how to model and predict these uncertainties. Second, the modeling approaches, differentiated between stochastic methods (e.g., queueing models, MDP) and deterministic methods (e.g., mathematical programming, statistical analysis) should be analyzed. Third, the applied solution method (e.g., exact solution methods, heuristics, and simulation based methods) is a further aspect of classification. As the ideas in the existing literature might inspire future researchers, modeling of uncertainties, modeling methods, and solution approaches are defined as three auxiliary classification methods and are summarized later.

This section is organized as follows. Section 2.3.1 reviews the articles focusing on ICU patient flow optimization and control. Section 2.3.2 is devoted to the topic of bed capacity management, while Section 2.3.3 discusses personnel planning. After that, the literature is reviewed based on the definition of uncertainty (Section 2.3.4), modeling methods, and solution approaches (Section 2.3.5).

2.3.1. ICU Patient Flow Optimization and Control

There are various ways to model patient flow in the ICU. Even for the definition of patient types within patient flow, there are discrepancies based on different optimization objectives and modeling methods. For example, most of the papers simply classify patients into two types: scheduled and unscheduled (or elective and emergency) [10, 17, 72]. However, Chan et al. [2] use patients' health conditions, and Kapadia et al. [51] use patient LOS as the classification criterion. To explain the ICU patient flow intuitively, we depict the patient flow based on the assumptions in most of the papers (Figure 2.6).

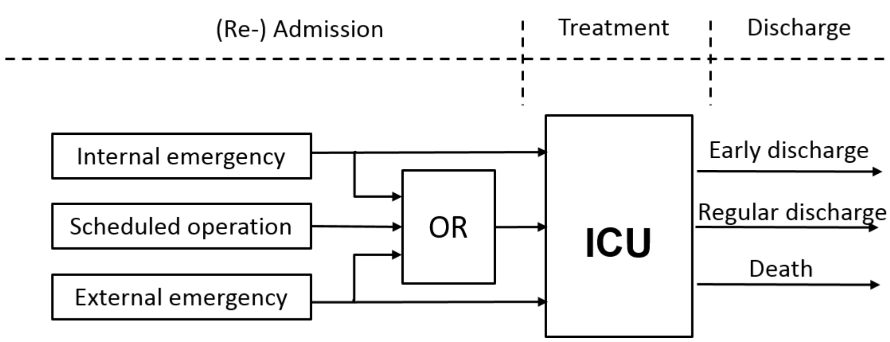


Figure 2.6: ICU patient flow (based on Litvak et al. [14])

Depending on the arrival pattern, patient types can be roughly grouped into three categories [14]. The first type consists of patients from *internal emergencies*, i.e., patients already admitted to the hospital in critical condition. For instance, patients who are already in a nursing ward but whose situation is deteriorating and who therefore need to receive special care in an ICU. This type includes the patients who are transferred in from the other care units. The second type consists of *external emergency* patients, mainly brought by ambulance. Both types of emergency patients may arrive with or without previous treatment in the OT department. There are also some patients transferred into the ICU from other hospitals because of medical reasons or capacity shortages. The third type of patients comes from a *scheduled surgery*. These patients undergo critical surgeries that require recovery in the ICU. Their time of arrival is relatively predictable. Patients with non-critical surgeries that have to be admitted to the ICU unexpectedly are considered internal emergency patients.

ICU patients usually require immediate care. As a consequence, it is not possible to make the new arrival patients queue for beds. When a patient arrives, the admission decision should be made according to the admission strategies and policies. Based on the three patient types, if there is no bed available, several different policies may be applied [14]. An internal emergency patient should be kept in the hospital mostly because it is not desirable to transport a critically ill patient. In some countries or states those transfers are even illegal if they are not beneficial to the patient [14]. Therefore, an “over-bed”, i.e., a temporary increase of bed capacity, could be created for these patients. For the scheduled surgery patients, the surgery may be canceled. The external emergency patients might be rejected (transferred to another hospital) or transferred to other ICUs.

The length of stay of ICU patients is influenced by the admission and discharge policies. After an admission, a patients receives medical treatment and stays in the ICU for a period that depends on the progress of the patients’ recovery and on the hospital’s treatment policies. Usually patients stay in the ICU until they reach a less critical health status and can be transferred to a general ward. This may be denoted as regular discharge. Another form of discharge is the death of a patient. Verburg et al. [73] compare eight different modeling methods for LOS, and find that models using survivors and non-survivors separately performed better than models without any differentiation of patients. However, Kim et al. [71] discuss the influencing factors of LOS distributions. They conclude that patients who died during their hospital stay show similar LOS patterns than those who recovered. Furthermore, most of the papers we reviewed in the field of OR/MS group both discharge types together, because both processes happen naturally and are not influenced by capacity limitations. If space for new arrivals is required, a so-called early discharge may be performed. Early discharge describes the situation that a patient , who is healthy enough to be moved, is transferred from the ICU to another unit (e.g., another ICU, a step-down unit or a regular ward) ahead of schedule [17]. In this case, capacity limitations influence the discharge decision. However, some researchers found that early discharges increase the readmission rate [15]. Making the discharge decision in light of possible readmissions (and related health issues) is another important topic for patient flow optimization.

Researchers have contributed many valuable ideas with regard to patient flow control. In the following sections, we discuss the papers for patient flow optimization in detail.

Long-term level. Topics for optimizing the patient flow at the long-term level concern both admission and discharge strategies. Admission control strategies determine, e.g., whether higher priority should be given to new arrival patients or to existing patients, or whether a reservation of capacities for emergency patients should be implemented. Discharge strategies, e.g., whether early discharges should be allowed, should also be considered. Additionally, the service-mix planning problem should also be discussed at this level. ICUs within a hospital could cooperate and pool capacities, if the medical requirements are fulfilled. This area has not been covered in the literature yet to the best of our knowledge.

Medium-term level. Admission control is an important approach to optimize the ICU patient flow. During “rush hours” in ICUs, patients who might benefit from intensive care are declined because the ICU has no spare capacity when all ICU beds are occupied or some beds are reserved for more critical patients. Naturally, rejections can possibly deteriorate patient health conditions, and therefore should be avoided. Kortbeek & Van Dijk [74] and van Dijk & Kortbeek [75] propose a queueing model to study the relationship between the rejection probability of patients and the capacity in the ICU. They prove and numerically illustrate the upper and lower bounds of the rejection probability using Erlang loss expressions. Li et al. [76] develop an MDP model to study admission policies. They separate patients into those who can be prematurely discharged and those who cannot. As a result, they find a threshold for the number of available beds reserved for more critically ill patients. If the number of available beds is lower than the threshold, the request of a patient will be rejected. Otherwise, the patient will be accepted. When patients arrive at a fully-booked ICU, early discharge of a current patient is an alternative to the rejection of an arriving patient. The fact that emergency patients are rarely rejected limits the impact of admission policies. However, patients might be hold in or transferred to other units, such as the ED or PACU, before being moved to the ICU. As elective patients are usually in less critical conditions than emergency patients, their surgeries are rescheduled, delayed, or canceled more frequently. Kim & Horowitz [63] propose daily elective surgery quotas in conjunction with a 1- or 2-week scheduling window to improve the scheduling process of elective surgeries and thus to improve the performance of both the ICU and the OT. Kolker [65] analyzes means to smooth the surgery schedule and ICU patient flows. He uses the work order leveling method as well as a simulation model to determine the maximum number of elective surgeries per day that should be scheduled. The objective is to reduce diversion to an acceptably low level in an ICU with fixed bed capacity. In addition, some researchers focus on improving the performance of the ICU without deteriorating the elective patients' position. Kim et al. [62] minimize the number of canceled surgeries by reserving beds for the exclusive use of elective surgery patients. Using a simulation model based on historical ICU data, a flexible bed allocation method is shown to be efficient in smoothing the ICU patient flow as well as keeping good service quality.

Reviews of specific cases have demonstrated that ICUs currently lack systematic criteria to help physicians make decisions based on quantitative information. Shmueli et al. [77] examine the impact of denied ICU admissions on mortality rates among patients who have been deferred. They use APACHE II[4] to score the acuity level of the patients and measure how ICU admissions decrease mortality rates. Based on their suggested ICU admission criteria, they compare three different admission policies with the goal of maximizing the expected incremental number of lives saved in the ICU. The first policy is simply first come first served (FCFS). The second policy sets a constant minimum benefit threshold – only when the excepted patient admission benefit exceeds the threshold the patient can be admitted into the ICU. In the third policy, the minimum benefit threshold depends on the level of occupied beds. After evaluating the three policies with real data from a medical center, the authors find that the third one performs best. Nevertheless, it only achieves a marginal improvement over the second policy. Kim et al. [71] also look at this problem. They analyze the patient outcomes for all patients admitted by the emergency department. The impact of ICU admissions on patient outcomes is analyzed by estimating the cost of denying ICU care. Both observed and unobserved factors that might influence the outcomes are considered.

When the ICU is fully or almost fully occupied, speeding up the treatment process is another approach to manage capacity efficiently, as shown by KC & Terwiesch [15]. They conclude that patient transport times can be accelerated in times of high hospital congestion. They study the effect of ICU occupancy levels on patient LOS and find that congested ICUs tend to speed up the treatment of their patients and implement an early discharge policy, which may increase readmission rates. Additionally, the LOS of readmitted patients can be significantly longer. As readmission rates are typically higher for patients with higher acuity, the authors suggest early discharge policies for patients with lower acuity if capacity needs to be freed. Dobson et al. [17] and Chan et al. [2] both model the discharge

[4] APACHE II (Acute Physiology and Chronic Health Evaluation II): a severity-of-disease classification system [192], one of several ICU scoring systems. It is applied within 24 hours of admission of a patient to an ICU: an integer score from 0 to 71 is computed based on several measurements; higher scores correspond to more severe disease and a higher risk of death.

process using a MDP and propose early discharge policies. The differences are that Dobson et al. [17] do not consider readmission and transferring patients based on their predicted remaining LOS in the ICU, while Chan et al. [2] make early discharge decisions based on costs, including both the patient related health costs (e.g., readmission and mortality) and the system related costs (e.g., resources and effect on other patients).

Optimizing and analyzing the treatment process and patient LOS in ICUs helps to improve ICU patient flows. Akkerman & Knip [47] combine Markov chain theory and simulation to analyze the patient LOS. They define the state space as the number of days a patient stays in the hospital and obtain the transition probabilities by creating an empirical distribution based on available data. They calculate the theoretical number of beds required on each day to study the relationship of the ICU and ward capacities, and they find a more efficient distribution of beds between hospital wards and ICUs to reduce surgery waiting time. They also suggest that idle beds in wards might be used as ICU beds. Griffiths et al. [78] model the ICU patient flow as a queuing network and analyze the arrival and LOS distributions based on real data. They find that a reduction of the time spent in the ICU could greatly impact capacity utilization. From a treatment perspective, Kapadia et al. [51] show that the Markovian approach can be used to investigate the relation between changes in the severity of illnesses. Adeyemi et al. [55] study LOS as well as the relationship of the LOS, gestation age and birth weight in neonatal care units by retrospective analysis. Similarly, Demir et al. [56] analyze the LOS and the patient pathways in a neonatal care unit applying three different methodologies to reduce inefficiencies, improve the patient experience, and reduce cost. Nathanson et al. [53] find that data envelopment analysis (DEA) can be used to predict patients performances early in their stay.

The papers discussed in the previous paragraphs focus on one specific part of the ICU patient flow (i.e., *admission, treatment/LOS, discharge and readmission*). However, there are also publications that model and optimize patient flow as a whole and incorporate interactions between different stages. In Kim et al. [64], the authors consider the admission and discharge processes jointly. The admission process is basically a FCFS policy. When a patient arrives at a congested ICU, the possibility of early discharge is checked first, and then the decision about canceling a surgery is considered. In Lowery & Arbor [70], the model is improved to incorporate early discharge behavior. Hagen et al. [68] continue in this direction. They build a simulation model that aims to capture all possible cases of patient flow after admission. The authors evaluate three different policies: smoothing the operating schedule for elective patients, prioritizing admissions by excepted LOS or severity of patients, and early discharge. They conclude that prioritizing admissions could considerably reduce delays for critical cases while increasing the average waiting time for all patients. In addition, they find that early discharges can raise readmission and mortality rates.

Short-term level. After setting up strategies and policies for the admission and discharge process, the implementation of these policies is considered at the short-term level. Admission and discharge scheduling problems have not been widely studied, even though they are important for decision making in ICUs. Wharton [50] classifies patients into different risk groups and develops a queueing model to predict the proportion of patients from each risk category. They find that this proportion is a function of the size of the ICU, the number of risk categories, the mean arrival rates, and the LOS.

Concluding Section 2.3.1, we find that there have been significantly more researches at the medium-term level than at the other two levels. Most papers discuss *admission, treatment/LOS,* and *discharge policies*. By comparison, there are no papers on long-term and only one on short-term decisions. The reason for this discrepancy might be that there are different situations in different ICUs or hospitals at different times, making it difficult to generate a common strategy for operational criteria. For instance, to define long-term admission strategies, such as "do never create over-beds" or "always reserve 70% of all beds for emergency patients", would not hold for all hospitals in general. Even if it is possible to develop such a strategy, it is difficult to validate it. Therefore, it seems to be more attractive to researchers to develop optimized patient flow models based on existing capacity limitations than to create a long-term patient flow strategy. Challenges for future research are how to overcome this difficulty and generate robust long-term strategies for ICU planning and how to efficiently implement the strategies and policies on the short-term operational level.

2.3.2. Bed Capacity Management

In the previous section, we review papers aiming to optimize patient flow through efficient use of ICU capacity, often combined with the constraint to ensure service quality. These papers usually assume the size of the ICU to be fixed except for the possibility of temporarily setting up a few extra beds. In this section, we summarize modeling approaches that aim to optimize bed capacity utilization by using capacity dimensioning (long-term level), regional coverage (long-term level), bed allocation (medium-term level), and bed relocation problems (short-term level).

Long-term level. The main topic in long-term bed capacity planning is the determination of the required ICU bed capacity. The central objectives in ICU capacity optimization are high utilization rates and high service quality. Obviously, these two goals are of conflicting nature. In earlier research, the average bed occupancy level was commonly used as a criterion for capacity planning. Green [13] criticizes this method, examines real-life data, and applies queueing theory to estimate bed availability in ICUs. She finds that nearly 90% of ICUs have insufficient capacity, and suggests that hospitals need to plan capacity based on service quality, i.e., the ability to place patients in appropriate beds in a timely fashion, rather than on target occupancy levels. Terwiesch et al. [79] point out that the average waiting time increases dramatically at higher utilization levels. Additionally, there are nonlinear relationships between the number of beds, mean occupancy level, and the number of transferred patients [72]. Therefore, the tradeoff between utilization and service quality levels should be considered.

There are several papers that analyze capacity dimensioning problems based on case studies. In earlier research, the average expected number of patients, the average LOS, and the target occupancy levels are commonly used to calculate the number of ICU beds needed. Lamiell [80] studies the relationships between bed capacities, average patient admission rates, average patient LOSs, utilization rates, and overfill rates in ICUs. Costa et al. [81] point out that the methods using average expected numbers would cause congestions at peak time. They set up a model at the individual patient level, collect detailed patient information, and apply data analysis and mathematical modeling to estimate the number of required ICU beds. The performance is evaluated using three examples in different clinical situations and is found to provide efficiency improvements. Barado et al. [59] develop a simulation model based on statistical analysis of the daily bed occupancy in an ICU, which can be used as a reliable sizing and capacity analysis tool. Masterson et al. [67] study the ICU sizing problem and analyze the performance of bed mix policies in a military ICU. McManus et al. [66] calculate the accurate bed requirements in a busy ICU and implement a sensitivity analysis of the model to changes in unit size. Troy & Rosenberg [57], Zilm & Hollis [58], and Yergens et al. [82] all use simulation to determine the requirement for ICU beds for special cases. The first two studies focus on surgery patients, while the third paper discusses the requirements for a tertiary care hospital (e.g., specialized cancer centers). Aiming to accurately determine the number of required ICU beds, Lowery [69] states that it is more cost effective to use the capacity in similar units as overflow beds to meet demand during peak times.

Another important topic at the long-term level is the bed mix policy in different care units. Cahill & Render [60] investigate the ICU bed utilization rate of a medical center and discover consistently high levels. They suggest that additional telemetry beds could be created, the respiratory care unit could share the work load with the ICU, and ICU swing beds could be assigned in the emergency room area. Cochran & Bharti [29, 30] and Cochran & Roche [28] treat all the inpatient beds (ward, PACU, IMC etc.) as a system and try to create better balanced bed utilization across the system while allowing differing local utilization targets. Besides the cooperation on the department level, the cooperation of different hospitals to provide regional coverage is another option [14] .

Medium-term level. Based on the long-term decisions, a variety of medium-term capacity management problems can be considered. Examples include bed allocation optimization with respect to patient types. After proposing the idea of (long-term) regional coverage, Litvak et al. [14] propose a medium-term cooperative solution for the ICU capacity where several hospitals in a region jointly reserve a small number of beds for regional emergency patients. They present a mathematical model for calculating the number of reserved regional beds for any given acceptance rate and prove that the regional cooperation could help to achieve a high acceptance level with a smaller number of beds. This results in an improved service for all patients. They use a special analytical approach similar to the overflow models in telecommunication systems. The number of patients transferred to hospitals outside the region is minimized, while a sufficient amount of ICU beds for planned surgeries is maintained. Similarly, Asaduzzaman et al. [54] develop a model to determine the number of beds required in a neonatal unit to achieve desired levels of service, which is measured by the probability of admission refusal and overflow to temporary care. The model is also based on the overflow mechanism in telecommunication networks.

Shahani et al. [83] simulate patient flow alterations in several different scenarios with detailed data analysis. They find that increasing ICU capacity can significantly reduce deferral and transfer rates.

From an economic perspective, increasing the occupancy level is an important goal. However, there is a trade-off between occupancy levels and service quality, which is discussed in several medical papers. We refer to Tierney & Conroy [21], who provide a literature review discussing 16 papers from a medical perspective. They explore how the occupancy level is measured, reported, and interpreted, and they investigate optimal ICU occupancy levels.

Short-term level. Main topics of bed capacity management at the short-term level include patient to bed assignments and relocation problems. In most existing papers, all beds in the ICU are assumed to be equal. In fact, we found no relevant research discussing assignment decisions involving different bed types in intensive care. However, when patient isolation in case of infective diseases is considered, bed type assignment will be an interesting topic to analyze.

Contrary to the patient flow topic, bed capacity planning topics are mainly discussed at the long-term level. Especially capacity dimensioning and regional coverage are discussed. However, short-term level topics such as patient to bed assignments and bed relocation are not regarded. Managing the

ICU bed capacity dynamically is a potential topic for future research. Furthermore, isolation or semi-isolation beds are necessary in order to reduce infection rates. Therefore, future research should address bed capacity management that considers the isolation concept.

2.3.3. Personnel Planning

Besides the expensive specialized bed capacity, personnel planning requirements in ICUs are high in both quantity and quality. The ICU is one of the most extensively staffed departments in hospitals. According to the legislation in some states in the U.S., the minimum nurse to patient ratio in the ICU can be as high as one to two compared to one to eight in regular wards [2]. Moreover, due to the critical condition of ICU patients, the staff has to be highly skilled. Staff costs contribute to over 50% of the total ICU expenditures [10]. Therefore, the staff in the ICU needs to be efficiently managed. There is a large number of papers on nurse scheduling problems and physician scheduling problems in hospitals [84, 85]. Due to the high level of uncertainty, staff scheduling in ICUs is particularly complex. However, the staff (including both nurses and physicians) scheduling problem in ICUs has not attracted much attention.

Long-term level. Similar to the capacity management section, staff dimensioning is discussed at the long-term level. The determination of the number of rostered nurses and physicians falls into this category. Because ICU staff is typically highly qualified and thus expensive (e.g., ICU nurses are more expensive than regular nurses due to additional training), overstaffing should be avoided. In comparison, understaffing might lead to critical situations due to the acuity of patient's health. Therefore, staff dimensioning problems are important to tackle this trade-off.

In an early article, Hashimoto et al. [61] provide a case report for a personnel planning simulation in a 12-bed medical/cardiac ICU. They evaluate the ICU workload and personnel planning system under different staffing levels. The financial concerns, quality of care issues, and staff working preferences are part of the evaluation. In more recent studies, staff scheduling in ICUs is not considered in isolation, but is discussed in combination with admission control or capacity management problems. In most of the aforementioned papers, the number of staffed beds is considered to be the bottleneck resource for patient admissions, but Griffiths et al. [10] provide a new perspective. They propose that the beds and staff could be treated separately, because compared to staff, the bed itself is not a critical and expensive resource. Therefore, they assume that the number of beds can be increased, but the number of nurses is limited. Their model assumes that all patients for intensive care therapy are admitted. They present a model to determine the number of rostered nurses as well as the required number of supplementary nurses per shift to minimize the nurse scheduling cost. In their paper, the uncertainty of individual patient arrival patterns and the LOS are both included in the model. In general, the combined optimization of nurse and physician staffing is missing from the literature.

Medium-term level. Personnel planning on the medium-term level discusses shift scheduling problems, i.e., the assignment of nurses and physicians to work patterns. Shift scheduling in the ICU is special compared to other departments. The workload is most intense at night, therefore the conventional roster patterns cannot offer ideal matching between staffing and workload [86]. However, shift scheduling problems in the ICU are only discussed in Duraiswamy et al. [87]. The authors simulate a 20 bed ICU and predict the performance of staffing levels that result from different schedules. Based on the distribution of patient arrivals, the required nursing care workload is determined for each patient, and then the total number of nursing hours in a shift is calculated.

Short-term level. At the short-term level, the staff to shift and nurse to patient assignment problem is discussed. The topic of assigning patients to nurses is proposed by Mullinax & Lawley [52]. They classify patients according to their health condition. At the beginning of each shift, the head nurse assigns the nurses to groups of patients. The authors develop a detailed neonatal acuity system that quantifies the nursing workload and use integer linear program to assign patients to nurses while balancing nurse workload. The model is solved with a heuristic approach utilizing the fact that most nurseries are divided into a number of physical zones. The performance of the heuristic is tested in ten case studies.

ICU personnel planning papers cover all three decision levels including dimensioning, shift scheduling, and nurse to patient assignment. Compared to nurse rostering problems and physician scheduling problems in general, ICU personnel planning problems show a higher level of complexity due to uncertainties and more complicated cost structures. In contrast with the other topics discussed in this review, the personnel planning problem has not been well studied yet, especially during the last ten years. In particular, online planning of physicians and nurses is an interesting area for future research due to the significance of patient demand uncertainty. Additionally, quick-response adjustments of staff assignment is a topic with high potential for future research. We conclude that promising future research topics exist on the long-term, medium-term, and the short-term level.

2.3.4. Modeling of Uncertainties

There are mainly two parts of uncertainties in the ICU modeling: the unpredictable arrival pattern and the LOS of the patients. Early studies employ expected values to represent the distributions of arrival patterns and the LOS. In contrast, more recent studies classify patients into different groups and estimate the distribution of their arrival patterns and LOS. In this section, we review the grouping of patients and the modeling of uncertainties within the literature. Additionally, we summarize the detailed methods from the papers in Table A2 in the Appendix.

Patient grouping. Typically, before discussing the distributions of arrival patterns and LOSs, patients are grouped based on a variety of different methods. The commonly used method is to group patients according to the following arrival properties. The patient can either be an emergency patient or a patient with a scheduled surgery (see Figure 2.4). The arrival time of an emergency patient is not perfectly predictable, and emergency patients should usually not be rejected at the ICU. In contrast, the arrival time of scheduled patients is often times deterministic. This type of patient is typically not critically ill, and the cancelation or delay of the surgery can be optional. Thus, patients may be divided into unscheduled and scheduled patients, or in other words, emergency and elective patients [10, 17, 66, 68, 72, 81]. Litvak et al. [14] propose a regional coverage strategy especially for the emergency patients. Therefore, they divide the emergency patients into two sub-groups, which are internal and regional emergency patients.

More specifically than the general arrival pattern, some papers group patients according to the department where they are transferred from. Cahill & Render [60] and Akkerman & Knip [47] group patients into inpatients and outpatients. Further, in Kim's papers [63, 64], the inpatients are differentiated into patients from the wards and the OT (2 sub-groups: emergency and elective). Kortbeek & Van Dijk [74] and van Dijk & Kortbeek [75] group patients according to whether they are transferred from the OT or not, which allows them to investigate the effect of the limited ICU capacity and its interaction with the OT. Kolker [65] considers patients that are transferred from other hospitals. Barado et al. [59] implement a detailed classification method that uses 8 groups: elective surgery, emergency department, surgical wards, medical wards, emergency surgery, trauma without surgery, trauma with surgery, and other. Similar to Barado et al. [59], Asaduzzaman et al. [54], Adeyemi et al. [55] and Cochran & Bharti [30] classify patients according to the departments they are admitted to.

The patient LOS plays an important role when measuring the utilization of bed capacity. Therefore, patient grouping methods can be based on different LOSs. Dobson et al. [17] optimize the discharge process using an MDP, and the states of the MDP are defined as the remaining LOSs of each patient in the ICU. Thus, the patients are classified into different groups based on the predicted LOS. A similar approach can be found in Griffiths et al. [78].

The modeling methods are diverse due to different research objectives. Some papers focus on efficient resource usage, while others focus on service quality and patient conditions. In the latter, patients are grouped based on their health condition [2, 50, 76, 88].

Arrival pattern and LOS. ICU management is a stochastic problem. Therefore, deterministic modeling methods that represent the patients' arrival pattern and LOS with deterministic values (e.g., expected values) do not fit well. In the literature, most of the papers use theoretical distributions based on historical data to model the arrival pattern and LOS. The distributions used for modeling arrival patterns and LOS are summarized in Table 2.4: Number of papers per modeling method of uncertainties. The vertical axis illustrates the distribution of arrival patterns and the horizontal axis illustrates the possible distribution of LOS. In total, there are 37 papers modeling uncertainties with some assumed distributions, and the remaining 15 papers do not discuss arrival or LOS distributions. The Poisson distribution is most commonly used to describe the patients' arrival pattern. 60% of the publications prefer to use a stationary Poisson distribution, while 27% of the papers choose other distributions such as empirical discrete probability distributions, weekly periodic arrival patterns, and so on. In comparison, the distributions of LOS are more diverse. About 57% of the papers use geometric and exponential distributions, but the lognormal distribution, Weibull distribution, and other distributions are also used.

		Patient LOS distribution								
		Geometric	Exponential	Weibull	Log-normal	Triangle	Gamma	Phase type	Other	Total
Arrival pattern	Stationary Poisson	1	15	-	3	1	-	1	1	22
	Non-stationary Poisson	1	1	-	1	-	1		1	5
	Other	-	3	2	2	-	-	1	2	10
	Total	2	19	2	6	1	1	2	4	37

Table 2.4: Number of papers per modeling method of uncertainties

Although actual arrival patterns to the ICUs never resemble a standard distribution [65], the Poisson distribution is widely used in research because of its mathematical convenience [13, 14, 54, 60, 65, 66, 68, 70, 76–78]. As patient arrivals vary not only from day to day but also within a day, the non-stationary Poisson distribution is applied in several papers [10, 17, 63, 67, 69] to improve modeling accuracy.

Although some researchers argue that there are no suitable distributions to model the LOS in an ICU [73] and the LOS is not reliably predictable for individual patients [89], most of the papers we review apply theoretical distributions. There are multiple choices for LOS distributions. In the papers using MDP and queueing, the LOS is usually modeled as an exponential or geometric distribution because of their memoryless property [2, 13, 14, 17, 29, 54, 66, 71, 76, 77]. The exponential distribution is also applied to model the LOS in regression modeling papers [55, 90, 91]. However, KC & Terwiesch [15] argue that the memoryless property is not realistic, because resources might exhibit increased service rates when system load is high. They suggest the usage of the Weibull distribution as it is commonly used in the biostatistics literature to model durations for patient recovery. According to empirical data analysis, the lognormal distribution could fit well in some cases [2, 67, 70, 83]. Cahill & Render [60] believe that Triangle distributions are another appropriate choice. Demir et al. [56] use a phase type distribution to model the LOS in neonatal care.

The aforementioned papers all fit common distributions for patient arrival patterns and the LOS of all patients. However, different types of patients show different characteristics. For example, scheduled surgeries always take place on weekdays, and the LOS of elective patients are relatively shorter than that of emergency patients. Consequently, more realistic results can be achieved by distinguishing between different patient groups.

Therefore, many studies model the arrival pattern and LOS for different patient groups [59, 63, 64, 69, 74, 75, 81]. The emergency patients are usually described by Poisson arrival patterns and exponentially or lognormally distributed LOS. The modeling of elective patients differs in the literature. Dobson et al. [17] and Griffiths et al. [10] demonstrate that the weekly periodic arrival patterns work well. Barado et al. [59] apply discrete empirical probability distributions to fit the properties of elective patients. In Lowery [69], scheduled arrivals are generated on the scheduled surgery day. Kim et al. [64] study a 14-bed ICU in a six month period and conclude that the arrival rate from the scheduled surgeries is not a Poisson process and the LOS for them is not exponentially distributed. Both conclusions cast doubt on standard queueing model assumptions.

Except for few literature reviews and data analysis papers, all reviewed publications model uncertainties based on specific objectives. Therefore, the patient grouping logics vary. Based on different grouping methods, different distribution types are used to model uncertainties. We further analyze whether different ICU types or different geographical regions lead to different models in the literature. However, no significant effect can be determined. Therefore, we conclude that the selection of grouping methods are based on the research objectives and not on ICU type or geographical region. A comparison of data of different ICU types and geographical regions could lead to interesting findings. Another goal for future research should be to more realistically model the uncertainties using nonhomogeneous distributions.

2.3.5. Modeling Methods and Solution Approaches

In this section, we summarize the modeling methods as well as the solutions approaches within a two dimensional framework. In Table 2.5: Number of papers per modeling method & solution approach, the horizontal axis illustrates the modeling methods while the vertical axis indicates the solution approaches. The two review articles are excluded in this table. As we cover a broad variety of topics, the modeling methods are diverse and include both deterministic and stochastic approaches. The corresponding solution approaches include mathematical exact solutions, heuristics, and simulation based solution approaches. We find that 80% of the papers employ stochastic methods to model the ICU management problem. Meanwhile, 56% of the papers use simulation to solve the problem. Demir et al. [56] apply three different modeling methods (stochastic modeling, statistical modeling, and system dynamics simulation) to study the relationship between LOS and the performance in a neonatal care unit. More detailed information can be found in Table A3 in the Appendix. As depicted in Figure 2.7, we detect some trends with regards to the application of different solution approaches. It is clear that simulation based solution has been the most widely used approach during the last 35 years, but a growing number of researchers have used exact solution approaches since 2000.

		Modeling methods					
		Stochastic methods			**Deterministic methods**		**Total**
		Queueing	MDP (Markov chain)	Process analysis	Mathematical programming	Statistical analysis	
Solution approaches	Exact solution	8	2	-	2	7	19
	Heuristics	-	3	-	1	-	**4**
	Simulation based	12	1	15	1	-	**29**
	Total	20	6	15	4	**7**	**52**

Table 2.5: Number of papers per modeling method & solution approach

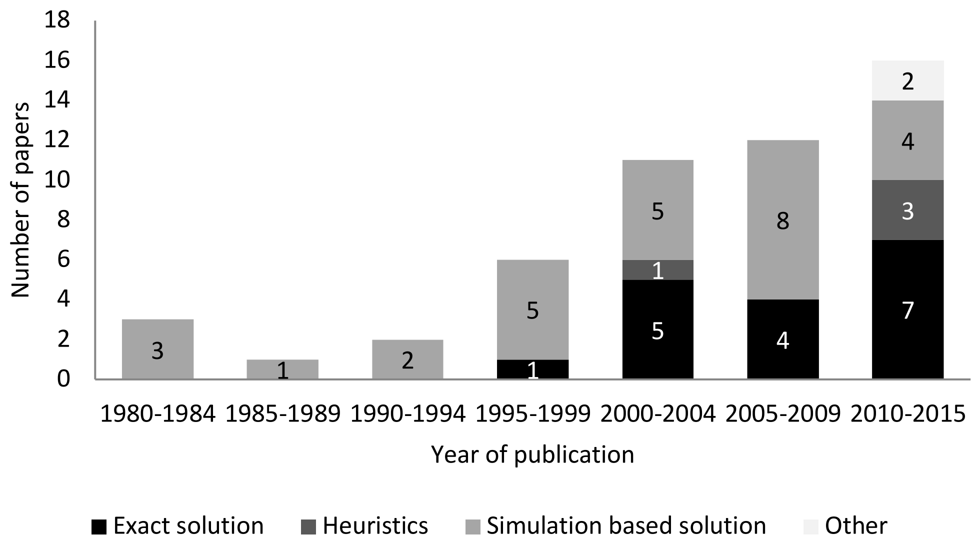


Figure 2.7: Number of papers per solution approach and publication year.

Stochastic methods. Because of the stochastic properties of ICUs, as mentioned, 80% of the papers apply stochastic methods in ICU management problems. These models can be generally classified as queueing, MDP/Markov chain, and stochastic process analysis methods. Figure 2.8 illustrates trends of stochastic modeling methods. From the 1980s until the middle of the 1990s, stochastic process analysis papers were common. However, during the recent 5 years, there have been no stochastic process analysis publications at all. Meanwhile, MDP modeling has started to receive attention. From 1995, queueing methods play a quite important role in modeling the ICU problems.

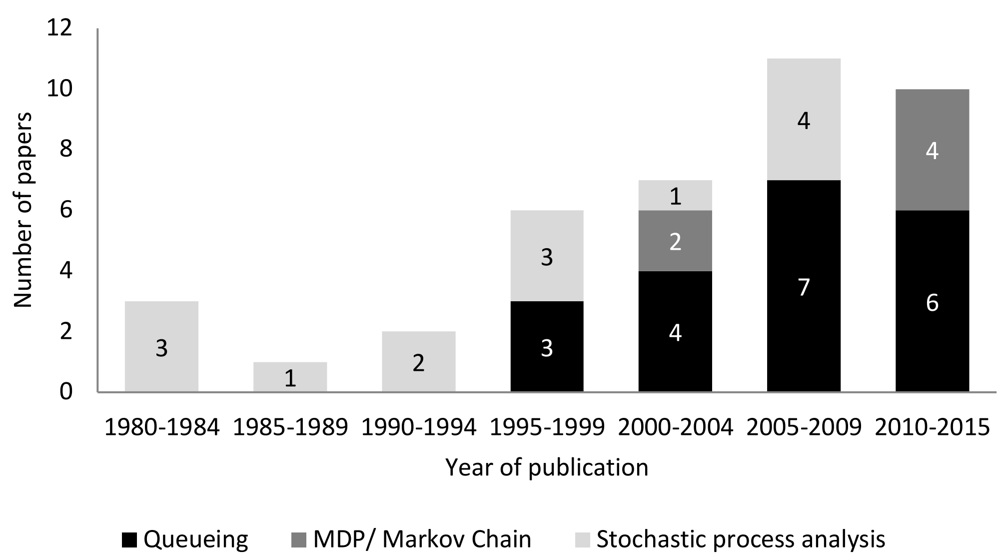


Figure 2.8: Number of papers per stochastic modeling method and publication year.

Queuing theory uses closed mathematical formulas to describe a number of predetermined simplified models of real processes [65]. Queueing models[5] can be applied to capacity planning problems, e.g., McManus et al. [66] present an M/M/c/c queueing model, and Ridge et al. [72] apply an M/M/S/∞/∞/NPRP queuing model (NPRP stands for non-preemptive) to determine ICU bed requirements. Cochran & Bharti [29, 30] and Cochran & Roche [28] use queueing networks to balance bed unit utilization. Other papers in capacity planning are Asaduzzamanet al. [54], Barado et al. [59], Costa et al. [81], Litvak et al. [14], and Ridge et al. [72]. Queuing models are also used to optimize the patient flow. Griffiths et al. [78] model the system as M/H/c/∞/FIFO, and a M/M/s multi server system is applied in Kim et al. [64]. Both papers study the patient admission control problem. Meanwhile, Chan & Yom-Tov [88] propose an Erlang-R queueing model for patient discharge decisions. Green [13] adopts an M/M/s queueing model with time homogeneous Poisson arrivals and exponential LOSs to estimate delays in the ICU. Other applications in patient flow optimization may be found in Kim & Horowitz [63], Hagen et al. [68], Wharton [50], and Yang et al. [49]. Queueing models are not easy to solve, especially for large and complex systems. Eight articles included in this review use exact analytical solution approaches [13, 50, 54, 74, 75, 77, 78, 88], while the others analyze the process by simulation approaches.

MDP and Markov chains have rarely been used in ICU modeling. There are three MDP and three Markov chain models included in our review. Kapadia et al. [51] focus on modeling patient flow through different stages in the recovery process. A stationary distribution is applied to predict the patient LOS. A similar method is used in Akkerman & Knip [47] and Demir et al. [56]. Although both Chan et al. [2] and Dobson et al. [17] demonstrate that discrete time MDP could work quite well in making discharge decisions, they define the states of the MDP differently. Chan et al. [2] set the state to be the number of different types of patients in the ICU to reflect the aggregated situation in the ICU. In contrast, Dobson et al. [17] prefer to focus on the individual patient's health condition and define the states to be the remaining LOS of each patient in the ICU. Li et al. [76] apply MDP to study the admission decisions. Their state setting is similar as in Chan et al. [2], as they incorporate the number of different types of patients in the ICU and the number of available beds. Markov chains may be solved with exact analytical approaches. MDPs are usually quite complex to solve, and either approximate dynamic programming or other heuristics can be applied. Chan et al. [2] employ approximation algorithms to solve it. Dobson et al. [17] develop a new aggregation-disaggregation

[5] Kendall's notation is used to characterize queueing systems [193]. It is given by A/S/c/K/N/D, where A indicates the distribution of inter-arrival time, S indicates the distribution of service time, C indicates the number of servers, K indicates the system capacity, N indicates the size of the customer population, and D indicates the queue's discipline.

algorithm and demonstrate the superior computational efficiency of their approach over the Gauss-Seidel iterative method. Li et al. [76] establish lower and upper bounds to evaluate the admission policy and adjust the parameter settings to improve the policy. Simulation based solution approaches can also be used for MDP, but there is no paper with such an application.

Queueing models depend on the assumption of Poisson arrivals, and MDPs always assume a memoryless LOS, which can be considered a limitation of these two methods. Simulation helps to support analytical results to overcome this shortcoming [69]. Simulation is more flexible and offers much more freedom regarding the assumptions of the particular types of arrival processes and LOSs. In the ICU management literature, simulation is also widely used in most topics, and it is also used as a solution approach for queueing models [10, 60, 65].

Deterministic methods. Because the stochastic features of patient arrival patterns as well as LOSs cannot be ignored when modeling the process, deterministic methods, such as mathematical programming, statistical analysis, are not suitable for many problems in ICU planning. However, deterministic methods fit well for some topics, such as personnel scheduling and performance analysis. Figure 2.9 shows the trends of deterministic modeling methods.

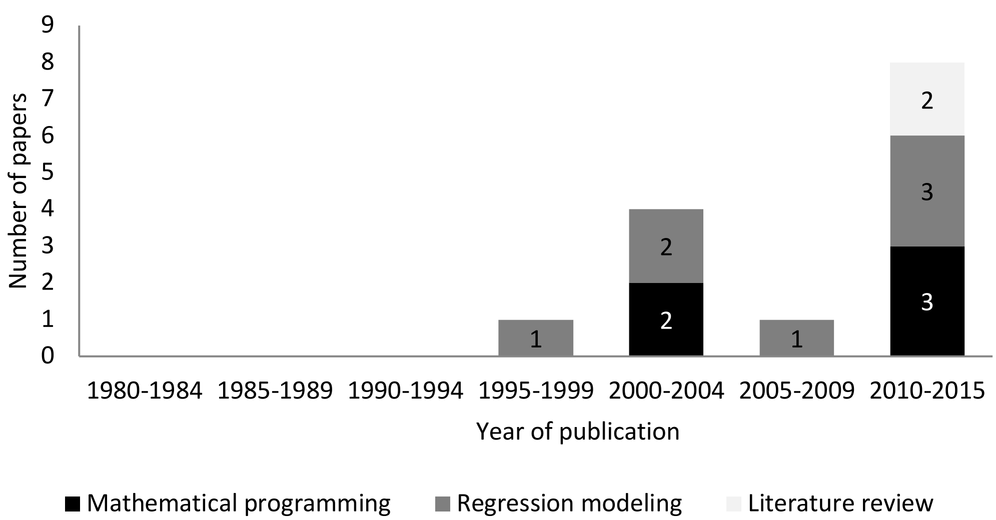


Figure 2.9: Number of papers per deterministic modeling method and publication year.

In general, mathematical programming methods, including linear programming, integer programming, mixed-integer programming, and nonlinear programming, are typically applied in nurse and physician scheduling problems. Mullinax & Lawley [52] develop a mixed-integer programming approach for nurse scheduling to achieve better workload balance in a neonatal ICU and solve the model using a heuristic approach. Another application of mathematical programming is performance analysis.

Nathanson et al. [53] propose that DEA can be applied to evaluate the health condition of ICU patients. They evaluate each patient individually by calculating an efficiency score based on the patient's ability to maximize the output for a given set of physiologic inputs. Patients with high efficiency scores are found to have a better chance of making a full recovery than similarly injured patients that are considered inefficient. The model is solved by an exact numerical approach as is common with DEA models. Kim et al. [71] analyze and compare the performance of admission strategies. They provide objective metrics that can be used by physicians in the ED and ICU to decide which patients to admit to the ICU from the ED. Empirical data analysis, dynamic programming, and simulation methods are applied.

Statistical analysis also plays a role in many topics. Adeyemi et al. [55] retrospectively analyze the LOS in a neonatal care unit. The results show that the LOS for different levels of care (ICU, High Dependency Care, and Special Care) are positively correlated, and that birth weight and gestation age are negatively correlated with the LOS. KC & Terwiesch [15] perform econometric statistical analysis of discharge and readmission processes. They find that the early discharge of patients with lower acuity frees up capacity, but readmission rates of patients with higher acuity increase. Kim et al. [62, 64] study the admission process by analyzing hospital data to detect characteristics of patient arrival patterns and LOS. Shmueli et al. [90] analyze hospital data using statistical methods to demonstrate the effect of ICU treatments on patient survival rates. They compare the effect of admitted and rejected patients and discuss the implications for an optimal admission policy. Roumani et al. [92] compare five different data mining and statistical techniques to construct a predictive model for patient conditions at discharge.

Deterministic models are usually simpler, less data intensive, and easier to solve. Stochastic models tend to be more realistic, but they require more effort on finding the solution for a complex system. Because of the specific features of the ICU management, stochastic models with efficient solution approaches should be developed in future research. However, deterministic models might still be relevant in staffing problems for ICUs with relatively stable demand patterns.

2.4. Future Research Agenda

Although a significant body of literature in ICU management exists, there are still some open issues. Following the structure of this review, we summarize the potential research topics in three streams: the interaction with the other departments in the hospital patient flow, the topics of ICU management (single department problems), and the modeling and solution approaches.

In general, as discussed in Section 2.2, ICU planning plays a significant role in the hospital patient flow. Thus, the problem of coordinating decisions between the ICU and connected departments is an interesting topic for future research. First, efficient and effective ICU admission control decisions should be made based on information from other departments, e.g., ED, OR, and wards. Second, other departments are able to make better decisions when considering ICU capacity levels, e.g., when scheduling elective surgeries.

In ICU management problems as presented in Section 2.3.1, 2.3.2, and 2.3.3 (single department problems), many topics contain interesting opportunities for future research. We define three major categories, which are important but less investigated by now. First, the beds in the ICU are usually assumed to be unified. However, differentiation between different levels of care might help to optimize the capacity. Intermediate intensive beds suited for less critically ill patients could be integrated in the ICU. These beds require fewer resources than ICU beds and can be used for relatively stable patients. In this case, the optimal share of intermediate beds in the ICU should be studied. Another option in differentiation of ICU beds is to reserve isolation beds for infective patients. In those cases, dynamic adjustments of the bed mix is an interesting topic. Second, bed capacity planning and personnel scheduling should be combined. Nurses and beds are two connected aspects of the ICU capacity. When bed capacity changes, the number of nurses should be adjusted accordingly on a strategic level. However, nurses and physicians may be scheduled independently on the short-term level. Third, there may be several different ICUs in different specialties in a hospital. Based on medical and organizational constraints, sizing and pooling of ICU capacities in multi-disciplinary ICUs could be a topic for future research. Besides the three categories mentioned above, topics on the long and the short-term level in patient flow optimization and control (such as admission and discharge strategy and planning, service mix planning, etc.), and the medium and the short-term level in bed capacity planning (such as bed allocation and relocation, and patient to bed assignment), as well as ICU staff scheduling problems are rarely discussed. Examples include the setup of patient admission rules for single hospitals and for groups of hospitals, methods to optimize the holistic ICU patient flow from admission to discharge, the definition of admission policies based on both the diagnosis of patient health status and the ICU capacity utilization, and the development of online scheduling regimes of nurses and quick-response adjustments to unforeseen events.

From a modeling and solution approach perspective, progress has been made during the last years. OR/MS methods have matured in the field of production and service operations. Some of those methods could be introduced and adapted to ICU management problems. For example, Asaduzzaman et al. [54] and Litvak et al. [14] adapt overflow models of telecommunication systems with multiple streams of telephone calls to the regional ICU planning problem. Although common production and service systems are typically more homogenous than the ICU, the adaptation of methods already applied in other fields for ICU problems is an attractive direction of future research. In the early stage of the ICU research, simple distributions, even average values, have been applied to describe patient arrival patterns and LOSs. Articles from recent years model the uncertainties more realistically, e.g., by grouping patients and using different distributions for each group, and applying non-homogenous distributions. However, there is still room for improvement in the future. For instance, patient's health statuses and minimum service quality in the ICU could be considered as objectives. Regarding the modeling methods, simulation is widely used for describing and modeling complex systems. The combination of optimization and simulation may add value for decision support tools. Moreover, stochastic modeling methods, such as MDP, are attracting attention. However, modeling the system with MDPs has limitations concerning the distribution of the arrival process. To overcome this weakness is a task for future research. The existing studies with MDPs optimize either admission or discharge processes in isolation. Optimizing the holistic patient flow in an MDP is an interesting topic for the future. Finally, efficient solution approaches for complex stochastic models should be developed and tested using real data.

2.5. Conclusion

ICU management mainly focuses on using capacity efficiently while ensuring an adequate quality of care for critical patients. The ICU management problem is important not only because of the critical medical effects on ICU patients, expensive capacities, and unpredictable demands, but also because the ICU has great influence on hospital-wide patient flow. OR/MS plays a significant role in balancing the demands and supplies under resource limitations.

In this paper, we provide the first structured and comprehensive review of ICU problems in OR/MS. The importance of the ICU for hospital patient flow is introduced in Section 2.2 based on 18 papers, which discuss the interactions between the ICU and upstream and downstream departments. In Section 2.3, the relevant 52 papers for the ICU management problem are discussed based on a new framework. Two dimensions are included in the framework: the time horizon of decisions (*long-term decision, medium-term decision, and short-term decision*) and the addressed research topics (*patient flow optimization, bed capacity management, and personnel planning*). Additionally, we highlight the assumptions of uncertainties, which is one of the key points in modeling the ICU management problems. Furthermore, the modeling methods and solution approaches are classified and discussed in detail.

Our search logic is based on keywords such as “ICU”, “planning”, and “scheduling”, and forward and backward search. We set the focus on OR/MS journals and include papers from medical journals only if keywords related to OR/MS methods could be found. We do not include all medical papers to maintain a clear focus on quantitative models and methods.

Based on the analysis of existing papers, we notice that there are still research gaps to be studied in future. The future research topics are discussed along three streams: First, the integration of ICU decisions in the hospital patient flow, e.g., improving the ICU admission control decisions process based on information from other departments or enabling better decision making in other departments considering ICU capacity levels. Second, ICU management problems, e.g., sizing and pooling all corresponding ICU capacities in a region or combining bed capacity planning and personnel scheduling. Third, modeling methods and solution approaches, e.g., applying non-homogenous distributions to model uncertainties or exploring more efficient solution approaches.

Nowadays, OR/MS is becoming more and more visible in ICU management. It will be worthwhile to pay close attention to future research in this area and to contribute to this highly relevant field in healthcare management.

3. Managing Admission and Discharge Processes in Intensive Care Units

Co-authors:

Andreas Fügener

Faculty of Management, Economics and Social Sciences, University of Cologne, Germany

Jochen Gönsch

Mercator School of Management, University of Duisburg-Essen

Jens O Brunner

Faculty of Business and Economics, University of Augsburg

Manfred Blobner

Clinics for Anaesthesiology, Technical University of Munich, Klinikum Rechts der Isar

Published in:

Health Care Management Science (2021), DOI: 10.1007/s10729-021-09560-6

Abstract:

The intensive care unit (ICU) is one of the most crucial and expensive resources in a health care system. While high fixed costs usually lead to tight capacities, shortages have severe consequences. Thus, various challenging issues exist: When should an ICU admit or reject arriving patients in general? Should ICUs always be able to admit critical patients or rather focus on high utilization? On an operational level, both admission control of arriving patients and demand-driven early discharge

of currently residing patients are decision variables and should be considered simultaneously. This paper discusses the trade-off between medical and monetary goals when managing intensive care units by modeling the problem as a Markov decision process. Intuitive, myopic rule mimicking decision-making in practice is applied as a benchmark. In a numerical study based on real-world data, we demonstrate that the medical results deteriorate dramatically when focusing on monetary goals only, and vice versa. Using our model, we illustrate the trade-off along an efficiency frontier that accounts for all combinations of medical and monetary goals. Coming from a solution that optimizes monetary costs, a significant reduction of expected mortality can be achieved at little additional monetary cost.

3.1. Introduction

The intensive care unit (ICU) is one of the most crucial and expensive resources in the health care system [93]. Specialized equipment and highly skilled staff provide special care to the most severe and acute patients, leading to significant costs. In the US, costs for intensive care represent about 16.9%–38.4% of total hospital cost, which amounts to 5.2%–11.2% of national health expenditures [8, 94]. In order to cut costs, hospitals have aggressively reduced ICU beds [95]. As a consequence, the demand exceeds the capacity on a regular basis. Limited resources and increasing demand lead to overcrowding in many ICUs. As a result of this, Boyd and Evans [96] expect a shortfall of intensivist hours in the United States of 22% by 2020, and that this shortfall will increase to 35% by 2030.

ICU processes contain various uncertainties, which increase the difficulty of ICU management [97]. For example, the patient arrival pattern is hard to predict. Patients may be directly admitted to the ICU, arrive spontaneously after problems during a scheduled surgery, or transfer from the emergency department (ED), if necessary with a stopover in the operating room [3]. Among the patients in the ICU, the degree and severity of the disease as well as its subsequent treatment vary significantly. Furthermore, these health conditions will change during the stay in the ICU rapidly and unexpectedly. Thus, the length of stay (LOS) of an individual patient is hard to predict [89].

Patients in need of ICU care are critically ill by definition. Most patients' life-threatening conditions have to be treated immediately because delayed ICU admission is associated with higher probability of mortality and additional resource expenditure ([98–100]). When an additional patient unexpectedly needs intensive care treatment in a hospital with a congested ICU, there are two options – both associated with a major loss of time until sufficient treatment can be initiated. First, the patient could be transferred to another department or even another hospital with available ICU capacity. Until then, the situation might lead to patients being treated in the ED [100]. Second, a patient currently staying in the ICU is discharged earlier than planned to make space for the new patient. KC and Terwiesch [15] suggested such practice when the system load is high. Early discharge, however, requires bridging strategies including respective facilities. Many ICUs, e.g., provide an intervention room to stabilize the patients' conditions and bridge for a short time until the bed is made available. Another option aims at surgical ICU patients expanding their treatment within the operation theatre, e.g., in the OR or in the recovery room. Such bridging approaches, however, do not substantially resolve the congestion of the respective ICU ([2, 4, 5]). Neither of these options is desirable, because the morbidity and mortality of patients might increase [15]. Furthermore, patient pathways connect the ICU to other units inside and outside the hospital [101]. Decisions made in the ICU also influence upstream departments and downstream departments [3]. Capacity shortages in the ICU can also cause congestion of the patient flow within the entire hospital, e.g., by blocking transferals from the ED. Additionally, overloaded staff and decreased revenues are other possible negative effects. Thus, making good admission and discharge decisions is crucial to managing ICU capacities efficiently and simultaneously ensuring a high service quality.

In many ICUs, including the case study hospital, a myopic strategy (that is, only considering direct and immediate effects) of patient admission and discharge control is applied: As long as free beds are available, any new arriving patient is admitted. In case of capacity shortages, different myopic policies (such as the early discharging of existing patients or the rejection of the arriving patient) are applied to minimize the direct negative consequences that are typically evaluated based on the judgment of the ICU physicians. Strategies applied in practice are discussed in several papers [2, 14, 49, 63]. Although these myopic strategies are easy to implement, they have shortcomings. For example, when the last available bed is assigned to a patient who might also be diverted or delayed, the next arriving patient who cannot either be diverted or delayed will cause an issue. The American College of Critical Care Medicine defines and regularly updates guidelines on ICU admission, discharge, and triage decisions [102]. They identify the prioritization of patients and management of scarce ICU resources as an open issue: Instead of providing a clear recommendation, they conclude that "further research is needed on all aspects of rationing critical care resources to narrow the current gaps in allocating scarce resources".

To help answer these questions, we consider optimal ICU admission and discharge policies in an analytical model. It shows that capacity allocation and rationing issues are central and at the heart of important operational questions: When should an ICU admit or reject arriving patients? Should ICUs reserve capacity in order to be able to admit critical patients most of the time or rather focus on high utilization? Should an arriving patient be admitted, although this necessitates prematurely discharging another? Obviously, both admission control of arriving patients and demand-driven early discharge of currently residing patients are operational decisions and should be considered simultaneously. Naturally, when employing additional staff in the ICU, more patients can be treated. But the fixed cost of staffing will also be increased, and the training cost should be considered as well. Finally, the above-mentioned bridging approaches are an inherent precondition of such solutions. Actually, more staff and adhoc available facilities can be assumed as additional ICU capacity. In our model, we assume that the capacity of the ICU (both beds and staffing irrespectively the labelling) is fixed. To do so, we use the stylized model of ICU admission and discharge visualized in Figure 3.1. As in Litvak et al. [14], patient arrivals may be differentiated between three types: The first type includes patients following *elective surgeries*, where the nature of the surgery typically requires intensive care. The arrival times of these patients depend, prima facie, on the surgical schedule. However, the uncertainties in operating room scheduling [32] as well as in the surgery process [103] make the time of arrival at the ICU stochastic. The second type comprises *internal emergency patients*. These patients have already been admitted to the hospital, and unexpectedly require intensive care. Typical examples are routine surgical procedures which become more complex and lead to the patient now requiring intensive care, or readmissions following early discharges from the ICU. The final type of arriving patients describes *external emergency patients*, who are mostly brought in by ambulance.

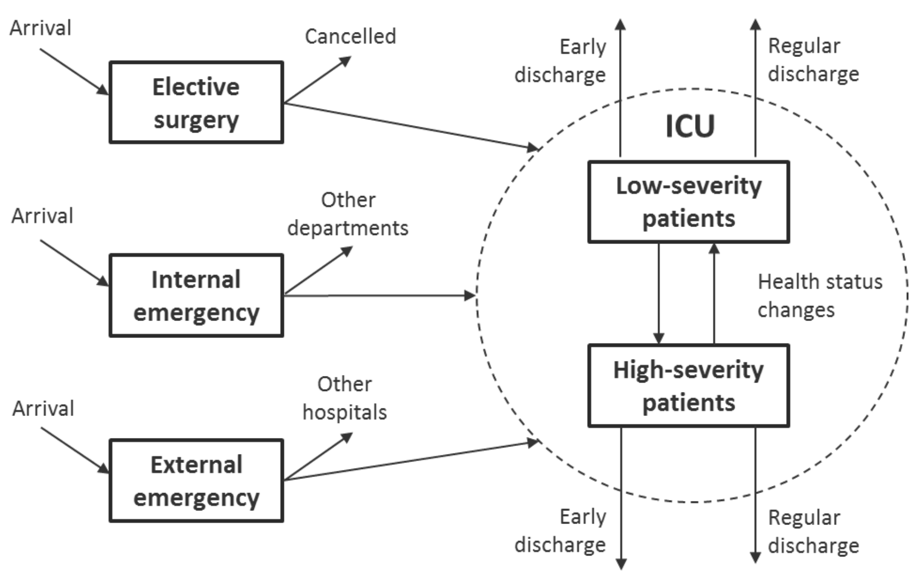


Figure 3.1: The patient flow in the ICU

Surgical patient need ICU treatment based on their preoperative health status, invasiveness and extent of surgery and the degree of surgical organ and tissue trauma. The prediction is complex and characterized by a considerable amount of uncertainty. This dilemma is reflected in the fact that 70%

of all deaths in hospital occur in normal wards rather than intensive care units [Pearse RM. Lancet 2012; 380: 1059-65]. Due to the increasing use of Big Data analyses in healthcare, machine learning algorithms have turned out to be superior to traditional scores in prediction accuracy.[Kagerbauer S. Anästh Intensivmed 2020; 61: 85–9] Recently, Jauk et al. suggested a highly accurate postoperative risk prediction model for ICU admission.[Jauk S. Stud Health Technol Inform 2019; 264: 173-7] Even if advanced models accurately predict the probability of postoperative ICU need on a personalized level, even exact prediction alone lacks operational benefit. Key question is, if a hospital provides for each of n scheduled patients with a predicted (even exact) individual probability p_i for ICU treatment n or $n*p_i$ beds? Since probabilities for ICU need usually are right skewed distributed, the answer to this question is critical. Nevertheless, any request for postoperative ICU treatment is decided before elective surgery and, therefore, part of the scheduling process for major elective surgery. Unexpected cancellation of the reservation always leads to re-scheduling of surgery. Actually, the surgery-related factors cannot be estimated sufficiently before surgery is completed. They major contribute to the severity of ICU treatment, which therefore has to be estimated at admittance to the ICU and re-evaluated each treatment day.

In case of in-hospital emergency, an already hospitalized patient will be transferred to the ICU – in case of gestation, emergency care has to be provided in another department of the hospital. These patients' medical history is given in the hospital chart. But it does not hold true for the current cause of the deterioration of the condition. Importantly, even during the recent COVID pandemic such patients are not transferred to other hospitals. Instead, other ICU patients in more stable conditions are transferred.

External emergency patients with a request for ICU treatment are at its best discussed with the out-hospital emergency team resulting in working diagnoses and an estimate of the worst complication to be averted. Actually, their physical status is unknown before admittance at the ICU. In case of gestation of the ICU, ambulances are diverted to other hospitals with available ICU capacity.

Therefore, almost every ICU patients' health status reliably determined ad admission to the ICU but not earlier. Our model dichotomizes the grade of severity, which maps the current human-based decision algorithm best: high-severity and low-severity patients. High-severity patients are characterized by a more critical condition, going along with a longer expected LOS compared to low-severity patients. During their stay at the ICU, the health status of high-severity patients may improve to the low-severity status, and the conditions of low-severity patients might also worsen. Moreover, both types of patients are regularly discharged from the ICU; in case their condition further improves, they are transferred to another unit, or in case of death. Please note that the model can be extended straightforwardly to include more arrival types that enable a more differentiated advanced estimation of a patient's health status by considering, for example, "safe" electives (e.g. hand surgery, young people) and "risky" electives (e.g. heart surgery, elderly), with different probabilities of the patient's status being high-severity. The options of admission and discharge control are admitting or rejecting an arriving patient, and early discharging an existing patient ("early discharge"). Both rejection and early discharge result in negative effects to patients and hospitals, both from a medical and a monetary point of view.

In this paper, we employ a discrete time Markov decision process (MDP). This modelling approach is standard in comparable stochastic dynamic problems with subsequent, interdependent decision opportunities. The objective is to minimize the negative consequences of capacity shortages. Denied admissions and early discharges are penalized. We evaluate the policy resulting from the MDP in two case studies capturing different management objectives – a medical and a monetary perspective – based on real-world data from a large German teaching hospital. The results show that the optimal policy from our MDP model can considerably reduce the negative effects from a medical perspective – the mortality due to capacity shortages may be reduced by 21% in our case study compared to myopic policies. In contrast, myopic policies mimicking intuitive decisions seem to work well from a monetary perspective. However, both perspectives are not aligned and may lead to considerably different decisions and results. Focusing on monetary instead of medical goals, for instance, leads to an increase of expected mortality of nearly 50%. To illustrate the trade-off between both perspectives, we draw an efficiency frontier that includes a representative sample of combinations of medical and monetary goals. We discuss the impact of different combinations of cost parameters on solutions and on the robustness of our model in case of over- or underestimation of cost parameters.

Our approach provides a novel contribution in two directions: First, it enables an analytical demonstration of the trade-offs between medical and monetary goals when designing admission and discharge policies in ICUs. The impact of different goals is large, and deciding on the percentage of resources to be spent on intensive care is of great societal importance. Second, our model provides optimal holistic policies combining admission and discharge decisions in an ICU based on realistic assumptions. Those policies may lead to direct implications for ICU management, such as reserving a certain number of beds for internal emergencies, or diverting ambulances if a certain threshold of critical patients is currently in the ICU. The policies our stylized model produces are of a low complexity level, which means that they can be printed out and be directly used by ICU managers. Thus, there are no requirements on certain information systems that have to be in place in order to implement such policies in practice.

The remainder of the paper is organized as follows. After reviewing the literature on ICU admission and discharge problems in Section 3.2, we describe the problem and present the MDP model in Section 3.3. Section 3.4 explains the data for the case studies. Section 3.5 contains the results of the case studies. We describe the optimal policies of a medical and a monetary objective, analyze their performance, and briefly discuss strategic implications. We perform sensitivity analyses in Section 3.6, including an efficiency frontier discussion that looks at combinations of medical and monetary goals by considering 32 different scenarios with different combinations of cost parameters, and a study on the robustness of our model to over- and underestimation of cost parameters. Finally, Section 3.7 concludes the paper.

3.2. Related Literature

ICU admission and discharge control problems have been studied both by medical and management scholars. Several papers in medical journals (mostly based on retrospective empirical analyses) demonstrate that both delayed admission and demand-driven early discharge result in negative medical outcomes. Chalfin et al. [104] state that patients should be admitted to the ICU as soon as possible, as rejections or delays lead to undesirable consequences. There are plenty of studies discussing the effects of early discharge and readmission in the medical literature. The researchers agree that patients discharged early face additional risks of health deterioration, which might lead to readmission to the ICU. A few studies indicate that these patients tend to have higher mortality than first-time admitted patients [105–107]. To monitor the time to readmission, Helm et al. [108] estimate a readmission density function in order to optimize a post-discharge monitoring schedule and staffing plan. Furthermore, Chrusch et al. [109] conclude that high utilization levels of ICUs may increase readmission rates and mortality rates. Iapichino et al. [16] agree that higher occupancy levels (indicating higher severity levels) lead to higher mortality rates. Consistent with those studies, Bouneb et al. [110] find that bed availability is a main driver for ICU refusals, and that these refusals lead to an increase in mortality; Louriz et al. [111] report an increase of mortality levels of around 10pp (percentage points) in case of refused ICU admission.

Operations research/management science plays an important role in identifying ways to manage ICU capacity efficiently and in ensuring desired levels of service quality. An overview of the related literature concerning ICU management problems published since 1980 can be found in Bai et al. [3].

Several papers discuss the patient flow in ICUs by applying empirical approaches. KC and Terwiesch [15] analyze discharge and readmission processes with econometric statistical methods. They demonstrate that early discharging ICU patients leads to higher ICU readmission rates. Focusing on patients admitted via the ED, Kim et al. [71] evaluate the effect of ICU admissions on patient outcomes by analyzing a large dataset. They conclude that the admission probability is strongly impacted by ICU capacities – the probability of being admitted significantly decreases with increasing ICU utilization. They demonstrate that admitting patients has preferable outcomes; for instance, readmissions or transfers can be significantly decreased. Thus, admission policies might have a considerable impact on patient outcomes. Based on their empirical findings, they model the admission control problem as a discrete version of the Erlang loss model, similar to Shmueli et al. [77], and apply a simulation to estimate the benefit of alternative admission policies. A threshold rule that leads to admission of patients based on the health status and the remaining free capacities shows promising results – the benefits of applying such a policy clearly exceed those of creating an additional bed. Hu et al. [112] focus on ICU admission decisions of internal emergency patients using a data set of 21 hospitals. While they find that early admissions of internal emergencies can significantly reduce negative medical consequences such as mortality, admitting patients proactively can also congest ICUs, leading to an increase of early discharges. A study focusing on the effects of occupancy levels on ICU LOS is carried out by Long and Mathews [113]. They divide the time a patient occupies an ICU bed in a real "service time", where care is provided, and a "boarding time", where patients are basically ready to leave but wait to be discharged. This boarding time correlates with occupancy levels of both hospital wards and the ICU – it increases with increasing ward occupancy, and decreases with increasing ICU occupancy. Interestingly, the effect of high ward occupancy seems to overweight the effect of high ICU occupancy, as in those situations, long boarding times are observed. Miedaner and Sülz [114] study 18 German neonatal intensive care units to analyze whether the ICUs should have a narrow focus and admit a homogeneous patient cluster or whether they should admit a pool of patient clusters. With an empirical study, they found that the organizational units providing services for complex patients should not have a narrow focus, but should rather provide services for related patient segments. .

In the analytical domain, queueing theory and Markov models are the methods mostly applied to ICU admission and discharge control problems. Three of these models apply different variations of queuing theory: Griffiths et al. [78] model the ICU admission control problem as an M/H/c/∞/FIFO (first-in-first out) model, and similarly, Kim et al. [64] apply an M/M/c multi-server system to analyze admission control processes. Shmueli et al. [77] apply a similar M/M/c model to compare myopic first-come-first-served policies to those where only patients with a certain incremental benefit are admitted. They demonstrate that higher rejection rates can lead to preferable medical outcomes. Finally, Chan and Yom-Tov[88] set up an Erlang-R queueing model to make discharge decisions.

MDP plays an important role not only for ICUs, but also various other hospital departments, such as operating rooms, EDs, and inpatient wards. Barz and Rajaram [115] use an MDP for admission control in a hospital. To accept emergency patients under multiple resource constraints, they decide whether to accept or reject elective patients. Approximate dynamic programming-based heuristics are used to solve the model. Samiedaluie et al. [116] study the admission policies in a neurology ward by an infinite horizon dynamic programming approach. Multiple types of patients are classified based on their medical characteristics. The large scale case study solved by approximate dynamic programming (ADP) prove that the optimal policies can reduce the overall deterioration in patients' health status. Zonderland et al. [117] develop an MDP-based decision support tool to schedule the admission of elective and semi-urgent surgeries, considering the capacity of operating rooms. Similarly, Yang et al. [118] optimize the admission policy for surgery patients considering capacity constraints in the surgical ICU. The patients are grouped based on the surgeon performing the surgery. They apply a heuristic solution method to solve the MDP. Even in regular wards, hospitals face the problem of insufficient capacity. Thompson et al. [95] manage ward capacity by transferring patients between different floors in the hospital. To optimize floor choice, they develop and implement an MDP-based decision support system. Gocgun and Puterman [119], Gupta and Lei [120] and Yu et al. [121] apply an MDP appointment scheduling model to optimize the utilization of medical resources, and also solve it using approximate dynamic programming. Li et al. [122] apply dynamic programming as well to schedule limited resources to a large number of jobs. Xie et al. [123] implement a nested policy based on dynamic programming solutions to schedule the appointments for a medical diagnostic facilities.

Four papers using Markov models in an ICU context are most closely related to our work. Dobson et al. [124] use a Markov chain model to evaluate ICU performance of an exogenously given, intuitive decision rule. They model time as discrete days and define patients by their remaining LOS, which, they argue, is in reality deterministic for most patients. If a patient arrives at a full ICU, the one with the shortest remaining LOS (even possibly the new arrival) is discharged early. Chan et al. [2], Li et al. [76], and Li et al. [5] use finite horizon, discrete time MDPs to derive optimal ICU policies. More specifically, Chan et al. [2] consider a planning horizon of one week and use a state space containing the number of patients of several types that are in the ICU. These patient types are defined by their expected initial LOS that is determined by a patient's condition when he/she enters the ICU. Patients do not change their type and the types have different probabilities for a regular discharge in one time period. In each time period, the decision problem is whether and which patient to discharge early. They only briefly discuss rejections of external emergency patients, but suggest in their outlook the consideration of ICU admission decisions to enable a more holistic view. In contrast, Li et al. [76]

study ICU admission decisions with a planning horizon of one day. Because of this short horizon, they assume that patients' health conditions do not change and there are no regular discharges. They distinguish two patient types based on the initial health status. The health condition of type 1 patients is more severe (diagnosed with sepsis, respiratory failure, or problems with the central nervous system) and they are always admitted, even if a (healthier) type 2 patient must be discharged early because the ICU is full. Type 2 patients may be admitted if there are free beds. However, if a type 2 patient is first admitted and later early discharged, it would have been better not to admit him/her. Thus, the decision problem considered is whether to accept an arriving type 2 patient given the current state of the ICU. The authors show that a threshold-type policy is optimal, that is, type 2 patients are only admitted if a certain number of beds is free and that this threshold decreases over time. However, this decrease is obviously an artefact of the artificially limited planning horizon. Controlling for start- and end-of-horizon effects, a time-homogeneous problem probably features a stationary solution. Threshold-based policies are often observed in real-life ICU decision making. Li et al. [5] focus on the maximization of the survival benefits by optimizing the ICU planning with early discharge from an engineering perspective. Their classification of patients follows Li et al. [76]. Although a longer time horizon of ten days is considered, patients still do not change their health status. Unfortunately, there are some disconnects between text and model (e.g. the probability of any health status change is independent of the ICU occupation), which may be caused by the need for simplifications to enable the analytical derivation of structural properties. Surprisingly, the optimal policy derived implies some situations where only the less critical type 2 patients are admitted, but the more critical type 1 patients are rejected.

The papers that are most connected to our work are Kim et al. [71] in the empirical literature, and Li et al. [76] and Li et al. [5] in the modeling literature. We see our approach as complementary to Kim et al. [71]. While they analyze a huge dataset to derive information on admission policies and consequences of those, our approach analytically models such policies. Contrary to Kim et al. [71], our model does not focus on patients admitted via the ED only, but also includes patients with scheduled surgeries or internal emergencies. Compared to the last two papers mentioned above, our model is based on less restrictive assumptions to capture important problem characteristics. In particular, our state space contains patients' current health status (high or low severity), and, thus, we consider health status changes while staying in ICU: some patients recover and some get worse. Furthermore, in line with Chan et al. [2], we derive the probability of regular discharges from the current patient mix in the ICU. Finally, none of the papers discussed above considers the effects of medical and monetary goals, and the underlying trade-off decisions.

3.3. Problem Description and Model Formulation

Both admission control and demand-driven early discharge decisions are included in the MDP model. In case a patient arrives at a congested ICU, there are two possible options: to reject the new patient and to discharge an existing patient early to make room for the new patient. However, both options can lead to negative consequences. Therefore, our objective is to find the optimal decision policy in order to minimize negative consequences of capacity shortage, which can be assessed from a medical or monetary perspective.

3.3.1. Problem Setting

We model the problem as a (stationary) discrete time Markov decision process as illustrated in Figure 3.2. The objective is finding the admission and discharge policy that minimizes average total cost.

We assume an infinite time horizon, and define a time period small enough that at most one patient arrives within each time period. The sequence of events is as follows: At time $t \in \{1, 2\ \dots\}$, time period t begins and all information indexed with t is available. The ICU with a total capacity of B beds is occupied by $x_{j,t}$ low-severity ($j = 1$) and high-severity ($j = 2$) patients and a new patient of type $i \in \{1,2,3\}$ (elective surgery, internal emergency, and external emergency) just arrived. If no patient arrived, we set $i = 0$. Now, admission of this patient and early discharge of an existing patient are decided. Please note that deciding on an arriving patient is for illustration purposes only. These decisions are equivalent to the hospital deciding in advance what it would do with an arriving patient. In practice, electives as well as external emergencies would not arrive and be rejected but would rather be cancelled or diverted in advance.

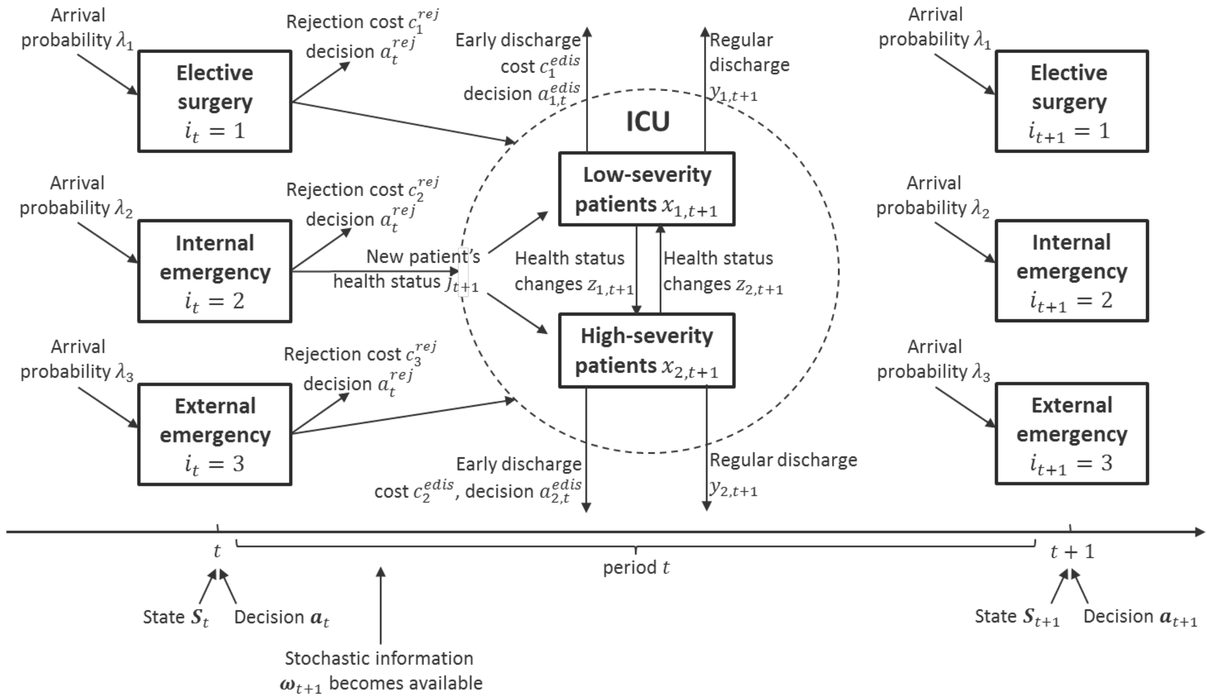


Figure 3.2: Sequence of events

The decisions are captured by the binary action vector $\boldsymbol{a}_t = (a_t^{rej}, a_{1,t}^{edis}, a_{2,t}^{edis})$ whose elements indicate rejecting the arriving patient and early discharging of a low-severity patient or a high-severity patient, respectively. Rejecting a type i arrival leads to penalty costs of c_i^{rej} and early discharging an existing patient of health status j costs c_j^{edis}. The bed preparation time for new patients is relatively short, so we ignore it for new patients as it is frequently did in the other ICU modelling papers ([2, 5, 17]). Therefore, an early discharge can make room for a new patient. As mentioned before, only after a new patient is admitted to the ICU, his/her health status $j_{t+1} \in \{1,2\}$ becomes known. This is because if the patient was not admitted, for example, the ambulance would be diverted and we would never know about that. Furthermore, both types of patients can be regularly discharged.

Moreover, patients' health status may change. A number $y_{j,t+1}$ of patients are regularly discharged and $z_{j,t+1}$ patients change their status from j to $3-j$. Technically speaking, the new information $\boldsymbol{\omega}_{t+1} = (j_{t+1}, y_{1,t+1}, y_{2,t+1}, z_{1,t+1}, z_{2,t+1}, i_{t+1})$ becomes available. If the current patient is admitted, his/her health status j_{t+1} is observed. The information also includes the possible arrival of the next patient i_{t+1}.

In the following, we present the elements of the MDP model in detail. All parameters and variables of the model are listed in Table 3.1.

3.3.2. State, Action and Policy

We use the pre-decision state which captures the state of the system immediately before a decision is taken. The *state* $\boldsymbol{S}_t = (x_{1,t}, x_{2,t}, i_t)$ at the beginning of time period t is defined by three elements: the number of low-severity ($x_{1,t}$) and high-severity ($x_{2,t}$) patients in the ICU as well as the type of the arriving patient i_t. We set $i_t = 0$ if there is no arrival and assume that while the arrival type is known at arrival, the health status can only be diagnosed when the patient is admitted at the ICU.

The *action vector* $\boldsymbol{a}_t = (a_t^{rej}, a_{1,t}^{edis}, a_{2,t}^{edis})$ consists of binary elements indicating whether the arriving patient is rejected ($a_t^{rej} = 1$), as well as whether a low-severity patient ($a_{1,t}^{edis} = 1$) or a high-severity patient ($a_{2,t}^{edis} = 1$) is discharged early. We are interested in a decision rule or *policy* π that gives a best action $\boldsymbol{a}_t$ for every state $\boldsymbol{S}_t$. Thus, the action is a function of the state: $\boldsymbol{a}_t = A^\pi(S_t)$.

Patient indices	i:	arriving patient's type ($i = 1$: elective surgery; i =2, internal emergency; $i = 3$, external emergency; $i = 0$, no arrival)
	j:	index for health status: ($j = 1$, low-severity; $j = 2$, high-severity)
Cost parameters.	c_i^{rej}:	rejection cost of a patient type $i \in \{1, 2, 3\}$ (when the request arrives, only the arrival type is known)
	c_j^{edis}:	early discharge cost of patient with health status $j \in \{1, 2\}$
	$\boldsymbol{c} = (c_1^{rej}, c_2^{rej}, c_3^{rej}, c_1^{edis}, c_2^{edis})$:	cost vector
Distribution parameters	$\mathbf{h} = [h_{i,j}]_{2\times3}$:	probabilities that a type i patient has health status j if admitted
	p_j^{dis}:	probability that a given patient of status j is regularly discharged in one period
	p_j^{cha}:	probability that a given patient of status j changes the health status in one period
	λ_i:	probability that a type i patient arrives, no arrival with probability $\lambda_0 = 1 - \sum_i \lambda_i$
Other param.	B:	total capacity of ICU (number of beds)
	T:	length of time horizon (index $t \in \{1, \dots, T\}$)
Action variables	a_t^{rej}:	binary decision variable indicating whether to reject the arriving patient ($a_t^{rej} = 1$)
	$a_{j,t}^{edis}$:	binary decision variable indicating whether to early discharge a patient with health status j ($a_{j,t}^{edis} = 1$)
	$\boldsymbol{a}_t = (a_t^{rej}, a_{1,t}^{edis}, a_{2,t}^{edis}) = A^\pi(S_t)$:	action vector decided at time t with policy π
State variables	$x_{j,t}$:	number of patients with health status j in ICU at time t
	i_t:	the arrival type of the new patient at period t
	$\boldsymbol{S}_t = (x_{1,t}, x_{2,t}, i_t)$:	state vector at period t

Stochastic information	j_{t+1}:	health status of new patient (a patient of type i has health status j with probability $h_{i,j}$, known if admitted in t)
	$y_{j,t+1}$:	number of regular discharges of type j patients during period t. $y_{j,t+1} \sim B\left(x_{j,t} - a_{j,t}^{edis}, p_j^{dis}\right)$
	$z_{j,t+1}$:	number of patients of type j patients who change their health status during period t. $z_{j,t+1} \sim B\left(x_{j,t} - a_{j,t}^{edis}, p_j^{cha}\right)$
	i_{t+1}:	the arrival type of the new patient at period $t+1$ (a patient is of type i with probability λ_i)
	$\boldsymbol{\omega}_{t+1} = (j_{t+1}, y_{j,t+1}, z_{j,t+1}, i_{t+1})$:	vector of information that becomes available at the end of period t

Table 3.1: Parameters and variables of the MDP model

3.3.3. Stochastic Events, Transformation Function and Transition Probabilities

Stochastic events include four parts. During period t, the information $\boldsymbol{\omega}_{t+1} = (j_{t+1}, y_{1,t+1}, y_{2,t+1}, z_{1,t+1}, z_{2,t+1}, i_{t+1})$ becomes available. If patient i_t was admitted to the ICU, his/her health status $j_{t+1} \in \{1, 2\}$ becomes known. A number $y_{j,t+1}$ of patients is regularly discharged and $z_{j,t+1}$ patients change their health status, for $j = 1, 2$. Finally, a new patient i_{t+1} might arrive.

The new state S_{t+1} at the beginning of the next period $t+1$ is a function of the previous state $\boldsymbol{S}_t$, the action $\boldsymbol{a}_t$ and the new information $\boldsymbol{\omega}_{t+1}$. It is given by the following *transformation function*, which could be easily generalized to more health statuses (please note that $\operatorname{sgn}(i_t)$ takes a value of 1 if a patient arrives ($i_t > 0$), and 0 if no patient arrives ($i_t = 0$)):

$$\boldsymbol{S}_{t+1}(S_t, \boldsymbol{a}_t, \boldsymbol{\omega}_{t+1}) =$$
$$\begin{pmatrix} x_{1,t} + \operatorname{sgn}(i_t) \cdot (1 - a_t^{rej}) \cdot (2 - j_{t+1}) - a_{1,t}^{edis} - y_{1,t+1} - z_{1,t+1} + z_{2,t+1}\,, \\ x_{2,t} + \operatorname{sgn}(i_t) \cdot + \left(1 - a_t^{rej}\right) \cdot (j_{t+1} - 1) - a_{2,t}^{edis} - y_{2,t+1} - z_{2,t+1} + z_{1,t+1}, \\ i_{t+1} \end{pmatrix}. \quad (1)$$

Both the patients' arrivals and LOS contain uncertainties that are difficult to model. Although some researchers argue that there are no suitable distributions to model the arrival pattern and LOS in the ICU [73], and especially the LOS is not reliably predictable for individual patients [89], most papers in the literature apply theoretical distributions. We apply memoryless distributions for arrival rates and lengths of stay, an assumption that proved to be suitable in the literature [3].

Regarding the elements of $\boldsymbol{\omega}_{t+1}$, we model the following dependencies and distributions:

- The new patient's health status j_{t+1} depends on his/her type i_t. Parameters $h_{i,j}$ give the probability that a type i patient has health status j if admitted and are grouped into a matrix $\mathbf{h} = [h_{i,j}]_{3\times 2}$. In case a patient is not admitted, the status is meaningless and, technically, an arbitrary one realizes.

- The number of regular discharges $y_{j,t+1}$ depends on the number of type j patients in the ICU, that is, $x_{j,t} - a_{j,t}^{edis}$. Each patient is regularly discharged (including transfers and events of death) with p_j^{dis}, independently from the other patients. Thus, the number of regular discharges $y_{j,t+1}$ follows a binomial distribution: $y_{j,t+1} \sim B\left(x_{j,t} - a_{j,t}^{edis}, p_j^{dis}\right)$.
- Analogously, the number of patients who change their status $z_{j,t+1}$ depends on the number of patients in the ICU as well, that is, also $x_{j,t} - a_{j,t}^{edis}$. Each patient's health status improves or deteriorates with probability p_j^{cha} in one period, independent of the other patients. Thus, $z_{j,t+1}$ follows a binomial distribution: $z_{j,t+1} \sim B\left(x_{j,t} - a_{j,t}^{edis}, p_j^{cha}\right)$. In addition, the sum of regular discharges $y_{j,t+1}$ and patients who change their status $z_{j,t+1}$ cannot exceed the number of patients in the ICU, that is, $x_{j,t} - a_{j,t}^{edis} \geq y_{j,t+1} + z_{j,t+1}$
- Finally, with probability λ_i, a new patient of type i arrives and with probability $\lambda_0 = 1 - \sum_i \lambda_i$, there is no arrival. We group these into the parameter vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$.

Thus, the stochastic distributions are described by the following set of parameters: $\mathbf{h}, p_j^{dis}, p_j^{cha}, \boldsymbol{\lambda}$.

3.3.4. Cost Function and Value Function

The one-step *cost function* $C(\boldsymbol{S}_t, \boldsymbol{a}_t)$ captures the cost of decision $\boldsymbol{a}_t$ in state $\boldsymbol{S}_t$:

$$C(\boldsymbol{S}_t, \boldsymbol{a}_t) = \begin{cases} c_{i_t}^{rej} \cdot a_t^{rej} \cdot \operatorname{sgn}(i_t) + \sum_j c_j^{edis} \cdot a_{j,t}^{edis} & , if\ x_{1,t}, x_{2,t} \geq 0 \ \wedge \ x_{1,t} + x_{2,t} \leq B \\ \infty & , otherwise \end{cases}. \quad (2)$$

Rejecting a type i arrival leads to penalty costs of c_i^{rej} and discharging an existing patient of health status j early costs c_j^{edis}. Note that the first line of (2) refers to feasible states. The second line prevents that an action is chosen that leads to an infeasible state via costs of infinity (for example, more than B beds occupied). We group the cost parameters into the vector $\boldsymbol{c} = (c_1^{rej}, c_2^{rej}, c_3^{rej}, c_1^{edis}, c_2^{edis})$. The "costs" in this model are an abstract concept, and its implications depend on the "cost" perspective applied. For instance, costs could be defined to be the negative effects to the patient health condition, or lost profits from a monetary perspective. Now, we can define the objective function. As we minimize average total cost, this is

$$V_t(\boldsymbol{S}_t) = \min_{\pi} \lim_{T\to\infty} \frac{1}{T} \mathbb{E}_{\boldsymbol{\omega}} \sum_{t=1}^{T} C(\boldsymbol{S}_t, A^{\pi}(S_t)) \quad (3)$$

with $\boldsymbol{S}_t = \boldsymbol{S}_{t+1}(S_t, A^{\pi}(S_t), \boldsymbol{\omega}_{t+1})$.

3.4. Model Input: Medical and Monetary Perspective of Admission and Discharge Consequences

Based on historical data from a large German teaching hospital and the current literature, we estimate model parameters, namely a set of distribution parameters and the cost parameter vector $\boldsymbol{c}$. The used patient-related data is either anonymized data or aggregated data, not requiring any patient informed consent in accordance to the European General Data Protection Regulation (EU directive - 2016/679). In Subsection 3.4.1, we analyze patient arrivals and the evolution of their health status (corresponding to the lengths of their stays) and derive the distribution parameters. In Subsection 3.4.2, we consider the cost vector $\boldsymbol{c}$ based on two different objectives of optimization, namely a medical and a monetary perspective.

3.4.1. Analysis of Historical Arrival and LOS Data

We obtained three months' worth of data concerning patient arrivals and discharges within an ICU of a large German teaching hospital. There are in total $B = 35$ beds in this ICU, and 514 patients were admitted during this time period. For each patient, we know his/her arrival type and LOS. Arrivals are highly fluctuating and range from 1 to 12 patients per day. The utilization level is high (about 95%).

We define the length of a time period as one hour, so that we can assume that there is at most one arrival per period. In the following, we shortly sketch how we obtained the required parameters $\mathbf{h}, p_j^{dis}, p_j^{cha}, \boldsymbol{\lambda}$ from real-life data.

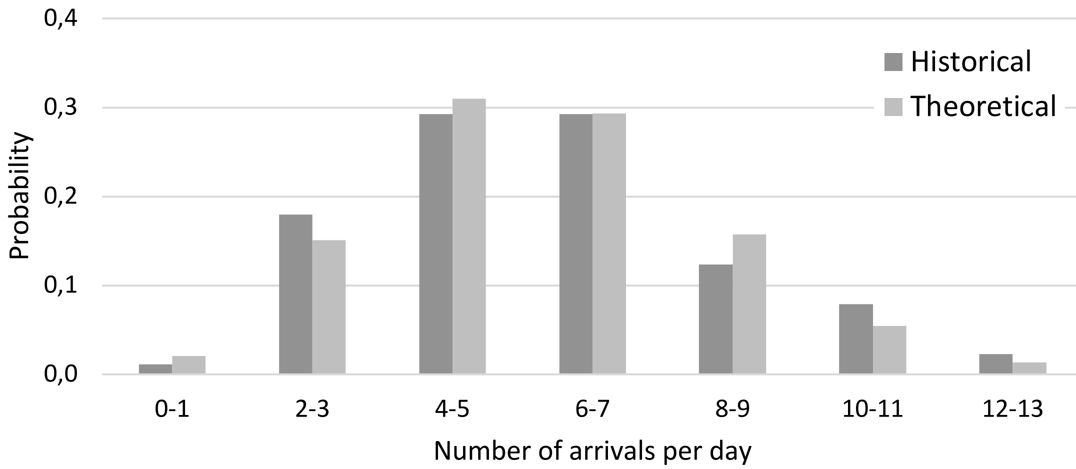


Figure 3.3: Comparison of historical and theoretical arrival process

3.4.1.1. Arrival Process

We analyze the historical data and conclude that Poisson processes are adequate to describe the arrival process for all three patient types. However, in the historical data, we know only the number of admitted patients without any records on the number of rejected patients. According to the literature [125–127], the percentage of patients being denied admission to the ICU ranges between 20% and 60%. Consistent with Mc Manus et al. [66], who analyzed rejection rates in relation to ICU utilization, we increase the historical admission numbers by a factor of 1.25 (that is, assuming an admission rate of 0.8) to calculate the arrival probabilities. In each time period, there can be an arrival of an elective surgery, an internal emergency patient, an external emergency patient, or no arrival. As we have the arrival type in the data, we can directly calculate the following probabilities:

- $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3) = (0.088, 0.153, 0.059)$ and, accordingly, $\lambda_0 = 0.700$.

A visual comparison of historic arrivals and the theoretical predictions (before increasing the parameters by 1.25) is illustrated in Figure 3.3.

3.4.1.2. Health Status Evolution

In a first step, we directly determined the empirical distribution of the LOS for each patient type from the data (solid lines in Figure 3.4). In a second step, we calibrated the stochastic model outlined in Section 3.3.3 to these distributions. Figure 3.4 shows a good fit between the historical (solid lines) and the theoretical (dashed lines) LOS distributions.

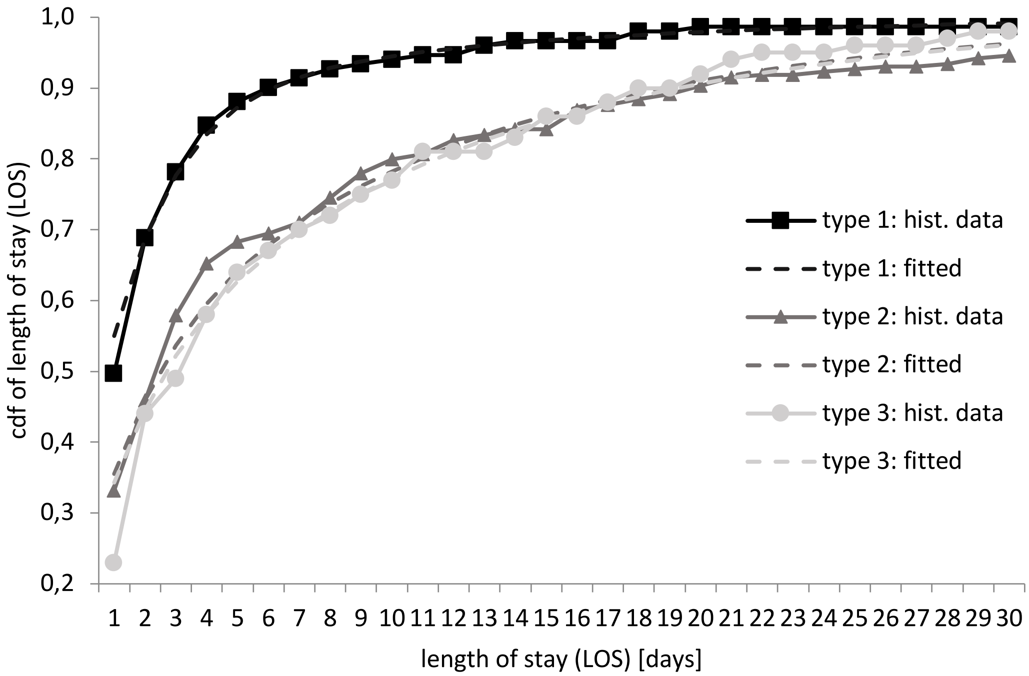


Figure 3.4: Comparison of historical and theoretical LOS distribution

More technically, we used a grid search to choose the probabilities $\mathbf{h}, p_j^{dis}$, and p_j^{cha} for each patient type such that the resulting probability distribution function of the LOS distribution most closely resembles the empirically observed one. Distance was defined as the sum of the absolute distances for each day. In doing so, we again assume that all patients are admitted (as only this is contained in our data) and no early discharges occurred. We obtained the following distribution parameters:

$$\mathbf{h} = \begin{bmatrix} 0.9980 & 0.002 \\ 0.5426 & 0.4574 \\ 0.5141 & 0.4859 \end{bmatrix}, p_j^{dis} = \begin{bmatrix} 0.0177 \\ 0.0024 \end{bmatrix}, p_j^{cha} = \begin{bmatrix} 0.0019 \\ 0.0014 \end{bmatrix},\ \boldsymbol{\lambda} = (0.088, 0.153, 0.059).$$

Thus, the share of high-severity patients depending on the arrival type is as follows: $h_{1,2} = 0.2\%$ of elective surgery patients, $h_{2,2} = 45.74\%$ of internal emergency patients, and $h_{3,2} = 48.59\%$ of the external emergency patients are high-severity patients. Note that most elective surgeries result in low-severity patients.

The probability of a regular discharge in the next period is $p_1^{dis} = 1.77\%$ for a low-severe patient and $p_2^{dis} = 0.24\%$ for a high-severe patient. In our case study, the capacity of the ICU is $B = 35$ beds. No matter how many low-severity patients are in the ICU, the probability of regularly discharging more than three low-severity patients is below 0.3%. Therefore, to simplify the solution of the MDP in this case study, we only consider three or less regular discharges in each time period, that is $y_{1,t+1} \leq 3$. With the same logic, we find that the probability of regularly discharging more than one high-severity patient is below 0.3%. Therefore, we can assume that at most one high-severity patient is regularly discharged, i.e. $y_{2,t+1} \leq 1$. This considerably reduces the number of state transitions to consider without simplifying too much. Of course, these simplifications depend on ICU size and the probabilities. If ICU is orders of magnitude bigger, then considering 2 or more simultaneous discharges may be necessary. But based on our knowledge, we feel this assumption should be widely applicable.

The probability that a low-severity patient worsens to high-severity is $p_1^{cha} = 0.19\%$, while the probability that a high-severity patient improves to low-severity is $p_2^{cha} = 0.14\%$. Again, we analyze the probabilities for all possible numbers of health status changes. For example, it can be shown for our data set that the probability of $z_{2,t+1}$ health status changes from high- to low-severity is highest if the ICU is full of high-severity patients ($x_{2,t} - a_{2,t}^{edis} = 35$) and decreases in $z_{2,t+1}$ for our data. For $z_{2,t+1} = 2$, it is only 0.2%. On the contrary, when $x_{1,t} - a_{1,t}^{edis} = 35$, the probability of $z_{1,t+1} = 2$ is 0.1%. Thus, to simplify the computation of the state transitions, we assume that at most 1 patient of each type ($z_{j,t+1} \leq 1$) will change the health status during one time period.

3.4.2. Definition of Costs

Our model minimizes the costs, that is, negative consequences of capacity shortages within the ICU. Obviously, there is no global definition of negative consequences. In the following, we define two possible perspectives: A medical perspective that minimizes the increase of mortality rates, and a monetary perspective that minimizes the negative effects on hospital profits due to lost revenues and additional costs. This offers the opportunities to discuss the value of our MDP approach compared to myopic heuristics in both perspectives, the consequences of optimizing the medical perspective on monetary performance indicators and vice versa, and possible structures of systems where both perspectives are aligned.

Contrary to lengths of stays and arrival rates as discussed in Section 3.4.1, data on the direct consequences of capacity shortages on mortality rates or hospital profits are typically not available. Besides, they depend on the specific case: mortality rates depend on the level of care and the patient mix. Effects on profits depend on the reimbursement system and contractual specifications. For our case study, we chose the following approach: Regarding the medical perspective, we derive realistic ranges from the literature, and discuss the value for our case study with the case hospital's ICU manager. Regarding the monetary perspective, we rely on the Diagnosis Related Groups system of Germany for the year 2017, the same year which is relevant for our hospital case study. This system publishes cost components based on 1,144 diagnosis groups covering 2.5 million patients. Thus, possible revenues and costs for treatments covering surgeries or intensive care can be derived. In case of rejecting internal emergencies or discharging patients early, additional nursing care might be required. Here, we rely on full cost averages for nurses in Germany. As specified in the previous section, ICU management may choose five possible actions – each with an associated cost – to deal with capacity shortages, depending on the type of an arriving patient and the patients within the ICU. In the online appendix, we discuss how we derived the values for the cost vector $\boldsymbol{c} = (c_1^{rej}, c_2^{rej}, c_3^{rej}, c_1^{edis}, c_2^{edis})$ for the medical ($\boldsymbol{c}_{med}$) in Appendix B.2.1 and the monetary perspective ($\boldsymbol{c}_{mon}$) in Appendix B.2.2. The selected cost vectors for the medical and the monetary perspective are as follows:

- $\boldsymbol{c}_{med} = \left(c_{1,med}^{rej}, c_{2,med}^{rej}, c_{3,med}^{rej}, c_{1,med}^{edis}, c_{2,med}^{edis}\right) = (1\text{ pp}, 15\text{ pp}, 3\text{ pp}, 2\text{ pp}, 10\text{ pp})$.
- $\boldsymbol{c}_{mon} = \left(c_{1,mon}^{rej}, c_{2,mon}^{rej}, c_{3,mon}^{rej}, c_{1,mon}^{edis}, c_{2,mon}^{edis}\right) =$

$(9{,}200\text{ €}, 5{,}800\text{ €}, 4{,}100\text{€}, 700\text{ €}, 6{,}500\text{ €})$.

Please note that a) the choice of parameter values might differ from hospital to hospital, and that b) some parameters might not be determined accurately. In the online appendix, we provide a detailed sensitivity analysis based on different parameter combinations (Appendix B.5.1) and parameter misspecification (Appendix B.5.2).

3.5. Case Study: Implications of Admission and Discharge Policies

We tested the performance of our MDP-based approach and a myopic benchmark mimicking intuitive decision making in practice for both the medical and monetary perspective. In particular, we implemented the following two approaches:

1) *MDP^o* ($o \in \{\text{med}, \text{mon}\}$) is the optimal policy following from the MDP approach described in Section 3.3 optimizing the medical or monetary perspective, that is with the medical ($\boldsymbol{c}_{med}$) or monetary ($\boldsymbol{c}_{mon}$) cost vector. Technically, we used a finite horizon approximation with horizon T to calculate the stationary policy from the value function $V_t(\boldsymbol{S}_t) = \min_{\boldsymbol{a}_t}\left\{\frac{1}{T-t+1} \cdot C(\boldsymbol{S}_t, \boldsymbol{a}_t) + \frac{T-t}{T-t+1} \cdot \mathbb{E}_{\boldsymbol{\omega}_t} V_{t+1}\big(\boldsymbol{S}_{t+1}(S_t, \boldsymbol{a}_t, \boldsymbol{\omega}_{t+1})\big)\right\}$ with the boundary condition $V_{T+1}(\boldsymbol{S}_{T+1}) = 0$. This MDP was first solved via backwards induction (see Appendix B.1), which involves the calculation of $\frac{(B+1)(B+2)}{2} \cdot 4$ states per time period, that is 2,664 states for $B = 35$. The time horizon was sufficiently large such that the ICU is in a steady state in the first time periods and, therefore, the optimal policy reported here does not depend on the time period. In our experiments, we used

$T = 168$ and observed time-independent actions for the first 5 time periods. The runtime for this one-week horizon was about 30 minutes without parallelization, which is negligible, as the optimization needs to be run only once for a set of parameters.

2) *Myopic^{o}* ($o \in \{\text{med}, \text{mon}\}$) is the benchmark policy following from an intuitive, hands-on approach. This policy mimics the decision making in practice and reflects the status quo in our case study hospital. It only takes immediate costs into account. For example, Chan et al. (2012) applied a similar myopic heuristic. More precisely, as opposed to the minimization of immediate ($C(\boldsymbol{S}_t, \boldsymbol{a}_t)$) and future costs $\left(\mathbb{E}_{\boldsymbol{\omega}_t} V_{t+1}\left(\boldsymbol{S}_{t+1}(S_t, \boldsymbol{a}_t, \boldsymbol{\omega}_{t+1})\right)\right)$ in the MDP's value function (3), *Myopic* minimizes only $C(\boldsymbol{S}_t, \boldsymbol{a}_t)$. In our setting, this results in the following simple decision rules:

Decision rule 1: If no patient arrives, do nothing.

Decision rule 2: If there are free beds, accept any arriving patient without early discharging.

Decision rule 3: If there is no free bed, select the cheapest alternative among rejecting the arriving patient or early discharge of a low-/high-severity patient from the ICU.

All policies were evaluated by simulation for a one-year horizon comprising 8,760 1-hour time periods with randomly generated arrivals, health status changes, etc. To eliminate start-of-horizon effects, we simulated an additional 1,000 time periods before this evaluation horizon because preliminary tests showed that after about 600 time periods, start-of-horizon effects were not visible any more. The values reported are averages over 1,000 simulation runs. Wherever appropriate, we also state the 95% confidence intervals of these means. The experiments were implemented using JAVA version 8 and ran on a computer with 3.20GHz CPU, 12 GB RAM, and 64-bit Windows 7.

3.5.1. Policies Resulting from the Medical Perspective

Figure 3.5 shows an overview of the policies resulting from *Myopic^{med}* (upper row) and *MDP^{med}* (lower row). For each possible state $\boldsymbol{S} = (x_1, x_2, i)$, it shows the action $\boldsymbol{a} = (a^{rej}, a_1^{edis}, a_2^{edis})$ taken. The columns represent the type of the arriving patient ($i \in \{1,2,3\}$ for elective surgery, internal emergency, and external emergency, respectively). If no patient arrives, no action is taken, as discharging one of the existing patients early only has negative consequences and can be done later, if necessary. The axes represent the occupancy of the ICU. The vertical axis is the number of low-severity patients and the horizontal axis is the number of high-severity patients in the ICU. For example, the lower right square represents an ICU full of high-severity patients ($x_1 = 0, x_2 = 35$). Now, four policies representing the relevant combinations of actions exist (all other actions are dominated):

1. The patient is admitted, and no patient is early discharged (light shade).
2. The patient is admitted, and a low-severity patient is early discharged (medium shade).
3. The patient is admitted, and a high-severity patient is early discharged (bright shade with "+").
4. The patient is rejected, and no patient is early discharged (dark shade).

Remember that the relationship of the rejection and early discharge costs in this scenario is $c_{i=1}^{rej} < c_{j=1}^{edis} < c_{i=3}^{rej} < c_{j=2}^{edis} < c_{i=2}^{rej}$. The *Myopic*^{med} policy is quite similar for all arrival types. In all cases, patients are admitted as long as empty beds exist (light shade below the diagonal). In case of a fully occupied ICU (represented by the diagonal), the action depends on the arriving patient: Elective surgeries ($i = 1$) will be always cancelled, internal emergencies ($i = 2$) will always be admitted leading to early discharges (if possible, of low-severity patients), while external emergencies ($i = 3$) will be admitted if low-severity patients can be early discharged. In case only high-severity patients are in the ICU the external emergency patient will be rejected.

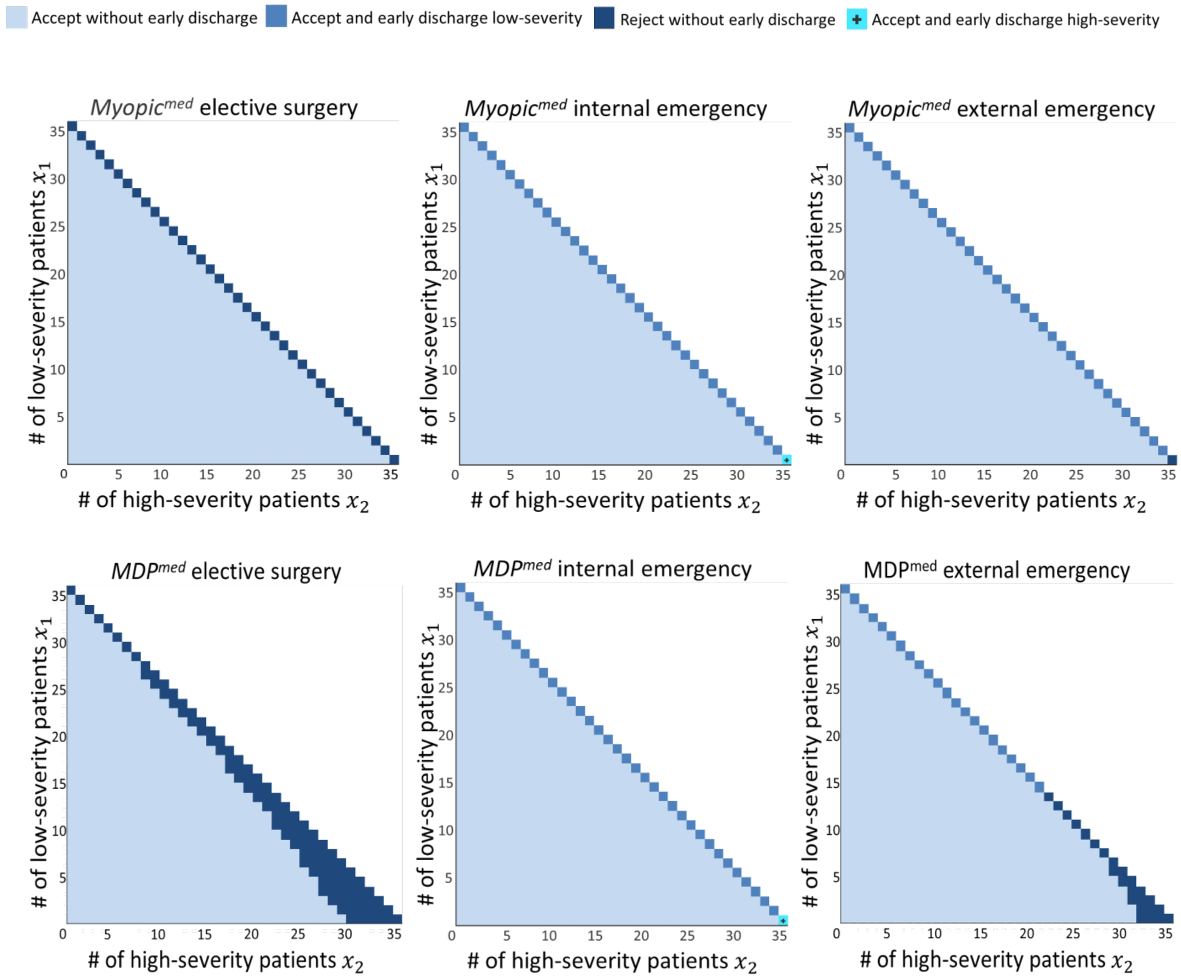


Figure 3.5: Comparison of the policies of *Myopic*^{med} (upper row) and *MDP*^{med} (lower row)

The *MDP*^{med} policy is identical for internal emergencies ($i = 2$), which are again always admitted. However, it differs for scheduled elective surgery patients ($i = 1$) and external emergencies ($i = 3$). In case the ICU contains many high-severity patients, these patients are rejected even if free beds exist. The effect is more pronounced for elective surgeries, from whom a free bed is already reserved even if only eight high-severity patients are in the ICU ($x_1 = 26, x_2 = 8$). In the most extreme case, with only high-severity patients in the ICU, elective surgeries will be cancelled if no more than nine free beds exist, and external emergencies will be rejected if no more than four free beds exist. This is driven by the high cost of early discharging a high-severity patient and her/his longer expected stay in the ICU. Thus, even with some free beds, MDP does not ‘risk’ having a full ICU in the future and rather rejects an elective surgery.

3.5.2. Policies Resulting from the Monetary Perspective

Second, we compare the myopic and MDP policies based on the monetary cost perspective (*Myopic*^{mon} and *MDP*^{mon}). The relationship of the rejection and early discharge costs in this scenario is $c_{j=1}^{edis} < c_{i=3}^{rej} < c_{i=2}^{rej} < c_{j=2}^{edis} < c_{i=1}^{rej}$. Note that the rejection cost of an elective surgery patient is now the most expensive cost, while it was lowest in the medical perspective.

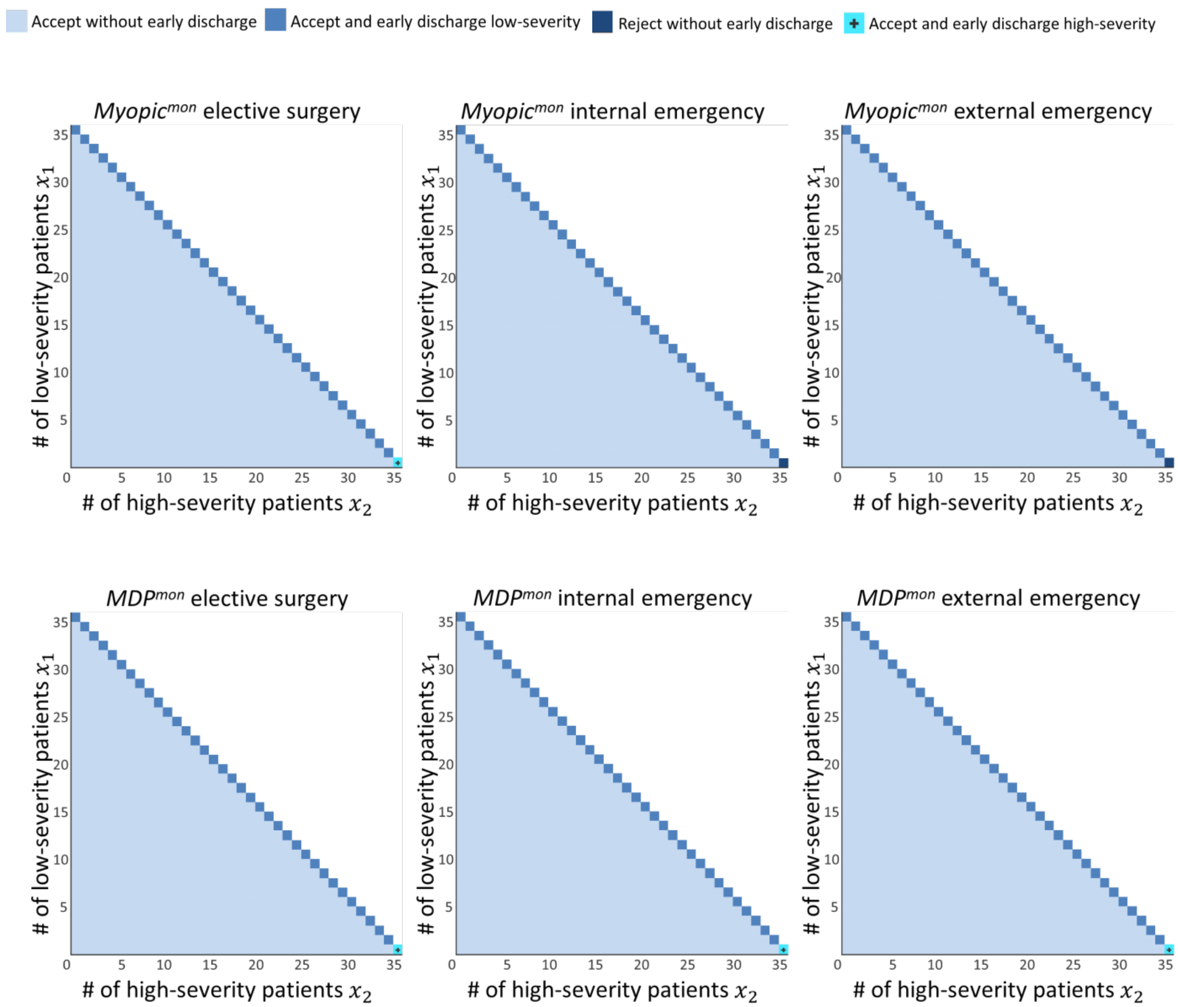


Figure 3.6: Comparison of the policies of *Myopic*^{mon} (upper row) and *MDP*^{mon} (lower row)

Both *Myopic*^{mon} and *MDP*^{mon} yield almost identical policies in this scenario (Figure 3.6). If possible, all patients are admitted. If the ICU is full, the least critically ill patient is discharged early. The only exception is the case where only high-severity patients are in a full ICU, and an internal emergency ($i = 2$) or an external emergency patient ($i = 3$) arrives. While *Myopic*^{mon} rejects this patient, *MDP*^{mon} admits the patient and discharges a high-severity patient early.

3.5.3. Performance Analysis of Admission and Discharge Policies

In this subsection, we analyze the performance of the policies described above over a one-year horizon using a simulation with 1,000 runs. On average, 2,628 patients arrived at the ICU during this year. We first describe ICU occupancy for MDP^{med} and MDP^{mon} using heatmaps and then share performance indicators like utilization and compare them to *Myopic*. The heatmap in Figure 3.7 shows the relative frequency of the ICU occupancy using MDP^{med}. Again, the vertical axis denotes the low-severity patients and the horizontal axis denotes the high-severity patients. Obviously, the ICU is never close to empty (the white area) and in the majority of time, around 1 to 10 low-severity and 23 to 34 high-severity patients are in the ICU (The 0.0 in the figure illustrates a probability below 0.1%). Overall, 82% of the patients in the ICU have the high-severity status, although they only account for roughly 20% of all admitted patients. This reflects the fact that high-severity patients stay longer. The key take-away here with regard to the interpretation of the policies shown in Figure 3.5 is that an ICU full of high-severity patients (almost) never happens, but a full ICU with high-severity and low-severity patients is quite common (27.7%). There is frequently (41.3 %) a high utilization level with only 1 or 2 free beds. Thus, the bed-reserving property by rejecting elective surgery patients that distinguishes MDP^{med} from $Myopic^{med}$ is clearly relevant. However, bed-reserving by rejecting external emergencies does not create a large effect.

The heatmap in Figure 3.8 shows that a situation where MDP^{mon} admits the patient and early discharges a high-severity patient actually occurs in some events (in about 2.0% of the time periods, lower right square). Using MDP^{mon}, the ICU is fully occupied in 54.6% of the time periods, more than twice as often as in the medical perspective. The ICU is almost full (1 or 2 free beds) in 33.4% of the time periods, thus, having more than 2 free beds is quite rare. We skip the visualization of the heatmaps for the myopic policies. As can be seen in Table 3.2, the resulting utilization levels are quite similar to MDP^{mon}. Table 3.2 summarizes performance indicators for both approaches and perspectives together with their 95% confidence intervals. First, we discuss the monetary perspective, where myopic policies perform well, before we turn to the medical perspective, where the MDP policies lead to significant improvements

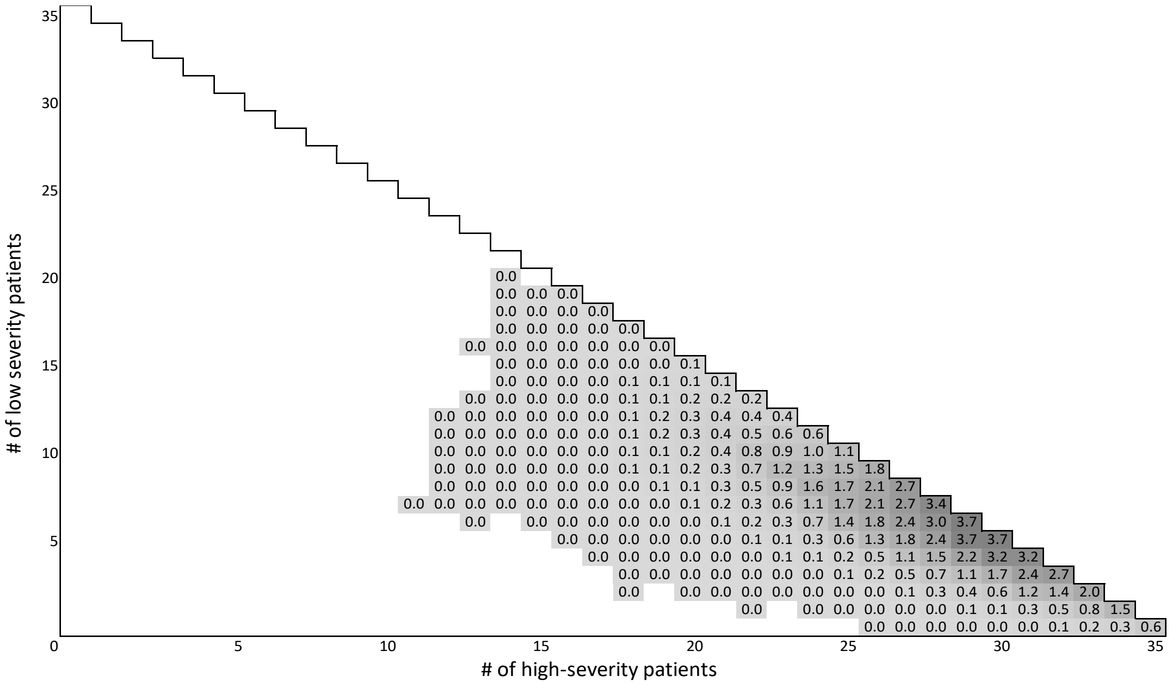


Figure 3.7: Frequency of ICU occupancy for MDP^{med} ([%]; empty/white: never observed after warm-up)

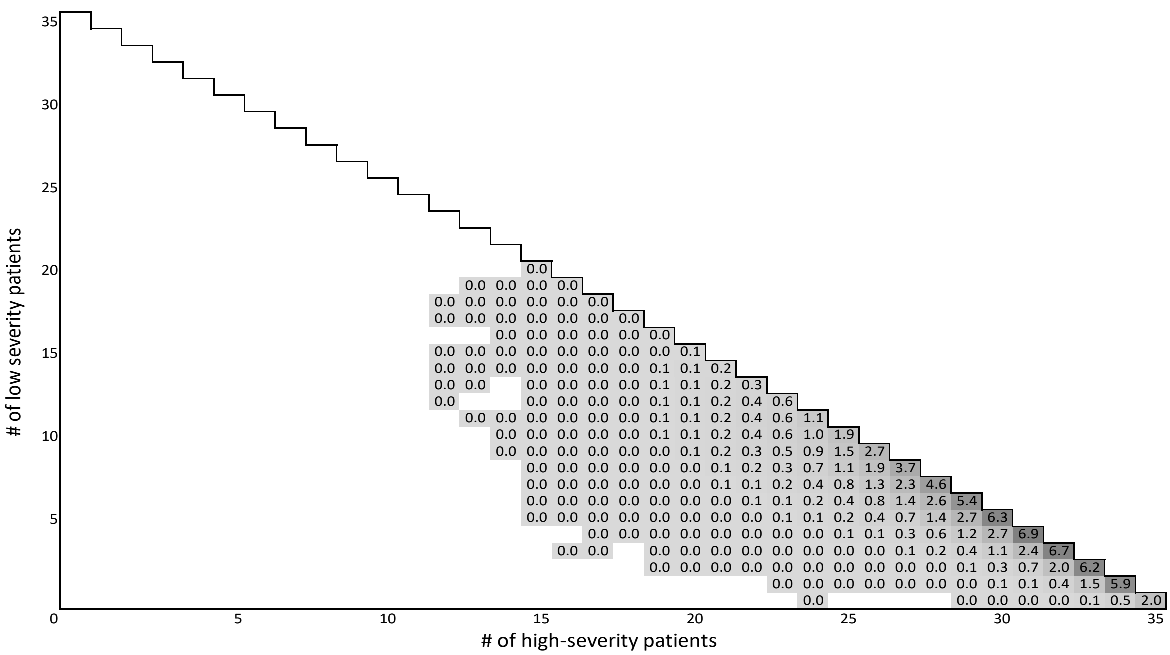


Figure 3.8: Frequency of ICU occupancy for MDP^{mon} ([%]; empty/white: never observed after warm-up)

In the monetary cost setting, the early discharge of low-severity patients has much lower costs compared to rejecting patients. Thus, both MDP^{mon} and $Myopic^{mon}$ admit all patients and early discharge low-severity patients if necessary. Even though both policies are nearly identical (see Section 3.5.2), MDP^{mon} outperforms $Myopic^{mon}$ regarding monetary cost by 7.8%.

Optimized Goal	Approach	Medical cost [pp]	Monetary cost [€]	Utilization rate [%]	Rejection rate [%]	Early discharge rate [%]
Medical	MDP^{med}	1,931±197	7,160,950±433,651	94.7±0.6	32.6±1.9	17.8±2.5
	$Myopic^{med}$	2,453±280	4,259,490±420,612	97.4±0.5	14.5±1.3	38.6±3.6
Monetary	MDP^{mon}	2,855±319	1,143,772±156,391	97.4±0.5	0±0	47.0±3.6
	$Myopic^{mon}$	3,172±412	1,239,946±186,341	98.4±0.4	2.0±0.7	46.5±3.3

Table 3.2: Comparison of performance indicators of $Myopic$ and MDP (pp: percentage points)

However, this changes for the medical perspective. Here, MDP^{med} reserves more capacity for critical patients, and starts rejecting scheduled surgery and external emergency patients if too many high-severity patients are treated in the ICU (see Section 3.5.1). Thus, the average utilization is considerably lower compared to $Myopic^{med}$ (94.7% versus 97.4%). Moreover, the effects differ: While the average increase in mortality due to capacity shortages is 2,453 pp per year for $Myopic^{med}$, using MDP^{med} decreases this figure to 1,931 pp. This means a reduction in additional annual mortality by 21% – thus, on average, 5.2 patients will die less every year due to the MDP policy.

Comparing the two objectives for the MDP policies (MDP^{mon} and MDP^{med}), the differences are striking. Applying the monetary perspective, the one-year mortality due to capacity shortages rises from 1,931 pp to 2,855 pp, but the lost profits decrease from 7.1 million € to 1.1 million Euro. That is, the difference of those two objectives is losing around 9.2 patients' lives against losing 6 million €. The reason for the mismatch between the medical and the monetary perspective lies within the reimbursement system.

In Figure 3.9, we plot the five events' costs with the dimensions medical costs (additional mortality rate in pp) on the vertical axis and monetary costs (lost profits in Euro) on the horizontal axis. Medical and monetary consequences are aligned if they have a linear relationship. In our case, there is a major mismatch with regard to rejecting elective surgery patients. This action is at the same time the most favorable from a medical perspective and the least favorable from a monetary perspective. A smaller mismatch occurs with the rejection of external emergencies. To induce that hospitals who are profit maximizers also maximize medical quality, reimbursement systems should make sure that these mismatches are eliminated. Here, one might consider decreasing the monetary costs of rejecting external emergency patients and especially elective surgery patients, while the monetary costs of early discharging patients or rejecting internal emergency patients might be increased.

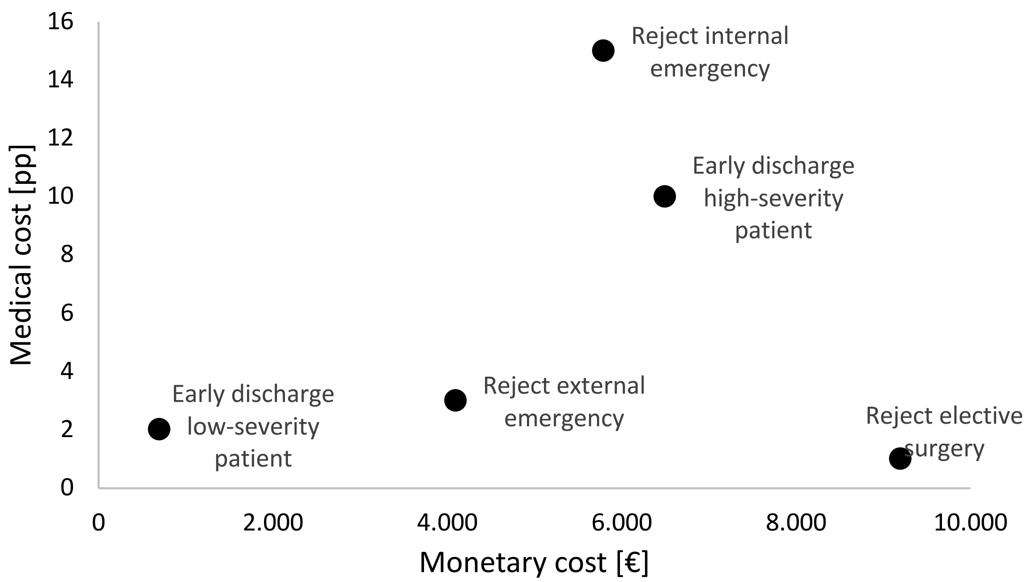


Figure 3.9: Medical and monetary costs for all five events

The results of our approach vary if different parameters are selected. As mentioned in the previous section, we provide a detailed sensitivity analysis based on different parameter combinations (Appendix B.5.1) and parameter misspecification (Appendix B.5.2) in the online appendix.

3.6. Scenario Analysis

The admission and discharge policies discussed in the previous section were based on two sets of cost estimations for our case hospital. In this section, we take a broader look by varying capacities and costs in simulation (1,000 simulation runs each). We consider two different variations of scenarios: First, we derive strategic implications by changing the available number of ICU resources: What are the benefits (costs) of adding (removing) one additional bed? Second, we drop the assumption of either considering purely the medical or the monetary perspective, and allow combinations of both. Based on 20 costs settings consisting of linear combinations of medical and monetary costs, we derive an efficiency frontier. In the online appendix, we consider two more variations. In Appendix B.2.1, we consider different cost parameter settings for the medical costs by increasing or decreasing each cost parameter by 50 percent, leading to 32 additional scenarios. Thus, we can demonstrate the value of our model based on different cost settings. In Appendix B.2.2, we analyze the impact of estimation errors on performance by comparing policies based on those 32 additional cost scenarios, while the costs of our case studies represent the ground truth. This sensitivity analysis demonstrates the cases where our model performance is robust, and those where wrong cost estimations lead to severe consequences. Please note that the appendix relates to classical sensitivity analyses. As high uncertainty is typically associated with the estimation of medical consequences, we concentrate on medical costs in these analyses. We do not perform a dedicated sensitivity analysis for the arrival rates, as this is to some extend already captured by the variation of ICU capacity.

3.6.1. Strategic Implications

Besides calculating optimal admission and discharge policies, the model can also serve to obtain insights on a strategic level, for example, regarding the dimensioning of ICU capacity. The model computes admission and discharge policies to minimize costs (that could be medical or monetary) based on a given capacity. Thus, by varying this capacity, we can estimate the benefits or costs of capacity changes. To this end, we run our model to determine the optimal policies for both the medical and the monetary perspective for capacities of 30 to 40 beds, and apply the simulation to report the medical (increase in mortality rate) and the monetary (lost profits) results in Appendix B.3. The impact of capacity changes is quite linear. If the medical perspective is optimized, decreasing the capacity to 30 beds leads to an additional mortality of 700 pp – meaning that, on average, 7 more patients die due to capacity shortages. In case the monetary performance is optimized, the additional mortality increases to 1,128 pp, resulting in more than 11 additional mortalities. The additional monetary opportunity losses to the hospital due to a decrease from 35 to 30 beds are about 0.6 million Euros (monetary perspective optimized) up to 1.3 million Euros (medical perspective optimized). Consistent with these results, the utilization is slightly increased from 94.7% to 94.9% in the medical setting, and increased from 97.4% to 98.2% in the monetary setting. Increasing the capacity from 35 to 40 beds leads to improvements: The number of additional lives saved ranges from 6.1 (medical perspective optimized) to 9.7 (monetary perspective optimized), and the reduction of monetary opportunity losses ranges from 0.4 million Euros (monetary perspective optimized) to 1.7 million Euros (medical perspective optimized). The utilization drops from 94.7% to 93.8% in the medical setting and from 97.4% to 95.8% in the monetary setting. It is interesting to see that the positive effects seem to be stronger for the perspective that is not optimized – in cases with fewer capacity shortages, the trade-off between medical and monetary consequences seems to disappear. However, fixed costs clearly increase with capacity. Besides costs for new equipment, different (non-continuous) staffing requirements must be considered.

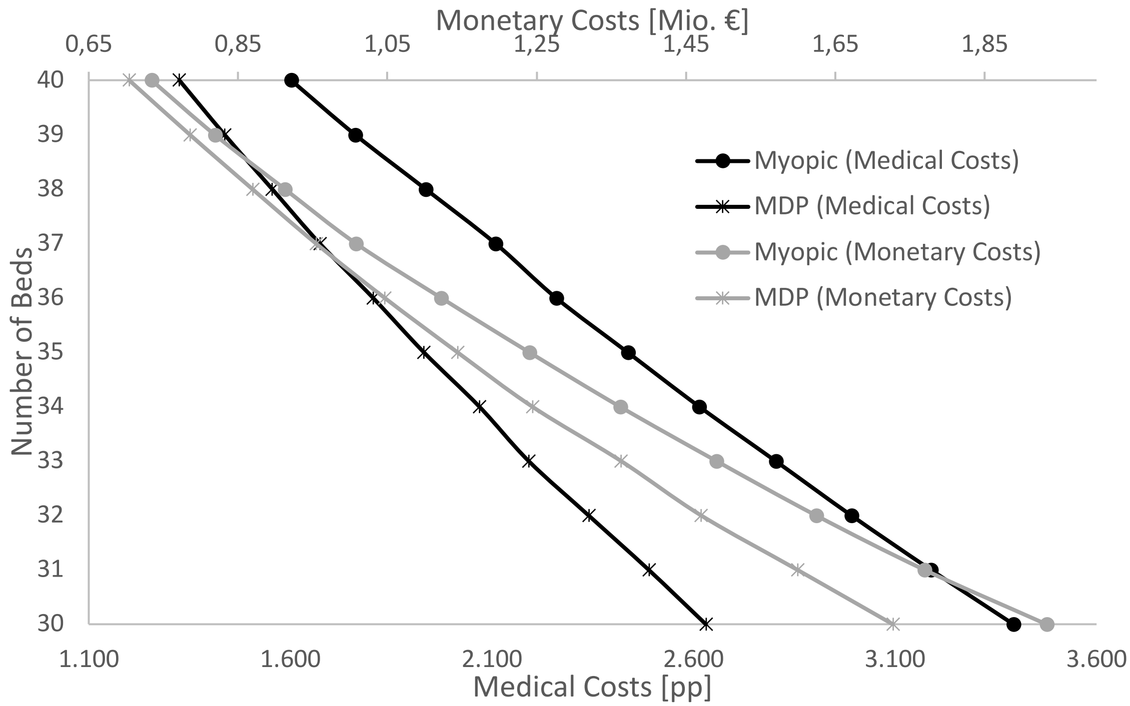


Figure 3.10: The capacity saved by MDP

Furthermore, we can also change the point of view and ask how many beds we can save by switching from myopic to MDP policies (Figure 3.10). More precisely, the same cost level can be obtained with fewer beds. Especially when focusing on the medical perspective, the MDP saves between two and five beds compared to the myopic policy at the same cost level. For instance, implementing the myopic policy in the ICU with 34 beds results in an expected mortality due to capacity shortages of around 2,600 pp. The same figure is achieved when implementing the MDP policy with a capacity of only 30 beds. Focusing on monetary goals, the differences are less pronounced. Thus, the approach may help to provide valuable input when making capacity dimensioning decisions. Besides strategic level planning, when making operational decisions, such as closing beds in the ICU because of staff shortages, it can help as well to show the resulting consequences in both medical and monetary perspectives.

3.6.2. Trade-off between Medical and Monetary Costs

The case study in Section 3.5 demonstrated that different cost perspectives lead to different policies. While we previously focused on either the medical or the monetary perspective, we now honor the fact that these are two extreme cases. In reality, this is a problem with two objectives which are both simultaneously important to the decision maker. Thus, we now numerically construct an efficient frontier (line with “x” in Figure 3.11) that contains policies which are optimal for a certain weighted combination of the medical and monetary perspectives. Note that in order to have comparable figures, we rescale monetary costs by a factor of 1/1,000 in this analysis. All points to the right/above the frontier are inefficient because at least one perspective can be improved without worsening the other. As our cost function is linear, this frontier can be easily constructed by considering convex combinations of the two cost perspectives' parameters. To obtain points on this frontier, we first calculated 11 cases with different weightings of the two perspectives by increasing the relative weight for medical costs from 0 (case 0) to 1 (case 10) in ten steps of 0.1. In order to capture all parts of the efficient frontier, we additionally inserted 10 non-equidistant cases between the aforementioned (e.g. case 8a with a weight of 0.825 for medical costs). For each case, the resulting optimal policy is evaluated in simulations as before, and medical as well as monetary costs are recorded. The policy is quite insensitive to the weights in some areas (e.g. cases 0, 1, and 2 with relative weights of medical costs of 0 to 0.2), resulting in very similar cost values for these cases. In other areas, a small change in weights (e.g. cases 8e and 8f with relative weights for medical costs of 0.8915 and 0.8916, respectively) results in a change in the policy with big effects on costs. More information and resulting policies are given in Appendix B.3.

As its weight increases from case 0 to case 10, the resulting total medical costs decrease from 2,855 to 1,931, which means that, on average, 9.2 patients die less due to capacity shortages. The reverse is also true: If we put stronger emphasis on the monetary perspective, monetary costs decrease, while mortality increases. As usual, this trade-off is not linear. When starting to increase the weight on medical costs in the first nine cases, the medical perspective can be improved at relatively low monetary costs: By decreasing the medical costs from 2,855 pp to 2,445 pp (saving around four patients, that is, 44% of the potential decrease of medical costs) the monetary costs only increase from 1.14 to 1.51 million Euro (costing around 370,000 €, that is, 6% of the potential increase of monetary costs). After case 9, improving the medical perspective gets more expensive – now, the policies start to reject external emergencies and scheduled surgery patients. Switching from case 9 to case 10 will save in expectation one life (decrease of medical costs from 2,069 to 1,931), and will lead to a monetary cost increase of 3.2 million Euro (increase of monetary costs from 3.98 Million Euro to 7.16 Million Euro).

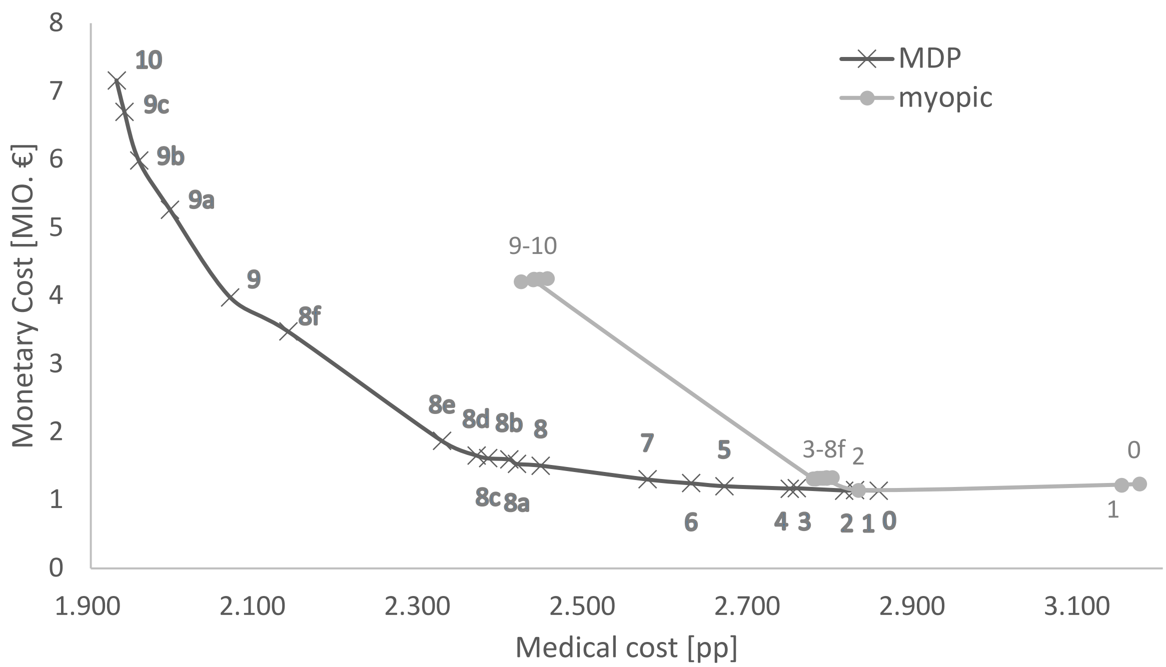


Figure 3.11: The trade off between medical and monetary cost

In addition to the MDP policies, Figure 3.11 also contains the results of the myopic policies (line with circles). These policies myopically decide on weighted costs. Obviously, the policy rarely changes as the weights vary. Here, the policies only change if the order of costs change. When we compare the performance of the MDP and the myopic policy in all 20 cases, the potential of the MDP solution becomes obvious: Considering case 9 with a myopic policy as a benchmark (remember that the myopic policy resembles policies used in practice), we could either reduce mortality by around four patients (from 2,453 pp to 2,069 pp), combined with a slight reduction of costs by moving to case 9 with the MDP solution, or reduce costs by 2.75 million Euro (from 4.26 to 1.51 million Euro), keeping the medical costs constant, by moving to case 8 with the MDP solution. We draft the resulting policies in Appendix B.4. The most noticable change is that the MDP starts to reserve beds by deferring external emergencies (starting from case 2), and by cancelling scheduled surgeries (especially in case 8a, where the high monetary costs of such policies are not considered) when moving from monetary- to medical-oriented policies.

According to the sensitivity analysis based on different parameter combinations (Appendix B.5.1), our model has significant benefits in most of the considered test cases, there are a few cases where the MDP does not reserve beds, and its use does not lead to a considerable improvement compared to a myopic policy. The sensitivity analysis of parameter misspecification (Appendix B.5.2) shows that erroneous estimation of cost parameters may indeed lead to dramatic results. The worst impact on medical costs was observed for combinations of overestimation of rejection costs and underestimation of the cost of early discharges, while results are otherwise relatively robust.

3.7. Conclusion and Future Research

Congestion problems in ICUs lead to dramatic negative effects on patients' health. Both rejections of arriving patients and early discharges of existing patients lead to worse outcomes. This paper proposes a method to define admission and early discharge policies that minimize these negative consequences. Our approach applies a discrete-time Markov decision process that is solved to optimality for realistic instances. We demonstrate that by minimizing the medical consequences, the approach significantly outperforms a myopic policy as applied by most hospitals in practice. Besides, we demonstrate that different objectives lead to different policies. If, for example, monetary profits are optimized, the medical outcome is strongly affected. We extend this logic to develop an efficiency frontier covering medical and monetary perspectives, and thereby contribute to the ongoing discussion on the trade-off between medical quality and monetary costs. We further provide robustness checks and situations in medical perspective where our approach is sensitive to cost changes and to cost estimation errors. Our model provides particularly high potential in cases with low medical costs for rejecting external emergencies and high costs for early discharging low-severity patients. It is relatively robust against underestimation of rejection costs for scheduled surgeries and external emergencies and overestimation of early discharge cost of high-severity patients. However, the opposite case, that is, overestimation of rejection costs and underestimation of early discharge costs, leads to inferior results.

Various applications of our approach exist. The major one is a framework to develop recommendations for admission and discharge control on a tactical decision level. One could, for example, use it to develop simple guidelines. Such decision rules may have the form that if a certain number of high-severity patients are treated in the ICU, no more elective surgeries will be scheduled that require post-operative ICU treatment, or define occupancy levels where coordinating units are informed to divert ambulances with external emergencies to other hospitals. The policies as we illustrate them could be printed out and the ICU manager could have them as a poster in the ICU – no additional information systems would be required. An additional application is to use the approach as decision support for capacity dimensioning on strategic and operational decision level. It provides insights on the consequences of capacity shortages, and allows decision makers to consider different objectives within the admission and discharge policies. In both cases, our approach has direct managerial applications.

We believe that managing ICU admissions and discharges is of great importance, and has large potential for future research. To focus on the trade-off between medical and monetary goals, and to allow easy implementation of our proposed policies in practice, we aggregated situations (e.g. day and night shift) and patient types. From a modeling point of view, adding a higher level of complexity could be of interest – even though this might reduce the ease of implementation. Possible extensions include a time-dependent arrival and discharge process (e.g., discharge at specific time of the day, arrival rates vary on different time slots and weekday), a more detailed clustering of patient types (e.g., cluster patients according to the specific symptoms and objective criteria), and the modelling of re-admissions or delayed admissions of rejected or discharged patients. These extensions may lead to a more complex model that cannot be solved to optimality. Thus, approximation schemes such as approximate dynamic programming may be necessary. Another approach may be to consider variations of the setting. For example, our results have shown that in a full ICU, there is usually at least one low-severity patient. If the medical perspective is optimized, there is even at least 3 with a probability of about 85%, while this figure is lower for the monetary perspective. Thus, if inferior beds with a lower level of care for low severity patients are considerably cheaper, a multi-tiered ICU should be considered. Last but not the least, the control policies implemented in the ICU might influence the other departments as well, and the interdependencies of the ICU on the rest of the hospital is an interesting topic to study.

4. Simulation and Evaluation of ICU Management Policies

Co-authors:

Jens O Brunner

Faculty of Business and Economics, University of Augsburg

Steven Gerstmeyr

Faculty of Mathematics and Natural Sciences, University of Augsburg

Published in:

Proceedings of the 2020 Winter Simulation Conference (2020),

DOI: 10.1109/WSC48552.2020.9383958

Abstract:

The intensive care unit is one of the bottleneck resources in the hospital, due to the fact that the demand grows much faster than the capacity. The pressure on intensive care unit managers to use resources efficiently and effectively increases. Therefore, optimal management policies are required. In this work, we evaluate eleven commonly referred policies from the literature and compare their performance by nine key performance indicators in different perspectives, such as utilization, patient health status, and profit of the hospital. The 30 most frequently occurring patient paths, based on the practical dataset of more than 75,000 patient records from a German teaching hospital, are simulated. According to our results, increasing the capacity and treating the patients in well-equipped

intermediate care units performed better in the medical perspective, while the early discharge policy performs well when the capacity is limited. Furthermore, the COVID-19 scenario could be integrated into the model.

4.1. Introduction

Intensive care unit (ICU) is one of the most crucial resources in hospital [3]. It provides constant, close monitoring to the most critical patients by specially trained staff. ICU is different from other regular hospital wards in its above average staff-to-patient ratio and prior access to advanced medical resources and equipment. The ICU is considered among the most expensive medical resources in a hospital, as it generally consumes more than 13% of a hospital's total budget [8]. Additionally, as the aging population is growing rapidly, the demand of ICUs grew 62% faster than the number of available beds [128]. Moreover, ICUs play the bottleneck role in a hospital-wide patient flow [14]. Once the admission of a patient-in-need is restricted by capacity, correlated up- and downstream departments in the hospital will be influenced. Therefore, an efficient capacity management is necessary.

From the economic perspective, a high utilization of the ICU is one of the goals of hospital managers in terms of capacity management. On the other hand, the increasing utilization level leads to a decreasing service quality, to not only the ICU, but also the up- and downstream departments, such as over-beds (a non-staffed bed which is forcefully brought into operation [14]), surgery cancellation or postponement, and demand-driven discharge (premature diverting inpatients out). One challenge to the ICU management is to find an appropriate balance between maintaining high utilization level of ICU beds and providing sound treatment quality as well as an appropriate service level. Furthermore, optimizing the total patient flow adds more difficulty.

Plenty of different ICU capacity management policies are discussed [3]. Early discharging current patients in the ICU to the downstream departments in order to create space for the new patients [2, 4], denying the admission requests from the upstream departments [5], and rescheduling the operations [6, 7] are the most frequently discussed policies to manage the ICU capacity. They are supposed to work well, based on their own key performance indicators (KPIs). However, what are the performances of these policies in the same scenario? Which ones are better when considering different KPIs? How are the upstream and downstream departments influenced by the ICU management policies? Currently, we haven't found a paper to answer these questions.

Therefore, the patient flow centered by the ICU is simulated to evaluate the performances of eleven different management policies. Nine KPIs are compared, covering different aspects, as well as different departments. The simulation model is based on a dataset with 75,934 patient records in the year 2015 from one of the largest teaching hospital in south Germany. The dataset covers 1,215 beds in general wards, 45 emergency beds, 20 IMC beds, and 30 ICU beds. The 30 most commonly occurring patient paths are integrated from the dataset in the simulation model. The results clearly indicate that the introduction of control policies has a positive impact on patient status, length of stay (LOS), and cost when comparing to the baseline case (first-come-first-served policy, FCFS). The parameters of our model can be flexibly adjusted to the parameters from different hospitals. Therefore, it can work as managerial reference to practice. The managers can choose proper policies according to their goals (specific KPIs).

The remainder of the paper is organized as follows: After the literature review, our contribution is summarized in Section 4.2. In Section 4.3, the detailed descriptions of the eleven management policies and the definitions of the KPIs are presented. The simulation model is discussed in Section 0. Section 4.5 follows with an analysis of the results. Finally, in Section 4.6, the work is concluded and the potential future research directions are highlighted.

4.2. Related literature

Simulation is an extremely suitable method for hospital planning, in particular due to the complexity of the real-world situation [35]. There are many papers applying simulation for the optimization of different departments in the healthcare system, for example, operating room scheduling [40, 129, 130], emergency department scheduling [131], and of course ICU planning (as described in the following).

There are a lot of scientific papers with the intention of finding better solutions for ICU capacity planning [3]. Most of the ICU simulation papers try to figure out the ideal number of beds in ICUs [57–60, 72, 80, 81]. Another stream is applying simulation for staff scheduling [61, 87]. In addition to patient care and nurse monitoring, Villa et al. [36] also examined the LOS of patients. All these objects can be improved through better ward designing, capacity planning, and management of patient flows. Furthermore, there are many papers simulating the operating theatres scheduling policies [6, 33, 34, 47, 49, 57, 132]. In these papers, ICUs are not the primary target, but are nevertheless considered in the optimization models or simulations.

Looking at the papers, it is obvious that a gap in comparing different control policies for admission to ICUs exists, especially when considering the complete patient flow centered around the ICU. Mahmoudian-Dehkordi and Sadat [133] present a detailed work on the effects of management guidelines for ICUs. They simulated a disaster with extraordinary emergency patient arrivals. In their paper, they describe eleven different control policies, six of which are ultimately used in the simulation. A simulation with a simultaneous implementation of all six control guidelines was carried out. Kim et al. [71] also examine various admission guidelines in their simulation study. Different policies are always applied in the appropriate cases based on observed measurements. They are able to achieve better results than the clinic's own regulation. However, the simulation model is very limited with the admission and discharge rates of patients. Both are constants in the study. This limits the evaluation of the influencing factors and reduces the meaningfulness of the effects of the control guidelines in different scenarios.

Our work extends the existing research by evaluating different ICU control policies not only inside the ICU, but in the complete patient flow (including emergency department (ED), intermediate care unit (IMC), ward, and operating theatre (OT)). Rather than focusing on one specific type of patients like Mahmoudian-Dehkordi and Sadat [133] did, we simulate 30 different patient paths covering almost all types of patients. Furthermore, medical and monetary KPIs are used to evaluate the performance. On top of that, our model is flexible to adjust the parameters. Therefore, it can be easily applied in different hospital settings.

4.3. ICU Control Policies and KPIs

Patient flows connect the ICU with other departments inside and outside the hospital. All the decision policies implemented in the ICU also influence the up- and downstream departments [3]. Vice versa, control policies applied in the up- and downstream departments can also help to optimize the utilization and service quality in the ICU. For example, IMC is logistically located between the ICU and the general ward. It can work as a physically independent unit or as a dedicated section, incorporated within the ICU [134]. Therefore, increasing the capacity in the IMC is an option to relieve the stress from the ICU. Additionally, rescheduling surgeries in the OT is another frequently applied method.

4.3.1. Control Policies

In order to figure out the best performance using simulation, various control policies are evaluated. When the ICU is fully occupied, the patients will be arranged according to the policies. Any of the policies result in different consequences, such as the patient's health condition, the occupancy at other departments, etc. Most hospitals have no regulated process when it comes to making decisions on patient transfers, since it usually depends on the situation, such as the current utilization of the ICU. Our cooperating hospital has a list of patients who are less critical. They can be early discharged when it is necessary. Sometimes when the ICU is full, the new patients are immediately rejected and transferred to the next station in their paths (after the ICU). We model the FCFS policy as the baseline case. The control policies evaluated in the simulation and the respective implementation procedures are described below.

P1. Increasing ICU capacity: Increasing the capacity is supposed to be a simple and intuitive policy to overcome the capacity shortage. For instance, in the current COVID-19 crises the ICU capacity in Germany has been doubled. Many researchers suggest that increasing the capacity is necessary [135, 136]. However, higher fixed costs for additional beds and instruments will incur and more staff will be needed. In the long run, inefficiency might be another problem [71]. De Bruin et al. [137], however, show that increasing the capacity of expensive beds in ICUs is ultimately more cost-effective than investing in cheaper beds at general wards.

P2. Early discharging the patient with the lowest remaining LOS to the next station in the path: If a patient urgently needs to be admitted into the ICU due to his/her critical condition, another patient in a more stable condition can be discharged from the ICU and transferred to the next department. However, an early discharge is proven to cause a higher readmission rate and to increase mortality [4, 33, 135]. As customary applied in hospitals, the patient with the expected lowest remaining LOS is transferred. This has the advantage that the patient is already in a good condition compared to the other patients in the ICU and therefore the negative consequences are minimized.

P3. Early discharging the patient with the highest remaining LOS to the next station in the path: Complementary to the previously described policy P2, the patient with the highest remaining LOS is discharged prematurely. This policy is inspired by a popular rule in manufacturing, which gives the shortest processing time (i.e., short LOS) highest priority [49]. It reduces occupancy in the long run, as the other patients remaining in the ICU will be regularly discharged soon. However, the fact that the patient would be expected to have a long LOS ahead of him/her suggests that his/her state of health is not even stable at all. This control policy can have serious negative consequences for the early discharged patients.

P4. Early discharging the patient with the highest LOS in the ICU to the next station in the path: This control policy could compensate for disadvantages of the policy P3 where the remaining LOS has been considered. The patient benefits from the fact that he/she has already been treated the longest time in the ICU.

P5. Early discharging the patient with the lowest remaining LOS to IMC: In this policy, if there is no capacity for a new patient left in the ICU, the patient with lowest expected remaining LOS is discharged early to the IMC, regardless of the next station his/her patient path.

P6. Early discharging the patient with the lowest remaining LOS to home: In the special case that all departments are completely busy, a relatively stable patient can be discharged straight back home. A lack of monitoring of the health state can lead to a high possibility of readmission, which is not only undesired to the health of the patient, but also inefficient to the system in a long run.

P7. Boarding in IMC or ward: If the ICU is fully utilized, boarding the new patient in the IMC or the general wards is another option. The patient waits in IMC or ward until a bed in the ICU becomes available. However, depending on the urgency and condition of the patient, this can lead to negative health consequences. This control policy might also lead to queuing in the IMC and general wards. Because the ICU patients are normally not allowed to wait for a long time for the ICU bed, after more than two periods (6 hours per period), he/she will be transferred to another hospital with free capacities.

P8. Increasing IMC capacity: Similar to the first policy P1, the number of beds in the IMCs can also be increased. In this policy the fixed costs, such as personnel and equipment costs, are not as high as in the ICU. Likewise, this rule is only effective in combination with other control policies that involve IMC and general wards.

P9. Rescheduling of planned surgeries: If a patient's surgery is planned to be performed on a certain date, a transfer to the ICU afterwards is also planned in advance. If the utilization level is already high in the ICU, this surgery can be postponed to keep a certain number of buffer beds free for emergencies. Kolker [65] find a reduction in patient deflections from 1.5 percent to 10 percent when up to four operations are postponed on a single day.

P10. Transferring to IMC: Since IMCs are roughly classified between ICUs and general wards in terms of care and costs, moving the new patient to the IMC is supposed to be an acceptable emergency solution in case of a full ICU. In this policy, the patient will receive the entire treatment in the IMC, without the intent of going back to ICU later on.

P11. Transferring to another hospital: Normally transferring the potential ICU patient is not desired [14]. However, when the health status permits a transfer, it might be a better option than waiting for an available ICU bed.

In order to generalize our model and make it more adjustable to different scenarios, some other more complex policies, which are based on the personalized condition of each patient, are not taken into account. For instance, the policy might consider the disease or the required treatment (e.g., cancer, cardiovascular disease, obstetrics etc.).

4.3.2. Performance Indicators

Providing care is very complex and cannot be evaluated by a single indicator (see Table 4.1). The utilization level, the service quality, as well as the economic perspectives are all needed to consider. Therefore, several KPIs are measured in our work. First of all, the utilization rates of the individual departments, which are the most frequently applied metric in practice, are monitored. To prevent hospital employees from being overstaffed or understaffed, patients from waiting, and the beds from idling, the utilization rates should be in an appropriate range. Although we agree that the LOS in the hospital itself has no direct correlation with patient satisfaction [138], a shorter average LOS goes hand in hand with a better efficiency of the respective hospital [139]. Likewise, a longer LOS is associated with a greater likelihood of complications [140]. Therefore, in our model, the average LOSs of the entire hospital stay and of the ICU stay are recorded. For the next indicators we look at patients that are discharged from or even rejected right before the ICU. We want to measure how many patients are influenced by the policies. Therefore, we define two KPIs. The first KPI is the number of patients rejected by the ICU. The number of early discharges is defined as the second KPI. What is naturally a top priority in the healthcare system and needed to be accurately recorded and evaluated is the health condition of the patient. The health state is quantified on a scale from zero to one hundred. It is similar to the Quality-Adjusted Life Year (QALY) approach, which represents values between zero and one [141]. Proper treatments increase the value of health state, and meanwhile delayed treatments lead to lower value. If the score is less than 5 when the patient leaves the hospital, he/she is declared to be dead. The increased number of deaths, which are caused by the management policies, is considered as another evaluation criterion. This number is independent of the mortality that occurs in the hospital anyway (average 2.65% of all discharges). A last key aspect to be evaluated is profit. The hospital can lose revenue if patients are rejected from the ICU or have to leave the hospital to early. In this monetary perspective, profit and revenue loss are considered as separated indicators. However, neither profit nor revenue is known in our practical dataset. By means of appropriate literature, we have obtained the parameters and summarize them in Table 4.2 [3, 142–149].

ID	KPIs	ID	KPIs	ID	KPIs
K1	Mean LOS – system	K4	Rejections	K7	Increased deaths
K2	Mean LOS – ICU	K5	Early discharges	K8	Profit
K3	Mean utilization – ICU	K6	Health	K9	Revenue loss

Table 4.1: Summary of the KPIs.

	Profit (/day)		Revenue loss (Euro)
ICU stay	75	Reject normal patient	6,000
ED stay	15	Reject emergency patient	4,950
IMC stay	38	Early discharge	3,600
Ward stay	13		

Table 4.2: Estimated profit and revenue loss.

4.4. Simulation Model by Anylogic

A discrete event simulation model is developed on the platform AnyLogic, which is a multi-method programming and simulation environment. The details of the input data, simulation model, and case study are discussed in this section. In order to keep the simulation as realistic as possible, but at the same time as generic as possible, distributions of arrival times and LOSs in different units are fitted by RStudio (see Table 4.3). All admissions follow Poisson distributions and the most of the LOSs can be modelled by Lognormal distributions.

Parameter	Admission (per hour)				LOS (day)			
	Weekday daytime	Weekday night	Weekend daytime	Weekend night	ICU	Ward	ED	IMC
Distribution	Poisson	Poisson	Poisson	Poisson	Logn	Weibull	Logn	Logn
Average	15.78	4.41	6.36	4.03	3.23	0.82	1.50	3.74

Table 4.3: Summary of the parameters.

In the dataset, each patient has a personal ID during his/her stay. The detailed information of each event (time stamps, departments transferred from and to) in the hospital stay are recorded. Based on the data, the paths for each patient are created and analyzed. In order to achieve the most precise and meaningful results of various management policies, the structure of the simulation model is kept as realistic as possible, but at the same time as minimalistic as possible. To this end, not all of the patient paths from the dataset are taken into account. The 30 most commonly occurring paths (Figure 4.1), which cover more than 95% of all patients, are simulated in our model.

1	47.62%	Admin	Ward	Discharge					
2	30.89%	Admin	ED	Ward	Discharge				
3	9.95%	Admin	ED	Discharge					
4	2.65%	Admin	ED	ICU	Ward	Discharge			
5	1.93%	Admin	Ward	ICU	Ward	Discharge			
6	1.73%	Admin	ED	Ward	ICU	Ward	Discharge		
7	0.94%	Admin	ED	ICU	Discharge				
8	0.46%	Admin	Ward	ICU	OT	Ward	Discharge		
9	0.36%	Admin	ED	ICU	OT	Ward	Discharge		
10	0.27%	Admin	ICU	Discharge					
11	0.25%	Admin	ICU	Ward	Discharge				
12	0.23%	Admin	Ward	OT	Ward	Discharge			
13	0.23%	Admin	ED	Ward	ED	Ward	Discharge		
14	0.21%	Admin	ED	Ward	ICU	Discharge			
15	0.19%	Admin	ED	OT	Ward	Discharge			
16	0.16%	Admin	ED	ICU	Ward	ICU	Ward	Discharge	
17	0.14%	Admin	ED	Ward	ICU	Ward	ICU	Ward	Discharge
18	0.14%	Admin	Ward	ED	Ward	Discharge			
19	0.12%	Admin	Ward	ICU	Discharge				
20	0.11%	Admin	ED	Ward	ICU	OT	Ward	Discharge	
21	0.11%	Admin	Ward	ICU	Ward	ICU	Ward	Discharge	
22	0.08%	Admin	ED	Ward	OT	Ward	Discharge		
23	0.07%	Admin	ED	ICU	OT	Discharge			
24	0.06%	Admin	ICU	OT	Ward	Discharge			
25	0.04%	Admin	ED	Ward	ICU	Ward	ICU	Discharge	
26	0.04%	Admin	OT	ICU	OT	Ward	Discharge		
27	0.03%	Admin	OT	ICU	Ward	Discharge			
28	0.03%	Admin	OT	Discharge					
29	0.03%	Admin	ED	ICU	Ward	ICU	Discharge		
30	0.02%	Admin	Ward	ICU	Ward	ICU	OT	Ward	Discharge

Figure 4.1: The 30 most frequently occurring patient paths.

To model the patient flows, a recursive network (see Figure 4.2) is set up that uses the patient paths in Figure 4.1. The patient (agent) enters the system with a specified patient type, which is either a normal or an emergency patient. Then based on the empirical probabilities the admission station is determined. The LOS in each department depends on distributions given in Table 4.3. When a patient arrives to a fully occupied ICU, different control guidelines, i.e., the eleven management policies P1 to P11, are applied. When the other units are full, the patient waits for a certain period of time and is then transferred to a downstream department (e.g., from IMC to ward). Finally, the patient leaves the hospital.

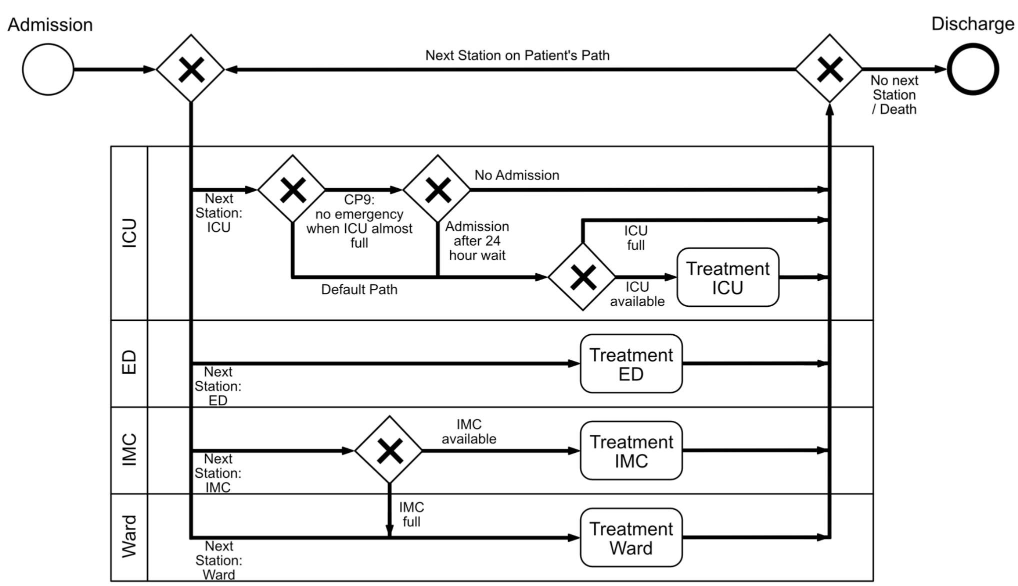


Figure 4.2: The flow chart of the simulation model.

4.5. Result Discussion

Eleven different management policies P1 to P11 are compared and parameter variations are investigated. The FCFS policy, which is used widely in practice, is marked as policy zero (P0). It is used as the baseline case to evaluate the relative performance of the other policies. The simulation period length is set to be 6 months, and 100 replications are implemented for each run.

4.5.1. Results from the Default Settings

In this part, the parameters of patient arrivals and LOSs are kept the same as summarized in Table 4.3. Except P1 (increasing ICU capacity) and P8 (increasing IMC capacity), the capacities in the other policies are constants as well (1,215 beds in general wards, 45 emergency beds, 20 IMC beds, and 30 ICU beds). The ICU and IMC capacities are increased 20% in P1 (36 ICU beds) and P8 (24 IMC beds). The dashboard in Figure 4.3 shows box plots for each KPI and control policy. The performance indicators of the polices are recorded from the simulation and summarized in each histogram. The best management policy of each KPI can be found in the upper right corner. For instance, the shortest system LOS (K1) results from applying P11, and the highest profit is obtained by P10. We assume when the utilization level is lower than 0.9, then the higher the better. The table below shows how often a policy is evaluated as the best one. The same information is given in the frequency chart below for a quick evaluation.

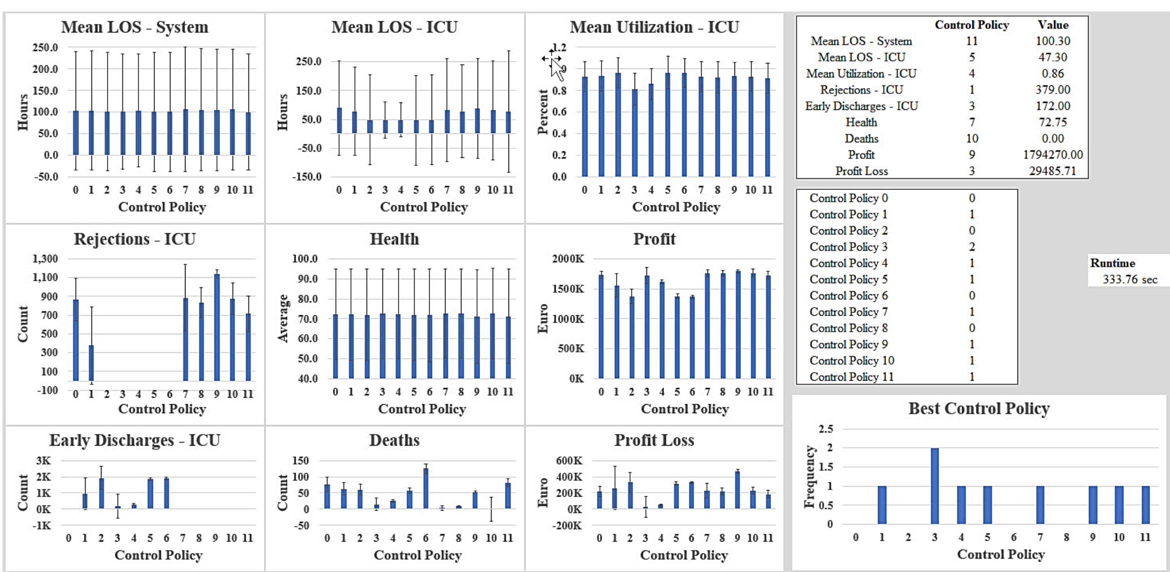


Figure 4.3: Dashboard of the simulation results.

From the histograms in Figure 4.3, we can see that no policy is dominant in all KPIs. From a service quality perspective, the two most important criteria to be evaluated are the average patient health condition (K6) and the increased number of deaths caused (K7). When looking at the increased number of deaths (K7), transferring the patients to IMC (policy P10) is optimal. If the ICU is fully utilized, no additional deaths incur. In this case, instead of waiting and then eventually being rejected, or early discharging other patients, the patients who arrive at a peak time are treated directly to the IMC, which can provide lower level of care. Similarly, boarding and queuing the patients in the IMC (P7 and P8) lead to the best performance in average patient health condition (K6). On the contrary, early discharging the patient with lowest remaining LOS to home (P6) and transferring patients to other hospitals (P11) put the patients in higher risk and result in extra deaths, even more than the baseline case (P0). Rescheduling of planned surgeries (P9) is evaluated to be the worst by looking at the health condition (K6). Let the patients waiting longer for a surgery might not cause additional death, but the health condition of these delayed patients will be deteriorated .

Besides the above discussed criteria, the number of patients influenced by the policies (e.g., K4 and K5) is important. For instance, early discharging the patient with highest remaining LOS (P3) can release more capacity in the long run, and therefore prevents more patients from early discharge (K5). Furthermore, early discharging the patient with the highest LOS (P4), which is widely implemented in practice, results in relative lower number of early discharging as well. Increasing ICU capacity (P1) leads to the lowest rejections (K4). This is naturally due to the increased number of beds in the ICU, which ultimately enables to accept more patients. Three of the remaining early discharge policies (P2, P5, and P6) influenced more than 1,800 patients in our case, which is around 200% of the baseline case (P0). In the monetary perspective, rescheduling surgeries (P9) results in the highest profit (K8) without a large advantage compared to P7, P8 and P10. Early discharging the patient with highest remaining LOS (P3) leads to the lowest revenue loss (K9), which is consistent with K5. As a general trend, profit (K8) is positively correlated with the LOS in the system (K1), while revenue loss is correlated with the number of influenced patients (K4 and K5).

In general, the results indicate that the introduction of management policies has a positive impact on patient status, LOS and/or costs. Since the assessment of the effects of control policies is carried out by comparing simulation results under the same general conditions, and thus comparability is given, valid decisions can be made. Finally, the managers can also make the decisions based on their own priorities and weighting of the KPIs.

4.5.2. Performance under Different Parameters

In order to figure out the influence of parameter variations on the results, different ICU and IMC capacities, as well as different combinations of profit and revenue loss for each management policy are simulated in this section.

4.5.2.1. Policy Performances under Different Capacities

We model the ICU capacity with 80%, 100%, and 120% (thus 24, 30 and 36 beds) and the IMC capacity with 100% and 120% (20 and 24 beds). In most of the cases, when increasing capacities, the total LOS in the hospital and in the ICU both increases. Only by boarding in the IMC or the ward (P7), the total LOS in the hospital decreases, because the patient recovers sooner. By P9, P10, and P11, the patients are delayed or transferred, so that there are sometimes free capacities in the ICU which cause the average LOS in the ICU to decrease. The LOS in the ward changes slightly with the increasing capacities. At the same time, the utilization levels of the ICU and IMC decrease. Having more capacity is beneficial to the service quality and the monetary KPIs. For instance, if the capacities in both ICU and IMC increase, by early discharging the patient with the highest remaining LOS (P3), the total increased deaths can be reduced by 25%, from 96 to 72, meanwhile, the revenue loss can be reduced to 0 from 0.3 million (M).

Consequently, different policies are preferable under different scenarios (Figure 4.4). In general, P3 and P10 perform good in almost all the cases we simulated. Specifically, if the manager focuses on the medical outcomes, transferring the patients to the IMC (P10) is a good choice to implement. Similarly, in case of ICU shortages increasing the capacity in the IMC can improve the performance. Additionally, the differences between increasing IMC beds and increasing ICU beds are not significant, especially when comparing to the increased costs. Therefore, the importance of the IMC should not be underestimated. Besides, in the ICU capacity shortage (decreasing the capacity) scenario, early discharging the patient with the highest LOS (P4) is worth to be considered. When the capacities in both ICU and IMC are increased, then early discharging the patient with the highest remaining LOS (P3) should be preferred. In particular, when the capacity shortages are small, most of the patients can receive proper treatment. Please remember that this policy might lead to severe negative consequences to the few early discharged patients.

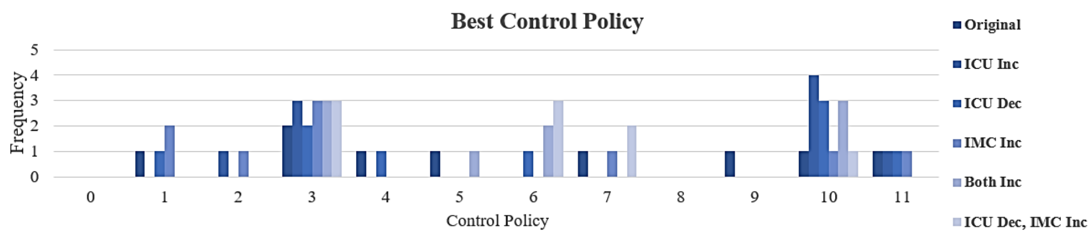


Figure 4.4: Frequency of best policy under different capacity assumptions for ICU and IMC.

4.5.2.2. Policy Performances under Different Monetary Settings

According to section 4.3.2, four parameters of profit (the profits in ICU, ED, IMC, and wards) and three parameters of revenue loss (the cost to reject an normal case, to reject an emergency case, and to early discharge a patient) are used in the monetary perspective. When defining these parameters, we refer to several published papers. Therefore, it's practical to figure out how these parameters influence the performance of the different policies. As a reminder, if parameter variations are implemented for profit (K8) and revenue loss (K9), the best policies under the other KPIs (K1 to K7) are not influenced. In the variation of both K8 and K9, each parameter has two possible values, which are the original one (factor 1.0) and 150% (factor 1.5) of the original value. All possible combinations (2^4=16 combinations for K8 and 2^3=8 combinations for K9) are evaluated. Considering the parameter variation for profits, the policies P1, P3, P7, and P10 are selected to be the best performing ones, i.e., the other policies are never chosen (see Table 4.4). Surprisingly, when increasing the profits of ICU and wards together, the FCFS policy (P0) performs the best. However, P0 is never selected by the other KPIs in any other setting.

Profit settings				Best	Profit settings				Best
ICU	ED	IMC	Ward	Policy	ICU	ED	IMC	Ward	Policy
1	1	1	1	P10	1.5	1	1	1	P10
1	1	1	1.5	P3	1.5	1	1	1.5	P0
1	1	1.5	1	P7	1.5	1	1.5	1	P1
1	1	1.5	1.5	P1	1.5	1	1.5	1.5	P1
1	1.5	1	1	P10	1.5	1.5	1	1	P10
1	1.5	1	1.5	P3	1.5	1.5	1	1.5	P3
1	1.5	1.5	1	P7	1.5	1.5	1.5	1	P3
1	1.5	1.5	1.5	P1	1.5	1.5	1.5	1.5	P3

Table 4.4: Results of policy performance under different parameters of profit.

When comparing different combinations of parameter variations for revenue loss, P3 performs always the best. The total revenue loss depends on the cost for each rejected or early discharged patient and the total number of patients affected. Because of discharging the most serve patients with longest remaining LOS (P3), much more capacities are released. Therefore, the total number of affected patients is less than 50% compared to the other policies. As a result, we see low sensitivity to the parameters' settings. Of course, when the difference between the parameters are much larger, the best policy might change.

4.6. Conclusion

Health care is facing great challenges to make processes more efficient and meanwhile provide better service to patients. The management of the ICU, which is one of the most critical departments in terms of patient status and patient flow, also tries to provide better service and reduce the mortality rate. In particular during COVID-19, the effective and efficient management is of utmost importance. Our simulation model allows a comprehensive evaluation of eleven different management policies for controlling ICU admissions when facing capacity shortages.

In comparison to the baseline case running on a FCFS rule, we show that any management policy is superior regardless of the evaluation criteria. Increasing the capacities of the ICU is obviously beneficial to all the patients but depend on structural circumstances. Generally speaking, it makes sense to discharge patients early when the ICU is fully loaded, i.e., all KPIs indicate an advantage. This ensures that most of the patients receive the appropriate level of care in the ICU. It remains to be decided whether the patient with the highest remaining, the highest elapsed, or the lowest remaining length of stay should be discharged. The simulation results with the data from the University Hospital Augsburg suggest the early discharge of patients with highest remaining LOS, especially in the peak time, when considering the overall impact instead of individual patient's consequences.

Currently, the COVID-19 is spreading all over the world. This disease is severe because of the high infection rate and the speedy spreading. Implementing efficient policies to manage ICU capacities is critical. The COVID-19 scenario can also be simulated by our model. With roughly approximated parameters, it can be shown that at the beginning of pandemic, postponing the scheduled surgery is an option. In the long run, increasing the ICU as well as IMC capacities should be definitely implemented. In extreme cases, early discharging the patients having the lowest survival probability is cruel but efficient.

With the attempt to model the problem as realistic as possible, we rely on the 30 most frequently occurring patient paths based on real data. However, there is still room for improvements. The type of disease, the type of treatment and its impact on the patient are possible variables for a future consideration. Furthermore, the readmission of the patients is not considered in our model which might be another direction to investigate. Additionally, the performance of each single policy is evaluated one by one. The combination of different management policies also offers the potential to make hospital processes even more effective. Last but not the least, to generalize the findings of our model, an application to other hospital settings might be necessary.

5. Conclusions and Outlook

5.1. Conclusions

This dissertation presents three articles on optimizing ICU management decisions. Each of these article focus on a different level of the research problem. Starting from a structural literature review around all the ICU management topics, we understand clearly the current research trends and gaps. Afterwards, a tactical-level research question on optimizing the ICU decision is discussed, focusing only on the ICU department. The third article shows a bigger picture. The down- and upstream departments in the hospital patient flow are all included in the discussion. Different methodologies are applied in different articles, and case studies with real-world datasets are implemented in the last two articles.

The first article provides the first structured and comprehensive review of ICU problems in OR/MS. The relevant papers covering the topics "importance of the ICU for hospital patient flow" and "the ICU management problems" are discussed based on a new framework. Two dimensions are included in the framework: the time horizon of decisions (*long-term decision, medium-term decision, and short-term decision*) and the addressed research topics (*patient flow optimization, bed capacity management, and personnel planning*). Furthermore, the modeling methods and solution approaches are classified and discussed in detail. Based on the analysis of existing papers, the future research topics are also addressed along three streams.

According to the research gaps discussed in the first article, the second one proposes a method to define admission and early discharge policies that minimize these negative consequences. A discrete-time MDP is applied and solved to optimality for realistic instances. By minimizing the medical consequences, the approach demonstrated to significantly outperform a myopic policy used by most hospitals in practice. Besides, it is found that different objectives lead to different policies. An efficiency frontier covering medical and monetary perspectives is developed, contributing to the ongoing discussion on the trade-off between medical quality and monetary costs. Robustness checks and situations in medical perspective where our approach is sensitive to cost changes and cost estimation errors are provided. Various applications of our approach exist. The major one is a framework to develop recommendations for admission and discharge control on a tactical decision level. An additional application is to use the approach as decision support for capacity dimensioning on strategic and operational decision levels. It provides insights into the consequences of capacity shortages, and allows decision-makers to consider different objectives within the admission and discharge policies.

The third article provides a broader perspective on the hospital patient flow instead of the single ICU department. Different management policies are compared based on different KPIs by a simulation study. In comparison to the baseline case running on an FCFS rule, it is shown that any management policy is superior regardless of the evaluation criteria. The 30 most frequently occurring patient paths, based on the practical dataset of more than 75,000 patient records from a German teaching hospital, are simulated.

5.2. Future research directions

We believe that managing ICU admissions and discharges are of great importance and have considerable potential for future research. From a modeling point of view, adding a higher level of complexity, generality, and flexibility could be of interest. Thus, developing and implementing a widely applicable decision support framework, which integrats the prediction and optimization together to make the ICU decisions, would be one of the most valuable research directions, especially with the support of big data techniques. This section tries to demonstrate several possible research ideas following this direction. In section 5.2.1, machine learning (supervised learning) methods are discussed to predict riskiness based on personalized data (patient records, diagnosis, arrival time, comorbidities, etc.). This approach does not require predetermined groups of patients with associated risk scores and length of stay forecasts. In section 5.2.2, a deep reinforcement learning (DRL) approach is proposed for finding an optimal solution based on the prediction. This approach is supposed to be a flexible modeling method with complex assumptions, such as non-homogenous arrivals and discharges, an infinite planning horizon, and different hospital settings and requirements. It could make operational level capacity management decisions, as well as personalized level decisions. In section 5.2.3 further research directions are presented.

5.2.1. Data-driven Cost Predictions

With the digitized data acquisition progress in the last decades, machine learning (ML) techniques gradually changed the practice in the healthcare area [150]. Because of the advantages of continuous accuracy improvement, up-to-date information acquisition, diagnostic and therapeutic error prevention, currently, ML is developed to solve various healthcare or medical problems. In general, it is frequently applied to analyze the structured and unstructured data with classical support machine and neural network, as well as modern deep learning approaches [151]. Specifically, the various application area of ML in healthcare covers patient engagement and adherence, administrative activities, the diagnosis and detection, treatment recommendations, outcome prediction, and prognosis evaluation of many different types of disease, such as cancer, neurology, cardiology, and so on.

The applications of ML in the ICU management area are diverse. Parts of the ML papers focus on medical treatment and diagnostics [152–155]. Several articles [156–159] discuss the prediction of sepsis in the ICU with different approaches. Groups of researchers implement supervised learning or deep learning algorithms to predict different disease outcomes [61, 119, 160–164]. At the same time, some other papers focus on the prediction of important performance indicators in the ICU, for example, mortality rate [158, 165–168], length of stay[169], and readmission risk [170]. Two papers are relevant to our topic, which focus on the admission and discharge decision-making process. Krämer et al. (2019) apply supervised learning to predict urgency, which can help diagnose whether a patient should be admitted to emergency or elective care. By this method, inappropriate admissions can be significantly reduced. McWilliams et al. (2019) present a random forest and a logistic classifier to detect the patients who are ready for discharge from the ICU.

According to the current research trends, the ML algorithms could also be applied in our topic to predict the individual cost of different policies. As discussed in Chapter 3, the objective of MDP model is to minimize the negative consequences of the capacity shortages in the ICU, which depends on the cost of rejection c_m^{rej} and early discharge c_n^{edis}. However, there is no explicit definition of the cost. It could be either the increased mortality rates in the medical perspective, or the lost revenue in the monetary perspective [173]. Similar to other researchers [2, 174, 175], we can propose a particular cost setting to balance the tradeoff between individual-level cost (mortality risk) and system-level cost (LOS), as well as the myopic cost (congestion) and the future cost (readmission risk).

$$c_i = f(p_i^{mor}, p_i^{cap}, l_i)$$

where i is the index of patient type or individual patient, p_i^{mor} is the predicted mortality risk, p_i^{cap} is the possibility that a bed is predicted to be occupied in the future, and l_i estimates the LOS.

The first part p_i^{mor} indicates the mortality risk resulting from the action. It can be estimated in real-time based on the patient's records (medical indicators and patient information) when making a decision. Therefore, we can say that it takes the patient's evolution into consideration. As keeping a patient one extra day can reduce mortality risk by nearly 6% [176], along with other parameters, the mortality risk is also a function of the spent LOS (which equals 0 when calculating p_i^{mor} for new arrival patients). p_i^{cap} and l_i are metrics indicating the system congestion. For instance, if a patient suffers a pretty high mortality risk and will occupy the ICU bed for a relatively short time, she could be admitted without hesitation. However, if the high risky patient stays in the ICU for a long time (e.g., longer than one month), while another lower risky patient is expected to leave sooner, it might be challenging to decide which one to be admitted. Therefore, we take the possible future ICU bed occupancy (the LOS l_i) into the cost. p_i^{cap} indicates the possibility that the capacity will be occupied. Specifically, regarding the new arrival patient, $p_i^{cap} = 1$ shows this patient will definitely use the capacity if we admit her. When early discharging a patient, p_i^{cap} represents the total expected readmission rate. All these three parameters for each individual patient can be predicted by classical ML algorithms in real-time, using the patient records as the features.

5.2.2. Deep Reinforcement Learning Supported Decision Making

Besides the classical supervised and unsupervised learning in data analysis, reinforcement learning (RL) is another appealing subfield of ML, empowering the system's sequential behavioral decision-making. Currently, RL is maturely used in many practical areas, such as playing games, autonomous driving, and controlling a robot. As having the features of making decisions with sampled, evaluative and delayed feedback simultaneously, RL is a suitable technique for finding efficient policies in the healthcare area, such as making diagnostic and treatment decisions [173]. Asoh et al. [174] model the medical records of diabetes with inverse RL, which can simulate each patient's future and therefore evaluate each treatment. Zhao et al. [175] present an RL approach to find the optimal individualized treatment for lung cancer patients. Gaweda et al. [176] also work on the individualization of drug administration in the treatment of renal anemia by an RL-based approach. Meanwhile, RL algorithms are used in sepsis treatment policies [177, 178]. Furthermore, many other works regarding RL applications in healthcare are structured summarized by Yu et al. [121].

DRL algorithms fit pretty well with our ICU admission decision-making problem. It can provide more flexibility than our MDP model presented in Chapter 3. For instance, the length of each period t can be determined based on the hospitals' requirements, which could be 24 hours, 12 hours, 1 hour, and even continuous decision making. It means that the ICU managers could make the admission and early discharge plans either for the next period at the beginning of each day or each shift, or each time when a new patient arrives. To generalize the model for different hospital settings, DRL can release constraints of patient grouping methods. The arrival patients could be classified into M types. They can be grouped according to the level of emergency (scheduled or emergency) [179], the source of arrivals (different apartments), the health conditions (DRG), or other reasonable classification methods. Similarly, the admitted patients can be classified into N groups. We assume the properties of the patients in the same group are identical. In the extreme case, each patient could be treated as a special group. Figure 5.1 illustrates the possible assumptions with DRL comparing to Figure 3.2.

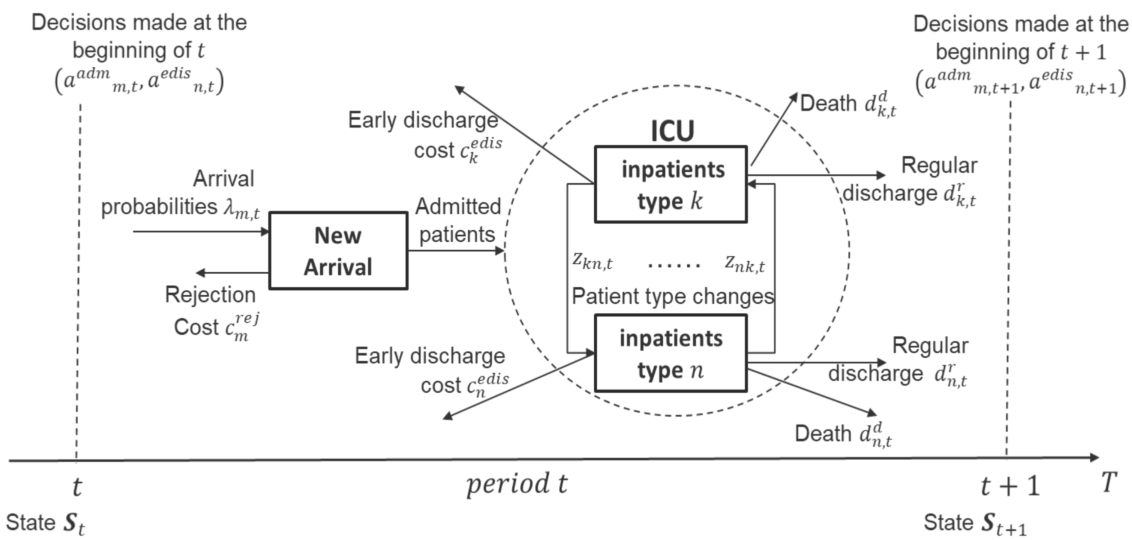


Figure 5.1 The ICU decision process by DRL

However, some difficulties should be considered when implementing DRL. The biggest challenge could be overfitting, which is a common issue in machine learning, especially DRL. It happens when the dataset is not large enough, and the features are not sufficient. Fixing overfitting and finding a good set of hyper parameters are always essential topics in DRL.

5.2.3. Other Research Directions

Still, there are several potential points for future research. Regarding modeling, multiple departments management is an exciting topic to discuss. It is normal to have many different ICUs in one hospital. Optimal allocation and reallocation of patients between different ICUs might reduce the negative effects rather than directly rejecting or early discharging patients. Regarding the DRL algorithms, it is valuable to reevaluate with additional datasets to overcome the overfitting problem and to verify the flexibility and applicability in different hospital settings. Because of the dynamic stochastic features of the ICU management problem, a robust decision-making method is also worthy to study. Last but not least, a decision tool that practitioners could easily implement will definitely be a big step to go.

Appendix A

Table A1 Framework for ICU management

	Patient flow in ICU			Bed capacity management	Personnel planning
	Admission	Treatment/LOS	Discharge & Readmission		
Long-term				[13, 14, 28–30, 48, 57–60, 66, 67, 69, 79–82]	[10, 61]
Medium-term	[49, 62–65, 68, 71, 72, 74–78, 90, 180]	[47, 51, 53, 55, 56, 68, 78, 91, 92, 180]	[2, 15, 17, 64, 70, 88, 92, 180]	[14, 21, 54, 83]	[87]
Short-term			[50]		[52]

Table A2 Modeling methods of uncertainties

Publication	Type of ICUs	Country	Patient grouping methods	Arrival pattern	LOS
Adeyemi [55]	Neonatal ICU	UK	3 groups: ICU, HDU, or SCU	-	Exponentially distribution
Akkerman & Knip [47]	Cardiac ICU	The Netherlands	2 groups: outpatients and inpatients	-	Empirical distribution
Asaduzzaman et al. [54]	Neonatal ICU	UK	3 groups: ICU, HDU, or SCU	Poisson	Exponentially distribution
Barado et al. [59]	Multidisciplinary ICU	Spain	8 groups: elective surgery, emergency department, surgical wards, medical wards, emergency	Elective patients and other patients: empirical discrete probability distributions Other groups: Poisson distribution	For each group: non-normal regression model, based on the Lognormal and Log logistic distribution families

			surgery, trauma without surgery, trauma with surgery, other		
Cahill & Render [60]	Medical/Cardiac ICU	US	2 groups: outpatients and inpatients	Exponential distribution	Triangle distribution
Chan & Yom-Tov [88]	Not specified	US	2 groups: critical and non-critical	-	
Chan et al. [2]	Not specified	US	m groups: based on the health condition	At most one new patient arrive in each time slot with probability λ (known prior to the optimal discharge policy)	Memoryless initial LOS: geometric distribution From empirical data: lognormal distribution Extension: phase type LOS distribution
Cochran & Bharti [20]	Neonatal ICU	US	m groups: classified based on department	-	Exponential
Cochran & Bharti [30]	Neonatal ICU	US	m groups: classified based on department	-	Exponential
Cochran & Roche [28]	Neonatal ICU	US	m groups: classified based on department	-	Exponential
Costa et al. [81])	Not specified	UK	2 basic groups: emergency and elective The emergency patients are divided into 6 sub-groups according to patient's source [i.e. operating theatres, general ward, specialist areas (radiology, delivery suite) and other hospital]	Different for each group	Different for each group, e.g., lognormal distribution for patients from HDC
Demir et al. [56]	Neonatal ICU	UK	-	-	Phase-type
Dobson et al. [17]	Not specified	US	2 groups according to arrival pattern: scheduled and unscheduled 2 groups according to LOS: short and long	Weekly periodic arrival pattern	Geometric distribution
Duraiswamy et al. [87]	Medical ICU	US	-	-	Common distribution
Green [13]	Not specified	US	-	Poisson distribution	Exponential distribution
Griffiths et al. [10]	Not specified	UK	2 groups: emergency and elective	Arrival rate has weekly and hourly trends, thus inter arrival times within each hour are assumed to follow negative exponential distribution, with mean time depending on the particular hour of the day.	For elective patient, no suitable distribution was found.

Griffiths et al. [78]	Not specified	UK	2 or more groups: according to LOS	Negative exponential distribution for inter-arrival time	Two (or more) negative exponential phases in series
Hashimoto et al. [61]	Medical /cardiac ICU	US	-	-	-
Kapadia et al. [51]	Pediatric ICU	US	4 groups: based on the severity of illness	-	Using Markov chain approaches to model the LOS
KC & Terwiesch [15]	Cardiac ICU	US	-	-	Weibull distribution Exponential distribution
Kim et al. [64]	Multidisciplinary ICU	China	4 groups: ward, accident & emergency, OT emergency, OT elective	Convolution of Poisson processes for the first three groups, resulting in a Poisson process	Different for each group, LOS assumed as the service rate
Kim et al. [62]	Multidisciplinary ICU	China	4 groups: ward, accident & emergency, OT emergency, OT elective	-	-
Kim & Horowitz [63]	Multidisciplinary ICU	China	4 groups: ward, accident & emergency, OT emergency, OT elective	First three patient types (ward, A&E, OT emergency): Poisson arrival; Elective only arrive on weekdays	First three patient types (ward, A&E, OT emergency): exponential distribution; Elective: tripartite Gamma density
Kim et al. [71]	15 different ICUs	US	1 group is considered: the patients from ED	Constant arrival and departure rate	Exponential distribution
Kolker [65]	Multidisciplinary ICU	US	4 groups: ED, other hospitals, inpatient units, and OR (3 subgroups: emergency, add-on and elective surgeries)	Practical: non-ordinary distribution with after effect Analytical: Poisson distribution	-
Kortbeek & Van Dijk [74]	Not specified	The Netherlands	2 groups: OT patient (patients visit ICU after having undergone surgeries), direct patients (enter ICU directly without surgeries)	2 independent Poisson distributions	Surgery time: exponential distribution LOS in the ICU: non exponential
Lamiell	Not specified	US	-	-	-
Rokni Lamooki et al. [180]	Not specified	Iran	No relevant uncertainty modeling	-	-
Hagen et al. [68]	5 different ICUs	US	2 arrival groups: scheduled and emergency 9 groups categorized by severity and LOS	Exponential distribution	-
Li et al. [76]	Not specified	China	2 groups: more critical patients who cannot be prematurely discharged, and patients who can be prematurely discharged	Poisson distribution	-

Litvak et al. [14]	Not specified	The Netherlands	3 groups: elective, internal trauma, regional trauma	Poisson distribution, different arrival rate of different hospitals	Exponentially distribution
Lowery [69]	Surgical/medical/ Coronary ICUs	US	3 groups: scheduled arrivals through operating room; emergency through operating room; emergency direct admissions	Scheduled arrivals: on appropriate day of the week, at the appropriate time and in appropriate operating rooms Emergency arrivals: a distribution based on historical data	A distribution based on historical data
Lowery & Arbor [70]	Surgical/medical/ Coronary ICUs	US	m groups: patients in different ICUs	Exponential distribution	Lognormal distribution
Masterson et al. [67]	Multidisciplinary ICU	US	4 groups: patients in different ICUs (medical, surgical, cardiac, trauma)	Non-stationary Poisson distribution	Lognormal distribution
McManus et al. [66]	Medical/Surgical	US	2 groups: scheduled, unscheduled	Poisson distribution	Exponential distribution
Mullinax & Lawley [52]	Neonatal ICU	US	No relevant uncertainty modeling	-	-
Nathanson et al. [53]	Neuro trauma ICU	US	No relevant uncertainty modeling	-	-
Ridge et al. [72]	Not specified	UK	2 groups: emergency, planned	-	Negative exponential distribution or Weibull distribution
Roumani et al. [92]	Not specified	US	No relevant uncertainty modeling	-	-
Shahani et al. [83]	Not specified	UK	-	-	Lognormal distribution
Shmueli et al. [77]	Not specified	Israel	-	Poisson distribution	Exponential distribution
Shmueli et al. [90]	Not specified	Israel	-	Poisson distribution	Exponential distribution
Shmueli & Sprung [91]	Not specified	Israel	-	Poisson distribution	Exponential distribution
Terwiesch et al. [79]	Not specified	US	No relevant uncertainty modeling	-	-
Tierney & Conroy [21]	Not specified	Australia	No relevant uncertainty modeling	-	-
Troy & Rosenberg [57]	Surgical ICU	Canada	No relevant uncertainty modeling	-	-
van Dijk & Kortbeek [75]	Not specified	The Netherlands	2 groups: OT patient (patients visit ICU after having undergone surgeries), direct patients (enter ICU directly without surgeries)	2 independent Poisson distributions	Surgery time: exponential distribution LOS in the ICU: non exponential
Wharton [50]	Coronary ICU	UK	m groups: based on different risk levels	Random or Poisson distribution	Exponential distribution
Williams et al. [48]	Cardiac ICU	US	-	-	-
Yang et al. [49]	Cardiothoracic Surgical ICU	US	-	-	-
Yergens et al. [82]	Not specified	Canada	-	-	-
Zilm & Hollis [58]	Surgical	US	-	-	-

Table A3 Modeling methods and solution approaches

		Modeling methods					
		Stochastic methods			Deterministic methods		Other
		Queueing	MDP (Markov chain)	Stochastic process analysis	Mathematical programming	Statistical analysis	Literature review
Solution approaches	Exact solution	[13, 50, 54, 74, 75, 77, 78, 88]	[51, 56]		[53, 180]	[15, 55, 56, 62, 90–92]	
	Heuristics		[2, 17, 76]		[52]		
	Simulation based solution	[14, 28–30, 49, 59, 63, 64, 66, 68, 72, 81]	[47]	[10, 48, 56–58, 60, 61, 65, 67, 69, 70, 80, 82, 83, 87]	[71]		
	Other						[21, 79]

Appendix B

Appendix B.1: Solution Approach

The standard approach to compute a Markov decision process is to simply solve the recursive value function backwards in time. That is, starting with T, $V_t(\boldsymbol{S}_t)$ is solved for all $\boldsymbol{S}_t$ using the previously calculated values of $V_{t+1}(\boldsymbol{S}_{t+1})$. This process may suffer from 3 curses of dimensionality (see, for example, Powell [181] Chapter 1): the size of the state space, the size of the action space, and the computation of the expectation.

- The size of the state space is $(\frac{(B+1)(B+2)}{2}) \cdot 4$ and, thus, grows quadratic in B. This is no issue, given that real-world ICUs are between 10 and 50 beds in size.
- Likewise, the action space with its three binary dimensions is limited to $2^3 = 8$ actions.
- The calculation of the expectation is a bit more complicated, as it depends on ICU occupancy. The number of possible state transitions (possible realizations of $\boldsymbol{\omega}_{t+1}$) is bounded by $4 \cdot \frac{B^2}{2} \cdot \frac{B^2}{2} \cdot 2 = 2B^4$. Note that the number of state transitions can be considerably reduced by neglecting transitions with very low probabilities (like 10 patients out of 10 improving from high-severity to low-severity in one period).

More specifically, a pseudo code for the standard backward dynamic programming algorithm is as follows:

Step 0. Initialize the terminal contributions $V_{T+1}(\boldsymbol{S}_{t+1})\ \forall\ S_{t+1}$, set $t = T$

Step 1. Calculate $V_t(\boldsymbol{S}_t) = \min_{\boldsymbol{a}_t}\{C(\boldsymbol{S}_t, \boldsymbol{a}_t) + \mathbb{E}_{\boldsymbol{\omega}_t} V_{t+1}\big(\boldsymbol{S}_{t+1}(S_t, \boldsymbol{a}_t, \boldsymbol{\omega}_{t+1})\big)\}\ \forall\ \boldsymbol{S}_t$.

Step 2. If $t > 0$, decrement t and go to step 1. Else stop.

Appendix B.2: Estimation of cost parameters

B.2.1. Medical perspective

From a medical point of view, we chose the absolute increase of mortality rates as the single performance indicator and cost component. Please note that the following numbers are percentage points (pp), that is, differences of percentages. If, for example, due to an action the mortality rate of a patient increases from 10% to 11%, we denote the increase as 1 pp. In this paragraph, we discuss changes in mortality rates due to the five possible actions mentioned above and state the assumed parameter values for our case study. After obtaining ranges from the literature, we discussed the concrete cost parameter with the ICU manager of the case hospital, and selected the values that fit best for our case hospital. However, those parameters contain some uncertainty. Thus, we discuss the impact of different parameter choices and of estimation errors of those parameters in Section 6, a more detailed sensitivity analysis can be found in the online appendix.

- $c_{1,med}^{rej} = 1$ pp. Rejection of an *elective surgery patient* typically results in rescheduling, that is, delaying the surgery, or scheduling it at another hospital. There is little literature on the medical consequences of cancelling surgeries in general. However, for certain orthopedic surgeries, there is data on mortality rates available. A widely investigated example is hip fracture surgeries, where some studies find no significant effect of delays on mortality rates [182], while others (for example, Shiga et al. [183]) detect systematic increases of mortality rates. Shiga et al. [183] performed a meta-analysis on hip fracture reports. They report an average short-term mortality rate of 7% for non-delayed surgeries, which increases to 10% for surgeries being delayed. Nyholm et al. [184] show that delaying surgeries of proximal femoral fracture leads to increases of mortality rates between 1 and 4 pp. Thus, the increase of mortality rate due to delays reported in the literature covers a range between 0 and 4 pp. Cancer surgeries with a high risk of complications, such as esophagectomy, whipple procedure, or cystectomy are changing rapidly due to new surgical approaches, improved surgical training, and oncological supportive treatment (Sabra et al. [185], Yibulayin et al. [186]). Therefore, we decided to use data of an index surgical procedure with stable mortality and ICU admission rates, the hip and the proximal femoral fracture, of which the surgical technique and the affected population did not change as much over the last decade. While orthopedic procedures that are often performed with old patients rather provide an upper bound of negative consequences of cancelling scheduled surgeries, they do account for a meaningful share of cases that are relevant to ICU management. In our case study, we assume an absolute increase of the mortality rate of 1 pp due to rejection of an elective surgery patient.

- $c_{2,med}^{rej} = 15$ pp. *Internal emergencies* cover patients who are treated within the hospital when their medical condition unexpectedly deteriorates. They could be located at a regular ward, an operating theater, or an emergency department. Rejecting these patients at the ICU typically means they need to be treated within a regular ward using extra nursing capacities. The studies of Kim et al. [62] and Kime et al. [64] claim that rejecting these patients increases mortality rates by more than 20 pp from around 30% to more than 50%. Checkley [187] and Iapichino et al. [188] report that internal emergency patients whose admission is initially denied show mortality rates that are 10 pp above those who are directly admitted (Checkley [187] reports 27% mortality for admitted versus 37% for denied admission, Iapichino et al. [188] 28% versus 39%). In our case study, we assume an absolute increase of the mortality rate of 15 pp due to rejection of an internal emergency patient.
- $c_{3,med}^{rej} = 3$ pp. Rejecting *external emergency* patients typically means informing a central coordination center that no emergency patients can be treated, so that emergency ambulances will be directly diverted to other hospitals. Emergency ambulances which have already arrived at the hospital might be sent away. In all cases, external patients will experience a delay in their treatment, which leads to an increase of mortality rates. Chalfin et al. [104] report mortality increases of 2 to 5 pp due to delayed ICU admissions (mortality in the ICU from 8.4% to 10.7%, mortality during the total stay from 12.9% to 17.4%), while Singer et al. [189] find that increased boarding times at emergency departments, which could be caused by ICU rejections, lead to an increase in mortality rates of around 2 pp (mortality over all patients from 2.5% to 4.5%). In our case study, we assume an increase of the mortality rate of 3 pp due to rejection of an external emergency patient.
- $c_{1,med}^{edis} = 2$ pp. Chrusch et al. [109] assume that a congested ICU provokes *early discharges of low-severity patients*. They find a higher level of re-admissions (around 4 pp higher) and higher mortality rate for re-admitted patients (21.3% against 0.3% for patients in the wards who were not re-admitted). Thus, the increase of mortality of the least critically ill patients due to readmissions was close to 1 percentage point. Furthermore, the ICU mortality for readmitted patients is slightly higher than the one for a primary stay (21.3% against 19.0%). In our case study, we assume an absolute increase of the mortality rate of 2 pp due to the early discharge of a low-severity patient.

- $c_{2,med}^{edis} = 10$ pp. There are few studies analyzing the effects of *early discharges of high-severity patients*. Some studies consider the mortality of high-severe versus low-severe patients: Smith et al. [190] analyze the effects on mortality based on the health status at the time of discharge. They show that patients discharged with a high criticality index exhibit mortality rates that are about 18 pp higher compared to patients discharged with a low criticality index (21.4% compared to 3.7%). Daly et al. [135] show that discharge of patients with high-severity leads to mortality rates of 25% compared to 4% for less risky patients. For high-severity patients, these mortality rates strongly exceed those of staying within the ICU. Chan et al. [2], for example, note that within the ICU, the difference between high-severity patients and low-severity patients is around 10 pp (14.6% compared to 4.2%). Obviously, these figures cannot be matched precisely, as not all high-severity patients will become low-severity patients if they are not early discharged. In our case study, we assign an absolute increase of the mortality rate of 10 pp due to the early discharge of a high-severity patient.

B.2.2 Monetary perspective

Rising cost pressure on hospitals increases the importance of a monetary perspective. ICUs are typically not profit centers, but decisions taken in the ICU might largely impact a hospital's profitability. From a monetary perspective, we consider the costs in our model to be the profit loss for the hospital due to rejections or early discharges. Obviously, these numbers heavily depend on the reimbursement model of the relevant health care system. As previously stated, we use data from the German DRG (Diagnosis Related Group) system as of 2017 [191]. Since most hospital costs are largely fixed costs, we consider the lost revenues as a proxy for lost profits, and only add additional costs if appropriate. Thus, our values can be seen as upper bounds for lost profits. As in the medical perspective, these parameters need to be adapted for each specific hospital. While the exact numbers vary a lot, we believe that the relation between those values are similar among various hospitals and health systems. In this paragraph, we explain the logic we applied to obtain our cost parameters for the five possible actions. Please note that for clarity and to avoid pseudo-accuracy, we round all values to the nearest 100 €.

- $c_{1,mon}^{rej} = 9{,}200$ €. If an *elective surgery patient* is rejected, the surgery will be cancelled or rescheduled. Thus, the hospital might either lose the profit for this patient (if rescheduled at another hospital), or for a similar patient (if rescheduled at the same hospital). The latter is because the operating theatre is typically the main bottleneck, and using another surgery slot results in scheduling one patient less. Thus, rejecting a planned surgery patient results in losing the average reimbursement of one patient with treatments in both the operating room and the ICU. According to the German DRG system, this value is around 9,200 €.

- $c_{2,mon}^{rej} = 5{,}800$ €. Rejecting an *internal emergency patient* has several monetary implications. First, if the patient is not treated at an ICU, no extra charges for ICU treatment can be billed. Second, additional nursing capacities need to be booked in order to secure adequate treatment on a regular ward. Third, additional costs such as legal costs in case of negative incidents may occur. We neglect the latter because they are often covered by insurances. Based on the German reimbursement system and nursing costs, the opportunity costs for the first component is around 1,050 €, and the extra nursing for the expected length of stay (7.3 days) amounts to 4,720 €. Thus, the total value is around 5,800 €.
- $c_{3,mon}^{rej} = 4{,}100$ €. Rejecting an *external emergency patient* typically leads to diverting the patient to another hospital. Thus, the revenue for this patient is lost and we set the cost parameter to the average reimbursement of patients with any ICU treatment, leading to a value of 4,100 €.
- $c_{1,mon}^{edis} = 700$ €. The costs for *early discharging a low-severity patient* are difficult to estimate. In our hospital, most early discharged patients require some additional supervision from a nurse. Thus, we consider the costs of about one day (half of the expected length of stay of a low-severity patient) of extra nursing, resulting in a cost parameter of 700 €.
- $c_{2,mon}^{edis} = 6{,}500$ €. The cost for *early discharging high-severity patients* are computed as follows: Similar to the rejection of internal emergency patients, extra nursing has to be paid for during the expected remaining length of stay in the ICU of the patient. Due to the considerably longer period of extra care (approximately 9.5 days, half of the expected length of stay of a high-severity patient) compared to low-severity patients, this cost parameter adds up to 6,500€. Again, we do not consider legal costs.

Appendix B.3: Effects of Changing Number of Beds (Section 3.6.1)

Objective optimized	Medical				Monetary			
Approach	MDP		Myopic		MDP		Myopic	
Costs eval. # Beds	Medical	Monetary	Medical	Monetary	Medical	Monetary	Medical	Monetary
30	2,630	8,479,086	3,393	5,596,936	3,983	1,727,001	4,674	1,932,752
31	2,489	8,243,499	3,189	5,324,028	3,748	1,599,217	4,329	1,769,349

32	2,339	7,994,341	2,991	5,057,786	3,506	1,470,294	4,018	1,624,208
33	2,191	7,703,195	2,803	4,785,807	3,297	1,363,047	3,730	1,490,802
34	2,068	7,444,602	2,614	4,501,967	3,061	1,244,085	3,444	1,361,448
35*	1,931	7,160,950	2,436	4,245,759	2,855	1,143,772	3,172	1,239,946
36	1,805	6,859,778	2,259	3,964,459	2,650	1,046,240	2,900	1,121,906
37	1,672	6,490,779	2,108	3,723,968	2,453	954,235	2,637	1,007,819
38	1,554	6,124,937	1,935	3,451,939	2,259	869,142	2,413	912,560
39	1,436	5,824,626	1,760	3,161,005	2,067	785,368	2,188	818,636
40	1,324	5,456,733	1,601	2,895,646	1,877	704,123	1,978	734,045

Table B1: Results for different ICU sizes (* is the base case)

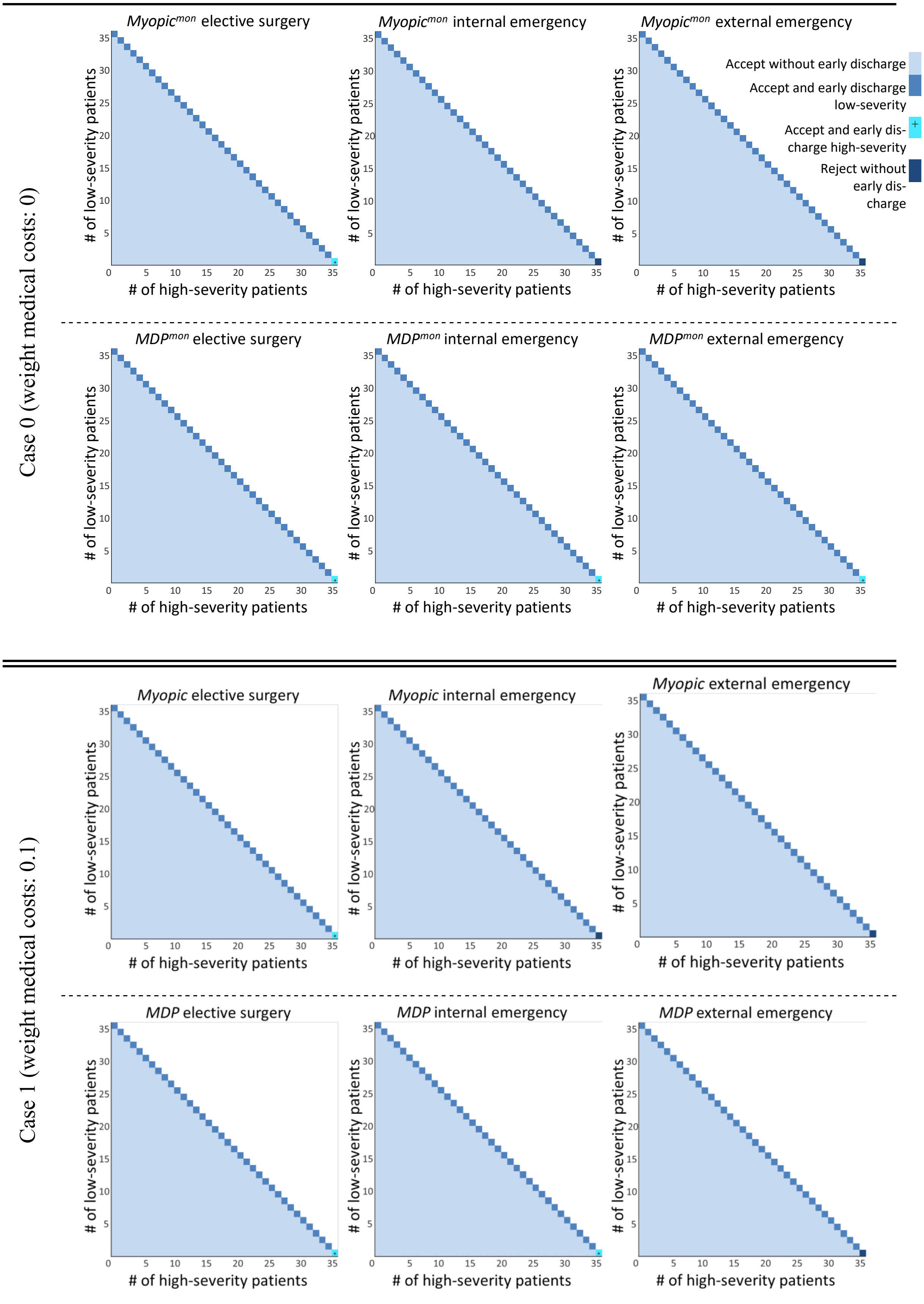
Appendix B.4: Trade-off between Medical and Monetary Costs (Section 3.6.2)

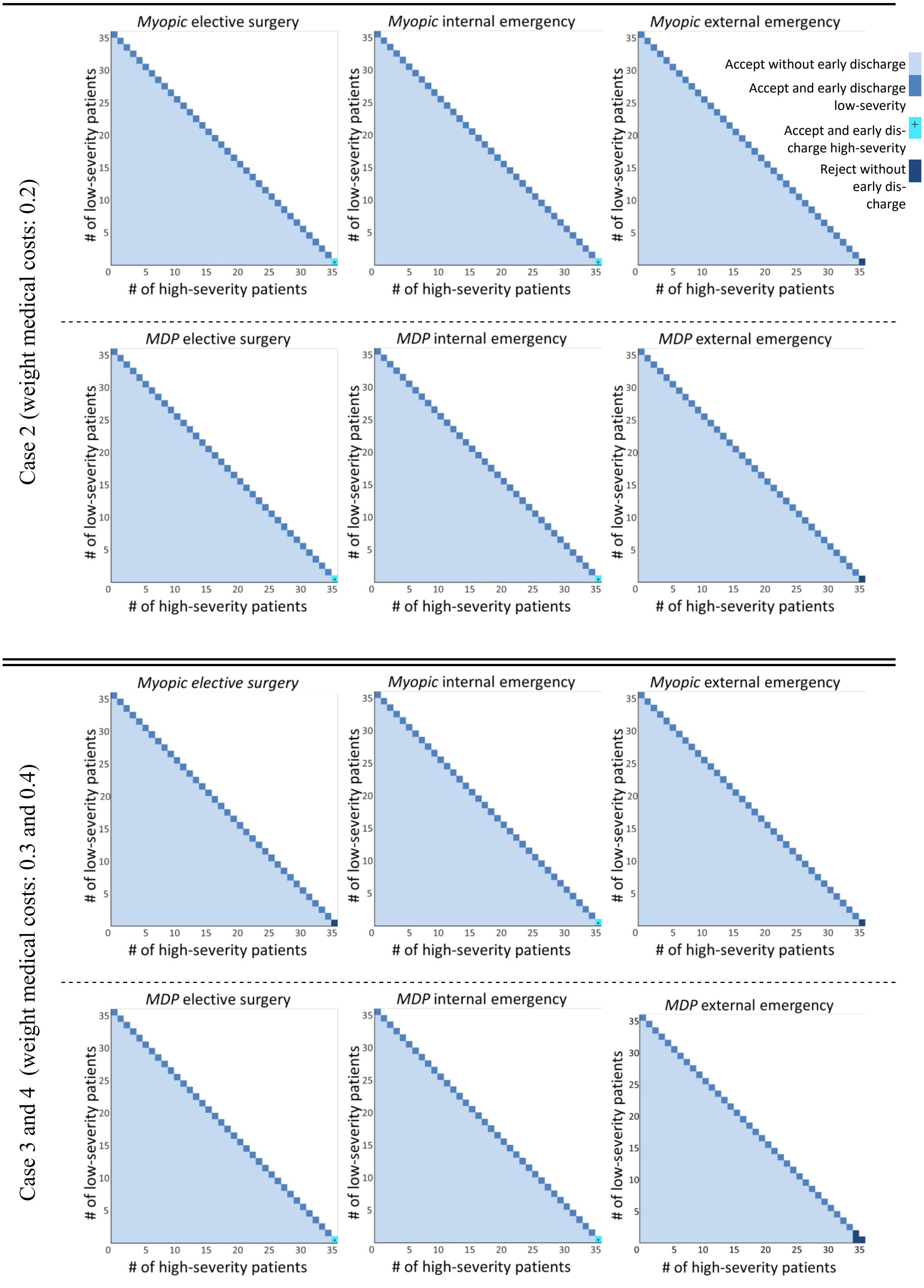
The cost settings of the 20 cases in the sensitivity analysis (the base cases of medical (case 10) and monetary (case 0) perspectives are included) are presented in Table B2. To have medical and monetary costs on a comparable level, we denote monetary costs in units of thousand Euro.

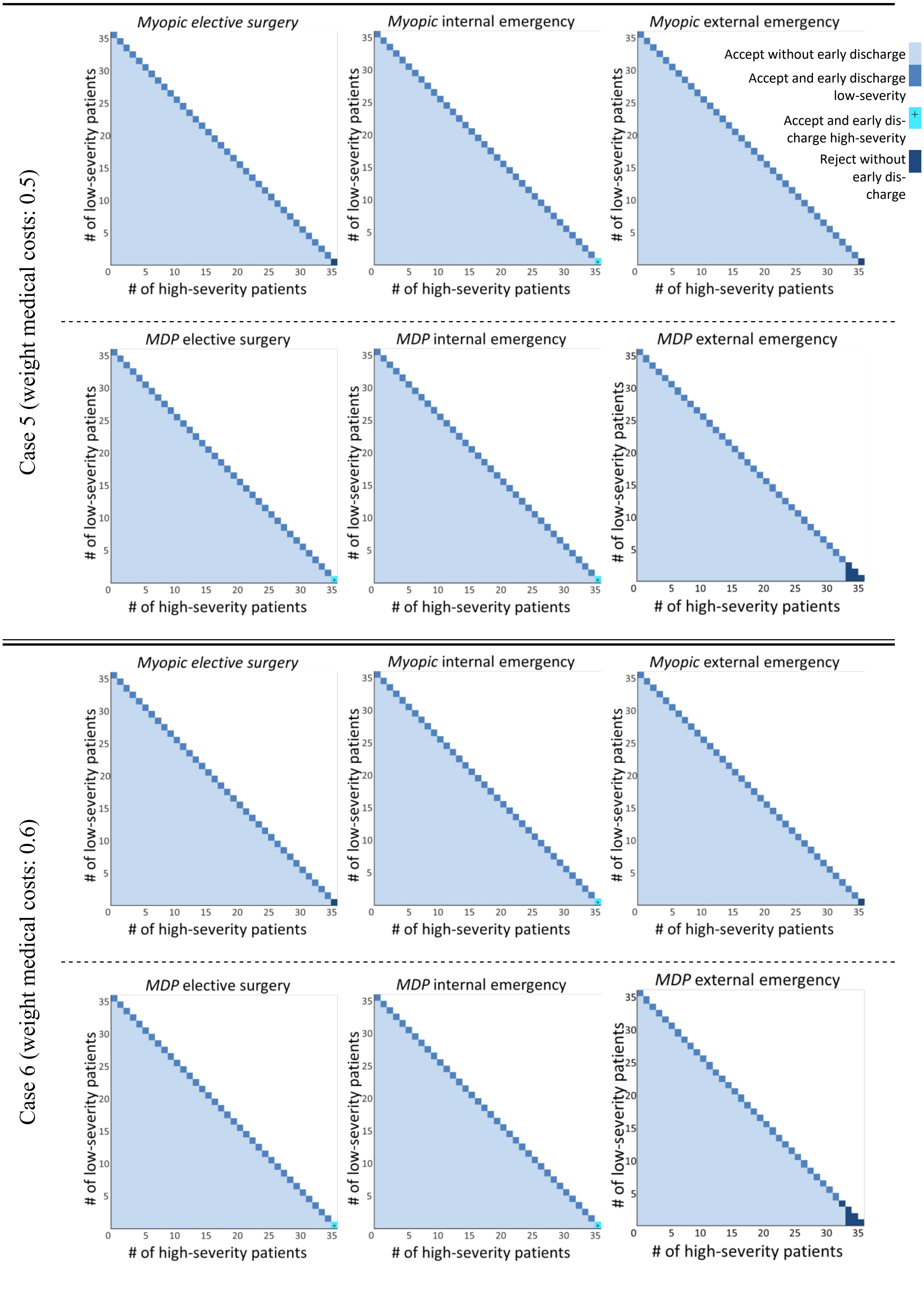
Case	Weight_med	Weight_Mon	$c_{i=1}^{rej}$	$c_{i=2}^{rej}$	$c_{i=3}^{rej}$	$c_{j=1}^{edis}$	$c_{j=2}^{edis}$
0	0	1	9.2	5.8	4.1	0.7	6.5
1	0.1	0.9	8.38	6.72	3.99	0.83	6.85
2	0.2	0.8	7.56	7.64	3.88	0.96	7.2
3	0.3	0.7	6.74	8.56	3.77	1.09	7.55
4	0.4	0.6	5.92	9.48	3.66	1.22	7.9
5	0.5	0.5	5.1	10.4	3.55	1.35	8.25
6	0.6	0.4	4.28	11.32	3.44	1.48	8.6
7	0.7	0.3	3.46	12.24	3.33	1.61	8.95
8	0.8	0.2	2.64	13.16	3.22	1.74	9.3
8a	0.825	0.175	2.435	13.39	3.1925	1.7725	9.3875
8b	0.85	0.15	2.23	13.62	3.165	1.805	9.475
8c	0.875	0.125	2.205	13.85	3.1375	1.8375	9.5625
8d	0.89	0.11	1.902	13.988	3.121	1.857	9.615
8e	0.8915	0.1085	1.8897	14.0018	3.11935	1.85895	9.62025
8f	0.8916	0.1084	1.88888	14.00272	3.11924	1.85908	9.6206
9	0.9	0.1	1.82	14.08	3.11	1.87	9.65
9a	0.925	0.075	1.615	14.31	3.0825	1.9025	9.7375
9b	0.95	0.05	1.41	14.54	3.055	1.935	9.825
9c	0.975	0.025	1.205	14.77	3.0275	1.9675	9.9125
10	1.0	0	1	15	3	2	10

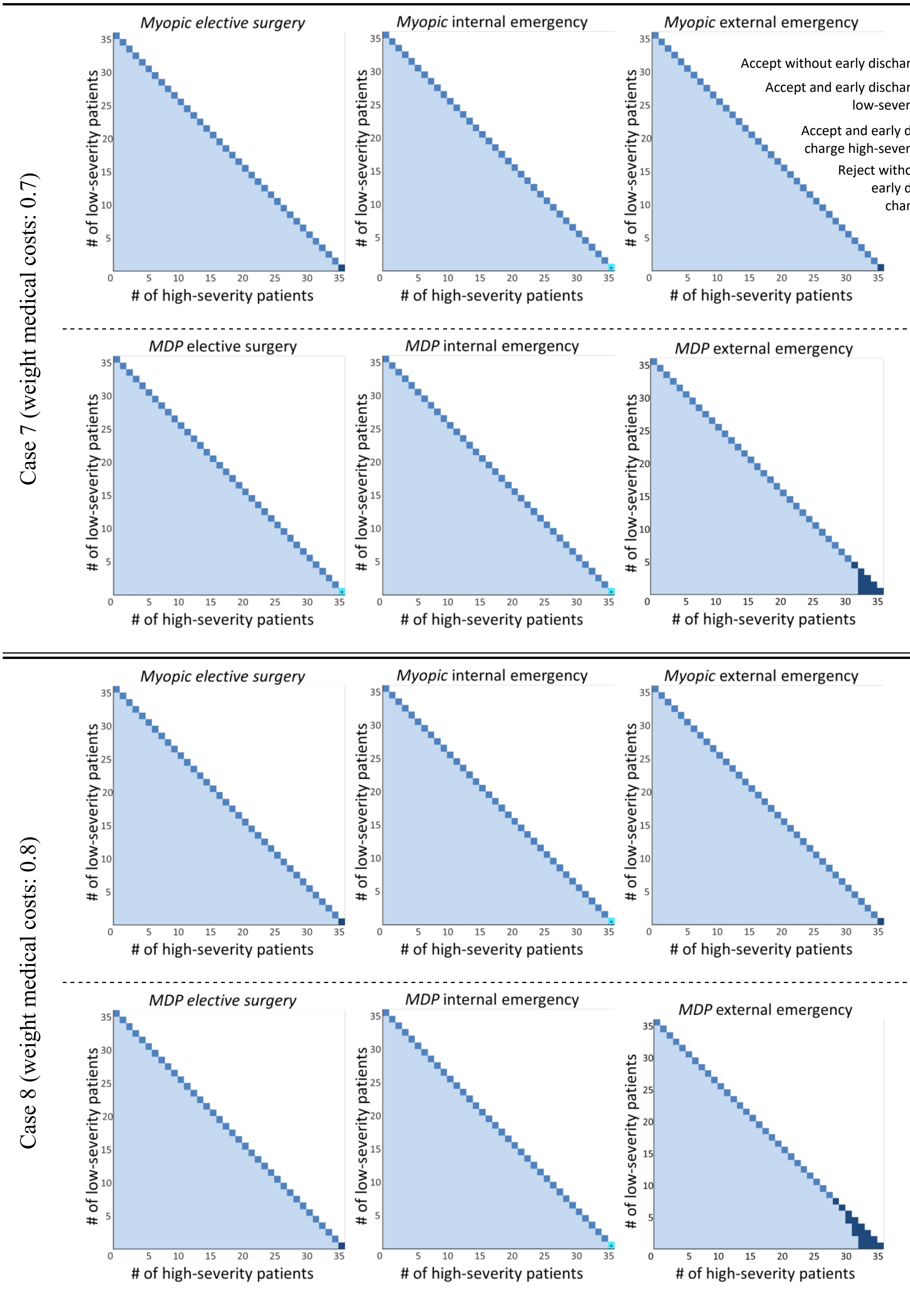
Table B2: Weighted cost settings in sensitivity analysis

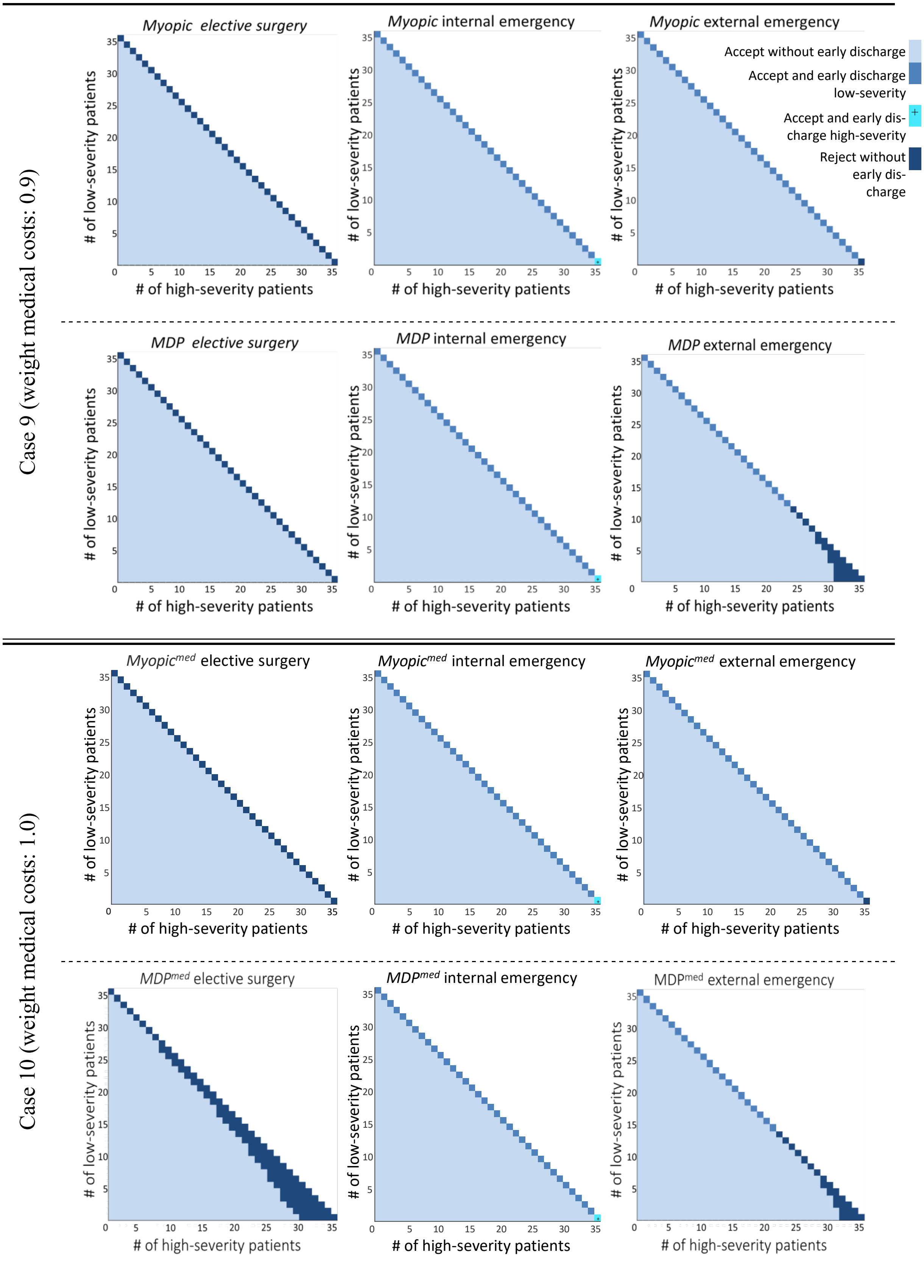
In the following, we document the myopic and MDP policies of the cases with the weight of medical costs ranging between 0 and 1 using steps of 0.1. This corresponds to cases zero to ten of Table B2.


Case 0 (weight medical costs: 0)
Myopic^mon elective surgery
Myopic^mon internal emergency
Myopic^mon external emergency
Accept without early discharge
Accept and early discharge low-severity
Accept and early dis- charge high-severity
Reject without early dis- charge
MDP^mon elective surgery
MDP^mon internal emergency
MDP^mon external emergency
Case 1 (weight medical costs: 0.1)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
MDP elective surgery
MDP internal emergency
MDP external emergency
# of low-severity patients
# of high-severity patients


Case 2 (weight medical costs: 0.2)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
Accept without early discharge
Accept and early discharge low-severity
Accept and early discharge high-severity
Reject without early discharge
MDP elective surgery
MDP internal emergency
MDP external emergency
Case 3 and 4 (weight medical costs: 0.3 and 0.4)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
MDP elective surgery
MDP internal emergency
MDP external emergency
# of low-severity patients
# of high-severity patients


Case 5 (weight medical costs: 0.5)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
Accept without early discharge
Accept and early discharge low-severity
Accept and early discharge high-severity
Reject without early discharge
MDP elective surgery
MDP internal emergency
MDP external emergency
Case 6 (weight medical costs: 0.6)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
MDP elective surgery
MDP internal emergency
MDP external emergency
# of low-severity patients
# of high-severity patients


Case 7 (weight medical costs: 0.7)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
Accept without early discharge
Accept and early discharge low-severity
Accept and early discharge high-severity
Reject without early discharge
MDP elective surgery
MDP internal emergency
MDP external emergency
Case 8 (weight medical costs: 0.8)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
MDP elective surgery
MDP internal emergency
MDP external emergency
# of low-severity patients
# of high-severity patients


Case 9 (weight medical costs: 0.9)
Myopic elective surgery
Myopic internal emergency
Myopic external emergency
Accept without early discharge
Accept and early discharge low-severity
Accept and early dis- charge high-severity
Reject without early dis- charge
MDP elective surgery
MDP internal emergency
MDP external emergency
Case 10 (weight medical costs: 1.0)
Myopic^{med} elective surgery
Myopic^{med} internal emergency
Myopic^{med} external emergency
MDP^{med} elective surgery
MDP^{med} internal emergency
MDP^{med} external emergency
# of low-severity patients
# of high-severity patients

Appendix B.5: Sensitivity Analysis

B.5.1 Sensitivity analysis I: Variation of parameters

In the following, we perform a sensitivity analysis to check whether the superiority of the MDP approach is robust against variations of the problem parameters. To do so, we use the medical perspective and decrease and increase all cost parameters by 50%. To limit the number of settings, and to concentrate on large variations, we don't include the base case settings in the study. Combining the two possible values for each of the five cost parameters in a full factorial design yields the 32 test cases given in Table B3. The cases are described using a sequence of five "+" and "-" (column two) that denote the parameters increased and decreased. The first three symbols indicate the rejection costs for elective surgery, internal emergency, and external emergency, while the remaining two symbols indicate the early discharging costs for low- and high-severity patients. For example, case 3 is denoted by "(---;+-)" because the costs of rejecting type 1, 2, and 3 patients (elective surgery, internal emergency, and external emergency, respectively) are decreased, the cost of early discharging a low-severity patient are increased, and the costs of early discharging a high-severity patient are decreased. For each test case, we determined the MDP and myopic policy, and calculate the improvement of MDP's performance in comparison to the myopic policy applying simulation analysis. The average improvement was 27%, with improvements in the test cases ranging from just under 2% (case 21, (+-+;--)), up to almost 50% (e.g. case 15, (-++;+-)). The standard deviation was 14%, and in only five cases (case 21, 22, 29, 30), we observed improvements below 10% (Table B3, columns three to five).

Cases		Costs (Correct Estimation) (Appendix B.4.1)			Costs (Base Case is true) (Appendix B.4.2)			Relative cost increase due to estimation error	
ID	Cost Settings	MDP	Myopic	Improvement	MDP	Myopic	Improvement	MDP	Myopic
1	- - - - -	964	1,221	21.04%	1,929	2,443	21.04%	-0.11%	-0.41%
2	- - - - +	1,003	1,478	32.14%	2,006	2,956	32.14%	3.88%	20.51%
3	- - - + -	1,166	2,126	45.17%	2,198	2,117	-3.85%	13.84%	-13.71%
4	- - - + +	1,255	2,204	43.07%	2,030	2,276	10.81%	5.12%	-7.22%
5	- - + - -	1,142	1,266	9.79%	2,284	2,444	6.55%	18.28%	-0.37%
6	- - + - +	1,388	1,571	11.65%	2,502	2,999	16.59%	29.57%	22.28%
7	- - + + -	1,520	2,902	47.63%	3,023	2,426	-24.63%	56.57%	-1.11%
8	- - + + +	2,168	3,175	31.72%	2,083	2,978	30.06%	7.88%	21.42%
9	- + - - -	969	1,214	20.23%	1,937	2,428	20.23%	0.31%	-1.00%
10	- + - - +	1,017	1,592	36.12%	1,925	2,434	20.90%	-0.29%	-0.76%
11	- + - + -	1,164	2,129	45.33%	2,195	2,120	-3.52%	13.66%	-13.57%
12	- + - + +	1,255	2,228	43.68%	1,981	2,115	6.32%	2.60%	-13.79%
13	- + + - -	1,137	1,273	10.70%	2,273	2,456	7.43%	17.74%	0.12%

14	-	+	+	-	+	1,437	1,641	12.43%	2,111	2,443	13.59%	9.32%	-0.41%
15	-	+	+	+	-	1,519	2,920	47.99%	3,020	2,443	-23.63%	56.39%	-0.42%
16	-	+	+	+	+	2,179	3,305	34.08%	1,959	2,447	19.94%	1.44%	-0.26%
17	+	-	-	-	-	1,145	1,426	19.75%	2,289	2,808	18.47%	18.54%	14.46%
18	+	-	-	-	+	1,165	1,703	31.59%	2,323	3,334	30.32%	20.31%	35.92%
19	+	-	-	+	-	1,458	2,432	40.06%	2,448	2,110	-16.00%	26.78%	-13.97%
20	+	-	-	+	+	1,635	2,518	35.05%	2,060	2,277	9.50%	6.70%	-7.19%
21	+	-	+	-	-	1,434	1,461	1.81%	2,869	2,789	-2.85%	48.57%	13.71%
22	+	-	+	-	+	1,710	1,769	3.36%	3,081	3,323	7.29%	59.54%	35.47%
23	+	-	+	+	-	2,047	3,285	37.70%	3,235	2,443	-32.44%	67.55%	-0.41%
24	+	-	+	+	+	2,908	3,543	17.92%	2,071	2,983	30.57%	7.26%	21.61%
25	+	+	-	-	-	1,152	1,415	18.53%	2,305	2,785	17.24%	19.35%	13.53%
26	+	+	-	-	+	1,179	1,800	34.48%	2,281	2,792	18.29%	18.15%	13.83%
27	+	+	-	+	-	1,458	2,490	41.44%	2,448	2,119	-15.53%	26.78%	-13.62%
28	+	+	-	+	+	1,630	2,541	35.85%	2,021	2,114	4.38%	4.66%	-13.83%
29	+	+	+	-	-	1,429	1,462	2.30%	2,858	2,791	-2.39%	47.99%	13.78%
30	+	+	+	-	+	1,753	1,831	4.27%	2,663	2,779	4.18%	37.89%	13.28%
31	+	+	+	+	-	2,047	3,281	37.63%	3,235	2,440	-32.58%	67.54%	-0.53%
32	+	+	+	+	+	2,899	3,654	20.66%	1,933	2,436	20.66%	0.08%	-0.70%
Base case						1,931	2,453	21.28%	1,931	2,453	21.28%	0.00%	0.00%

Table B3: Result of sensitivity analyses 1 and 2

In the following, we describe the cases where the potential of the MDP is particularly high, and those where the additional value from an MDP is negligible. In general, the main difference is that the MDP policies may reserve beds, while the myopic policies do not (ignoring future effects, it is always dominant to admit a patient when free capacities exist). Thus, the MDP potential is relatively low if the optimal policy does not reserve any beds, that is, when rejection costs are high, and early discharge costs are low. The biggest impact seems to result from changing the costs for deferring ambulances (rejecting external emergencies) and for early discharging low-severity patients: In the cases with high costs for rejecting external emergencies and low costs for early discharging low-severity patients (cases 5, 6, 13, 14, 21, 22, 29, 30), the relative improvement of the MDP is on average 7%, while this figure increases to 41% for the cases with low costs for external emergencies and high costs for early discharging low-severity patients (cases 3, 4, 11, 12, 19, 20, 27, 28). Changing the high-cost parameters (rejecting internal emergencies, early discharging high-severity patients) has little impact on the potential of our model.

We conclude that while our model has significant benefits in most of the considered test cases, there are a few cases where the MDP does not reserve beds, and its use does not lead to a considerable improvement compared to a myopic policy. When rejection costs outweigh the early discharge costs (at least for low-severity patients), myopic policies in fact are similar to MDP policies, resulting in basically no difference.

B.5.2 Sensitivity analysis II: Robustness against cost misspecifications

As we have already seen in Section 3.4, the medical cost of ICU decisions can be hard to estimate. Thus, an approach that crucially depends on exactly knowing parameter values may perform very poorly in reality if cost parameters are inaccurately estimated or change over time. To investigate this, we assume that the real medical cost parameters still follow our base case. However, the hospital erroneously believes them to be one of the 32 test cases from the previous Subsection 3.6.3 and follows the MDP policies from the corresponding test case. This policy is then evaluated using the real (i.e., base case) cost parameters and the result is compared to the optimal MDP policy for the base case (Table B3, columns six to ten). Thus, the policies and the resulting simulations (number of admissions, rejections and early discharges) for each case are the same for Sections 3.6.4 and 3.6.3. However, we use the cost settings in column two of Table B3 to calculate the costs in Section 3.6.3 (Table B3, columns three and four), while we use the base cost settings in Section 3.6.4 (Table B3, columns six and seven).

Over all 32 cases of biased cost settings, the cost bias leads to an increase of costs of 22% compared to the MDP using the correct cost parameters. There are cases with only little deviation, but also some cases with increases of more than 50%. In the following, we are going to identify critical and less critical misinterpretations of costs and the related implications on policies. The critical cases have in common that rejection costs for scheduled surgeries and external emergencies are overestimated, and that early discharge costs for high-severity patients are underestimated. The average additional costs for these cases (cases 21, 23, 29, 31) amount to 58%, while they are only 3% higher for cases where rejection costs for scheduled surgeries and external emergencies are underestimated and early discharge costs for high-severity patients are overestimated (cases 2, 4, 10, 12). The rationale for those critical cases is that the MDP does not reserve enough beds, and high-severity patients are discharged early.

In contrast to MDP policies, a myopic decision maker is more robust against erroneous estimation of costs – the additional costs per scenario are only around 4%. This is not surprising, as the baseline costs for the myopic case were much higher, and since fewer different policies exist, the possibility to differ is lower. Confronted with biased costs, the MDP still outperforms myopic policies by around 7% (average of column 8 in Table B3). We now discuss the cases where the myopic policy considerably outperforms the MDP: The most relevant criteria are the estimated costs for early discharging high-severity patients – in case of underestimation, the MDP leads to 4% higher costs compared to the myopic policies (all cases with odd indices), while the MDP reduces total cost compared to myopic policies by 17% when these costs are overestimated (all cases with even indices). Especially in combination with overestimation of rejection costs, leading to a relatively high utilization of the ICU (again, in these cases, patients will not be rejected, both considering MDP and myopic policies), the MDP policies discharge high-severity patients early instead of low-severity patients when these costs are underestimated, leading to strong increases of mortality rates.

We conclude that erroneous estimation of cost parameters may indeed lead to dramatic results. The worst impact on medical costs was observed for combinations of overestimation of rejection costs and underestimation of the cost of early discharges, while results are otherwise relatively robust. Please note that when all costs are scaled (e.g. case 1 and case 32), the MDP policy does not change. The small difference in the results between the policy from case 1 or case 32 and the baseline policy (-0.11% and -0.70%) is due to the stochastic nature of simulations and lies within the error margin.

Bibliography

1. Bai J, Fügener A, Schoenfelder J, Brunner JO (2016) Operations research in intensive care unit management: A literature review. Health Care Manag Sci 1–24. https://doi.org/10.1007/s10729-016-9375-1
2. Chan CW, Farias VF, Bambos N, Escobar GJ (2012) Optimizing intensive care unit discharge decisions with patient readmissions. Oper Res 60 (6):1323–1341. https://doi.org/10.1287/opre.1120.1105
3. Bai J, Fügener A, Schoenfelder J, Brunner JO (2018) Operations research in intensive care unit management: a literature review. Health Care Manag Sci 21:1–24. https://doi.org/10.1007/s10729-016-9375-1
4. Dobson G., Hasija S., Pinker EJ. (2011) Reserving capacity for urgent patients in primary care. Prod Oper Manag 20:456–473. https://doi.org/10.1111/j.1937-5956.2011.01227.x
5. Li X, Liu D, Geng N, Xie X (2019) Optimal ICU Admission Control with Premature Discharge. IEEE Trans Autom Sci Eng 16:148–164. https://doi.org/10.1109/TASE.2018.2827664
6. Chow VS, Puterman ML, Salehirad N, et al (2011) Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. Prod Oper Manag 20:418–430
7. Price C, Golden B, Harrington M, et al (2011) Reducing Boarding in a Post-Anesthesia Care Unit. Prod Oper Manag 20:431–441
8. Halpern NA, Pastores SM (2010) Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 38:65–71. https://doi.org/10.1097/CCM.0b013e3181b090d0
9. Rapoport J, Teres D, Zhao Y, Lemeshow S (2003) Length of Stay Data as a Guide to Hospital Economic Performance for ICU Patients. Med Care 41:386–397
10. Griffiths JD, Price-Lloyd N, Smithies M, Williams JE (2005) Modelling the requirement for supplementary nurses in an intensive care unit. J Oper Res Soc 56:126–133
11. Barrett M, Smith M, Elixhauser A, et al (2011) Utilization of Intensive Care Services, 2011. Healthc cost Util Proj 15:1–14
12. Pinker E, Tezcan T (2013) Determining the Optimal Configuration of Hospital Inpatient Rooms in the Presence of Isolation Patients. Oper Res 61:1259–1276
13. Green L (2002) How many hospital beds? Inquiry 39 (4):400–12. https://doi.org/10.5034/inquiryjrnl_39.4.400
14. Litvak N, van Rijsbergen M, Boucherie RJ, van Houdenhoven M (2008) Managing the overflow of intensive care patients. Eur J Oper Res 185 (3):998–1010.

https://doi.org/10.1016/j.ejor.2006.08.021

15. KC DS, Terwiesch C (2012) An Econometric Analysis of Patient Flows in the Cardiac Intensive Care Unit. Manuf Serv Oper Manag 14:50–65
16. Iapichino G, Gattinoni L, Radrizzani D, et al (2004) Volume of activity and occupancy rate in intensive care units. Association with mortality. Intensive Care Med 30 (2):290–297. https://doi.org/10.1007/s00134-003-2113-4
17. Dobson G, Lee HH, Pinker E (2010) A model of ICU bumping. Oper Res 58 (6):1564–1576. https://doi.org/10.1287/opre.1100.0861
18. Hulshof PJH, Kortbeek N, Boucherie RJ, et al (2012) Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. Heal Syst 1:129–175
19. Lakshmi C, Appa Iyer S (2013) Application of queueing theory in health care: A literature review. Oper Res Heal Care 2:25–39
20. Hof S, Fügener A, Schoenfelder J, Brunner JO (2015) Case mix planning in hospitals: a review and future agenda. Health Care Manag Sci 1–14
21. Tierney LT, Conroy KM (2014) Optimal occupancy in the ICU. Austrilian Crit care 27:77–84
22. Ridley SA, Burchett K, Burns A, Gunning K (1999) A comparison of hospital and critical-care activity. Anaesthesia 54:521–528
23. Rosenberg AL, Watts C (2000) Patients readmitted to ICUs* : a systematic review of risk factors and outcomes. Chest 118:492–502
24. Strand K, Flaatten H (2008) Severity scoring in the ICU: A review. Acta Anaesthesiol Scand 52:467–478
25. Kramer AA, Zimmerman JE (2011) The relationship between hospital and intensive care unit length of stay. Crit Care Med 39:1015–22
26. McConnell KJ, Richards CF, Daya M, et al (2005) Effect of increased icu capacity on length of stay in the emergency department. Ann Emerg Med 47:471–478
27. Fügener A, Edenharter GM, Kiefer P, et al (2015) Improving Intensive Care Unit and Ward Utilization by Adapting Master Surgery Schedules. Anesth \& Analg Case Reports Forthcoming
28. Cochran JK, Roche K (2008) A queuing-based decision support methodology to estimate hospital inpatient bed demand. J Oper Res Soc 59:1471–1482
29. Cochran JK, Bharti A (2006) Stochastic bed balancing of an obstetrics hospital. Health Care Manag Sci 9:31–45
30. Cochran JK, Bharti A (2006) A multi-stage stochastic methodology for whole hospital bed planning under peak loading. Int J Ind Syst Eng 1:8. https://doi.org/10.1504/IJISE.2006.009048
31. Bowers J (2013) Balancing operating theatre and bed capacity in a cardiothoracic centre. Health Care Manag Sci 16:236–244
32. Fügener A, Hans EW, Kolisch R, et al (2014) Master surgery scheduling with consideration of multiple downstream units. Eur J Oper Res 239:227–236. https://doi.org/10.1016/j.ejor.2014.05.009
33. Marmor YN, Rohleder TR, Huschka T, et al (2011) A simulation tool to support recovery bed planning for surgical patients. Proc - Winter Simul Conf 1333–1339

34. Marmor YN, Rohleder TR, Cook DJ, et al (2013) Recovery bed planning in cardiovascular surgery: A simulation case study. Health Care Manag Sci 16:314–327

35. McHardy S, Kozan E, Cook D (2004) Modelling of intensive care units operating theatre in public hospitals. In: The Fifth Asia-Pacific Industrial Engineering and Management systems conference. pp 12–15

36. Villa S, Barbieri M, Lega F (2009) Restructuring patient flow logistics around patient care needs: Implications and practicalities from three critical cases. Health Care Manag Sci 12:155–165

37. Van Houdenhoven M, van Oostrum JM, Wullink G, et al (2008) Fewer intensive care unit refusals and a higher capacity utilization by using a cyclic surgical case schedule. J Crit Care 23:222–226

38. Van Oostrum JM, Van Houdenhoven M, Hurink JL, et al (2008) A master surgical scheduling approach for cyclic scheduling in operating room departments. OR Spectr 30:355–374

39. Adan IJBF, Vissers JMH (2002) Patient mix optimisation in hospital admission planning: a case study. Int J Oper Prod Manag 22:445–461

40. Adan I, Bekkers J, Dellaert N, et al (2011) Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. Eur J Oper Res 213:290–308

41. Allon G, Deo S, Lin W (2013) The Impact of Size and Occupancy of Hospital on the Extent of Ambulance Diversion: Theory and Evidence. Oper Res 61:544–562

42. Anderson D, Golden B, Jank W, Wasil E (2012) The impact of hospital utilization on patient readmission rate. Health Care Manag Sci 15:29–36

43. Anderson D, Price C, Golden B, et al (2011) Examining the discharge practices of surgeons at a large medical center. Health Care Manag Sci 14:338–347

44. Barz C, Rajaram K (2015) Elective patient admission and scheduling under multiple resource constraints. Prod Oper Manag 24 (12):1907–1930. https://doi.org/10.1111/poms.12395

45. Fügener A (2015) An Integrated Strategic and Tactical Master Surgery Scheduling Approach with Stochastic Resource Demand. J Bus Logist Forthcoming

46. Gartner D, Kolisch R (2014) Scheduling the hospital-wide flow of elective patients. Eur J Oper Res 233:689–699

47. Akkerman R, Knip M (2004) Reallocation of beds to reduce waiting time for cardiac surgery. Health Care Manag Sci 7:119–126

48. Williams S V. (1983) How many intensive care beds are enough? Crit Care Med 11:412–6

49. Yang M, Fry MJ, Raikhelkar J, et al (2013) A model to create an efficient and equitable admission policy for patients arriving to the cardiothoracic ICU. Crit Care Med 41 (21):414–22. https://doi.org/10.1097/CCM.0b013e31826a44d7

50. Wharton F (1996) On the risk of premature transfer from coronary care units. Omega 24:413–423

51. Kapadia AS, Chan W, Sachdeva R, et al (2000) Predicting duration of stay in a pediatric intensive care unit: A Markovian approach. Eur J Oper Res 124:353–359

52. Mullinax C, Lawley M (2002) Assigning patients to nurses in neonatal intensive care. J Oper Res Soc 53:25–35

53. Nathanson BH, Higgins TL, Giglio RJ, et al (2003) An exploratory study using data

envelopment analysis to assess neurotrauma patients in the intensive care unit. Health Care Manag Sci 6:43–55

54. Asaduzzaman M, Chaussalet TJ, Robertson NJ (2010) A loss network model with overflow for capacity planning of a neonatal unit. Ann Oper Res 178:67–76

55. Adeyemi S, Demir E, Chahed S, Chaussalet T (2010) Analysis of variability in neonatal care units: A retrospective analysis. 2010 IEEE Work Heal Care Manag 1–6

56. Demir E, Lebcir R, Adeyemi S (2014) Modelling length of stay and patient flows: methodological case studies from the UK neonatal care services. J Oper Res Soc 65:532–545. https://doi.org/10.1057/jors.2013.51

57. Troy PM, Rosenberg L (2009) Using simulation to determine the need for ICU beds for surgery patients. Surgery 146:608–620

58. Zilm F, Hollis RB (1983) An application of simulation modeling to surgical intensive care bed need analysis in a university hospital. Hosp Health Serv Adm 28:82–101

59. Barado J, Guergué JM, Esparza L, et al (2012) A mathematical model for simulating daily bed occupancy in an intensive care unit*. Crit Care Med 40:1098–1104

60. Cahill W, Render M (1999) Dynamic simulation modeling of ICU bed availability. In: Proceedings of the1999 Winter Simulation Conference. pp 1573–1576

61. Hashimoto F, Bell S, Marshment S (1987) A computer simulation program to facilitate budgeting and staffing decisions in an intensive care unit. Crit. Care Med. 15:256–259

62. Kim SC, Horowitz I, Young KK, Buckley T (2000) Flexible bed allocation and performance in the intensive care unit. J Oper Manag 18 (4):427–443. https://doi.org/10.1016/S0272-6963(00)00027-9

63. Kim SC, Horowitz I (2002) Scheduling hospital services:the efficacy of elective-surgery quotas. Omega 30:335–346. https://doi.org/10.1016/S0305-0483(02)00050-6

64. Kim SC, Horowitz I, Young KK, Buckley TA (1999) Analysis of capacity management of the intensive care unit in a hospital. Eur J Oper Res 115 (1):36–46. https://doi.org/10.1016/S0377-2217(98)00135-0

65. Kolker A (2009) Process modeling of ICU patient flow: Effect of daily load leveling of elective surgeries on ICU diversion. J Med Syst 33:27–40

66. McManus ML, Long MC, Cooper A, Litvak E (2004) Queuing theory accurately models the need for critical care resources. Anesthesiology 100:1271–1276

67. Masterson BJ, Mihara TG, Miller G, et al (2004) Using models and data to support optimization of the Military Health System: a case study in an intensive care unit. Heal Care Manag Sci 7:217–224

68. Hagen MS, Jopling JK, Buchman TG, Lee EK (2013) Priority queuing models for hospital intensive care units and impacts to severe case patients. AMIA Annu Symp Proc 2013:841–50

69. Lowery JC (1992) Simulation of a hospital's surgical suite and critical care area. In: Proc. 1992 Winter Simulation Conf., ACM, New York. New York, pp 1071–1078

70. Lowery JC, Arbor A (1993) Multi-hospital Validation of critical care simulation model. In: Proceedings of the 1993 Winter Simulation Conference

71. Kim SH, Chan CW, Olivares M, Escobar G (2015) ICU admission control: An empirical study of capacity allocation and its implication on patient outcomes. Manage Sci 61:19–38

72. Ridge JC, Jones SK, Nielsen MS, Shahani AK (1998) Capacity planning for intensive care units. Eur J Oper Res 105:346–355

73. Verburg IWM, De Keizer NF, De Jonge E, Peek N (2014) Comparison of regression methods for modeling intensive care length of stay. PLoS One 9 (10): https://doi.org/10.1371/journal.pone.0109684

74. Kortbeek N, Van Dijk MN (2007) An dimensioning intensive care units. In: Dr. Jörg Kalcsics; Nickel PDS (ed) Operations Research Proceeding. Springer Berlin Heidelberg, pp 291–296

75. van Dijk NM, Kortbeek N (2009) Erlang loss bounds for OT-ICU systems. Queueing Syst 63:253–280

76. Li J, Dong M, Zhao W (2015) Admissions optimisation and premature discharge decisions in intensive care units. Int J Prod Res 53:7329–7342. https://doi.org/10.1080/00207543.2015.1059520

77. Shmueli A, Sprung CL, Kaplan EH (2003) Optimizing admissions to an intensive care unit. Health Care Manag Sci 6:131–136

78. Griffiths JD, Price-Lloyd N, Smithies M, Williams J (2006) A queueing model of activities in an intensive care unit. IMA J Manag Math 17:277–288. https://doi.org/10.1093/imaman/dpi042

79. Terwiesch C, Diwas KC, Kahn JM (2011) Working with capacity limitations: operations management in critical care. Crit Care 15:308

80. Lamiell JM (1995) Modeling intensive care unit census. Mil. Med. 160:227–232

81. Costa AX, Ridley SA, Shahani AK, et al (2003) Mathematical modelling and simulation for planning critical care capacity. Anaesthesia 58:320–327

82. Yergens DW, Doig CJ, Sandham JD (1995) Computer simulation modeling for predicting ICU bed requirements in a tertiary care hospital. Clin Investig Med 18:suppl: B23

83. Shahani AK, Ridley SA, Nielsen MS (2008) Modelling patient flows as an aid to decision making for critical care capacities and organisation. Anaesthesia 63:1074–1080

84. Burke EK, De Causmaecker P, Vanden Berghe G, Van Landeghem H (2004) The state of the art of nurse rostering. J Sched 7:441–499

85. Erhard M, Schoenfelder J, Fügener A, Brunner J (2016) State of the Art in physician scheduling. Work Pap 1–35

86. Brown KL, Pagel C, Pienaar A, Utley M (2011) The relationship between workload and medical staffing levels in a paediatric cardiac intensive care unit. Intensive Care Med 37:326–333

87. Duraiswamy N, Welton R, Reisman A (1981) Using computer simulation to predict ICU staffing needs. J Nurs Adm 12:39–44

88. Chan C, Yom-Tov G (2011) Intensive care unit patient flow with readmissions : A state-dependent queueing network. In: 2011 MSOM Annual Conference Ann Arbor, Michigan

89. Zimmerman JE, Kramer AA, McNair DS, et al (2006) Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. Crit Care Med 34 (10):2517–29. https://doi.org/10.1097/01.CCM.0000240233.01711.D9

90. Shmueli A, Baras M, Sprung CL (2004) The effect of intensive care on in-hospital survival. Heal Serv Outcomes Res Methodol 5:163–174

91. Shmueli A, Sprung CL (2005) Assessing the in-hospital survival benefits of intensive care. Int J Technol Assess Health Care 21:66–72

92. Roumani YF, May JH, Strum DP, Vargas LG (2013) Classifying highly imbalanced ICU data. Health Care Manag Sci 16:119–128

93. Hyer NL, Wemmerlöv U, Morris JA (2009) Performance analysis of a focused hospital unit: The case of an integrated trauma center. J Oper Manag 27 (3):203–219. https://doi.org/10.1016/j.jom.2008.08.003

94. Coopersmith CM, Wunsch H, Fink MP, et al (2012) A comparison of critical care research funding and the financial burden of critical illness in the United States. Crit Care Med 40:1072–1079. https://doi.org/10.1097/CCM.0b013e31823c8d03

95. Thompson S, Nunez M, Garfinkel R, Dean MD (2009) OR practice--efficient short-term allocation and reallocation of patients to floors of a hospital during demand surges. Oper Res 57 (2):261–273. https://doi.org/10.1287/opre.1080.0584

96. Boyd O, Evans L (2016) The future workforce of our Intensive Care Units – Doctor, physician assistant or no-one? J Intensive Care Soc 17:186–190. https://doi.org/10.1177/1751143716638375

97. Bittencourt O, Verter V, Yalovsky M (2018) Hospital capacity management based on the queueing theory. Int J Product Perform Manag 67:224–238. https://doi.org/10.1108/IJPPM-12-2015-0193

98. Cardoso LTQ, Grion CMC, Matsuo T, et al (2011) Impact of delayed admission to intensive care units on mortality of critically ill patients: a cohort study. Crit Care 15:R28. https://doi.org/10.1186/cc9975

99. Hung S-C, Kung C-T, Hung C-W, et al (2014) Determining delayed admission to intensive care unit for mechanically ventilated patients in the emergency department. Crit Care 18:485. https://doi.org/10.1186/s13054-014-0485-1

100. Chan CW, Farias VF, Escobar GJ (2016) The impact of delays on service times in the intensive care unit. Manage Sci 63 (7):2049–2072. https://doi.org/10.1287/mnsc.2016.2441

101. Bai J, Gerstmeyr S, Brunner JO (2020) SIMULATION AND EVALUATION OF ICU MANAGEMENT POLICIES. In: Proceedings of the 2020 Winter Simulation Conference

102. Nates JL, Nunnally M, Kleinpell R, et al (2016) ICU Admission, Discharge, and Triage Guidelines

103. Fügener A, Schiffels S, Kolisch R (2017) Overutilization and underutilization of operating rooms - insights from behavioral health care operations management. Health Care Manag Sci 20:115–128. https://doi.org/10.1007/s10729-015-9343-1

104. Chalfin DB, Trzeciak S, Likourezos A, et al (2007) Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. Crit Care Med 35 (6):1477–1483. https://doi.org/10.1097/01.CCM.0000266585.74905.5A

105. Durbin CJ, Kopel R (1993) A case-control study of patients readmitted to the intensive care unit. Crit Care Med 21 (10):1547–53

106. Chen L, Martin C, Keenan S, Sibbald W (1998) Patients readmitted to the intensive care unit during the same hospitalization : Clinical features and outcomes. Crit Care Med 26 (11):1834–1841. https://doi.org/10.1097/00003246-199811000-00025

107. Chalfin D (2005) Length of intensive care unit stay and patient outcome: The long and short of it all. Crit Care Med 33 (9):2119–2120. https://doi.org/10.1097/01.CCM.0000178371.88891.26

108. Helm JE, Alaeddini A, Stauffer JM, et al (2016) Reducing hospital readmissions by integrating empirical prediction with resource optimization. Prod Oper Manag 25 (2):233–257. https://doi.org/10.1111/poms.12377

109. Chrusch CA, Olafson KP, McMillan PM, et al (2009) High occupancy increases the risk of early death or readmission after transfer from intensive care. Crit Care Med 37 (10):2753–

2758. https://doi.org/10.1097/CCM.0b013e3181a57b0c

110. Bouneb R, Mellouli M, Dardouri M, et al (2018) Determinants and outcomes associated with decisions to deny intensive care unit admission in Tunisian ICU. Pan Afr Med J 29:1–9. https://doi.org/10.11604/pamj.2018.29.176.13099

111. Louriz M, Abidi K, Akkaoui M, et al (2012) Determinants and outcomes associated with decisions to deny or to delay intensive care unit admission in Morocco. Intensive Care Med 38:830–837. https://doi.org/10.1007/s00134-012-2517-0

112. Hu W, Chan CW, Zubizarreta JR, Escobar GJ (2018) An Examination of Early Transfers to the ICU Based on a Physiologic Risk Score. Manuf Serv Oper Manag 20:531–549. https://doi.org/10.1287/msom.2017.0658

113. Long EF, Mathews KS (2018) The Boarding Patient: Effects of ICU and Hospital Occupancy Surges on Patient Flow. Prod Oper Manag 27:2122–2143. https://doi.org/10.1111/poms.12808

114. Miedaner F, Sülz S (2019) Boundaries of Focus and Volume: An Empirical Study in Neonatal Intensive Care. Prod Oper Manag 0:1–11. https://doi.org/10.1111/poms.13110

115. Barz C, Rajaram K (2015) Elective Patient Admission and Scheduling under Multiple Resource Constraints. Prod Oper Manag 24:1907–1930. https://doi.org/10.1111/poms.12395

116. Samiedaluie S, Kucukyazici B, Verter V, Zhang D (2017) Managing patient admissions in a neurology ward. Oper Res 65:635–656. https://doi.org/10.1287/opre.2016.1574

117. Zonderland ME, Boucherie RJ, Litvak N, Vleggeert-Lankamp CLAM (2010) Planning and scheduling of semi-urgent surgeries. Health Care Manag Sci 13:256–267. https://doi.org/10.1007/s10729-010-9127-6

118. Yang M, Fry MJ, Scurlock C (2015) The ICU will see you now: Efficient-equitable admission control policies for a surgical ICU with batch arrivals. IIE Trans 47:586–599. https://doi.org/10.1080/0740817X.2014.955151

119. Gocgun Y, Puterman ML (2014) Dynamic scheduling with due dates and time windows: An application to chemotherapy patient appointment booking. Health Care Manag Sci 17 (1):60–76. https://doi.org/10.1007/s10729-013-9253-z

120. Gupta D, Wang L (2008) Revenue Management for a Primary-Care Clinic in the Presence of Patient Choice. Oper Res 56:576–592. https://doi.org/10.1287/opre.1080.0542

121. Yu S, Kulkarni VG, Deshpande V (2019) Appointment Scheduling for a Health Care Facility with Series Patients. Prod Oper Manag 0:1–22. https://doi.org/10.1111/poms.13117

122. Li D, Ding L, Connor S (2019) When to Switch? Index Policies for Resource Scheduling in Emergency Response. Prod Oper Manag 0:1–22. https://doi.org/10.1111/poms.13105

123. Xie AJ, Zhuang W, Ang M, et al (2020) Analytics for Hospital Resource Planning — Two Case Studies. POMS. https://doi.org/https://doi.org/doi:10.1111/poms.13158

124. Dobson G, Lee H-H, Pinker E (2010) A Model of ICU Bumping. Oper Res 58:1564–1576. https://doi.org/10.1287/opre.1100.0861

125. Sprung CL, Geber D, Eidelman L a, et al (1999) Evaluation of triage decisions for intensive care admission. Crit Care Med 27 (6):1073–9

126. Azoulay E, Pochard F, Chevret S, et al (2001) Compliance with triage to intensive care recommendations. Crit Care Med 29 (11):2132–2136

127. Joynt G, Gomersall C, Tan P, et al (2001) Prospective evaluation of patients refused admission to an intensive care unit: Triage, futility and outcome. Intensive Care Med 27 (9):1459–1465. https://doi.org/10.1007/s001340101041

128. Howell MD (2011) Managing ICU throughput and understanding ICU census. Curr Opin Crit Care 17:626–633. https://doi.org/10.1097/MCC.0b013e32834b3e6e

129. Persson MJ, Persson JA (2010) Analysing management policies for operating room planning using simulation. Health Care Manag Sci 13:182–191. https://doi.org/10.1007/s10729-009-9122-y

130. VanBerkel PT, Blake JT (2007) A comprehensive simulation for wait time reduction and capacity planning applied in general surgery. Health Care Manag Sci 10:373–385. https://doi.org/10.1007/s10729-007-9035-6

131. Mielczarek B, Uziałko-Mydlikowska J (2012) Application of computer simulation modeling in the health care sector: A survey. Simulation 88:197–216. https://doi.org/10.1177/0037549710387802

132. Astaraky D, Patrick J (2015) A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. Eur J Oper Res 245:309–319. https://doi.org/10.1016/j.ejor.2015.02.032

133. Mahmoudian-Dehkordi A, Sadat S (2017) Sustaining critical care: using evidence-based simulation to evaluate ICU management policies. Health Care Manag Sci 20:532–547. https://doi.org/10.1007/s10729-016-9369-z

134. Plate JDJ, Leenen LPH, Houwert M, Hietbrink F (2017) Utilisation of Intermediate Care Units: A Systematic Review. Crit Care Res Pract 2017:8038460. https://doi.org/10.1155/2017/8038460

135. Daly K, Beale R, Chang RWS (2001) Reduction in mortality after inappropriate early discharge from intensive care unit: Logistic regression triage model. Br Med J 322:1274. https://doi.org/10.1136/bmj.322.7297.1274

136. Lyons RA, Wareham K, Hutchings HA, et al (2000) Population requirement for adult critical-care beds: A prospective quantitative and qualitative study. Lancet 355:595–598. https://doi.org/10.1016/S0140-6736(00)01265-4

137. de Bruin AM, Koole GM, Visser MC (2005) Bottleneck analysis of emergency cardiac in-patient flow in a university setting: an application of queueing theory. Clin Invest Med 28:316–317

138. Borghans I, Kleefstra SM, Kool RB, Westert GP (2012) Is the length of stay in hospital correlated with patient satisfaction? Int J Qual Heal Care 24:443–451. https://doi.org/10.1093/intqhc/mzs037

139. Herr A (2008) COST AND TECHNICAL EFFICIENCY OF GERMAN HOSPITALS: DOES OWNERSHIP MATTER? Health Econ 1131:1127–1131. https://doi.org/10.1002/hec

140. Cohen ME, Bilimoria KY, Ko CY, et al (2009) Variability in length of stay after colorectal surgery: Assessment of 182 hospitals in the national surgical quality improvement program. Ann Surg 250:901–907. https://doi.org/10.1097/SLA.0b013e3181b2a948

141. Drummond M, McGuire A (2011) Economic Evaluation in Health Care: Merging theory with practise.

142. Bertolini G, Confalonieri M, Rossi C, et al (2005) Costs of the COPD. Differences between intensive care unit and respiratory intermediate care unit. Respir Med 99:894–900. https://doi.org/https://doi.org/10.1016/j.rmed.2004.11.014

143. Correia ITDM, Waitzberg DL (2003) The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. Clin Nutr 22:235–239. https://doi.org/https://doi.org/10.1016/S0261-5614(02)00215-7

144. Encinosa WE, Bernard DM (2005) Hospital Finances and Patient Safety Outcomes. Inq J

Heal Care Organ Provision, Financ 42:60–72. https://doi.org/10.5034/inquiryjrnl_42.1.60

145. Kahn JM, Rubenfeld GD, Rohrbach J, Fuchs BD (2008) Cost savings attributable to reductions in intensive care unit length of stay for mechanically ventilated patients. Med Care 46:1226–1233. https://doi.org/10.1097/MLR.0b013e31817d9342

146. Lone NI, Walsh TS (2011) Prolonged mechanical ventilation in critically ill patients: epidemiology, outcomes and modelling the potential cost consequences of establishing a regional weaning unit. Crit Care 15:R102. https://doi.org/10.1186/cc10117

147. Ryan JL, Pracht EE, Orban BL (2019) Inpatient and emergency department costs from sports injuries among youth aged 5–18 years. BMJ Open Sport & Exerc Med 5:e000491. https://doi.org/10.1136/bmjsem-2018-000491

148. Sznajder M, Aegerter P, Launois R, et al (2001) A cost-effectiveness analysis of stays in intensive care units. Intensive Care Med 27:146–153. https://doi.org/10.1007/s001340000760

149. Williams RM (1996) Distribution of Emergency Department Costs. Ann Emerg Med 28:671–676. https://doi.org/https://doi.org/10.1016/S0196-0644(96)70092-6

150. Yu K-H, Beam AL, Kohane IS (2018) Artificial intelligence in healthcare. Nat Biomed Eng 2:719–731. https://doi.org/10.1038/s41551-018-0305-z

151. Jiang F, Jiang Y, Zhi H, et al (2017) Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol 2:230–243. https://doi.org/10.1136/svn-2017-000101

152. Cheng L-F, Prasad N, Engelhardt B (2018) An Optimal Policy for Patient Laboratory Tests in Intensive Care Units

153. Prasad N, Cheng LF, Chivers C, et al (2017) A reinforcement learning approach to weaning of mechanical ventilation in intensive care units. Uncertain Artif Intell - Proc 33rd Conf UAI 2017

154. Wang L, He X, Zhang W, Zha H (2018) Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation. Proc ACM SIGKDD Int Conf Knowl Discov Data Min 2447–2456. https://doi.org/10.1145/3219819.3219961

155. Sieben W, Schettlinger K, Kuhls S, Imhoff M (2006) Machine learning techniques in intensive care monitoring. 2–4

156. Burdick H, Pino E, Gabel-Comeau D, et al (2017) Evaluating a sepsis prediction machine learning algorithm using minimal electronic health record data in the emergency department and intensive care unit. bioRxiv 224014. https://doi.org/10.1101/224014

157. Nemati S, Holder A, Razmi F, et al (2018) An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med 46:547–553. https://doi.org/10.1097/CCM.0000000000002936

158. Rayan Z, Alfonse M, Salem A-BM (2019) Intensive Care Unit (ICU) Data Analytics Using Machine Learning Techniques. Int J "Information Theor Appl 26:69–82

159. Vellido A, Ribas V, Morales C, et al (2018) Machine learning in critical care: State-of-the-art and a sepsis case study. Biomed Eng Online 17:. https://doi.org/10.1186/s12938-018-0569-2

160. Kwong MT, Colopy GW, Weber AM, et al (2019) The efficacy and effectiveness of machine learning for weaning in mechanically ventilated patients at the intensive care unit: a systematic review. Bio-Design Manuf 2:31–40. https://doi.org/10.1007/s42242-018-0030-1

161. Xia J, Pan S, Zhu M, et al (2019) A Long Short-Term Memory Ensemble Approach for Improving the Outcome Prediction in Intensive Care Unit. Comput Math Methods Med

2019:. https://doi.org/10.1155/2019/8152713

162. Martínez-Agüero S, Mora-Jiménez I, Lérida-García J, et al (2019) Machine learning techniques to identify antimicrobial resistance in the intensive care unit. Entropy 21:1–24. https://doi.org/10.3390/e21060603

163. Tang F, Xiao C, Wang F, Zhou J (2018) Predictive modeling in urgent care: a comparative study of machine learning approaches. JAMIA Open 1:87–98. https://doi.org/10.1093/jamiaopen/ooy011

164. Meiring C, Dixit A, Harris S, et al (2018) Optimal intensive care outcome prediction over time using machine learning. PLoS One 13:1–19. https://doi.org/10.1371/journal.pone.0206862

165. Veith N, Steele R (2018) Machine learning-based prediction of ICU patient mortality at time of admission. ACM Int Conf Proceeding Ser 34–38. https://doi.org/10.1145/3206098.3206116

166. Kim SY, Kim S, Cho J, et al (2019) A deep learning model for real-time mortality prediction in critically ill children. Crit Care 23:1–10. https://doi.org/10.1186/s13054-019-2561-z

167. Krajnak M, Xue J, Kaiser W, Balloni W (2012) Combining machine learning and clinical rules to build an algorithm for predicting ICU mortality risk. Comput Cardiol (2010) 39:401–404

168. Overweg H, Popkes A-L, Ercole A, et al (2019) Interpretable Outcome Prediction with Sparse Bayesian Neural Networks in Intensive Care

169. Livieris IE, Kotsilieris T, Dimopoulos IF, Pintelas P (2018) Predicting length of stay in hospitalized patients using SSL algorithms. ACM Int Conf Proceeding Ser 16–22. https://doi.org/10.1145/3218585.3218588

170. Chen S, Kong N, Sun X, et al (2019) Claims data-driven modeling of hospital time-to-readmission risk with latent heterogeneity. Health Care Manag Sci 22:156–179. https://doi.org/10.1007/s10729-018-9431-0

171. Krämer J, Schreyögg J, Busse R (2019) Classification of hospital admissions into emergency and elective care: a machine learning approach. Health Care Manag Sci 22:85–105. https://doi.org/10.1007/s10729-017-9423-5

172. McWilliams CJ, Lawson DJ, Santos-Rodriguez R, et al (2019) Towards a decision support tool for intensive care discharge: Machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. BMJ Open 9:1–8. https://doi.org/10.1136/bmjopen-2018-025925

173. Yu C, Liu J, Nemati S (2019) Reinforcement Learning in Healthcare: A Survey

174. Asoh H, Shiro M, Akaho S, Kamishima T (2013) An Application of Inverse Reinforcement Learning to Medical Records of Diabetes. Eur Conf Mach Learn Princ Pract Knowl Discov Databases (ECML/PKDD ’13) 1–8

175. Zhao Y, Zeng D, Socinski MA, Kosorok MR (2011) Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. Biometrics 67:1422–1433. https://doi.org/10.1111/j.1541-0420.2011.01572.x

176. Gaweda AE, Muezzinoglu MK, Aronoff GR, et al (2005) Individualization of pharmacological anemia management using reinforcement learning. Neural Networks 18:826–834. https://doi.org/10.1016/j.neunet.2005.06.020

177. Raghu A, Komorowski M, Singh S (2018) Model-Based Reinforcement Learning for Sepsis Treatment

178. Komorowski M, Celi LA, Badawi O, et al (2018) The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 24:1716–1720. https://doi.org/10.1038/s41591-018-0213-5

179. Bai J, Fügener A, Gönsch J, et al (2020) Managing Admission and Discharge Processes in Intensive Care Units

180. Rokni Lamooki GR, Maleki F, Hajihosseini A (2014) A mathematical model for the admission process in intensive care units. Commun Nonlinear Sci Numer Simul 19:8–18

181. Powell WB (2011) Approximate Dynamic Programming: Solving the Curses of Dimensionality, 2nd ed. John Wiley & Sons, Inc, Hoboken, New Jersey

182. Grimes JP, Gregory PM, Noveck H, et al (2002) The effects of time-to-surgery on mortality and morbidity in patients following hip fracture. Am J Med 112 (9):702–709. https://doi.org/10.1016/S0002-9343(02)01119-1

183. Shiga T, Wajima ZI, Ohe Y (2008) Is operative delay associated with increased mortality of hip fracture patients? Systematic review, meta-analysis, and meta-regression. Can J Anesth 55 (3):146–54. https://doi.org/10.1007/BF03016088

184. Nyholm AM, Gromov K, Palm H, et al (2015) Time to surgery is associated with thirty-day and ninety-day mortality after proximal femoral fracture: a retrospective observational study on prospectively collected data from the Danish fracture database collaborators. J Bone Jt Surg 97 (16):1333–1119. https://doi.org/10.2106/JBJS.O.00029

185. Sabra A, Daayf F, Renault S (2012) Differential physiological and biochemical responses of three Echinacea species to salinity stress. Sci Hortic (Amsterdam) 135:23–31. https://doi.org/https://doi.org/10.1016/j.scienta.2011.11.024

186. Yibulayin W, Abulizi S, Lv H, Sun W (2016) Minimally invasive oesophagectomy versus open esophagectomy for resectable esophageal cancer: a meta-analysis. World J Surg Oncol 14:304. https://doi.org/10.1186/s12957-016-1062-7

187. Checkley W (2012) Mortality and denial of admission to an intensive care unit. Am J Respir Crit Care Med 185 (10):1038–1040. https://doi.org/10.1164/rccm.201202-0343ED

188. Iapichino G, Corbella D, Minelli C, et al (2010) Reasons for refusal of admission to intensive care and impact on mortality. Intensive Care Med 36 (10):1772–1779. https://doi.org/10.1007/s00134-010-1933-2

189. Singer AJ, Thode Jr HC, Viccellio P, Pines JM (2011) The association between length of emergency department boarding and mortality. Acad Emerg Med 18 (12):1324–1329. https://doi.org/10.1111/j.1553-2712.2011.01236.x

190. Smith L, Orts CM, O'neil I, et al (1999) TISS and mortality after discharge from intensive care. Intensive Care Med 25 (10):1061–1065

191. InEK GmbH (2017) G-DRG-System 2017. In: http://www.g-drg.de/G-DRG-System_2017

192. Knaus WA, Draper EA, Wagner DP, Zimmerman J (1985) APACHE II: A severity of disease classification system . Crit Care Med 13:818–829

193. Stewart W (2009) Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling

Curriculum Vitae

Education

09.2011- 02.2022	Doctoral Candidate
	Graduate School of Economics and Social Sciences (GESS)
	University of Mannheim, Germany
08.2008 - 07.2011	Master, Management Science and Engineering
	Tsinghua University, China
03.2009 - 10.2010	Master, Production Systems Engineering
	RWTH Aachen University, Germany
08.2004 - 07.2008	Bachelor, Mechanical Engineering and Automation
	Tsinghua University, China

Professional Experience

07.2020 - 10.2021	Research Associate, TUM School of Management, Germany
06.2013 – 06.2020	Research Assistant, University of Augsburg, Germany