

A Stylized Bed Allocation Model for a Turkish Private
Hospital with a Two Class Loss System

by

Nermin Elif Kurt

A Thesis Submitted to the
Graduate School of Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Industrial Engineering

Koç University

September 9, 2014

Koç University
Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Nermin Elif Kurt

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Assoc. Prof. E. Lerzan Örmeci

Assist. Prof. Pelin Gülşah Canbolat

Prof. Zeynep Akşin Karaesmen

Date: _____

To my parents

ABSTRACT

Hospital managers apply certain policies when developing daily schedules to admit patients to the hospital wards. Severeness of the patient, expected length of stay of the patient, expected number of incoming patients during the day or cost of the hospitalization are some factors that the manager considers during the decision making process. In this research, we develop a simplified model to understand the effects of admission, early discharge and overflow policies on the expected cost of the system. The model assumes that two hospital wards (main ward and overflow ward) serve two types of patients (severe and mild). The recovery time of a severe patient is represented by a two stage hypoexponential distribution whereas that of a mild patient follows an exponential distribution. The main ward serves both patient types, but the overflow ward can only accommodate mild patients. A mild patient may be discharged before his/her complete recovery upon an arrival of a severe patient. Early discharge decision is valid if there exists at least one type-2 patient in the system. We show that early discharge decision is not optimal unless all beds are occupied in the main ward. Furthermore, we establish the conditions which ensure that a patient type is “preferred” or “strongly preferred”, where patients of a preferred type are always accepted when there is at least one free bed and patients of a strongly preferred type are accepted even when the ward is full by early discharge decision. Finally, we verify our findings with a numerical analysis whose data is taken from a private hospital in Turkey.

ÖZETÇE

Hastane yöneticileri günlük planları belirlerken hastaları yataklarına kabul etmek için belirli politikalar uygular. Hastanın durumunun ciddiyeti ve beklenen yatış süresi, gün içinde beklenen hasta sayısı ya da yatışın maliyeti hastane yöneticilerinin karar alırken düşündükleri bazı kıstaslardır. Bu araştırmada, kabul, ret, erken taburcu ve ek koğuşa kabul politikalarının maliyet üzerindeki etkilerini anlamak için basitleştirilmiş bir model sunuyoruz. Modelde iki hastane koğuşı (ana ve ek koğuşı) iki durumdaki hastaya (ağır ve normal durumda) hizmet etmektedir. Ağır durumdaki hastanın iyileşme süreci iki aşamalı hipoeksponensiyel dağılımlı iken normal durumdaki hastaninkisi eksponensiyel dağılımlıdır. Ana koğuşı her iki durumdaki hastaya hizmet verirken ek koğuşı sadece normal durumdaki hastalara hizmet verir. Normal durumdaki hasta ağır durumdaki hastayı kabul etmek için iyileşme süreci tam olarak bitmeden taburcu edilebilir (erken taburcu politikası). Bu yüzden erken taburcu politikasının uygulanabilmesi için en az bir tane normal durumda hasta olması gerekir.

Bu çalışmada, erken taburcu politikasının ancak ana koğuştaki tüm yataklar dolduğunda yapılırsa en-iyi olcağını gösterdik. Ayrıca, hastaların “tercih edilen” veya “özellikle tercih edilen” olması için gerekli koşulları bulduk. Bir durumki hasta son boş yatağa kadar kabul ediliyorsa o durum “tercih edilen”dir. Koğuşı doluyken bile erken taburcu politikası ile kabul edilen ağır durumdaki hastalar “özellikle tercih edilen” olarak tanımlanırlar. Son olarak bütün bu bulgularımızı özel bir hastaneden aldığımız verilerle sayısal çözümleme ile destekledik.

ACKNOWLEDGMENTS

First I would like to thank my supervisor Assoc. Prof. Lerzan Örmeci for accepting me as an Ms. student, her encouragement, critical comments and corrections of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee, Prof. Zeynep Akşin Karaesmen and Assist. Prof. Pelin Gülşah Canbolat for critical reading of my thesis and their insightful comments.

Many friends have helped me stay sane through these difficult years. I am grateful to Bahadır Şahin for providing me a morale support that helps me in hard days of my research. I cannot think a world without him.

I am very lucky to have home-mates like Zehra Önen and Burcu Özçimen. I thank them for their delightful conversation and delicious coffee and cakes which make me feel at home. Moreover, they always share their valuable comments and allocate time for my questions about this research. Thank you for your support.

Most importantly, none of this would have been possible without the unconditional love and patience of my family. My caring, supportive and encouraging parents, Berrin and Abdullah, have sacrificed their lives for myself and give me a opportunity to have a great educational life. I love them so much.

Finally, I also thank TÜBİTAK for its financial support of our research project whose project number is 111 M 525.

TABLE OF CONTENTS

List of Figures	ix
List of Tables	x
Chapter 1: Introduction	1
Chapter 2: Literature Review	4
2.1 Operating Room Planning and Scheduling	5
2.2 Queuing Models	8
2.2.1 Appointment Scheduling	8
2.2.2 Loss Queuing Models	10
2.2.2.1 Admission planning and scheduling	10
2.2.2.2 Policy Modeling in ICUs	13
2.2.2.3 Admission control models using MDP	14
Chapter 3: PROBLEM DESCRIPTION AND MATHEMATICAL MODEL	17
3.1 Bed Management and Bed-Capacity Modeling	17
3.2 Markov Decision Model	20
3.3 Coupling	26
3.4 Effect of an Additional Job	27
3.5 Infinite Horizon Models	27
Chapter 4: Structure of the Optimal Policy	29

4.1	Preliminary Results	30
4.2	Existence of a Preferred Class	33
4.2.1	A sufficient condition for type 2 to be preferred	34
4.2.2	Sufficient conditions for type 1 to be preferred and strongly preferred	37
Chapter 5:	Numerical Analysis	55
5.1	Memorial Hizmet Hospital	55
5.1.1	Data Set and Current Condition	55
5.2	Data Preparation and Parameter Estimation	57
5.3	Base Case Model	60
5.3.1	Allowing overflow	60
5.3.2	No overflow	63
5.4	Sensitivity Analysis	65
5.4.1	Impact of c_r^1 and load	65
5.4.2	Impact of c_e and load	68
Chapter 6:	Conclusion	72
Appendix A:	Derivation of the Conditions where Type-1 Patients are Preferred and Strongly Preferred	80
Appendix B:	Hizmet Hospital	96
Appendix C:	Occupancy Rates of Hizmet Hospital	97
Vita		100

LIST OF FIGURES

3.1	Arrival and departure stream of type-1 and type-2 patients	21
4.1	Flowchart of the derivation of $\bar{d}_1$ and $\bar{d}_2$ expressions	43
4.2	Condition 5 and 6 when $\lambda_1\lambda_2 \leq \lambda_1(\lambda_1 + 2\lambda_2 + \mu_2) \leq \mu_1(\lambda_2 + \mu_2)$. . .	52
4.3	Condition 1-2-3 and 4 before (on the left) and after the arrangement (on the right) $\mu_2(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$	53
5.1	Flowchart of the data preparation	58
5.2	Differential cost (h) , for initial state (x_1, x_2)	61
C.1	Monthly average occupancy rates of surgical branches (Ward-1) and gynecology and obstetrics (Ward-2)	97
C.2	Monthly average occupancy rates of the empty ward (Ward-3) and internal branches (Ward-4)	98
C.3	Monthly average occupancy rates of cardiology and KVC (Ward-5) and urology and transplantation (Ward-6)	99

LIST OF TABLES

4.1	Preferred and strongly preferred conditions when $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$	50
4.2	Preferred and strongly preferred conditions when $\mu_2(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$	52
5.1	Department transactions of Memorial Hizmet Hospital	56
5.2	Parameter Estimations (Base case)	60
5.3	Decisions when a type-1 patient arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. (Base case with $\overline{M} = 25$)	62
5.4	Decisions when a type-2 patient arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. (Base case with $\overline{M} = 25$)	63
5.5	Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. (Base case with $\overline{M} = M$)	64
5.6	Average cost values when c_r^1 changes	65
5.7	Optimal actions when c_r^1 changes (S: Same as the base case)	65
5.8	Decisions when a type-1 patient arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. ($l = 60\%$ and $\frac{c_r^1}{c_r^2} = 1.5$)	66
5.9	Decisions when a type-1 patient arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. $\frac{c_r^1}{c_r^2} = 1.5$ and $l = 85\%$ (on the left-L), $l = 110\%$ (on the right-R)	67

5.10	Decisions when a type-1 patient arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. $\frac{c_r}{c_e} = 2$ and $l = 85\%$ (on the left-L), $l = 110\%$ (on the right-R)	68
5.11	Average cost values when c_e changes	69
5.12	Optimal actions when c_e changes (S: Same as the base case)	69
5.13	Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. ($c_e = 8$ and $l = 60\%$)	70
5.14	Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. ($c_e = 8$ and $l = 85\%$)	71
5.15	Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: ($\square$) Accept, (Δ) Accept with early discharge, ($\star$) Accept with overflow, ($\blacktriangle$) Reject. ($c_e = 8$ and $l = 110\%$)	71
B.1	Hizmet Hospital floor plans, their capacities and allowed patients in these floors	96

Chapter 1

INTRODUCTION

Operations research in healthcare can be divided into three main categories: medical decision making, public health and healthcare operations. Medical decision making generally focuses on patient-based problems. One of the objectives in this field is to assess the best decision about having a surgery, biopsy, etc. to maximize quality-adjusted life time of a certain patient. Public health is concerned with the health-care problem of the society. For instance, determining the most effective time period for screening a disease to maximize quality adjusted life years (QALY) can be one of the goals in the public health literature. On the other hand, healthcare operations focuses on managerial aspects. Handling the demand to obtain a stable occupancy rates is important in hospital management.

Growing and aging population raises the demand for healthcare. While demand is evolving, the hospital resources (e.g. beds, nurses, equipment, etc.) remain relatively constant (Denton, 2013, Ch. 1). As a consequence of the imbalance between high demand and inadequate supply, many patients face difficulties while accessing care.

Hospital beds are limited resources. Therefore, they may become a bottleneck of hospital systems when occupancy rates are high. In the United Kingdom, the Department of Health stated that if the bed-occupancy rates exceed 85%, there exist problems in handling both emergency and elective admissions (UK (2000)). Overcrowding in emergency departments, blockages and diversions, cancellations due to backlogs are some of these problems, which cause fluctuations in patient census (also known as occupancy rate or utilization). Controlling inpatient admissions prevents

these problems, therefore, the variability in occupancy rates decreases. Hence, hospitals develop their own admission and discharge policies.

In this research, we propose a Markov Decision Process (MDP) model to control inpatient admissions when the beds are limited. We observe the patient flow in a private Turkish hospital called Hizmet Hospital. We find the optimal policies that a manager should take to minimize expected cost induced by these policies. We assume that only two types of inpatients arrive to the hospital: severe and mild. We aim to develop insights on the control of the inpatient admissions to the wards and mild patient discharges.

Currently, capacity management problems are generally modeled with single server or parallel servers. The main contribution of this research is that, we study a bed management problem in which a severe patient has a two stage recovery process. We assume that a severe patient becomes a mild patient when he/she completes the first stage of his/her recovery. Hence, the recovery time of a severe patient is modeled as a two stage hypoexponential distribution, whereas the recovery of a mild patient takes an exponential time. Both mild and severe patients have their own arrival rates. Secondly, we allow mild patients to overflow to another ward (overflow ward) with a holding cost, whereas severe patients are allowed to stay in the main ward. Finally, by finding sufficient conditions we can determine that an incoming severe/mild patient is always accepted unless the wards are completely full, which we define as preferred, following the convention in the literature. We can always admit a severe patient even if all beds are occupied in the main ward by discharging a mild patient early. If this is the optimal action, then severe patients are called strongly preferred. All of these conditions give practical hospital admission policies, which can improve the healthcare delivery.

The rest of this thesis is organized as follows: A comprehensive literature review about operating room (OR) scheduling and planning, queuing models and certain techniques that are used in healthcare is presented in Chapter 2. In Chapter 3, we give

the background about bed capacity management to motivate our research and through this motivation we present our model. Some preliminary results and conditions where patients are preferred and strongly preferred are provided in Chapter 4. Chapter 5 implements the model using a dataset from Hizmet Hospital. Finally, the research is concluded in Chapter 6.

Chapter 2

LITERATURE REVIEW

In this section, we will be presenting related literature about healthcare operations and some techniques to analyze these models. In our study, we aim to develop insights on the admission control of patients to the wards. Most patients enter the wards after going through surgery. Hence, next section (Section 2.1) presents a literature review on OR planning and scheduling. We should note that, this subject is especially important since ORs are the highest cost and revenue resources (Macario et al. (1995)).

Modeling the patient flow is one of the key factors in healthcare operations. Patient flow covers all stages of medical care starting from taking an appointment to the departure from the hospital upon recovery. It is generally represented by a queuing or a process simulation model.

Discrete-event simulation (DES), or process model simulation, is used to represent the whole system behavior and to understand the system performance over time under internal and external resource constraints. It is also a tool used to model and evaluate the results of the healthcare delivery in hospitals. Thus, hospital administrators can find the efficiency of the current system, analyze the effects of possible changes in the system or design a new system through DES. Jun et al. (1999) is a review on DES for single or multi-facility healthcare clinics. The author reviews some DES models in which patient/OR scheduling and admission policies and patient routing especially in emergency rooms (ERs) are considered. Furthermore, some DES models about allocation of resources such as bed, room or staff are provided in Jun et al. (1999).

Since we study a queuing model, we will investigate queuing models in a detailed

way in Section 2.2.

2.1 Operating Room Planning and Scheduling

Magerlein and Martin (1978) review OR scheduling based on two different scheduling techniques: advance scheduling and allocation scheduling. In allocation scheduling, surgery time and OR room is assigned to the patient assuming that all patients are in the hospital and ready for the surgery. In advance scheduling, the patient is scheduled for upcoming days. Resource constraints are available OR time, bed and nurse. Major booking strategies of advance scheduling are blocked booking and non-block booking. In block booking, OR time is allocated to surgeons or surgical groups. In non-block booking, the first-come first-served policy is applied. Although Magerlein and Martin (1978) is not a recent review, it defines classification of techniques that are still in use.

Patient mix is a widely used concept in OR planning and scheduling. It refers to a group of patients who belong to different types of specialties. The aim of case mix planning is to choose the optimal patient mix and volume of each pathology to maximize profit under resource constraints of hospitals.

Beliën and Demeulemeester (2007) decompose OR scheduling in three different stages. The first stage determines how much available OR time is reserved to each surgical group. This phase is a part of case mix planning because in this stage capacity is allocated according to disease types, which affects quality of service. In the second stage, master surgical scheduling (MSS), which is a part of block scheduling, is developed. MSS is a cyclic timetable that determines the number and type of operating rooms available in specific hour blocks. Patient scheduling is done in the last stage. This phase determines the number of surgeries performed in a day, the starting time of each surgery and their sequence.

In the literature, modeling one stage is preferred to modeling all three stages together. Blake et al. (2002) place ORs with surgical group into a weekly timetable

to obtain master surgical schedule. They assign a whole block time to the surgical group instead of using block times. The problem is modeled with linear integer programming with prespecified target OR times for each surgical group, minimizing the weighted sum of difference between target block time and the given block time. The model is run with 97% of accuracy while more than four ORs exist in the hospital. The model is implemented in Mount Sinai hospital and the implementation is studied in Blake and Donald (2002).

The three stage definition is not a good categorization technique because of different definitions of categorization. Thus, Cardoen et al. (2010) review the concept of OR scheduling and suggest six different headings to meticulously categorize it. Initially, they focus on patient characteristics and show that in the literature, models deal with elective patients more than non-elective ones who undergo an urgent unexpected surgery. Elective patients are divided into two categories: inpatient (who stays in the hospital at least one night) and outpatient (who do not spend a night in hospital). Non-elective patients are divided into two categories: emergent patients cannot wait the surgery whereas surgery of urgent patients can be postponed for another hour during the day. Moreover, performance measures constitute another field to compare studies in the literature. Waiting time, utilization and overtime are the most used performance measures in the literature. Decision delineation is another category. Decisions about date (e.g. assign patient on 10th December), time (e.g. assign patient at 10:00 a.m.), OR (e.g. OR 1) or capacity (e.g. two hours of OR time) are in two levels: patient and surgeon. Patient level decisions are more popular in the literature.

Up to this point, we provided diverse review about OR planning and scheduling and identify how it can be categorized. In our model we consider bed number as a capacity constraint. Hence, we are moving to MSS problems in which beds are used as resource.

Van Oostrum et al. (2008) define MSS differently as a schedule that determines

a combination of operations for each OR-day. They focus on frequently performed surgery types, which have to be assigned in each cycle. They find a combination of operations for each OR in a day. The main target is to optimize the number of OR and OR resources, which counts not only beds as a resource but also other resources. They formulate the MILP model which has a stochastic OR capacity constraint. Two phase decomposition approach which levels bed requirements is formulated and heuristically solved. Numerical analysis is based on data taken from Erasmus Medical Center.

When bed resource constraint is part of MSS models, length of stay (LOS) distribution can also be considered. Beliën and Demeulemeester (2007) study a MSS by which bed occupancy is leveled. They use stochastic arrival and LOS. Performance measures are the daily expected bed occupancy, variance of the occupancy, expected bed shortage and daily shortage probability. The model has a non-linear objective function whose capacity constraint limits the OR blocks on each day and the demand constraint guarantees the specific number of OR block time for each surgeon. This problem minimizes total expected bed shortage. To solve the problem, non-linear objective function is linearized and obtains the mixed integer programming (MIP) problem. The problem is solved by heuristics such as the repetitive MIP heuristic and the quadratic MIP heuristic. Moreover, the main objective function is solved by meta-heuristic (i.e. simulated annealing). Although simulated annealing gives the best results, MIP heuristics are better in terms of ease of the problem adaptation and less computational complexity.

Beliën and Demeulemeester (2007), which uses bed as a resource constraint and provides a theoretical approach to master surgery scheduling, Beliën et al. (2006) present a simulation tool to visualize block scheduling effect on renewable resource consumption pattern. Renewable resources such as nurses, equipment and machines renew each period. The tool is implemented in the Belgian university hospital. Although the model in Beliën and Demeulemeester (2007) uses the stochastic approach, the model of Beliën et al. (2006) is deterministic and simpler but does not cover LOS.

Santib   ez et al. (2007) also use impact of resources on master scheduling. They interpret master surgical scheduling as block scheduling, modeled using MILP tools under demand for OR time as throughput and limited post-surgical resource constraints. The model is applied to hospitals in the British Columbia Health Authority. Patients belonging to different specialties show different LOS and can be inpatient or outpatients. Thus, utilization of resources is affected by surgery type. Adan et al. (2009) also maximize patient throughput under resource utilization to find MSS as in Santib   ez et al. (2007). Although Santib   ez et al. (2007) uses deterministic LOS, LOS is probabilistic in Adan et al. (2009).

2.2 Queuing Models

Queues in healthcare facilities (i.e. hospitals and clinics) form when patients wait for service because other patients are using the resources to complete their treatment. According to Kendall’s notation (Kendall (1953)), each queue is described by an arrival process, a service time distribution, number of servers. In healthcare literature, servers can be beds, doctors or ORs. Doctors are generally depicted as servers in appointment scheduling. ORs can be represented as a server in the area of OR planning and scheduling. Beds in wards or in intensive care units (ICUs) can be also modeled as servers in admission scheduling. Our model falls into this category.

Queuing models can be separated into two categories: queuing models and loss models (queuing models with no waiting room). A detailed investigation about queuing models is given in the appointment scheduling section.

2.2.1 Appointment Scheduling

Patients are assigned appointment times in appointment scheduling. Uncertainty in the system is a handicap for realistic appointment schedules Denton (2013). Queuing theory is generally used in appointment scheduling.

Cayirli and Veral (2012) and Gupta and Denton (2008) provide comprehensive reviews about appointment scheduling. Gupta and Denton (2008) state that appointment scheduling is used by primary care, specialty care and elective surgery. Primary physicians act as principal points of consultation in primary care. Special care focuses on complex diagnosis whereas elective surgery deals with specific procedure. The authors state that three factors affect the performance of appointment scheduling: (1) arrival and service variability, (2) the preferences of patients and providers, and (3) performance measures (e.g. patient waiting time and overtime).

Zeng et al. (2010) focus on scheduled patients with different no-show probabilities to make an appointment schedule with overbooking. Service time is assumed to be exponential, independent and identically distributed (i.i.d.) across patients. The day is divided into time slots and first come first served logic is applied to make appointments. The objective is to maximize expected profit, which is the difference between the expected revenue and the expected overbooking cost (i.e. physician overtime cost and waiting time cost). This objective function is difficult to solve unless it is multimodular and convex, and constraints have upper and lower bounds. If the objective function has those properties, a local search algorithm is used to solve the objective function to obtain a global optimal solution. The homogeneous patient case is multimodular; however, for heterogeneous patients it is not multimodular. Thus, heuristic methods are used to solve the objective function.

Green and Savin (2008) study appointment scheduling while appointment lengths are equal. They use the single server queuing system to represent the physician appointment system in which cancellations and no-shows occur. In such a system, patients have state-dependent probability of not being served and may re-attend the queue. The authors find the number of patients per physician through simulation. Their model can assign the number of patients more than capacity to handle no-shows. The model considers backlog-dependent cancellation rates and also provides upper and lower bound for backlog while the number of patients per physician is large.

Creemers and Lambrecht (2010) analyze appointment driven systems. A patient is initially in the waiting list queue (appointment making queue, AMQ) then at the inner queue on the appointment day (service facility queue, SFQ). Thus, two queues exist in such systems and each requires different modeling approaches. AMQ is a vacation queue and SFQ is the appointment system. The block appointment rule was selected for SFQ to manage the arrival of customers at the service facility. By combining two kinds of queues, the authors model a single appointment driven model which assesses performance measures. Performance measures are time spent on the waiting list, waiting time and servers out time. These measures determine the server capacity planning; i.e. how often, when and for how long a server should be online.

2.2.2 Loss Queuing Models

A loss queuing model is a queuing model with no waiting room. In these kinds of models, the main problem is due to refused admissions. Therefore, blocking probability is generally used as a performance measure. Queuing models are analyzed in different contexts: (1) time-varying arrival streams, (2) representation of ICUs, and (3) control of loss queuing models.

2.2.2.1 Admission planning and scheduling

Arrival streams to healthcare facilities vary by time. For example, generally most of the doctors do not work on weekends, therefore, we expect to see different arrival rates in weekends and weekdays. Time-dependent modeling of complex systems like hospitals is hard to achieve, so that most of the health care literature covers homogeneous arrivals. Bekker and Kooleman (2011), de Bruin et al. (2010) and Helm and Van Oyen (2011) observe impact of fluctuating arrivals on bed occupancies by using constant arrival rates. On the other hand, de Bruin and Bekker (2010) handle this problem by analyzing the time dependent loss queuing model.

Patients arrive at the hospital in two different ways: scheduled and unscheduled.

de Bruin et al. (2010) demonstrate that although the number of unscheduled patients can be modeled as a Poisson random variable, scheduled patients cannot. However, scheduled admissions can be described by the Poisson random variable, if they split into weekdays and weekends. Thus, the authors convert every arrival pattern into a Poisson process. They use the Erlang loss ($M/G/c/c$) model to describe capacity plan of smaller units to find the fraction of refused admissions and occupancy rates. They show that using fixed 85% occupancy rate is not feasible for smaller units. Moreover, the researchers develop a decision support system to find the current size of nursing units and number of beds allocated to each ward. The authors suggest that mixing patient flows or merging departments, in which beds are merged, can be applied to obtain an acceptable service level and an economically feasible occupancy rate in circumstances such as limited budget or high number of admissions. Pooling two care units reduces the number of beds from 29 to 22 where 5% blocking rate is achieved in both cases.

Fluctuations in the admission cause variation in bed occupancy and increase the number of rejected patients. Bekker and Kooleman (2011) focus on the impact of variability of admissions and length of stays on number of required bed. They use a heavy traffic limit theorem for $G/G/\infty$ queue to approximate processes which are not Poisson (i.e. scheduled arrivals), so that number of busy servers (non-idle beds) form a normal distribution. Hence, they observe the impact of fluctuations on bed occupancy. Moreover, they interpret Gini coefficient concept used in economics as the inequality in the LOS among patients. Peakedness of normal distribution depends on the gini-coefficient and variance of non-idle beds depends on the peakedness. As a remark, gini coefficient concept is also used in de Bruin and Bekker (2010) and de Bruin et al. (2010) as in Bekker and Kooleman (2011). Bekker and Kooleman observe that reducing variability in LOS becomes meaningful when there exist stable (elective) arrival stream. Another aim in this study is to determine an admission quota for scheduled patients. Given the weekly admission pattern, they find time-dependent

offered load per day and use this information to develop a quadratic programming model (QPM). This model finds an admission quota that minimizes the gap between the expected desired occupied number of beds.

Helm and Van Oyen (2011) focus on variations in the elective patient admissions. They study a stochastic model for the patient flow. The authors suggest a stochastic location model which describes location (bed) of a patient over time randomly. Poisson arrival process is used in the stochastic location model. Helm and Van Oyen use this model in the emergency census level in a ward. Arrivals to emergency wards form an independent non-stationary Poisson process. The authors combine the elective census process with the emergency census process. This total process is linked with the elective admission decision variables in the optimization model which finds the optimal type and number of patients under performance constraints.

de Bruin and Bekker (2010) analyze the impact of a time-dependent arrival pattern on the performance (workload and blocking probability) and bed capacity by using the $M_t/H/c/$ queuing model, where arrivals are coming from a time-dependent Poisson process and service times follow a hyperexponential distribution. The arrival rate is piecewise constant to show different time periods of the day. Capacity is restricted by the nursing staff or the number of beds. The infinite server queue approximations are used to determine performance measures (i.e. time-dependent offered load and blocking rates). The authors indicate that arrival patterns change with different time scales. Thus, daily and weekly arrival rate variations are used. For the daily variation pattern, two different arrival rates (one of them represents the rate during office hours and the other one represents the rate in the other times) and exponential LOS distribution is assumed. For different average LOS cases (4 days, 1.5 days and 1.5 hours), daily offered workload and fraction of refused admissions across day are examined for a fixed capacity level. Fluctuations across day gain a big impact while LOS decreases. Hence, if LOS is substantially larger than the arrival cycle length, fluctuations have a limited impact on the performance measures. On the other hand,

the arrival cycle length for the weekly arrival rate is similar with LOS. Therefore, fluctuations have negative impact on weekly offered workload and loss fraction. De Bruin and Bekker show that an increased variability in LOS stabilizes variability in refused admissions.

2.2.2.2 Policy Modeling in ICUs

KC and Terwiesch (2012) observe that the intensive care unit (ICU) is more expensive than wards for in-patient treatment and is one of the bottlenecks of the patient flow due to its high occupancy levels and costly equipment. ICU is the only recovery room for severe patients. Generally, severe patients cannot wait for an empty bed in ICU, so that ICU related problems are generally solved by using loss queuing models. Due to high occupancy levels of ICUs, to admit new-arrived patients, discharging policies are applied to already existing patients. However, early discharge increases the probability of complications which results in re-admission of early discharged patient to the hospital. Dobson et al. (2010) and Chan et al. (2010) study the so-called bumping policies, whereas re-admission is considered by Chan et al..

Dobson et al. (2010) analyze a demand-driven discharge policy by which patients who have the shortest remaining LOS are discharged to admit newly arrived patients to the ICU. Patients come to the ICU as scheduled and unscheduled. The arrival stream and the service time have uncertainty. If there is enough bed capacity, an arrived patient is assigned to this bed; otherwise, some of the patients are discharged early according to the bumping policy. The state of the system is the remaining LOS for each bed. Since the state space is too large, an aggregation-disaggregation method is used to reduce computational complexity. Performance measures (i.e. probability of being bumped, the expected number of remaining days after being bumped and the utilization of the ICU capacity) are calculated in computational experiments for different surgical schedule cases (i.e. 3 days, 5 days and 7 days) and for different properties of scheduled and unscheduled arrivals. According to the results, increasing bed

capacity decreases the probability of being bumped. Moreover, when the unscheduled arrival proportion increases, the demand driven discharge rate also increases. A three-day plan gives the patients more time for recovery, which results in clearing out the system, therefore the probability of being bumped decreases. This analysis shows that demand-driven discharges are unavoidable and affected by various parameters.

Instead of indexing according to remaining LOS to determine which patients should be discharged as in Dobson et al. (2010), the demand discharge policy of Chan et al. (2010) depends on reducing readmission load ($\frac{p_m}{\mu_m^R}$) where p_m is probability of readmission and LOS after readmission has geometric distribution with mean $(1/\mu_m^R)$. Chan et al. (2010) suggest an index policy to serve as a decision support tool. A greedy heuristic has been applied to find the policy. Due to the nature of greedy heuristic, this approach ignores the effect of future arrivals and LOS of current patients. The resulting discharge policy guarantees readmission load at worst $\hat{p} + 1$ times the optimal discharge strategy where $\hat{p}$ is the utilization ratio in which small $\hat{p}$ value means more available beds.

2.2.2.3 Admission control models using MDP

The last part we focus on under loss queuing models, is the control models which contain decision processes and stochastic parameters. Markov Decision Process (MDP) models are a frequent technique to analyze the decisions in queuing models in which stochastic parameters preserve the Markovian property. First we focus on event-based dynamic programming by which we can analyze the properties of an MDP model easily. Then, we review Helm et al. (2010), Örmeci and Van der Wal (2006) and Ulukus et al. (2011), which have studies on MDP models.

In event-based dynamic programming, the operators are defined to represent the events and the general model is constructed as a combination of the operators. Properties of the general model are determined by analyzing the properties of the operators. Koole (2006) studies the monotonicity behaviors through the operators and suggest

the optimal policies for the problems. Cil et al. (2009) investigate the impact of the system parameters on the optimal policy by analyzing the operators. Cil et al. modify some operators defined in Koole (2006) and present the properties preserved by each operator such as monotonicity or convexity.

Helm et al. (2010) study a MDP model for the admission control in hospitals. They divide incoming patients into three groups: emergent patients, scheduled patients, which can be canceled, and expedited patients, which are determined as a part of emergent patients but not really an emergent case. Therefore, they put expedited patients in a queue to wait call-in. They use uniformization to discretize the continuous time model and use event based dynamic programming for convexity analysis. The authors show that optimal threshold policies exist.

Örmeci et al. (2001) study a loss system with two types of jobs whose arrivals form a Poisson distribution and service times are exponential. They maximize the total expected discounted reward. Örmeci and Van der Wal (2006) generalize the finding in Örmeci et al. (2001) and present a model in which arrivals have a general distribution. They construct a semi MDP model to maximize the total expected discounted revenue and the long-run average revenue. Both of these models have two decisions: accept and reject. The optimal admission policies and the preferred classes under specific conditions are determined in both works. A preferred class is defined as a class which is always admitted to the system when a free server exists. They support their findings with a Markovian example, when arrivals have a general distribution.

Ulukus et al. (2011) also study a loss system with arrival of two types of jobs with Poisson distribution. They construct a discrete time MDP model to maximize the total expected discount return as in Örmeci and Van der Wal (2006). However, in Ulukus et al. (2011) another decision exists: termination of one type of job in the server. They find the optimal threshold policies, the preferred and strongly preferred classes for this model.

Our model is extended version of the model defined by Ulukus et al. (2011). Initial

difference is that the service time distribution of one class has a hypo-exponential distribution with parameters μ_1 and μ_2 while the other class has an exponential service time with μ_2 . The other distinction is that overflow of a specific type of patient to the other ward is permitted.

Chapter 3

PROBLEM DESCRIPTION AND MATHEMATICAL MODEL

3.1 *Bed Management and Bed-Capacity Modeling*

Occupancy rates are directly affected by arrival and departure streams. Doctors visit patients to evaluate their current circumstances in the morning hours. During these evaluations, doctors also decide to discharge patients. Patients cannot be discharged without their doctors' approval. This means that, hospital beds become available for new patients only after late morning hours. On the other hand, incoming in-patients generally arrive to hospital before the doctor evaluations (and so discharges) take place. This conflict between admissions and departures creates an admission problem for patient arrivals and discharges in congested hospitals. In order to eliminate this conflict, hospitals develop their own admission and discharge policies.

We observe the patient flow in Hizmet Hospital, which encounters similar problems as described above. In this hospital, each floor is specialized for one medical branch, e.g., gynecology, general surgery and cardiovascular surgery, and there exists an empty floor that is used when necessary. Although each patient has his/her own room in a floor, we refer to that floor as a “ward” in this study. Patients who have the same type of diseases stay in the same ward.

When a new patient arrives and registers at the hospital, the patient is assigned to a bed in the corresponding ward, if there is an empty bed. A patient can be also assigned to a bed in other wards, if there is no room in his/her own ward. However, there exist strict rules for this assignment (see Appendix B). For example, gynecology ward does not accept any patients from other departments because their nurses are

specialized. After the patient is directed to his/her room, identity bracelets are put and the patient fills a survey about his/her current condition. Their medical tests are conducted and if the results are in desired levels, he/she undergoes surgery.

The above flow is valid when there is an empty bed. If there is no empty bed, the most common action, which Hizmet Hospital takes, is to conduct medical tests while the incoming patient is waiting for an empty bed. Moreover, if a high number of inpatients is expected on a certain day, doctors are informed about the situation and encouraged to discharge their patients early if they are in relatively good health condition, which we call as “early discharge.” However, this can cause a relapse for the patient. Sometimes, even this policy cannot sufficiently decrease the congestion, and patients have to wait for beds. To handle this situation, Hizmet Hospital opens their empty ward (additional floor). To operate this ward they need to hire temporary nurses. However, according to hospital managers, finding new nurses is a hard task because the number of nurses is insufficient in the sector and their fees are high. Hence, opening this ward is regarded as a last resort. Furthermore, if there are no places left and an emergency patient comes, their priority is to stabilize the patient condition and to transfer him to another hospital. Stabilization procedure occurs in the emergency room, hence the patient does not occupy an in-patient bed. Rejecting an in-coming patient is also a valid option if his condition is not urgent. Nevertheless, it harms hospital reputation.

We model Hizmet Hospital’s bed management system and analyze the admission problems stated above. In our model, the system has $\overline{M}$ identical and parallel servers (beds). M of $\overline{M}$ beds belong to ward-1, and the remaining beds are in ward-2. Ward-1 is specialized for one type of disease (main ward) and Ward-2 is an overflow ward just like the empty ward in Hizmet Hospital as described above. We assume that, only patients who belong to ward-1 can overflow to ward-2 so that there is no flow into ward-2 from other wards.

We generalize patients who come to ward-1 and group them into two different

types: type-1 and type-2. We assume that type-1 represents more serious patients and their LOSs consist of two stages. In the first stage, a type-1 patient recovers to become a mild patient. (We use mild, light or normal patient interchangeably.) In second stage, a light patient recovers and leaves the hospital. Light patients are defined as type-2 patients. Hence, a type-1 patient becomes a type-2 patient after the first recovery stage. Moreover, a normal patient can also arrive to the hospital without being a type-1 patient.

Roberts (2012) studies the impact of overflow on LOSs for four different departments. There is no difference among three of the departments (Cardiology, Gastroenterology and Respiratory Medicine) in terms of LOSs. However, readmission rate of the overflow patients in cardiology and respiratory medicine departments are higher than that of non-outliers. On the other hand, overflow patients of general medicine stay shorter than patients who stay in their own ward. The author indicates that overflow patients receive diminished quality of care. Communication between overflow ward and the main ward can be difficult and lags may occur during the treatment of outliers.

With this motivation, we assume that a type-1 patient can only be admitted to beds in ward-1, since ward-1 is specialized for the disease under consideration. On the other hand, type-2 patients, who stay shorter than type-1 patients, can be admitted to both ward-1 and ward-2. However, being in ward-2 costs c_o for each type-2 patient, which corresponds to the cost of temporary nurses in Hizmet Hospital.

When a patient is rejected, if he is a severe patient (type-1), he is transferred to another hospital by Hizmet Hospital. If he is a normal patient, we assume that he goes to another hospital and the hospital loses patients' goodwill and/or the revenue to be generated from that patient. A rejection cost of c_r^i is incurred when a type- i patient is rejected. We assume that only patients with normal condition can be discharged early with a cost of c_e . We assume that $c_r^2 < c_e < c_r^1$. The consequence of rejecting a serious patient (type-1) can be significant since the health condition

of such a patient may deteriorate considerably which may call for a longer and/or more heavy treatment. Hence, the cost of rejecting a type-1 patient is taken as the highest cost. Furthermore, interrupting the treatment of a normal patient has more significant drawbacks than rejecting the patient because an unfinished remedy can cause more serious health problems than the original one, so that the same patient can be re-admitted to the hospital. On the other hand, the refused patient can receive his treatment from another healthcare facility. In that case the only cost is due to the loss of goodwill. At the end, terminating the remedy of a patient is more expensive than rejecting him/her.

3.2 Markov Decision Model

In this section, we develop a Markov Decision Process (MDP) model based on the above assumptions. The model is a two-class loss system with the decisions of admission, rejection, admission with early discharge and admission with overflow. Modeling hospital beds as a loss system, i.e., a queuing system with no waiting room, is a common approach in the literature, see e.g. de Bruin and Bekker (2010) and Helm et al. (2010).

We assume that a serious patient arrives to the system according to a Poisson process with rate λ_1 and the first stage recovery takes an exponential time with rate μ_1 . After the first stage, he/she becomes a type-2 patient. Recovery time of type-2 patient is an exponential distribution with rate μ_2 . Our two stage assumption for a severe type of patient, creates an hypoexponential distribution. In hypoexponential distributions, there exists a series of k exponential distributions and i^{th} exponential distribution has a rate of γ_i . For our case, $k = 2$ and γ_i is equal to μ_i for $i = 1, 2$.

The system can have a type-2 patient in two different ways: The first one is due to the two stage definition of type-1 patients as we discussed above. The second one is due to type-2 patients' own arrival process which is assumed to be a Poisson process with rate λ_2 .

$\overline{M}\}$.

Uniformization: Arrivals and departures occur in continuous time. Uniformization converts this continuous time system into a discrete time model (Lippman (1975)). To use this technique, we need to find the maximum rate out of any state. The next event can be:

- I. Arrival of a type-1 patient where interarrival times are exponentially distributed with rate λ_1
- II. Arrival of a type-2 patient where interarrival times are exponentially distributed with rate λ_2
- III. Recovery of a type-1 patient from serious to normal condition, which takes a exponential time with rate μ_1
- IV. Departure of a type-2 patient which takes a exponential service time with rate μ_2 .

We know that, if $X_1, X_2, \dots, X_n$ are independent exponential random variables with rates $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively. Then, $\min \{X_1, X_2, \dots, X_n\}$ is also an exponential random variable with rate $\lambda_1 + \lambda_2 + \dots + \lambda_n$. Since the events of our model are independent, the expected time until next event is exponential random variable with rate $\lambda_1 + \lambda_2 + \mu_1 x_1 + \mu_2 x_2$. Then, the maximum possible rate out of any state occurs when $x_1 = 0$ and $x_2 = \overline{M}$ since $\mu_1 \leq \mu_2$. Then, we can normalize the time unit so that $\lambda_1 + \lambda_2 + \overline{M}\mu_2 = 1$ and we can observe the system in exponential time intervals with rate one. In the discrete models at an observation epoch, a type- i patient arrives with probability λ_i , a type-1 patient becomes type-2 with probability $x_1\mu_1$, a type-2 patient leaves the system with probability $x_2\mu_2$ and with probability $\overline{M}\mu_2 - x_1\mu_1 - x_2\mu_2$ fictitious type-2 patients in empty beds leave the system. Fictitious terms are used to keep the rate in any state constant and do not cause any state change.

Value Function: We define value function $V_n(x_1, x_2)$ as the minimum expected cost for the system starting at state (x_1, x_2) . The initial condition is assumed to be zero (i.e. $V_0(x_1, x_2) = 0$) for mathematical convenience, and has no impact on the infinite-horizon problem. Moreover, $U_{n-1}(x_1, x_2, i)$ is the minimum expected cost in state (x_1, x_2) when a type- i patient arrives and n observation points remain in the horizon.

Actions: The model takes a decision when a type- i patient arrives. There are four different decisions to take: (1) Accept the type- i patient, (2) reject the type- i patient, (3) accept the type- i patient by overflowing one type-2 patient and (4) accept the type- i patient by early discharging a type-2 patient. By the early discharge decision we terminate the treatment of one type-2 patient. By the overflow of type-2 patient, one type-2 patient moves to an extra bed in ward-2.

Parameters:

M : Number of beds

$\overline{M}$: Total bed capacity (including the overflow beds)

c_r^i : Rejecting cost of type- i patient

c_e : Early discharge cost of a type-2 patient per patient

c_o : Cost of overflow of a type-2 patient per unit time per patient

λ_i : Arrival rate of type- i patient

μ_i : Service time of type- i patient

Model

$$\begin{aligned}
 V_n(x_1, x_2) = & \lambda_1 U_{n-1}(x_1, x_2, 1) + \lambda_2 U_{n-1}(x_1, x_2, 2) \\
 & + \mu_1 x_1 V_{n-1}(x_1 - 1, x_2 + 1) + \mu_2 x_2 V_{n-1}(x_1, x_2 - 1) \\
 & + \max\{0, x_1 + x_2 - M\} c_o + (\overline{M} \mu_2 - x_1 \mu_1 - x_2 \mu_2) V_{n-1}(x_1, x_2),
 \end{aligned} \tag{3.1}$$

The first two terms represent a type- i patient arrival with probability λ_i for $i = 1, 2$ and the expected cost function at a patient arrival epoch. U_n functions will be specified below. The third term represents the first stage in the recovery of a type-1 patient,

upon which a type-1 patient becomes a type-2 patient. The fourth term represents the discharge of a type-2 patient. When the total number of patients is more than M , a holding cost is incurred due to the overflow of patients, which is represented by the fifth term. The final term corresponds to the fictitious events.

The system takes an action, which may or may not change the state at the patient arrival epochs. When a type-1 patient arrives to the system, he/she can either be accepted or rejected with a cost of c_r^1 or accepted with early discharge of type-2 patient which causes a cost of c_e . When a type-2 patient arrives to the system, he/she can either be accepted or rejected with a cost of c_r^2 or accepted with discharge of a type-2 patient which causes a cost of c_e . When a type-2 patient arrives and early discharge decision is taken, the state stays in x and $c_e \geq c_r^2$, so that early discharge decision cannot be optimal when a type-2 patient arrives. Hence, we exclude this decision for type-2 arrivals. When a type- i patient is admitted, the state moves from x to $x + e_i$ for all $i = 1, 2$, where e_i is a two-dimensional unit vector. If a type- i patient is rejected, the state remains the same for all $i = 1, 2$. When type-2 patient is discharged early and type-1 patient is admitted, the state moves from x to $x + e_1 - e_2$. Depending on the number of type-1 and type-2 patients, the decisions that are possible change, which result in six cases for $U_{n-1}(x_1, x_2, i)$:

Case 1: $x_1 < M, x_2 > 0$ and $x_1 + x_2 < M$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= \min\{V_{n-1}(x_1 + 1, x_2), V_{n-1}(x_1, x_2) + c_r^1, V_{n-1}(x_1 + 1, x_2 - 1) + c_e\} \\ U_{n-1}(x_1, x_2, 2) &= \min\{V_{n-1}(x_1, x_2 + 1), V_{n-1}(x_1, x_2) + c_r^2\} \end{aligned} \quad (3.2)$$

Case 2: $x_1 < M$ and $x_2 = 0$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= \min\{V_{n-1}(x_1 + 1, x_2), V_{n-1}(x_1, x_2) + c_r^1\} \\ U_{n-1}(x_1, x_2, 2) &= \min\{V_{n-1}(x_1, x_2 + 1), V_{n-1}(x_1, x_2) + c_r^2\} \end{aligned} \quad (3.3)$$

Case 3: $x_1 < M$ and $M \leq x_1 + x_2 < \overline{M}$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= \min\{V_{n-1}(x_1 + 1, x_2), V_{n-1}(x_1, x_2) + c_r^1, V_{n-1}(x_1 + 1, x_2 - 1) + c_e\} \\ U_{n-1}(x_1, x_2, 2) &= \min\{V_{n-1}(x_1, x_2 + 1), V_{n-1}(x_1, x_2) + c_r^2\} \end{aligned} \quad (3.4)$$

Case 4: $x_1 = M$ and $x_1 + x_2 < \overline{M}$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= V_{n-1}(x_1, x_2) + c_r^1 \\ U_{n-1}(x_1, x_2, 2) &= \min\{V_{n-1}(x_1, x_2 + 1), V_{n-1}(x_1, x_2) + c_r^2\} \end{aligned} \quad (3.5)$$

Case 5: $x_1 < M$ and $x_1 + x_2 = \overline{M}$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= \min\{V_{n-1}(x_1, x_2) + c_r^1, V_{n-1}(x_1 + 1, x_2 - 1) + c_e\} \\ U_{n-1}(x_1, x_2, 2) &= V_{n-1}(x_1, x_2) + c_r^2 \end{aligned} \quad (3.6)$$

Case 6: $x_1 = M$ and $x_1 + x_2 = \overline{M}$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= V_{n-1}(x_1, x_2) + c_r^1 \\ U_{n-1}(x_1, x_2, 2) &= V_{n-1}(x_1, x_2) + c_r^2 \end{aligned} \quad (3.7)$$

Note that, although U_n functions of Case 1 and Case 3 are the same, we use different cases to represent them. In Section 4.1 we will show that, early discharge decision is not optimal when there exist at least one empty bed in the main ward. Hence, we will combine Case 1 and Case 2 in Chapter 4.

In the first case, there exists at least one empty bed in ward-1 and there exists at least one type-2 patient in the ward. Thus, we can take all possible actions when type- i patient arrives for all $i = 1, 2$. The second case is the same as the first one, except that the early discharge decision cannot be taken as the system does not have any type-2 patient. In the third case, ward-1 is full and there exist patients also in ward-2. This guarantees that there exists at least one type-2 patient. Thus, we can take all possible actions when a type- i patient arrives for $i = 1, 2$. In the fourth case, ward-1 is full of type-1 patients and ward-2 has not filled yet. Hence, an incoming

type-1 patient must be rejected and all possible actions are valid for incoming type-2 patient. In the fifth case, both ward-1 and ward-2 are full, however, there exists at least one type-2 patient in the system. Therefore, an incoming type-2 patient must be rejected, but an incoming type-1 patient may be accepted with early discharge of type-2 or may be rejected. In the last case, the only action is reject because ward-1 is full with type-1 patients and both wards are completely occupied.

3.3 Coupling

The sample path argument, also known as coupling, is a method which analyzes two systems with similar properties. We use sample path argument in most of the proofs. Two systems can be coupled in such a way that their arrival and service times are the same. Coupling arrival times allows two systems to have the same arrival pattern. Coupling service times yields the same type of patients to have the same service times. If the service time of a type-1 patient is coupled with that of a type-2 patient, the assumption $\mu_1 \leq \mu_2$ guarantees that if the type-1 patient finishes its first stage recovery time, then type-2 patient departs from the hospital with probability 1. In other words, with probability μ_1 both patients finish their service times and with probability $\mu_2 - \mu_1$ the type-2 patient leaves the system whereas type-1 remains as a type-1 patient.

Assume that system A is in state (x) and system B is in state (x') . We couple system A and system B in such a way that their arrival and service times are the same. To find an upper bound on the difference between system A and system B, which is defined as U , we let system B follow the optimal policy and system A follow policy π . This implies that $V_n(x) - V_n(x') \leq V_n^\pi(x) - V_n(x')$, where $V_n^\pi(x)$ is a non optimal value so that $V_n^\pi(x) \geq V_n(x)$. Then, we bound $V_n^\pi(x) - V_n(x')$ by U , which implies $V_n^\pi(x) - V_n(x') \leq U$. To do this, we characterize policy π in each specific situation. In order to find a lower bound on the difference between two systems, we follow the same procedure except that system A follows the optimal policy whereas

B follows policy π .

3.4 Effect of an Additional Job

We analyze the difference between the two systems, which is defined as $D_n(ij)(x)$. $D_n(i0)(x)$ is the expected cost of an additional type- i patient at state x when there exist n more transition periods. It is expressed as $V_n(x + e_i) - V_n(x)$. Moreover, $V_n(x + e_i) - V_n(x + e_j)$ represents the difference between the two systems if the first one starts in state x with an additional type- i patient, and the second system starts in state x with an additional type- j patient. Then, there exist four possible functions we should look at:

- i. $D_n(10)(x) = V_n(x + e_1) - V_n(x)$
- ii. $D_n(20)(x) = V_n(x + e_2) - V_n(x)$
- iii. $D_n(12)(x) = -D_n(21)(x) = V_n(x + e_1) - V_n(x + e_2)$

We can interpret the first two of these terms as the expected burden that a type- i patient brings when he/she is admitted to the system. Moreover, the last one represents the additional cost when a system takes an early discharge action in state $(x + e_2)$. Note that, since we assume for all $x \in S$, $V_0(x) = 0$, $D_0(10)(x) = D_0(20)(x) = D_0(12)(x) = 0$.

3.5 Infinite Horizon Models

In Chapter 4, we prove the properties of our model by induction for all finite n . The results which are proven for finite n are also true when $n \rightarrow \infty$. Hence, our results are valid when total expected discounted cost over an infinite horizon is minimized. Since the state space and the action space is finite, corresponding results are also valid for

a long-run average cost problem. Average cost problem is defined as,

$$J^\pi(x) = \lim_{N \rightarrow \infty} \frac{1}{N} E \left\{ \sum_{n=0}^{N-1} V_n(x) \right\}$$

where the average cost is $J^\pi(x)$ for a policy π in state x . When policies are unichain, $\min_\pi J^\pi(x)$ is a common value for all states. MDP is unichain, so there is only one $g^* = \min_\pi J^\pi(x)$.

Chapter 4

STRUCTURE OF THE OPTIMAL POLICY

In this chapter, we aim to investigate the structure of optimal policies. In the first section, we will present some preliminary properties of our model that are valid for all parameter sets. Then, we focus on special cases of parameters to derive the conditions which guarantee that there exist preferred and strongly preferred classes in our model.

The model introduced in chapter 3 can be considered as a variation of stochastic knapsacks. A stochastic knapsack has c parallel servers and no waiting room. When the admission control problem of a stochastic knapsack is analyzed, certain monotonicity properties are investigated:

- (a) Monotonicity of the value functions
- (b) Supermodularity of the value functions
- (c) Boundedness of $D_n(i0)(x)$.

Moreover, when termination decisions (corresponding to early discharge decisions in our model) are also allowed in stochastic knapsacks, it is shown that these decisions can be optimal only in a subset of the state space S (Ulukus et al. (2011)). In this section, we will show that properties (a) and (c) hold in our model, in addition to the non-optimality of early discharge decisions in a certain set of states.

Supermodularity of the value functions would establish the existence of an optimal threshold policy. Monotonicity of thresholds is another property to explore. However, this requires the convexity of the value functions, and can be shown for stochastic

knapsacks only under restrictive conditions on the system parameters (Örmeci et al. (2001)). In our model, convexity is required to show supermodularity due to the stages in the LOS of type-1 patients. Consequently, we cannot show the supermodularity of the value function. This phenomena is also mentioned in Ghoneim and Stidham (1985), who analyze a system with two stations in series. They state that if the number of servers in station-1 is more than or equal 2, it would be difficult to have supermodularity. Note that neither Stidham and Ghoneim nor we can provide a counter example for supermodularity.

4.1 Preliminary Results

An early discharge decision can be made only if there is a type-2 patient present in ward 1. In this section, we show that, early discharge decision is not optimal unless all beds are occupied in ward 1. This result simplifies the model. Secondly, the effect of an additional job is always positive. Moreover, the effect of early discharge ($D_n(12)(x)$) is always positive, which implies that $D_n(21)(x) < 0$.

Theorem 1. *The early discharge decision is never optimal in states (x_1, x_2) with $x_1 + x_2 < M$ upon a type-1 arrival, i.e. $V_n(x_1 + 1, x_2) \leq V_n(x_1 + 1, x_2 - 1) + c_e$.*

Proof. We use a sample path argument to prove this statement. Assume that system A is in state $(x_1 + 1, x_2 - 1)$ with an extra cost of c_e and system B is in state $(x_1 + 1, x_2)$, where $x_2 > 0$ and $x_1 + x_2 < M$. We couple system A and system B as described in Section 3.3. Assume that system A follows its optimal policy and system B imitates the decisions of system A except for the following cases, which preserves the difference between two systems:

- i. The additional type-2 patient recovers and leaves the system. Then the two systems couple. System A has an additional non-negative cost of c_e .
- ii. When a type-2 patient arrives, if system A accepts, system B rejects so that

the two systems couple. Rejection incurs an additional cost of c_r^2 for system B which is less than c_e by assumption. Thus, B is less costly.

- iii. When ward of system B (ward-1) is full, i.e., $x_1 + x_2 + 1 = M$ and $x_2 > 0$. If a type-1 patient comes, system A accepts and system B gives an early discharge decision with a cost of c_e . Then, the two systems couple. System A and B has the same cost c_e .

In all these cases system B has less (or equal) cost than system A. Therefore, the decisions of admitting with early discharge is always more costly than admitting without, so that the early discharge decision can never be optimal. $\square$

Remark: This setting ensures that system A and B couple before any of them uses the additional ward. Note that, this property does not hold when $x_1 + x_2 \geq M$. According to Theorem 1, we can combine case 1 (3.2) and case 2 (3.3) into case 1** and re-write the model as follows:

Case 1:** If $x_1 < M$ and $x_1 + x_2 < M$

$$\begin{aligned} U_{n-1}(x_1, x_2, 1) &= \min\{V_{n-1}(x_1 + 1, x_2), V_{n-1}(x_1, x_2) + c_r^1\} \\ U_{n-1}(x_1, x_2, 2) &= \min\{V_{n-1}(x_1, x_2 + 1), V_{n-1}(x_1, x_2) + c_r^2\} \end{aligned} \quad (4.1)$$

Other cases, (3)-(6), remain the same.

The following theorem states that our value function is monotonically increasing in x_1 and x_2 . By proving that the expected cost of an additional type- i patient, $D_n(i0)(x)$, is bounded by 0 from below. It states that having one more type- i patient in the system creates a burden to the system. Thus, the additional cost caused by an additional type- i patient is always positive.

Theorem 2. For all x with $x + e_i \in S$, the value function is monotonically increasing in x_i , i.e., $D_n(i0)(x) = V_n(x + e_i) - V_n(x) \geq 0$ for $i = 1, 2$.

Proof. We use a sample path argument to prove this statement. Assume that system A is in state $(x + e_i)$ and system B is in state (x) . We couple system A and system

B as described in Section 3.3. We let system B imitate the decisions of system A, which is always possible. Then, the difference between the two systems due to an arrival is the same except for the cases when system A is in a state (x_1, x_2) with $\overline{M} \geq x_1 + x_2 + 1 > M$. In these cases, system A will have an additional overflow cost. Note that an overflow cost for the additional patient will be incurred per time unit until the two systems couple.

For $i = 2$, if the additional type-2 patient in system A recovers, the two systems couple, where either the costs of the two systems are equal, or system A has a higher cost. For $i = 1$, when the additional type-1 patient in system A becomes a type-2 patient after the first stage recovery, then system A moves into state $(x_1, x_2 + 1)$ and system B is still in state (x_1, x_2) . This corresponds to the case with $i = 2$. Hence, in all service completions and arrivals, the inequality $V_n(x + e_i) - V_n(x) \geq 0$ holds. $\square$

The following theorem gives a lower bound on $D_n(12)(x)$. It states that having one more type-1 patient in the system costs more than having one more type-2 patient. Thus, the difference between the two systems which is caused by the first stage recovery of additional type-1 patient, is always positive.

Theorem 3. *For all $x + e_i \in S$, $D_n(12)(x) = V_n(x_1 + 1, x_2) - V_n(x_1, x_2 + 1) \geq 0$.*

Proof. We use a sample path argument to prove this statement. Assume that system A starts in state $(x + e_1)$ and system B starts in state $(x + e_2)$.

We couple the two systems as described in Section 3.3. When the additional type-1 patient completes first stage recovery and type-2 patient leaves the hospital with probability μ_1 , the difference between expected costs becomes $V_{n-1}(x + e_2) - V_{n-1}(x)$, which is greater than equal to zero by Theorem 2. There exists another possibility that, with probability $\mu_2 - \mu_1$, the type-2 patient leaves while the type-1 patient remains as type 1. Then, the difference between expected costs becomes $V_{n-1}(x + e_1) - V_{n-1}(x)$, which is greater than equal zero by Theorem 2. Hence, in all service completions the inequality is satisfied.

Assume that system A follows the optimal policy, and system B follows policy π . We define policy π such that system B imitates the decisions of system A, which is always possible. Then, there is no difference in the costs due to an arrival.

As a result, $V_n(x_1 + 1, x_2) - V_n(x_1, x_2 + 1) \geq 0$ holds. $\square$

From Theorem 3, we know that $D_n(12)(x) \geq 0$ for all $x \in S$ and for all n . By definition, for all $x \in S$ and for all n , $D_n(12)(x) = -D_n(21)(x)$ so that, $D_n(21)(x)$ can be at most zero.

4.2 Existence of a Preferred Class

In this section, we present sufficient conditions for each patient type to be preferred and strongly preferred for the special case with $\overline{M} = M$. A patient type (class) is preferred if its patients are always admitted whenever there exists an empty bed. Moreover, type-1 patients are called strongly preferred if they are always accepted with early discharge decision even if the system is full.

Let d_i be an upper bound on $D_n(i0)(x)$ for all $x \in S$, for all $n \geq 0$ and $i = 1, 2$, so that $d_i \geq \max_{x \in S} D_n(i0)(x)$. Moreover, we define d_{ij} as an upper bound on $D_n(ij)(x)$ for all $x \in S$, for all $n \geq 0$ and $(ij) \in \{(12), (21)\}$, so that $d_{ij} \geq \max_{x \in S} D_n(ij)(x)$. Then, we can set an upper-bound on d_{21} denoted by $\overline{d}_{21}$, using Theorem 3 as $\overline{d}_{21} = 0$. We will derive upper bounds on d_i and d_{12} , which are defined as $\overline{d}_i$ and $\overline{d}_{12}$, respectively.

As mentioned earlier, a type- i patient is admitted if $V_n(x) + c_r^i \geq V_n(x + e_i)$ or equally $D_n(i0)(x) \leq c_r^i$. Thus, type- i is preferred if $d_i \leq c_r^i$. Furthermore, a type-1 patient is admitted via terminating a type-2 patient's recovery process if $V_n(x_1 + 1, x_2 - 1) + c_e \leq V_n(x) + c_r^1$ or equally $D_n(12)(x) \leq c_r^1 - c_e$. Thus, type-1 is strongly preferred if $d_{12} \leq c_r^1 - c_e$.

In the following sections, we derive $\overline{d}_i$ values which should guarantee $\overline{d}_i \leq c_r^i$ condition if type- i is preferred. Moreover, we find $\overline{d}_{12}$ which should be less than or equal to $c_r^1 - c_e$ to derive the sufficient conditions when type-1 is strongly preferred.

4.2.1 A sufficient condition for type 2 to be preferred

In the first part of this section, we derive upper bounds on $D_n(20)(x)$. At the end, we obtain a theorem, which presents a sufficient condition for type-2 patients to be preferred.

Let system A be in state $(x+e_2)$ and system B be in state (x) . We couple these two systems as discussed earlier. Assume that system B follows the optimal policy, and system A imitates system B whenever possible, which keeps the difference between two systems bounded by d_2 . This imitation is possible except for the case where $x_1 + x_2 + 1 = M$, since system B may accept an incoming type-1 patient, but system A cannot. Thus, system B moves to state $(x + e_1)$ and system A preserves its state as $(x + e_2)$. Then, the difference between the two systems is bounded by $c_r^1 + d_{21}$. When a type-2 patient arrives, system B accepts it and moves to state $(x_1, x_2 + 1)$ and system A cannot accept it and stays in state $(x_1, x_2 + 1)$. Then, two systems couple so that the difference between two systems becomes c_r^2 . Note that, according to Theorem 1, early discharge is not an optimal action for system B because even if system A is full, system B is not. Consequently, we derive the following expression for an upper-bound on $D_n(20)(x)$:

$$\begin{aligned}
D_n(20)(x) &= V_n(x_1, x_2 + 1) - V_n(x_1, x_2) \leq V_n^\pi(x_1, x_2 + 1) - V_n(x_1, x_2) \\
&\leq \lambda_1 \max\{D_{n-1}(21)(x) + c_r^1, D_{n-1}(20)(x)\} + \lambda_2 \max\{D_{n-1}(20)(x), c_r^2\} \\
&\quad + \mu_1 x_1 (V_{n-1}(x_1 - 1, x_2 + 2) - V_{n-1}(x_1 - 1, x_2 + 1)) \\
&\quad + \mu_2 x_2 (V_{n-1}(x_1, x_2) - V_{n-1}(x_1, x_2 - 1)) \\
&\quad + \mu_2 (V_{n-1}(x_1, x_2) - V_{n-1}(x_1, x_2)) \\
&\quad + (M\mu_2 - \mu_1 x_1 - \mu_2(x_2 + 1))(V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1, x_2)) \\
&\leq \lambda_1 \max\{d_{21} + c_r^1, d_2\} + \lambda_2 \max\{d_2, c_r^2\} + (M - 1)\mu_2 d_2
\end{aligned} \tag{4.2}$$

where the first inequality comes from the policy system B follows, the second one is because of the coupling, and the last one is due to the definition of d_2 and d_{21} . From

this, we can extract the following inequality which ensures that d_2 is an upper bound on $D_n(20)(x)$ for all $x \in S$:

$$\lambda_1 \max\{d_{21} + c_r^1, d_2\} + \lambda_2 \max\{d_2, c_r^2\} + (M - 1)\mu_2 d_2 \leq d_2 \quad (4.3)$$

After a simplification by using uniformization, i.e., $\lambda_1 + \lambda_2 + M\mu_2 = 1$, inequality (4.3) becomes:

$$\lambda_1 \max\{d_{21} + c_r^1, d_2\} + \lambda_2 \max\{d_2, c_r^2\} \leq d_2(\lambda_1 + \lambda_2 + \mu_2) \quad (4.4)$$

We eliminate the d_{21} term by substituting its upper bound, i.e., $\bar{d}_{21} = 0$ into (4.4). Then, we have the following expression:

$$\lambda_1 \max\{c_r^1, d_2\} + \lambda_2 \max\{d_2, c_r^2\} \leq d_2(\lambda_1 + \lambda_2 + \mu_2) \quad (4.5)$$

Now, we aim to derive an expression for $\bar{d}_2$ in all possible cases. Since we assume $c_r^2 < c_r^1$, we have three possible cases in (4.5):

Case 1. $d_2 \leq c_r^2$

From the inequality (4.5):

$$\frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \leq d_2,$$

from which we can set:

$$\bar{d}_2 = \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2}. \quad (4.6)$$

Note that $\bar{d}_2$ is a valid upper-bound, if the following condition is satisfied:

$$\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1}. \quad (4.7)$$

Case 2. $c_r^2 \leq d_2 \leq c_r^1$

Similarly to Case 1, we define:

$$\bar{d}_2 = \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2}, \quad (4.8)$$

where $\bar{d}_2$ is a valid upper-bound, if the following condition is satisfied:

$$\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}. \quad (4.9)$$

Case 3. $c_r^1 \leq d_2$

From the inequality (4.5),

$$d_2 \geq 0$$

In this case, $\bar{d}_2 = 0$, which makes $c_r^2 \leq 0$. This is not true, since we assume that c_r^i is strictly positive. Hence, this case is not valid, i.e. d_2 cannot exceed c_r^1 .

We summarize these findings in the following lemma:

Lemma 1. *For all $n \geq 0$ and for all $x \in S$*

$$\begin{aligned} i. \text{ if } \frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}, D_n(20)(x) &\leq \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2} \text{ and,} \\ ii. \text{ if } \frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1}, D_n(20)(x) &\leq \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \end{aligned}$$

Proof. We prove Lemma 1 by induction. For $n = 0$, $D_0(20)(x) = 0$ so,

$$\text{I. } D_0(20)(x) = 0 \leq \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2} \text{ and,}$$

$$\text{II. } D_0(20)(x) = 0 \leq \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2}$$

because all parameters are nonnegative. The rest of the proof repeats the derivation presented above. $\square$

From Lemma 1, we have the result as:

Theorem 4. *Type-2 patients are preferred for all $n \geq 0$, if*

$$\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1}.$$

The intuition behind Theorem 4 is straightforward. Assume that, all beds except one are occupied. When a type-2 patient arrives, there exist two options: accept (System A) or reject (System B). If we accept the incoming type-2 patient, there exist two possible events that will change the difference between the two systems: either a type-1 patient arrives or the additional type-2 patient leaves. With probability $\frac{\lambda_1}{\lambda_1 + \mu_2}$ a type-1 patient arrives before the departure of the type-2 patient, which incurs a cost of c_r^1 due to the rejection of the incoming type-1 patient. The additional type-2 patient leaves with probability $\frac{\mu_2}{\lambda_1 + \mu_2}$ with no additional cost. Then, the expected cost due to type-2 patient's acceptance becomes $\frac{\lambda_1}{\lambda_1 + \mu_2} c_r^1$. Theorem 4 suggest that, if the rejection cost of type-2 patient is greater than the expected cost due to type-2 patient's acceptance, then it is always better to accept the type-2 patients.

Remark: Different coupling assumptions result in different sufficient conditions for type-2 patients to be preferred. Assume that system B follows the optimal policy but if the next transition is arrival of a type-1 patient, system A accepts the patients in the next decision epoch with early discharge of a type-2 patient. Then, the upper bound expression becomes $\lambda_1 \max\{c_e, d_2\} + \lambda_2 \max\{d_2, c_r^2\} \leq d_2(\lambda_1 + \lambda_2 + \mu_2)$ so that type-2 patients are preferred when $\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_e}$. It turns out that this condition provides a more strict bound than the one in Theorem 4 since $c_e \leq c_r^1$ (see also Section 5.3).

4.2.2 Sufficient conditions for type 1 to be preferred and strongly preferred

In the first part of this section, we show that if type-1 patients are strongly preferred, they are also preferred. Then, we derive upper bounds on $D_n(10)(x)$ and $D_n(12)(x)$ when type-1 patients are preferred or strongly preferred. We conclude with a theorem, which presents sufficient conditions for type-1 patients to be preferred and strongly preferred.

The following theorem establishes a relation between preferred and strongly pre-

ferred classes.

Theorem 5. *If type-1 patients are strongly preferred, then they are also preferred, i.e. if $d_{12} \leq c_r^1 - c_e$, then $d_1 \leq c_r^1$.*

Proof. We use sample path argument to prove this statement. Assume that system A is in state $(x + e_1)$ and system B is in state (x) . We couple system A and system B. Assume that $d_{12} \leq c_r^1 - c_e$. Under this assumption, showing $V_n^\pi(x + e_1) - V_n(x) \leq c_r^1$ yields $d_1 \leq c_r^1$. We let system A imitate the decisions of system B, which is always possible, except for the following cases where system A is full, i.e. $x_1 + x_2 + 1 = M$:

- i. When a type-1 patient arrives: if system B accepts the incoming patient, system A rejects the patient. Then, two systems couple and system A has an extra cost of $c_r^1 \leq c_r^1$.
- ii. When a type-2 patient arrives, if system B accepts the incoming patient and moves to state $(x_1, x_2 + 1)$, system A rejects the patient and stays in state $(x_1 + 1, x_2)$. Then, the difference between the two systems is $D_{n-1}(12)(x) + c_r^2 \leq d_{12} + c_r^2 \leq c_r^1 - c_e + c_r^2$ for all $x \in S$ and for all n . We assume that $c_r^2 < c_e < c_r^1$, therefore, $c_r^1 - c_e + c_r^2 \leq c_r^1$.
- iii. When the additional type-1 patient in system A becomes a type-2 patient after the first-stage recovery, then system A moves to state $(x_1, x_2 + 1)$ while system B stays in state (x_1, x_2) . Then, we should investigate this new state pair in a detailed way.
 - a. If the additional type-2 patient in the system A leaves, the two systems couple with no additional cost.
 - b. When a type-1 patient arrives, if system B accepts the incoming patient, system A rejects the patient. Hence, system A stays in state $(x_1, x_2 + 1)$ with cost of c_r^1 and system B moves into state $(x_1 + 1, x_2)$, so that the

difference between two systems is $D_{n-1}(21)(x) + c_r^1$. From Theorem 3, we know that $D_{n-1}(21)(x) \leq 0$. Then, the difference is at most c_r^1 which is less than or equal to c_r^1 .

- c. When a type-2 patient arrives, if system B accepts the incoming patient, system A rejects the patient with an additional cost c_r^2 . Then, the two systems couple and system A has an extra cost of $c_r^2 \leq c_r^1$ by definition.

□

Now, we will derive upper bounds on $D_n(10)(x)$ and $D_n(12)(x)$ for cases where type-1 patients are preferred or strongly preferred.

Let system A be in state $(x + e_1)$ and system B be in state (x) . We couple system A and system B as discussed earlier. Assume that system B follows the optimal policy, and system A imitates system B whenever possible, by which the difference between the two systems is bounded by d_1 . This imitation is always possible, except for the following cases with $x_1 + x_2 + 1 = M$:

- a. System B may accept the type-1 patient, but system A cannot. Thus, system B moves to state $(x + e_1)$ and system A preserves its state as $(x + e_1)$, so that the two systems couple. Then, the difference between two systems is bounded by c_r^1 .
- b. When a type-2 patient arrives, system B accepts it and moves to state $(x_1, x_2 + 1)$ and system A cannot accept it and stays in state $(x_1 + 1, x_2)$. Then, the difference between states becomes $D_n(12)(x) + c_r^2$.

Accordingly, we have the following expression for an upper-bound on $D_n(10)(x)$:

$$\begin{aligned}
D_n(10)(x) &= V_n(x_1 + 1, x_2) - V_n(x_1, x_2) \leq V_n^\pi(x_1 + 1, x_2) - V_n(x_1, x_2) \\
&\leq \lambda_1 \max\{c_r^1, D_{n-1}(10)(x)\} + \lambda_2 \max\{D_{n-1}(10)(x), D_{n-1}(12)(x) + c_r^2\} \\
&\quad + \mu_1 x_1 (V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1 - 1, x_2 + 1)) \\
&\quad + \mu_2 x_2 (V_{n-1}(x_1 + 1, x_2 - 1) - V_{n-1}(x_1, x_2 - 1)) \\
&\quad + \mu_1 (V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1, x_2)) \\
&\quad + (M\mu_2 - \mu_1(x_1 + 1) - \mu_2 x_2) (V_{n-1}(x_1 + 1, x_2) - V_{n-1}(x_1, x_2)) \\
&\leq \lambda_1 \max\{c_r^1, d_1\} + \lambda_2 \max\{d_1, d_{12} + c_r^2\} + \mu_1 d_2 + (M\mu_2 - \mu_1) d_1
\end{aligned} \tag{4.10}$$

where the first inequality comes from the policy system A follows, the second one is because of the coupling, and the last one is due to the definition of d_1 , d_2 and d_{12} . Then, d_1 is an upper bound on $D_n(10)(x)$ for all $x \in S$ and for all n if it satisfies the following inequality:

$$\lambda_1 \max\{c_r^1, d_1\} + \lambda_2 \max\{d_1, d_{12} + c_r^2\} + \mu_1 d_2 + (M\mu_2 - \mu_1) d_1 \leq d_1 \tag{4.11}$$

After the simplification by using uniformization, i.e., $\lambda_1 + \lambda_2 + M\mu_2 = 1$, inequality (4.11) becomes:

$$\lambda_1 \max\{c_r^1, d_1\} + \lambda_2 \max\{d_1, d_{12} + c_r^2\} + \mu_1 d_2 \leq d_1 (\lambda_1 + \lambda_2 + \mu_1) \tag{4.12}$$

From Theorem 5, type-1 patients must be preferred if they are strongly preferred. Since we want to find sufficient conditions which ensure that type-1 patients are preferred or strongly preferred, we derive expressions for d_1 only for these cases. We know that type-1 is preferred if $d_1 \leq c_r^1$. Then, we can replace the first maximization in inequality (4.12) with c_r^1 to eliminate cases where type-1 patients are not preferred. The new inequality is:

$$\lambda_1 c_r^1 + \lambda_2 \max\{d_1, d_{12} + c_r^2\} + \mu_1 d_2 \leq d_1 (\lambda_1 + \lambda_2 + \mu_1) \tag{4.13}$$

Since d_1 depends on d_{12} , we will find an expression which gives an upper-bound on $D_n(12)(x)$. Let system A be in state $(x + e_1)$ and system B be in state $(x + e_2)$. We couple system A and system B as described in Section 3.3. Assume that system B follows the optimal policy and system A imitates system B whenever possible, so that the difference between two systems is bounded by d_{12} . This imitation is always possible except for the case with $x_1 + x_2 + 1 = M$ and $x_2 = 0$, since in such a state system B may accept the arriving type-1 patient with early discharge decision, but system A cannot. Thus, system B moves to state $(x + e_1)$ and system A preserves its state as $(x + e_1)$, so that the two systems couple. Then, the difference between these systems is bounded by $c_r^1 - c_e$. The expression for upper-bound on $D_n(12)(x)$ becomes:

$$\begin{aligned}
D_n(12)(x) &= V_n(x_1 + 1, x_2) - V_n(x_1, x_2 + 1) \leq V_n^\pi(x_1 + 1, x_2) - V_n(x_1, x_2 + 1) \\
&\leq \lambda_1 \max\{D_{n-1}(12)(x), c_r^1 - c_e\} + \lambda_2 D_{n-1}(12)(x) \\
&\quad + \mu_1 x_1 (V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1 - 1, x_2 + 2)) \\
&\quad + \mu_2 x_2 (V_{n-1}(x_1 + 1, x_2 - 1) - V_{n-1}(x_1, x_2)) \\
&\quad + \mu_1 (V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1, x_2)) \\
&\quad + (\mu_2 - \mu_1) (V_{n-1}(x_1 + 1, x_2) - V_{n-1}(x_1, x_2)) \\
&\quad + (M\mu_2 - \mu_1 x_1 - \mu_2(x_2 + 1)) (V_{n-1}(x_1 + 1, x_2) - V_{n-1}(x_1, x_2 + 1)) \\
&\leq \lambda_1 \max\{d_{12}, c_r^1 - c_e\} + \lambda_2 d_{12} + (M - 1)\mu_2 d_{12} + (\mu_2 - \mu_1)d_1 + \mu_1 d_2
\end{aligned} \tag{4.14}$$

where the first inequality comes from the policy system B follows, the second one is because of the coupling and the last one is due to the definition of d_1 , d_2 and d_{12} . From this, we can extract the following inequality which ensures that d_{12} is an upper bound on $D_n(12)(x)$ for all $x \in S$ and for all n :

$$\lambda_1 \max\{d_{12}, c_r^1 - c_e\} + \lambda_2 d_{12} + (M - 1)\mu_2 d_{12} + (\mu_2 - \mu_1)d_1 + \mu_1 d_2 \leq d_{12} \tag{4.15}$$

After the simplification by using uniformization $\lambda_1 + \lambda_2 + M\mu_2 = 1$, inequality (4.15)

becomes:

$$\lambda_1 \max\{d_{12}, c_r^1 - c_e\} + (\mu_2 - \mu_1)d_1 + \mu_1 d_2 \leq d_{12}(\lambda_1 + \mu_2) \quad (4.16)$$

There are two possible cases coming from the maximization; from which we derive an upper-bound on d_{12} , $\bar{d}_{12}$.

Case a. $d_{12} \geq c_r^1 - c_e$

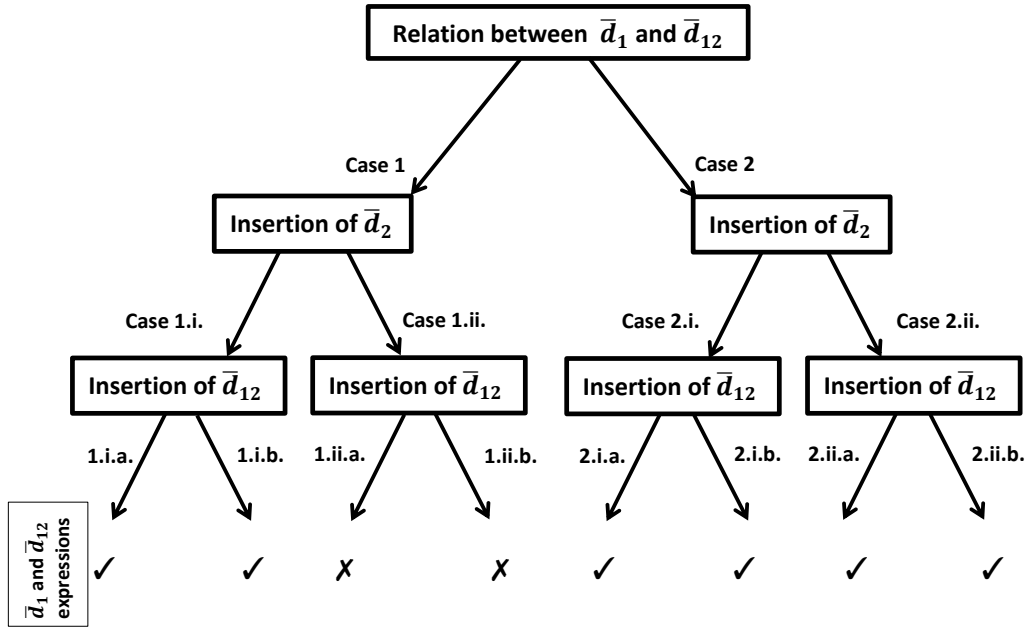
$$\bar{d}_{12} = d_1 \frac{\mu_2 - \mu_1}{\mu_2} + d_2 \frac{\mu_1}{\mu_2} \quad (4.17)$$

Case b. $d_{12} \leq c_r^1 - c_e$

$$\bar{d}_{12} = \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + d_1 \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2} + d_2 \frac{\mu_1}{\lambda_1 + \mu_2} \quad (4.18)$$

To obtain expressions for $\bar{d}_1$ and $\bar{d}_{12}$, we use inequalities (4.13), (4.17), (4.18) as well as the expressions for $\bar{d}_2$ in equations (4.6) and (4.8). In Figure 4.1, we show the flow of the derivation of $\bar{d}_1$ and $\bar{d}_{12}$. We first consider the cases emerging from (4.13). There exist two possible cases coming from the maximization in (4.13). Case 1 is $d_1 \geq d_{12} + c_r^2$ and case 2 is $d_1 \leq d_{12} + c_r^2$. Within each of these cases, we substitute the expression for $\bar{d}_2$ from Lemma 1, which brings two more cases. The first case (i.e. $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}$) is represented by list index (i) and the second one (i.e. $\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1}$) is represented by list index (ii), where type-2 patient is preferred. Then, within each of these cases, we eliminate the d_{12} term by inserting the upper bound values obtained from equation (4.17) and (4.18). The case coming from (4.17) is $d_{12} \geq c_r^1 - c_e$ and represented by list index (a), which corresponds to the case where type-1 patients are preferred. The case coming from (4.18) is represented by list index (b) corresponding to strongly preferred type-1 patients. Finally, we solve for $\bar{d}_1$ and $\bar{d}_{12}$ and check whether these values are valid in the range defined for each case. If we found feasible constraints on $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ for each case, we put a tick symbol in Figure 4.1, otherwise we depicted it by a cross symbol.

Now we present all the cases. The derivations can be found in Appendix A.

Figure 4.1: Flowchart of the derivation of $\bar{d}_1$ and $\bar{d}_2$ expressions

Case 1. $c_r^1 \geq d_1 \geq d_{12} + c_r^2$

From (4.13), we have:

$$d_1 \leq \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \mu_1} \quad (4.19)$$

We are going to eliminate d_2 term in inequality (4.19). According to Lemma 1, there exist two possible $\bar{d}_2$ values.

Case 1.i. $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}$

From Lemma 1, for this range of parameters $\bar{d}_2 = \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2}$. Substituting $\bar{d}_2$ in (4.19), we have

$$\bar{d}_1 = \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1. \quad (4.20)$$

Now we find $\bar{d}_{12}$ corresponding to these $\bar{d}_1$ and $\bar{d}_2$ expressions:

Case 1.i.a. $d_{12} \geq c_r^1 - c_e$

For this range of parameters $\bar{d}_{12}$ is defined in equation (4.17), in

which the unknowns are d_1 and d_2 . Inserting (4.20) and $\bar{d}_2$ in (4.17), results in:

$$\bar{d}_{12} = \frac{\lambda_1}{\lambda_1 + \mu_1} c_r^1 \quad (4.21)$$

where $\bar{d}_1$ and $\bar{d}_{12}$ are valid upper-bounds if the following conditions are satisfied:

Condition 1.

$$\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \quad (4.22)$$

and

$$\frac{\mu_1}{\lambda_1 + \mu_1} \leq \frac{c_e}{c_r^1}. \quad (4.23)$$

We obtain these conditions by inserting $\bar{d}_1$ and $\bar{d}_{12}$ values given by equations (4.20) and (4.21), respectively, into the conditions of Case 1.i.a: Condition $d_1 \geq d_{12} + c_r^2$ becomes (4.22) whereas condition $d_{12} \geq c_r^1 - c_e$ becomes (4.23). Condition $c_r^1 \geq d_1$, on the other hand, always holds true. We note that the right hand side of (4.22) is less than the left hand side of (4.23) for all possible parameter values, which agrees with our assumption, i.e. $\frac{c_r^2}{c_r^1} \leq \frac{c_e}{c_r^1}$.

Case 1.i.b. $d_{12} \leq c_r^1 - c_e$

For this range of parameters $\bar{d}_{12}$ is defined in equation (4.18), in which the unknowns are d_1 and d_2 . Substituting (4.20) and $\bar{d}_2$ in (4.18), results in

$$\bar{d}_{12} = \frac{\lambda_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1 + \mu_1 + \mu_2}{\lambda_1 + \mu_1} c_r^1 - c_e \right)$$

where $\bar{d}_1$ and $\bar{d}_{12}$ are valid upper-bounds if they satisfy the following conditions:

Condition 2.

$$\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \quad (4.24)$$

and

$$\frac{c_e}{c_r^1} \leq \frac{\mu_1}{\lambda_1 + \mu_1} \quad (4.25)$$

We did the same procedure we described above. It resulted with two inequalities: The first one comes from condition $d_{12} \leq c_r^1 - c_e$ and gives an upper bound on $\frac{c_e}{c_r^1}$ as in (4.25). The second one comes from condition $d_1 \geq d_{12} + c_r^2$ and consists of both $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$. We derived the upper bound on $\frac{c_r^2}{c_r^1}$ defined in (4.24) from these two inequalities by using linear algebra.

Case 1.ii. $\frac{c_r^2}{c_r^1} \geq \frac{\lambda_1}{\lambda_1 + \mu_2}$

For this range of parameters $\bar{d}_2 = \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2}$. Substituting $\bar{d}_2$ in (4.19), we have

$$\bar{d}_1 = \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)c_r^1 + \lambda_2\mu_1 c_r^2}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \quad (4.26)$$

Now we find $\bar{d}_{12}$ corresponding to these $\bar{d}_1$ and $\bar{d}_2$ expressions:

Case 1.ii.a. $d_{12} \geq c_r^1 - c_e$

For this range of parameters $\bar{d}_{12} = d_1 - \frac{\mu_1}{\mu_2}(d_1 - d_2)$, as in (4.17).

We repeated the same procedure we described above. However, when we checked $\bar{d}_1 \geq \bar{d}_{12} + c_r^2$ we found an upper bound on $\frac{c_r^2}{c_r^1}$ (i.e. $\frac{\lambda_1\mu_1(\lambda_2+\mu_2)}{\lambda_1\lambda_2\mu_1+\mu_2(\lambda_1+\mu_1)(\lambda_1+\lambda_2+\mu_2)}$) which is less than the lower bound $\frac{\lambda_1}{\lambda_1+\mu_2}$. Recall that this lower bound comes from the definition of d_2 . As a result, this case, which shows that type-1 patient is preferred, is not valid.

Case 1.ii.b. $d_{12} \leq c_r^1 - c_e$

This case should give conditions where type-1 patient is strongly preferred. Since we could not find any valid range for type-1 patient is preferred, from Theorem-5 this case can not produce any valid condition, as well.

Case 2. $c_r^1 \geq d_1$ and $d_{12} + c_r^2 \geq d_1$

From (4.13), we have:

$$d_1 \leq \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \lambda_2 + \mu_1} + d_{12} \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu_1}. \quad (4.27)$$

We are going to eliminate d_2 term in inequality (4.27). According to Lemma 1, there exist two possible $\bar{d}_2$ values.

Case 2.i. $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}$

For this range of parameters $\bar{d}_2 = \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2}$. Now we find $\bar{d}_{12}$ and $\bar{d}_1$ corresponding to this $\bar{d}_2$ expression.

Case 2.i.a. $d_{12} \geq c_r^1 - c_e$

For this range of parameters, $\bar{d}_{12}$ is defined in equation (4.17). Since (4.27) depends on d_{12} , we solve (4.27) and (4.17) together and insert $\bar{d}_2$. Then, $\bar{d}_1$ and $\bar{d}_{12}$ become:

$$\bar{d}_1 = \frac{\lambda_1}{\lambda_1 + \mu_2} \left(1 + \frac{\mu_2^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) c_r^1 + \frac{\lambda_2 \mu_2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} c_r^2$$

$$\bar{d}_{12} = \frac{\lambda_1}{\lambda_1 + \mu_2} \left(1 + \frac{\mu_2 (\mu_2 - \mu_1)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) c_r^1 + \frac{\lambda_2 (\mu_2 - \mu_1)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} c_r^2$$

where $\bar{d}_1$ and $\bar{d}_{12}$ are valid upper-bounds if the following conditions are satisfied:

Condition 3. I. If $\lambda_1 \lambda_2 \leq \mu_1 (\lambda_2 + \mu_2)$

$$\frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \leq \frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2} \quad (4.28)$$

$$\frac{\mu_1 \mu_2 (\lambda_1 + \lambda_2 + \mu_2) + \lambda_1 \lambda_2 \mu_1 - \lambda_1 \lambda_2 \mu_2}{(\lambda_1 + \mu_2)(\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2))} \leq \frac{c_e}{c_r^1} \quad (4.29)$$

II. If $\mu_2 (\lambda_2 + \mu_2) \leq \lambda_1 \lambda_2$

$$\frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \leq \frac{c_r^2}{c_r^1} \leq \frac{\mu_1 (\lambda_2 + \mu_2)}{\lambda_2 (\lambda_1 + \mu_2)} \quad (4.30)$$

$$\frac{\mu_1}{\lambda_1 + \mu_2} \leq \frac{c_e}{c_r^1} \quad (4.31)$$

After repeating the steps we described above, we obtain three inequalities: The first one is obtained while checking $c_r^1 \geq d_1$ and provides an upper bound on $\frac{c_r^2}{c_r^1}$. From the definition of d_2 in Lemma 1, we also have an upper bound on $\frac{c_r^2}{c_r^1}$, which is $\frac{\lambda_1}{\lambda_1 + \mu_2}$. Hence, we should find the minimum of those two upper bounds to extract an upper bound on $\frac{c_r^2}{c_r^1}$. Index (I) represents the case where the upper bound is $\frac{\lambda_1}{\lambda_1 + \mu_2}$ and index (II) shows the other case. The second inequality is obtained while checking $d_{12} + c_r^2 \geq d_1$ and provides a lower bound on $\frac{c_r^2}{c_r^1}$. The last inequality we faced with after following the steps consisted of both $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$. We derived the lower bound on $\frac{c_e}{c_r^1}$ by using simple algebra.

Case 2.i.b. $d_{12} \leq c_r^1 - c_e$

For this range of parameters $\bar{d}_{12}$ is defined in equation (4.18). We find $\bar{d}_1$ and $\bar{d}_{12}$ as:

$$\begin{aligned}\bar{d}_1 &= \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 + \lambda_2 \frac{(\lambda_1 + \mu_2)c_r^2 - \lambda_1 c_e}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \\ \bar{d}_{12} &= \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 + \frac{\lambda_2(\mu_2 - \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} c_r^2 \\ &\quad - \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} c_e\end{aligned}$$

where $\bar{d}_1$ and $\bar{d}_{12}$ are valid upper-bounds if the following conditions are satisfied:

Condition 4.

$$\frac{c_r^2}{c_r^1} \leq \min\left\{\frac{\lambda_1}{\lambda_1 + \mu_2}, \frac{\mu_2(\lambda_2 + \mu_2)}{\lambda_2(\lambda_1 + \mu_2)}\right\} \quad (4.32)$$

$$\frac{c_e}{c_r^1} \leq \frac{\mu_1}{\lambda_1 + \mu_1} \quad (4.33)$$

To find these conditions, we obtained three inequalities and all of them consisted of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$. We derived an upper bound on $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ from these three inequalities by using simple linear algebra.

From Lemma 1, we have also have an upper bound on $\frac{c_r^2}{c_r^1}$, which is $\frac{\lambda_1}{\lambda_1 + \mu_2}$. Hence, we should find the minimum of those two upper bounds to extract an upper bound on $\frac{c_r^2}{c_r^1}$.

Case 2.ii. $\frac{c_r^2}{c_r^1} \geq \frac{\lambda_1}{\lambda_1 + \mu_2}$

For this range of parameters $\bar{d}_2 = \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2}$. Now we find $\bar{d}_{12}$ and $\bar{d}_1$ corresponding to this $\bar{d}_2$ expression.

Case 2.ii.a. $d_{12} \geq c_r^1 - c_e$

For this range of parameter $\bar{d}_{12}$ is defined as in (4.17). We find $\bar{d}_1$ and $\bar{d}_{12}$ as:

$$\bar{d}_1 = (\lambda_1 c_r^1 + \lambda_2 c_r^2) \frac{\mu_1(\lambda_2 + \mu_2) + \mu_2(\lambda_1 + \lambda_2 + \mu_2)}{(\lambda_1 + \lambda_2 + \mu_2)(\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2))}$$

$$\bar{d}_{12} = \frac{\mu_2(\lambda_1 c_r^1 + \lambda_2 c_r^2)}{\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)}$$

where $\bar{d}_1$ and $\bar{d}_{12}$ are valid upper-bounds if the following conditions are satisfied:

Condition 5.

$$\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1} \leq \frac{\mu_1(\lambda_2 + \mu_2)^2}{\lambda_2(\mu_1(\lambda_2 + \mu_2) + \mu_2(\lambda_1 + \lambda_2 + \mu_2))} \quad (4.34)$$

and

$$\frac{\mu_1(\lambda_2 + \mu_2)}{\mu_1(\lambda_2 + \mu_2) + \mu_2(\lambda_1 + \lambda_2 + \mu_2)} \leq \frac{c_e}{c_r^1} \quad (4.35)$$

After following the steps, we obtained three inequalities: The first one is obtained while checking $c_r^1 \geq d_1$ and provides an upper bound on $\frac{c_r^2}{c_r^1}$ as given in (4.34). When we check $\bar{d}_{12} + c_r^2 \geq \bar{d}_1$, we obtain a lower bound on $\frac{c_r^2}{c_r^1}$ which is always smaller than $\frac{\lambda_1}{\lambda_1 + \mu_2}$. The last inequality contains both $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$. We derived a lower bound on $\frac{c_e}{c_r^1}$, which is determined in (4.35), from the first and the last inequalities by using simple linear algebra. As a remark, the upper bound we found for $\frac{c_r^2}{c_r^1}$ must be greater than the lower bound

on $\frac{c_r^2}{c_r^1}$, which comes from the definition of case 2.ii. However, this only happens when $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$. Hence, this case is not valid if the parameters do not satisfy this inequality.

Case 2.ii.b. $d_{12} \leq c_r^1 - c_e$

For this range of parameters $\bar{d}_{12}$ is defined as in (4.18). We find $\bar{d}_1$ and $\bar{d}_{12}$ as:

$$\begin{aligned}\bar{d}_1 &= \frac{(\lambda_1 + \mu_1 + \mu_2)(\lambda_1 c_r^1 + \lambda_2 c_r^2) + \lambda_1 \lambda_2 (c_r^1 - c_e)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \\ \bar{d}_{12} &= \frac{\mu_2(\lambda_1 c_r^1 + \lambda_2 c_r^2) + \lambda_1(\lambda_1 + \lambda_2 + \mu_1)(c_r^1 - c_e)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}\end{aligned}$$

where $\bar{d}_1$ and $\bar{d}_{12}$ are valid upper-bounds if the following conditions are satisfied:

Condition 6.

$$\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1} \leq \frac{\mu_1(\lambda_2 + \mu_2)^2}{\lambda_2(\mu_1(\lambda_2 + \mu_2) + \mu_2(\lambda_1 + \lambda_2 + \mu_2))} \quad (4.36)$$

and

$$\frac{c_e}{c_r^1} \leq \frac{\mu_1(\lambda_1 + \mu_2)(\lambda_2 + \mu_2) - \lambda_1 \lambda_2 \mu_2}{(\mu_1(\lambda_2 + \mu_2) + \mu_2(\lambda_1 + \lambda_2 + \mu_2))(\lambda_1 + \mu_2)} \quad (4.37)$$

To find these conditions, we obtained three inequalities and both of them consisted of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$. From these inequalities by using simple linear algebra, we derived two upper bounds on $\frac{c_r^2}{c_r^1}$ and we choose the minimum of these upper bounds as stated in (4.36). The lower bound on $\frac{c_r^2}{c_r^1}$ is obtained from the definition of $\bar{d}_2$. To derive (4.37), we use the lower bound on $\frac{c_r^2}{c_r^1}$ and the inequality coming from $d_{12} \leq c_r^1 - c_e$.

The upper bound we found for $\frac{c_r^2}{c_r^1}$ must be greater than the lower bound on $\frac{c_r^2}{c_r^1}$, which holds when $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$. When $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$ is satisfied, the right hand side of (4.37) is always positive. Moreover, the upper bound on $\frac{c_e}{c_r^1}$ must be greater than

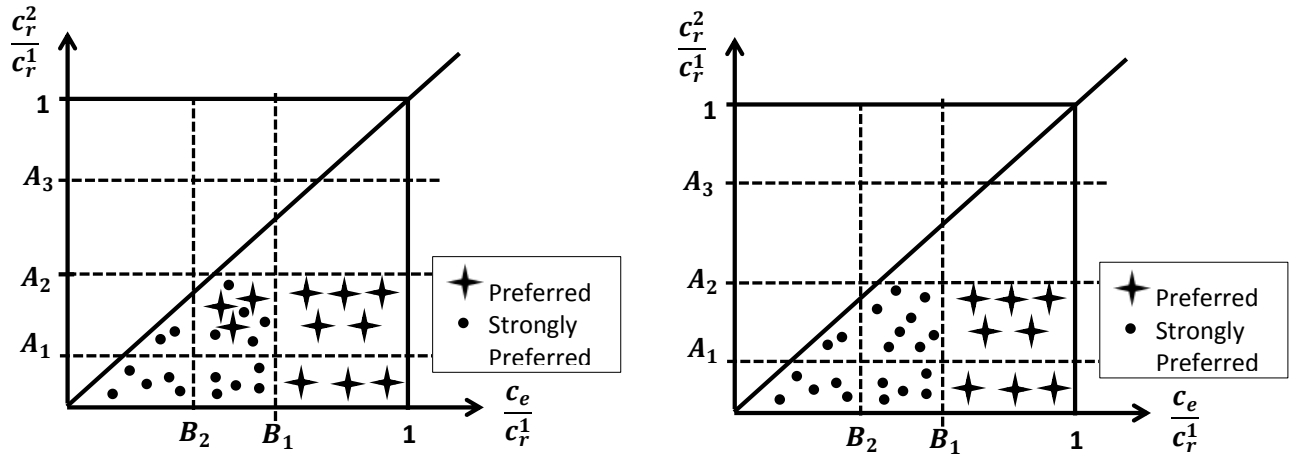
the lower bound on $\frac{c_r^2}{c_r^1}$. This case is valid when $\mu_1(\lambda_2 + \mu_2) \geq \lambda_1(\lambda_1 + 2\lambda_2 + \mu_2)$.

Preferred	Strongly Preferred
Condition 1 $\frac{c_r^2}{c_r^1} \leq A_1$ (4.22) $B_1 \leq \frac{c_e}{c_r^1}$ (4.23)	Condition 2 $\frac{c_r^2}{c_r^1} \leq A_1$ (4.24) $\frac{c_e}{c_r^1} \leq B_1$ (4.25)
Condition 3 $A_1 \leq \frac{c_r^2}{c_r^1} \leq A_2$ (4.28) $B_2 \leq \frac{c_e}{c_r^1}$ (4.29)	Condition 4 $\frac{c_r^2}{c_r^1} \leq A_2$ (4.32) $\frac{c_e}{c_r^1} \leq B_1$ (4.33)
Condition 5 $A_2 \leq \frac{c_r^2}{c_r^1} \leq A_3$ (4.34) $B_3 \leq \frac{c_e}{c_r^1}$ (4.35)	Condition 6* $A_2 \leq \frac{c_r^2}{c_r^1} \leq A_3$ (4.36) $\frac{c_e}{c_r^1} \leq B_4$ (4.37)

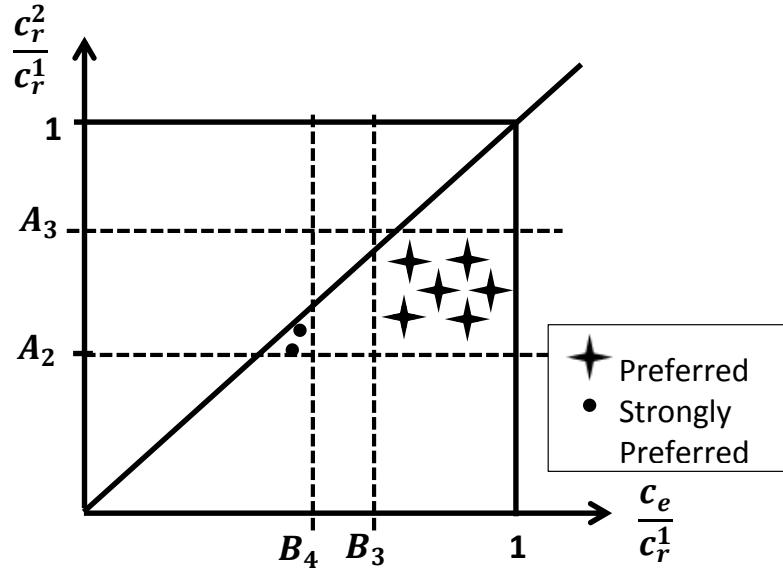
Table 4.1: Preferred and strongly preferred conditions when $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$

Table 4.1 summarizes our findings on preferred and strongly preferred classes when $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$. According to Table 4.1, (4.22) and (4.24) are the same inequalities and A_1 represents the right hand side of those inequalities. Moreover, inequality (4.23) and (4.25) are complement to each other, so that unless $\frac{c_e}{c_r^1}$ is less than B_1 , type-1 cannot be strongly preferred. Note that, the upper bound on $\frac{c_e}{c_r^1}$ must be greater than the lower bound on $\frac{c_r^2}{c_r^1}$. Hence, Condition 6 is valid when $A_2 \leq B_4$ which is possible when $\mu_1(\lambda_2 + \mu_2) \geq \lambda_1(\lambda_1 + 2\lambda_2 + \mu_2)$ holds. This parametric set is mathematically correct when departure rates are as high as arrival rates. However, this circumstance is not practically possible.

The following figures presents the conditions given in Table 4.1.



With some linear algebra, we can show that under the assumption of $\mu_1 \leq \mu_2$, $B_2 \leq B_1$ where B_1 is the left hand side of inequality (4.23) and B_2 is the left hand side of inequality (4.29). According to left hand side of Figure 4.2.2, since $B_1 \geq B_2$ there exists a region ($B_2 \leq \frac{c_e}{c_r^1} \leq B_1$ and $A_1 \leq \frac{c_r^2}{c_r^1} \leq A_2$) in which both preferred and strongly preferred conditions are valid. Hence, we can re-arrange Condition 3 and 4 by using Theorem 5. Then, the threshold between strongly preferred and preferred becomes B_1 . The right hand side of the Figure 4.2.2 shows Condition 1, 2, 3 and 4 after this re-arrangement.

Figure 4.2: Condition 5 and 6 when $\lambda_1\lambda_2 \leq \lambda_1(\lambda_1 + 2\lambda_2 + \mu_2) \leq \mu_1(\lambda_2 + \mu_2)$

With some linear algebra, we can show that that $B_4 \leq B_3$, where B_4 is the right hand side of inequality (4.37) and B_3 is the left hand side of inequality (4.35). Moreover, $B_4 \leq B_3 \leq B_1 \leq B_2$. According to Figure 4.2, there exists a region $(B_4 \leq \frac{c_e}{c_r^1} \leq B_3 \text{ and } A_2 \leq \frac{c_r^2}{c_r^1} \leq A_3)$ in which we do not know whether type-1 is preferred or strongly preferred.

Preferred	Strongly Preferred
Condition 1 $\frac{c_r^2}{c_r^1} \leq A_1$ (4.22) $B_1 \leq \frac{c_e}{c_r^1}$ (4.23)	Condition 2 $\frac{c_r^2}{c_r^1} \leq A_1$ (4.24) $\frac{c_e}{c_r^1} \leq B_1$ (4.25)
Condition 3 $A_1 \leq \frac{c_r^2}{c_r^1} \leq \overline{A_2}$ (4.30) $\overline{B_2} \leq \frac{c_e}{c_r^1}$ (4.31)	Condition 4 $\frac{c_r^2}{c_r^1} \leq \overline{A_2}$ (4.32) $\frac{c_e}{c_r^1} \leq B_1$ (4.33)

Table 4.2: Preferred and strongly preferred conditions when $\mu_2(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$

Table 4.2 shows preferred and strongly preferred conditions when $\mu_2(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$. For this set of parameters Condition 5 and 6 do not hold. According to Table 4.2, Condition 1 and 2 are the same as Table 4.1. The following figure presents the conditions given in Table 4.2.

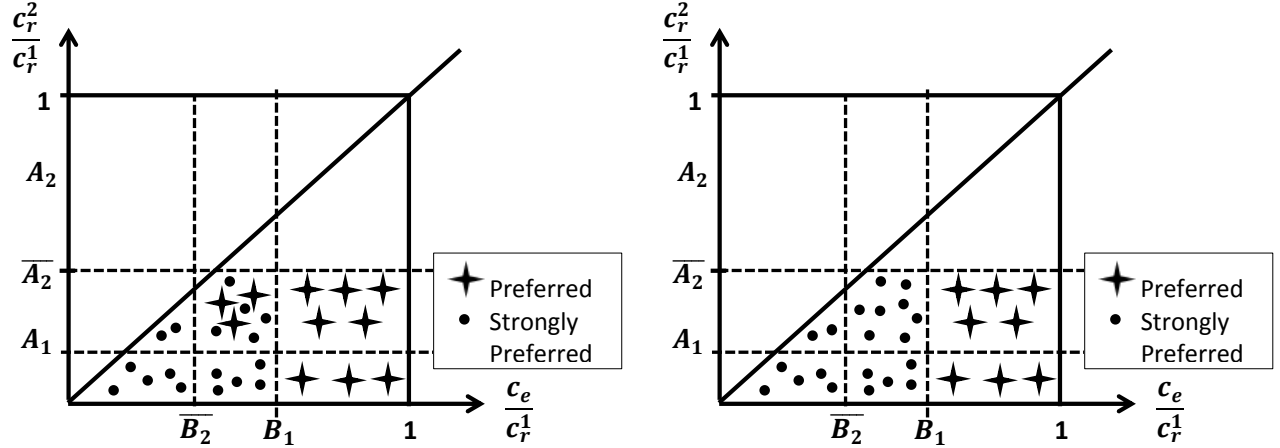


Figure 4.3: Condition 1-2-3 and 4 before (on the left) and after the arrangement (on the right) $\mu_2(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$

With some linear algebra, we can show that $\overline{B_2} \leq B_1$ where B_1 is the left hand side of inequality (4.23) and $\overline{B_2}$ is the left hand side of inequality (4.31). According to the left side of Figure 4.3, there exist a region ($\overline{B_2} \leq \frac{c_e}{c_r^1} \leq B_1$ and $A_1 \leq \frac{c_r^2}{c_r^1} \leq \overline{A_2}$) in which both preferred and strongly preferred conditions are valid. Hence, as we discussed above we can re-arrange the conditions using Theorem 5. The right hand side of the figure combines Condition 1, 2, 3 and 4 after this re-arrangement.

To sum up, if $\frac{c_e}{c_r^1} \leq \frac{\mu_1}{\lambda_1 + \mu_1} = B_1$, type-1 is strongly preferred. If we decrease λ_1 or increase μ_1 , B_1 increases. Hence, when the costs are fixed, if the arrival rate of a severe patient decreases or the expected LOS of the patient becomes shorter, hospital managers start to admit the incoming severe patients by discharging normal patients early. Moreover, type-1 is preferred when $\frac{c_r^2}{c_r^1} \leq \min\{\frac{\lambda_1}{\lambda_1 + \mu_2}, \frac{\mu_2(\lambda_2 + \mu_2)}{\lambda_2(\lambda_1 + \mu_2)}\}$ holds. When we fix the costs, if the arrival rates (λ_1 or λ_2) decrease or the expected LOS of a normal patient declines, hospital managers start to the admit incoming severe patients until

the last empty bed.

As a remark, when we re-arrange the conditions such a way that they satisfy our assumption, which is $\frac{c_r^2}{c_r^1} \leq \frac{c_e}{c_r^1}$, we can obtain more strict bounds on these conditions.

Chapter 5

NUMERICAL ANALYSIS

5.1 Memorial Hizmet Hospital

Memorial Hizmet Hospital has the capacity of 118 beds, 4 ORs and 16 ICUs. Furthermore, there exist 5 beds in the emergency room.

5.1.1 Data Set and Current Condition

Hizmet Hospital has high occupancy rates at certain times, especially in the mornings, and to cope with the bed allocation problem they plan to increase their capacity. However, we believe that this problem can be handled via effective use of beds. Initially, we observe the admission process at the hospital. When patients arrive to registration unit, they are assigned to empty beds. However hospital managers reported that sometimes there exist patients waiting for rooms at around 07:00 a.m. when the hospital does not have available beds. This motivates us to analyze the current system and present an admission policy model.

The hospital provided a data set over the period of 1 January 2011 to 6 October 2012 (645 days). The data includes personal information such as gender, age, patient ID as well as operational information such as registration date, bed assignment and discharge dates, floor, room numbers, surgery type, anesthesia type, surgery room, start and end time of surgery and initial diagnosis. However, among the surgery data, there is no distinction between emergency and planned surgeries. Moreover, the data set does not contain any decisions which are applied by managers of Hizmet Hospital to each patient.

The data set had a lot of noise which we should eliminate for a reliable analysis.

In a number of cases, there are replicated transactions, where the only difference is the initial diagnosis. The corresponding diagnoses are labeled as “null”. Moreover, some patients have negative LOS values because their arrival and departure dates were entered incorrectly. We deleted the records which were problematic because of a aforementioned reasons. This corresponds to about 5% of the whole data set, and resulted in 6599 transactions. We split these transactions according to departments as shown in Table 5.1, each of the departments has its own restrictions to assign its patients into wards as discussed in Chapter 3 (Appendix B). Moreover, we found the occupancy rates for each ward. In order to find the rate, for each day we sum the total number of occupied beds and divide this number with the ward’s capacity. Then, we take the time average of daily rates to find the occupancy rates for a month. For example, the monthly occupancy rate of ward-1 is around 60%. (See Appendix C).

Rank	Department	Number of transaction
1	Gynecology and obstetrics	3186
2	Urology	661
3	Otorhinolaryngology	610
4	Orthopedics	549
5	General Surgery	504
6	Cardiovascular Surgery	409
7	Cosmetic Surgery	238
8	Neurosurgery	155
9	Transplantation	120
10	Pediatric Surgery	90
11	Pediatrics	39
12	Dental Diseases	13
13	Other Departments	< 10

Table 5.1: Department transactions of Memorial Hizmet Hospital

The most “popular” department is gynecology and obstetrics. Therefore, we initially focused on this department. However, most of the patients are coming for giving birth and their LOSs are almost deterministic, since LOS is 2 days and 1 day for cesarean section and natural birth, respectively. Thus, this department does not

satisfy our model assumptions. The missing diagnoses, labeled as “null”, constitute more than half of the data in the first four departments. Consequently, we focus on general surgery department because the first diagnoses are better defined than the others.

5.2 Data Preparation and Parameter Estimation

Each patient is special and LOS of two patients differ even if they undergo the same procedure. Age and chronic diseases (diabetes, obesity, etc.) have an impact on the recovery process. To specify an appropriate LOS distribution, we select a subset of the data that corresponds to a certain group of patients for whom the patient characteristics are not very dominant. The recovery process of elderly people is longer and the chance of multiple chronic diseases is higher. Thus, we took 65 as an upper bound for age of patients, following the suggestions of medical and public authorities. Within the patients who are younger than 65, we select those who have a diagnosis of abdominal cavity or breast problems, because these patients are generally categorized as trauma patients and the recovery process of trauma patients are similar. After the elimination, we obtain 271 distinct transactions.

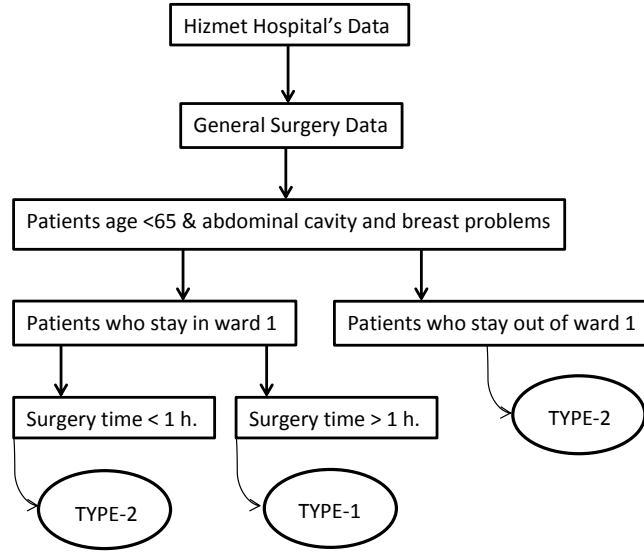


Figure 5.1: Flowchart of the data preparation

To classify the type-1 and type-2 patients, we consider the patients who stay in their own ward and the patients who do not (For a general surgery patient, this corresponds to floor (ward) 1). Our model assumes that patients who do not stay in their ward are mild (type-2) patients. Hence, we label the patients who stay out of their ward as type-2 patients. To classify the patients who stay in ward 1 as type-1 or type-2 patients, we use surgery durations.

When surgery duration increases, the probability of having complications also increases, therefore, longer surgery times yield longer LOS on the average. Thus, a severe patient can be identified according to his/her surgery duration. Doctors in Hizmet Hospital stated that for general surgery patients, surgeries that take longer than 1 hour generally indicate the possibility of more complications. Hence, patients whose surgery duration is less than 1 hour are assumed to be type-2 patients, and the remaining patients are classified as type-1. To test the claim that LOS depends on the surgery duration, we run a simple linear regression between these two variables for the 271 transactions. The corresponding p-value shows that the relationship is statistically significant for all plausible significance levels. Moreover, the R^2 value

is 37.4%, meaning that the 37.4% of the variability in LOS can be explained by the surgery durations. Hence, we can assume that surgery duration is a good indicator for the LOS. After this classification, we obtain that average LOS of type-1 and type-2 patients are 1.5 and 1.1 days, respectively.

There exist 109 type-1 patients and 162 type-2 patients (97 of them stay in ward-1 and 65 of them stay out of their ward). Moreover, the average LOS of general surgery is 1.46 days and the average occupancy rate is 60% as we mentioned before. Then, the average arrival rate is 5.4. ($\lambda = load \times \# of beds \times \mu$). Hence, arrival rate of type-1 patient is 2.2 and arrival rate of type-2 patient is 3.2. Note that, general surgery covers variety of initial diagnoses, however, definition of a type- i patient contains only some of these diagnoses. Hence, the average LOS of general surgery is not consistent with the average LOS of type- i patients.

To determine the values of the rejection costs, we assume that each of these parameters consists of two components, one is fixed and the other is incurred linearly over time. We let c be the cost of one-day stay for any patient, and refer to it as daily cost. Fixed cost refers to the surgery cost and changes with the type of the patient. So we have:

$$c_r^1 = c\left(\frac{1}{\mu_1} + \frac{1}{\mu_2}\right) + fixed\ cost_1$$

$$c_r^2 = \frac{c}{\mu_2} + fixed\ cost_2,$$

where we took daily and surgery cost values from Hizmet Hospital.

The cost of early discharge, on the other hand, takes the possible consequences of early discharge into account to reflect the risk of relapse. Accordingly, we set $c_e = c_r^2 + risk\ portion \times cost\ of\ risk$. We calculate the risk component from the data taken from Newby et al. (2000). We compute the expected life years that a patient loses due to an early discharge decision. Then, we multiply this value with cost effectiveness of a full health year (1 QALY=\$50,000). The currency we use is Turkish liras.

In order to calculate the cost of overflow, we took the cost of opening the spare

ward from Hizmet Hospital. Since c_o represents the cost per person per unit time, we divide the opening cost with the average number of patients who stay in the spare ward (floor 3) times average LOS of type-2 patients. The average number of patients who stay in ward-3 is found by multiplying the capacity of the spare ward with average occupancy rate of that ward. So that, $c_o = \frac{\text{opening cost}}{\text{average number of patients} \times \text{average LOS}} = \frac{1}{13 \times 0.3 \times 1.1} = 0.23$.

The following table shows the estimated values of all parameters for the base case:

Parameter	Estimated value
λ_1	2.2
λ_2	3.2
$\frac{1}{\mu_1}$	1.5
$\frac{1}{\mu_2}$	1.1
M	13
c_r^1	11.7
c_r^2	4.1
c_e	4.7

Table 5.2: Parameter Estimations (Base case)

5.3 Base Case Model

We use the relative value iteration to solve the long-run average cost problem. For the parameters given in Table 5.2, we focus on two different scenarios: allowing overflow and no overflow.

As a remark, we cannot compare our results with decisions of managers in Hizmet Hospital, because our dataset does not contain such information.

5.3.1 Allowing overflow

For this case, we permit overflows to the empty ward, so that $\overline{M} = 25$. (There exist 13 beds in the main ward and 12 beds in the overflow ward.) We took $c_o = 0.23$ and

all other parameters are same as Table 5.2. We find the differential cost value (h) as shown in Figure 5.2, where the average cost (g^*) is 0.041718.

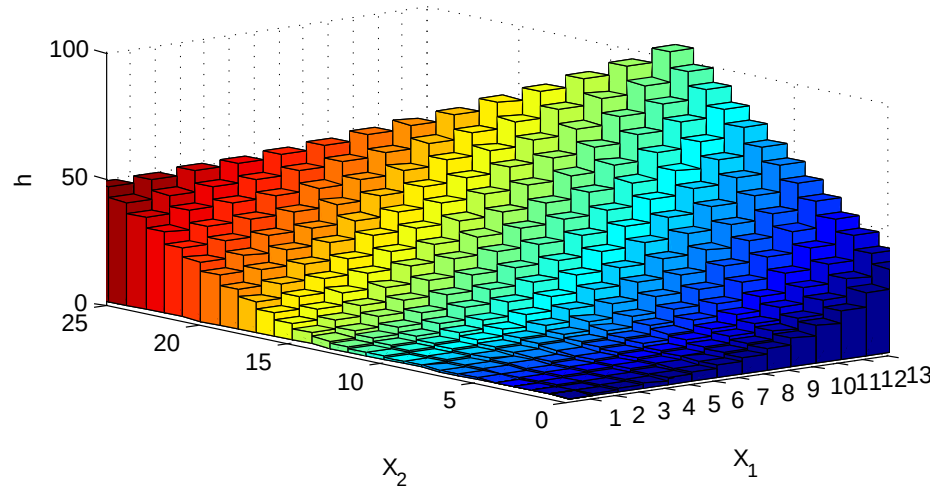


Figure 5.2: Differential cost (h), for initial state (x_1, x_2)

Figure 5.2 demonstrates that our value function is increasing in both x_1 and x_2 as we prove in Theorem 2. Moreover, we observe that having one more type-1 patient in the system costs more than having one more type-2 patient, as we prove in Theorem 3.

Table 5.3 and 5.4 show the optimal admission policies for type-1 and type-2 patients with the base case parameters. The number of type-1 patients is x_1 (left column) and the number of type-2 patients is x_2 (upper row).

According to Table 5.3, type-1 patients are always admitted to the main ward. When the main ward is full with type-1 and type-2 patients, type-1 patients are still admitted into the ward by placing type-2 patients to the overflow ward. If the number of patients is more than 22, the optimal policy is always to discharge type-2 patients to admit type-1 patients, because the expected overflowing cost becomes higher than the expected cost of early discharge.

According to Table 5.4, type-2 patients are always admitted to the main ward. However, at most 7 of beds in overflow ward are used for type-2 patients. There is a region where incoming type-2 patients are rejected. The reason for this rejection is to allocate beds for expected incoming type-1 patients, who are always admitted.

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	□	□	□	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	★	★	★	★	△	△	△
1	□	□	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	★	★	★	△	△	△	△	
2	□	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	★	★	△	△	△	△	△		
3	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	★	△	△	△	△	△	△			
4	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	△	△	△	△	△	△	△	△			
5	□	□	□	□	□	□	□	□	★	★	★	★	★	△	△	△	△	△	△	△	△	△				
6	□	□	□	□	□	□	□	★	★	★	★	△	△	△	△	△	△	△	△	△	△					
7	□	□	□	□	□	□	★	★	★	★	△	△	△	△	△	△	△	△	△							
8	□	□	□	□	□	★	★	★	△	△	△	△	△	△	△	△	△	△								
9	□	□	□	□	★	★	△	△	△	△	△	△	△	△	△	△	△									
10	□	□	□	★	△	△	△	△	△	△	△	△	△	△	△	△										
11	□	□	★	△	△	△	△	△	△	△	△	△	△	△	△											
12	□	△	△	△	△	△	△	△	△	△	△	△	△	△												
13	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲													

Table 5.3: Decisions when a type-1 patient arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. (Base case with $\bar{M} = 25$)

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	□	□	□	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	★	▲	▲	▲	▲	▲	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	▲	▲	▲	▲	▲	▲	▲	▲
2	□	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	★	▲	▲	▲	▲	▲	▲	▲	▲	▲
3	□	□	□	□	□	□	□	□	□	★	★	★	★	★	★	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
4	□	□	□	□	□	□	□	□	★	★	★	★	★	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
5	□	□	□	□	□	□	□	★	★	★	★	★	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
6	□	□	□	□	□	□	★	★	★	★	★	★	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
7	□	□	□	□	□	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
8	□	□	□	□	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
9	□	□	□	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
10	□	□	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
11	□	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
12	□	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
13	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★

Table 5.4: Decisions when a type-2 patient arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. (Base case with $\overline{M} = 25$)

5.3.2 No overflow

In this case, we do not permit overflows to the empty ward, so that $\overline{M} = M = 13$. After calculations g^* is 0.099251, and the optimal policies are as follows:

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	Δ
1	□	□	□	□	□	□	□	□	□	□	□	□	Δ	
2	□	□	□	□	□	□	□	□	□	□	□	Δ		
3	□	□	□	□	□	□	□	□	□	□	Δ			
4	□	□	□	□	□	□	□	□	□	Δ				
5	□	□	□	□	□	□	□	□	Δ					
6	□	□	□	□	□	□	□	Δ						
7	□	□	□	□	□	□	Δ							
8	□	□	□	□	□	Δ								
9	□	□	□	□	Δ									
10	□	□	□	Δ										
11	□	□	Δ											
12	□	Δ												
13	▲													

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	▲				
4	□	□	□	□	□	□	□	□	▲					
5	□	□	□	□	□	□	□	▲						
6	□	□	□	□	□	□	▲							
7	□	□	□	□	□	▲								
8	□	□	□	□	▲									
9	□	□	□	▲										
10	□	□	▲											
11	□	▲												
12	□	▲												
13	▲													

Table 5.5: Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: (□) Accept, (Δ) Accept with early discharge, (★) Accept with overflow, (▲) Reject. (Base case with $\bar{M} = M$)

Table 5.5 shows optimal actions when a type- i patient arrives in state x and $i = 1, 2$. According to Table 5.5, type-1 patients are always accepted even the ward is full. Hence, type-1 patients are strongly preferred for our base case. The base case parameter estimations do not satisfy the sufficient conditions we found for type-1 patients to be strongly preferred given in Table 4.2, so that these conditions cannot identify type-1 as strongly preferred. Hence, the conditions we found in section 4.2.2 are not very tight. Moreover, type-2 patients are always accepted. When we replace parameters with our base case estimations, the condition in Theorem 4 is not ensured either. However, when the tighter bound is used to derive the other sufficient condition for type-2 patients to be preferred, i.e. when the condition becomes $\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_e}$, the base case parameter estimations guarantee that type-2 is preferred.

Note that, when we compare overflow and no overflow cases, we observe that optimal actions are the same when the main ward is not full. The optimal actions for the last bed in the overflow ward (given in Table 5.3 and Table 5.4) are the same as the optimal actions for the last bed in the main ward when we do not allow overflow

(given in Table 5.5).

5.4 Sensitivity Analysis

In this section, we focus on the impact of c_r^1 and c_e on the optimal actions and the average cost value when c_r^2 is fixed. Then, we also consider the impact on higher load (l), changing by λ_1 and λ_2 values. All these analyses are done for $\overline{M} = M$.

5.4.1 Impact of c_r^1 and load

Average cost values decrease when c_r^1 decreases. Moreover, higher load causes higher number of transactions. Hence, the average cost values increase when load (l) increases (Table 5.6).

		$\frac{c_r^1}{c_r^2} = 1.5$	$\frac{c_r^1}{c_r^2} = 2$	$\frac{c_r^1}{c_r^2} = \frac{11.7}{4.7}$	$\frac{c_r^1}{c_r^2} = 4$
g^*	$l=60\%$	0.094492	0.099236	0.099251	0.099270
	$l=85\%$	0.262566	0.294234	0.294830	0.294931
	$l=110\%$	0.698205	0.822269	0.875955	0.879651

Table 5.6: Average cost values when c_r^1 changes

At each load value, the effect of the cost ratio ($\frac{c_r^1}{c_r^2}$) is limited, e.g. when this ratio is doubled from 2 to 4, the largest increment is only 7% for $l = 110\%$. On the other hand, 25% increment in the load changes g^* significantly at all levels.

		$\frac{c_r^1}{c_r^2} = 1.5$	$\frac{c_r^1}{c_r^2} = 2$	$\frac{c_r^1}{c_r^2} = 4$
Optimal Actions	$l=60\%$	Table 5.8	S	S
	$l=85\%$	Table 5.9-L	Table 5.10-L	S
	$l=110\%$	Table 5.9-R	Table 5.10-R	S

Table 5.7: Optimal actions when c_r^1 changes (S: Same as the base case)

Table 5.7 summarizes the optimal actions when c_r^1 changes. For our base case, $\frac{c_r^1}{c_r^2}$ is approximately 3. When we increase this fraction to 4, the optimal actions do not change. They are the same as Table 5.5. However, if this fraction becomes smaller, the optimal actions taken upon a type-1 arrival change as indicated in Table 5.7. On the other hand, the optimal actions for type-2 patients remain the same as base case for all c_r^1 values and load levels.

From Table 5.8, we can see the effect of c_r^1 when the load is fixed. If the cost of rejection for a severe patient is sufficiently small, the system prefers to reject the patient instead of accepting him when all beds are occupied. This rejection stream starts in states with a higher number of severe patients. When the load is higher, we see the same pattern. For $l = 85\%$ Table 5.9-L and 5.10-L and for $l = 110\%$ Table 5.9-R and 5.10-R show this pattern.

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	△
1	□	□	□	□	□	□	□	□	□	□	□	□	△	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	□	▲			
4	□	□	□	□	□	□	□	□	□	▲				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	□	▲						
7	□	□	□	□	□	□	▲							
8	□	□	□	□	□	▲								
9	□	□	□	□	▲									
10	□	□	□	▲										
11	□	□	▲											
12	□	▲												
13	▲													

Table 5.8: Decisions when a type-1 patient arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. ($l = 60\%$ and $\frac{c_r^1}{c_r^2} = 1.5$)

When $\frac{c_r^1}{c_r^2} = 2$, if utilization is 60%, higher LOS of type-1 patient can be handled. However, if utilization is 85% or more, higher LOS of type-1 patient can cause a higher burden. According to Table 5.10, when the ward is full, the system does not want to accept a type-1 patient when compared to the base case. If the system accepted the patient, it would face with the incoming type-2 patients during hospitalization time of type-1 patient and the system would reject them. It is not a problem when the utilization is 60%, because the arrival rates are low hence the number of type-2 patients who would be rejected is lower. As a result, when the load increases, the system starts to reserve more beds for type-2 patients via rejecting type-1 patients. This result confirms the management preference for patients staying shorter in Hizmet Hospital.

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	□	▲			
4	□	□	□	□	□	□	□	□	□	▲				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	□	▲						
7	□	□	□	□	□	□	▲							
8	□	□	□	□	□	▲								
9	□	□	□	▲	▲									
10	□	□	▲	▲										
11	□	▲	▲											
12	▲	▲												
13	▲													

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	▲	▲			
4	□	□	□	□	□	□	□	□	▲	▲				
5	□	□	□	□	□	□	□	▲	▲					
6	□	□	□	□	□	□	▲	▲						
7	□	□	□	□	□	▲	▲							
8	□	□	□	□	▲	▲								
9	□	□	▲	▲	▲									
10	□	▲	▲	▲										
11	▲	▲	▲											
12	▲	▲												
13	▲													

Table 5.9: Decisions when a type-1 patient arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. $\frac{c_r^1}{c_r^2} = 1.5$ and $l = 85\%$ (on the left-L), $l = 110\%$ (on the right-R)

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	△
1	□	□	□	□	□	□	□	□	□	□	□	□	△	
2	□	□	□	□	□	□	□	□	□	□	□	△		
3	□	□	□	□	□	□	□	□	□	□	△			
4	□	□	□	□	□	□	□	□	□	△				
5	□	□	□	□	□	□	□	□	△					
6	□	□	□	□	□	□	□	△						
7	□	□	□	□	□	□	△							
8	□	□	□	□	□	▲								
9	□	□	□	□	▲									
10	□	□	□	▲										
11	□	□	▲											
12	□	▲												
13	▲													

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	□	▲			
4	□	□	□	□	□	□	□	□	□	▲				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	□	▲						
7	□	□	□	□	□	□	▲							
8	□	□	□	□	□	▲								
9	□	□	□	□	▲									
10	□	□	□	▲										
11	□	□	▲											
12	□	▲												
13	▲													

Table 5.10: Decisions when a type-1 patient arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. $\frac{c_r^1}{c_r^2} = 2$ and $l = 85\%$ (on the left-L), $l = 110\%$ (on the right-R)

When $\frac{c_r^1}{c_r^2} = 4$ and $\frac{c_r^1}{c_r^2} = 2, l = 60\%$ type-1 is strongly preferred. Moreover, type-1 is preferred when $\frac{c_r^1}{c_r^2} = \{2, 4\}$ and $\frac{c_r^1}{c_r^2} = 1.5, l = 60\%$. However, the only parameter set that satisfies the condition derived by inequalities (4.32) and (4.33) in section 4.2.2 is when $\frac{c_r^1}{c_r^2} = 4, l = 60\%$. The remaining parameter sets satisfy none of the conditions we derive in section 4.2.2 because of the same reason we discussed in the base case. Moreover, type-2 is preferred for all these parameter sets, because $\frac{c_r^2}{c_e} = 0.87 \geq \frac{\lambda_1}{\lambda_1 + \mu_2}$ is satisfied when $\lambda_1 \leq 6.2$, which is valid for all values of l .

5.4.2 Impact of c_e and load

In order to observe the effect of c_e , we keep c_r^1 and c_r^2 same as Table 5.2. The effect of c_e on g^* is higher than that of c_r^1 as a comparison of Table 5.11 and 5.6.

		$c_e = 4.2$	$c_e = 4.7$	$c_e = 5$	$c_e = 8$
g^*	$l=60\%$	0.094599	0.099251	0.102042	0.129782
	$l=85\%$	0.282349	0.294830	0.302319	0.369454
	$l=110\%$	0.836427	0.822269	0.899672	1.003472

Table 5.11: Average cost values when c_e changes

		$c_e = 4.2$	$c_e = 5$	$c_e = 8$
Optimal Actions	$l=60\%$	S	S	Table 5.13
	$l=85\%$	S	S	Table 5.14
	$l=110\%$	S	S	Table 5.15

Table 5.12: Optimal actions when c_e changes (S: Same as the base case)

The difference between c_e and c_r^2 comes from the fact that there exists a probability of readmission when we discharge patients early. According to Table 5.12, even if we decrease the cost of early discharge, optimal actions do not change. They are the same as in Table 5.5. Hence, if the probability of re-admission is very low, the optimal actions are the same as the base case. Note that, $c_e = 5$ also does not change the optimal actions.

Type-2 is preferred for all the parameter sets except $c_e = 8$. We know that type-2 is preferred when $\frac{c_r^2}{c_e} = \frac{4.1}{8} = 0.51 \geq \frac{\lambda_1}{\lambda_1 + \mu_2}$ so that type-2 is preferred when $\lambda_1 \leq 0.96$, and we have $\lambda_1 \geq 0.96$ for all levels of l considered here. According to Table 5.13, 5.14 and 5.15, we reject type-2 patients even if there exists an empty bed. Moreover, when the load increases, the rejection region for an incoming type-2 patient increases.

When c_e is sufficiently high ($c_e = 8$), system starts to reject type-1 patients instead of accepting type-1 patients with early discharge decision when the system is full. Hence, type-1 is preferred and not strongly preferred. Moreover, system starts to

reject more type-1 patients when the load increases, similarly to the impact of c_r^1 . However, this is not to allocate beds for type-2 patients since type-2 is not preferred.

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	Δ
1	□	□	□	□	□	□	□	□	□	□	□	□	Δ	
2	□	□	□	□	□	□	□	□	□	□	□	Δ		
3	□	□	□	□	□	□	□	□	□	□	Δ			
4	□	□	□	□	□	□	□	□	□	Δ				
5	□	□	□	□	□	□	□	□	Δ					
6	□	□	□	□	□	□	□	Δ						
7	□	□	□	□	□	□	Δ							
8	□	□	□	□	□	Δ								
9	□	□	□	□	Δ									
10	□	□	□	▲										
11	□	□	▲											
12	□	▲												
13	▲													

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	□	▲			
4	□	□	□	□	□	□	□	□	□	▲				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	□	▲						
7	□	□	□	□	□	□	▲							
8	□	□	□	□	▲	▲								
9	□	□	□	▲	▲									
10	□	□	▲	▲										
11	□	▲	▲											
12	▲	▲												
13	▲													

Table 5.13: Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: (□) Accept, (Δ) Accept with early discharge, (★) Accept with overflow, (▲) Reject. ($c_e = 8$ and $l = 60\%$)

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	△
1	□	□	□	□	□	□	□	□	□	□	□	□	△	
2	□	□	□	□	□	□	□	□	□	□	□	△		
3	□	□	□	□	□	□	□	□	□	□	△			
4	□	□	□	□	□	□	□	□	□	△				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	□	▲						
7	□	□	□	□	□	□	▲							
8	□	□	□	□	□	▲								
9	□	□	□	□	▲									
10	□	□	□	▲										
11	□	□	▲											
12	□	▲												
13	▲													

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	□	▲			
4	□	□	□	□	□	□	□	□	□	▲				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	▲	▲						
7	□	□	□	□	□	▲	▲							
8	□	□	□	□	▲	▲								
9	□	□	□	▲	▲									
10	□	□	▲	▲										
11	□	▲	▲											
12	▲	▲												
13	▲													

Table 5.14: Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. ($c_e = 8$ and $l = 85\%$)

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	□	▲	
2	□	□	□	□	□	□	□	□	□	□	□	▲		
3	□	□	□	□	□	□	□	□	□	□	▲			
4	□	□	□	□	□	□	□	□	□	▲				
5	□	□	□	□	□	□	□	□	▲					
6	□	□	□	□	□	□	□	▲						
7	□	□	□	□	□	□	▲							
8	□	□	□	□	□	▲								
9	□	□	□	□	▲									
10	□	□	□	▲										
11	□	□	▲											
12	□	▲												
13	▲													

$x_1 \backslash x_2$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	□	□	□	□	□	□	□	□	□	□	□	□	□	▲
1	□	□	□	□	□	□	□	□	□	□	□	▲	▲	
2	□	□	□	□	□	□	□	□	□	□	▲	▲		
3	□	□	□	□	□	□	□	□	□	▲	▲			
4	□	□	□	□	□	□	□	□	▲	▲				
5	□	□	□	□	□	□	□	▲	▲					
6	□	□	□	□	□	□	▲	▲						
7	□	□	□	□	□	▲	▲							
8	□	□	□	□	▲	▲								
9	□	□	□	▲	▲									
10	□	□	▲	▲										
11	□	▲	▲											
12	▲	▲												
13	▲													

Table 5.15: Decisions when a type-1 patient (on the left) or a type-2 patient (on the right) arrives: (□) Accept, (△) Accept with early discharge, (★) Accept with overflow, (▲) Reject. ($c_e = 8$ and $l = 110\%$)

Chapter 6

CONCLUSION

In this research, we model a bed allocation problem with an MDP and find admission, early discharge and rejection policies to minimize the expected cost. We model two hospital wards: a main ward and an overflow ward. Severe (type-1) and mild (type-2) patients arrive to the wards according to a Poisson process and only type-2 patients can be admitted to the overflow ward. We assume that recovery time of a severe patient is a two stage hypoexponential distribution whereas recovery of a mild patient takes an exponential time. The main objective is to develop insights on the admission control of patients to the wards.

Initially, preliminary properties of the model are analyzed, which provides practical insights into hospital admission policies. Early discharge decision is valid if there exists at least one type-2 patient in the system. This decision is not optimal unless all beds are occupied in the main ward. Then, the effect of an additional patient is analyzed, which shows the monotonicity behavior of our model. The expected burden that a type- i patient brings when he/she admitted to the system is always positive. Hence, our value function is monotonically increasing. Moreover, accepting a type-1 patient via discharging a type-2 patient early is a burden.

Secondly, we focus on special cases of parameters to derive the conditions which guarantee that there exist preferred and strongly preferred classes in our model. These conditions are valid when we use only the main ward. We found that if the rejection cost of a type-2 patient is greater than the expected cost of an additional type-2 patient, then it is always better to accept the type-2 patient, where the expected cost of a type-2 patient consists of probability of an incoming type-1 patient times

the cost of early discharge. If type-1 is strongly preferred then it must be preferred. Hence, we focus on the cases where type-1 is preferred and obtain sufficient, but not necessary, conditions for being strongly preferred and preferred classes as shown in Table 4.1 and 4.2. All of these policies can improve healthcare delivery, because by controlling inpatient admissions we can diminish problems such as blockages and diversions, cancellations due to backlogs.

Our model is based on a Turkish private hospital, called Hizmet Hospital. Hizmet Hospital provide us a dataset over a period of 1 January 2011 to 6 October 2012. From the dataset we found the occupancy rates of the hospital wards as 60%. Although gynecology and obstetrics has the highest number of patient, we focus on the general surgery department because patient diagnoses in that department are well-recorded and LOS of the patients in that department is more random. When we apply our model to the dataset, we observe that type-1 patients are strongly preferred and type-2 patients are preferred. According to results of sensitivity analysis, when load increases, the system starts to reject type-1 patients to accept type-2 patient, whose LOS are less than type-1. Type-2 patient is always preferred for the cost values we choose. However, when we increase the early discharge value, we observe that type-2 is not preferred anymore. From here, we can conclude that optimal policies depend on both types of costs and the system load.

Our model can be extended in terms of overflow costs. Assuming overflow cost as a fixed cost is more realistic, because the system pays the cost when it opens the empty ward. Moreover, patients can be moved to another main ward as an overflow patient. Then, the variable overflow-cost term we presented in our model can be interpreted as cost of hospitalization in another main ward.

The model we propose in this research also has some limitations. First, it does not contain the fact that normal patients can become a serious patient. Moreover, serious patients can die in the first stage recovery, which is not covered in our model.

We state that early discharge increases the probability of complications which

results in re-admission of early discharged patient to the hospital. However, managers of Hizmet Hospital indicated that they did not prefer to discharge a patient who has a probability of complications. Hence, the model we propose in this thesis seems more appropriate for a department where all patients are severe like an ICU. Assume that all beds are unavailable in the ICU and an arriving patient who requires intensive care is admitted to the ICU by transferring one of the existing patients to a ward. Since the transferred patient also needs the intensive care, we can talk about the probability of complication in this case.

Moreover, rejection decisions are more vital when incoming patients are in severe condition. On the other hand, patients in Hizmet Hospital are generally elective patients who do not have critical conditions. Most of the patients who arrive to ER are in a severe condition. After triaging patients into two groups as mild and severe, ER doctors can decide on which policy they should use: accept, reject, accept with early discharge or use temporary beds to accept the patient. ERs have higher load, the recovery process of an ER patient is similar and the rejection cost of an urgent patient is more worthwhile, when compared to Hizmet Hospital.

For future work, one may wish to consider to use overflow ward to find preferred and strongly preferred conditions. One can also consider a different dataset, especially an ER or an ICU dataset in which early discharge and rejection decisions are more important than any other department.

BIBLIOGRAPHY

- Department of Health(UK). 2000. Shaping the future NHS: Long term planning for hospitals and related services: Consultation document on the findings of the national beds inquiry.
- Adan, I., J. Bekkers, N. Dellaert, and X. Vissers, J.and Yu 2009. Patient mix optimization and stochastic resource requirements: a case study in cardiothoracic surgery planning. *Health Care Management Science* 12(2), 129–141.
- Bekker, R. and P. M. Kooleman 2011. Scheduling admissions and reducing variability in bed demand. *Health Care Management Science* 14(3), 237–249.
- Beliën, J. and E. Demeulemeester 2007. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research* 176(2), 1185–1204.
- Beliën, J., E. Demeulemeester, and B. Cardoen 2006. Visualizing the demand for various resources as a function of the master surgery schedule: A case study. *Journal of Medical Systems* 30(5), 343–350.
- Blake, J., F. Dexter, and J. Donald 2002. Operating room managers use of integer programming for assigning block time to surgical groups: a case study. *Anesthesia and Analgesia* 94(1), 143–148.
- Blake, J. and J. Donald 2002. Mount sinai hospital uses integer programming to allocate operating room time. *Interfaces* 32(2), 63–73.
- Cardoen, B., E. Demeulemeester, and J. Beliën 2010. Operating room planning and

- scheduling: a literature review. *European Journal of Operational Research* 201(3), 921–932.
- Cayirli, T. and E. Veral 2012. Outpatient scheduling in health care: A review of literature. *Production and Operations Management* 12(4), 519–549.
- Chan, C., V. Farias, N. Bambos, and G. Escobar 2010. Maximizing throughput of hospital intensive care units with patient readmissions. *Operations Research* 33(1), 27–40.
- Cil, E., E. Örmeci, and F. Karaesmen 2009. Effects of system parameters on the optimal policy structure in a class of queueing control problems. *Queueing System* 61(1), 273–304.
- Creemers, S. and M. Lambrecht 2010. Queueing models for appointment-driven systems. *Annals of Operations Research* 178(1), 155–172.
- de Bruin, A. and R. Bekker 2010. Time dependent analyses for refused admissions in clinical wards. *Annals of Operations Research* 178(1), 45–65.
- de Bruin, A., R. Bekker, L. van Zaten, and G. Koole 2010. Dimensioning hospital wards using erlang loss model. *Annals of Operations Research* 178(1), 23–43.
- Denton, B. (Ed.) 2013. *Handbook of Healthcare Operations Management: Methods and Applications*. Springer.
- Dobson, G., H. Lee, and E. Pinker 2010. A model of icu bumping. *Operations Research* 58(6), 1564–1576.
- Ghoneim, H. and S. Stidham 1985. Optimal control of arrivals to two queues in series. *European Journal of the Operational Research* 21, 399–409.
- Green, L. and S. Savin 2008. Reducing delays for medical appointments: A queueing approach. *Operations Research* 56(6), 1526–1538.

- Gupta, D. and B. Denton 2008. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* 40(9), 800–819.
- Helm, J., S. AhmadBeygi, and M. Van Oyen 2010. Design and analysis of hospital admission control for operational effectiveness. *Production and Operations Management* 20(3), 359–374.
- Helm, J. E. and M. Van Oyen 2011. Design and optimization methods for elective hospital admissions. Technical report, Department of IOE, Univeristy of Michigan, Ann Arbor, MI.
- Jun, J. B., S. H. Jacobson, and J. R. Swisher 1999. Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society* 50(2), 109–123.
- KC, D. S. and C. Terwiesch 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing and Service Operations Management* 14(16), 50–65.
- Kendall, D. G. 1953. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals o Mathematical Statistics* 24(3), 338–354.
- Koole, G. 2006. Monotonicity in markov reward and decision chains: Theory and applications. *Foundations and Trends in Stochastic Sytems* 1(1), 1–76.
- Lippman, S. 1975. Applying a New Device in the Optimization of Exponential Queuing Systems. *Operations Research* 23(4), 687–710.
- Macario, A., T. Vitez, B. Dunn, and T. McDonald 1995. Where are the costs in perioperative care? analysis of hospital costs and charges for inpatient surgical care. *Anesthesiology* 83(6), 1138–1144.

- Magerlein, J. and J. Martin 1978. Surgical demand scheduling: a review. *Health Services Research* 13(4), 418–433.
- Newby, L. K., E. L. Eisenstein, R. M. Califf, T. D. Thompson, C. L. Nelson, E. D. Peterson, et al. 2000. Cost effectiveness of early discharge after uncomplicated acute myocardial infarction. *New England Journal of Medicine* 342(11), 749–755.
- Örmeci, E., A. Burnetas, and J. van der Wal 2001. Admission policies for a two class loss system.
- Örmeci, E. and J. Van der Wal 2006. Admission policies for a two class loss system with general interarrival times. *Stochastic Models* 22(1), 37–54.
- Roberts, C. 2012. The effect of outlying patients to clinically inappropriate wards. Master’s thesis.
- Santibáñez, P., M. Begen, and D. Atkins 2007. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Management Science* 10(3), 269–282.
- Ulukus, M., R. Gullu, and E. Ormeci 2011. Admission and termination control of a two class loss systems. *Stochastic Models* 27(1), 2–25.
- Van Oostrum, J., M. Van Houdenhoven, J. Hurink, E. Hans, G. Wullink, and G. Kazemier 2008. A master surgery scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum* 30(2), 355–374.
- Zeng, B., A. Turkcan, J. Lin, and M. Lawley 2010. Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities. *Annals of Operations Research* 178(1), 121–144.

Appendices

Appendix A

DERIVATION OF THE CONDITIONS WHERE TYPE-1 PATIENTS ARE PREFERRED AND STRONGLY PREFERRED

Case 1. $c_r^1 \geq d_1 \geq d_{12} + c_r^2$

Case 1.i. $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}$

For all cases starting with i., we will insert $\bar{d}_2 = \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2}$ when there exist an d_2 term in the equation.

Derivation of $\bar{d}_1$:

$$\begin{aligned} \bar{d}_1 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \mu_1} \text{ as in (4.19).} \\ &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} + \left(\frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2} \right) \frac{\mu_1}{\lambda_1 + \mu_1} \\ &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} \left(1 + \frac{\mu_1}{\lambda_1 + \mu_2} \right) \\ &= \frac{\lambda_1 (\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 \text{ as in (4.20).} \end{aligned}$$

We will use $\bar{d}_1$ for case 1.i.a and case 1.i.b.

Case 1.i.a. $d_{12} \geq c_r^1 - c_e$

Derivation of $\bar{d}_{12}$:

Initially lets find $d_1 - d_2$.

$$\begin{aligned} d_1 - d_2 &= \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 - \frac{\lambda_1}{\lambda_1 + \mu_2} c_r^1 \\ &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1 + \mu_1 + \mu_2}{\lambda_1 + \mu_1} - 1 \right) c_r^1 \\ &= \frac{\lambda_1 \mu_2}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} \bar{d}_{12} &= \frac{\mu_2 - \mu_1}{\mu_2} d_1 + \frac{\mu_1}{\mu_2} d_2, \text{ as in (4.17)} \\ &= d_1 - \frac{\mu_1}{\mu_2} (d_1 - d_2), \text{ insert } (d_1 - d_2) \text{ we found in (A.1)} \\ &= \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 - \frac{\mu_1}{\mu_2} \left(\frac{\lambda_1 \mu_2}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 \right) \\ &= \frac{\lambda_1}{\lambda_1 + \mu_1} c_r^1, \text{ as in (4.21).} \end{aligned}$$

Check whether $c_r^1 \geq d_1$ holds:

$$c_r^1 \geq? \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1$$

$$(\lambda_1 + \mu_1)(\lambda_1 + \mu_2) \geq? \lambda_1(\lambda_1 + \mu_1 + \mu_2),$$

which is only valid when $\mu_1 \mu_2 \geq 0$ and it is always true.

Check whether $d_1 \geq d_{12} + c_r^2$ holds:

$$\begin{aligned} d_1 - d_{12} &\geq? c_r^2 \\ d_1 - (d_1 - \frac{\mu_1}{\mu_2} (d_1 - d_2)) &\geq? c_r^2 \\ \frac{\mu_1}{\mu_2} (d_1 - d_2) &\geq? c_r^2 \\ \frac{\mu_1}{\mu_2} \left(\frac{\lambda_1 \mu_2}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 \right) &\geq? c_r^2, \end{aligned}$$

which is only valid when $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)}$.

Check whether $d_{12} \geq c_r^1 - c_e$ holds:

$$\begin{aligned} c_e &\geq? c_r^1 - d_{12} \\ c_e &\geq? c_r^1 \left(1 - \frac{\lambda_1}{\lambda_1 + \mu_1} \right), \end{aligned}$$

which is only valid when $\frac{\mu_1}{\lambda_1 + \mu_1} \leq \frac{c_e}{c_r^1}$.

Case 1.i.b. $d_{12} \leq c_r^1 - c_e$

Derivation of $\bar{d}_{12}$:

$$\begin{aligned}
 \bar{d}_{12} &= \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2}d_1 + \frac{\mu_1}{\lambda_1 + \mu_2}d_2, \text{ as in (4.18)} \\
 &= \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 \right) + \frac{\mu_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1}{\lambda_1 + \mu_2} \right) \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(c_r^1 - c_e + \frac{(\mu_2 - \mu_1)(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 + \frac{\mu_1}{\lambda_1 + \mu_2} c_r^1 \right) \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(\left(1 + \frac{(\mu_2 - \mu_1)(\lambda_1 + \mu_1 + \mu_2) + \mu_1(\lambda_1 + \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \right) c_r^1 - c_e \right) \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(\left(\frac{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2) + \mu_2(\lambda_1 + \mu_1 + \mu_2) - \mu_1\mu_2}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \right) c_r^1 - c_e \right) \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1 + \mu_1 + \mu_2}{\lambda_1 + \mu_1} c_r^1 - c_e \right)
 \end{aligned} \tag{A.2}$$

Check whether $c_r^1 \geq d_1$ holds:

Since d_1 value didn't change, c_r^1 is still greater than equal to d_1 .

Check whether $d_{12} \leq c_r^1 - c_e$ holds:

$$\begin{aligned}
 \frac{\lambda_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1 + \mu_1 + \mu_2}{\lambda_1 + \mu_1} c_r^1 - c_e \right) &\stackrel{?}{\leq} c_r^1 - c_e \\
 \left(1 - \frac{\lambda_1}{\lambda_1 + \mu_2} \right) c_e &\stackrel{?}{\leq} \left(1 - \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \right) c_r^1 \\
 \frac{\mu_2}{\lambda_1 + \mu_2} c_e &\stackrel{?}{\leq} \frac{\mu_1\mu_2}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1,
 \end{aligned}$$

which is only valid when $\frac{c_e}{c_r^1} \leq \frac{\mu_1}{\lambda_1 + \mu_1}$.

Check whether $d_1 \geq d_{12} + c_r^2$ holds:

It is obvious that when we rearrange (A.2), $\bar{d}_{12} = d_1 - \frac{\lambda_1}{\lambda_1 + \mu_2} c_e$.

Then,

$$\begin{aligned}
 d_1 &\geq^? d_{12} + c_r^2 \\
 d_1 &\geq^? d_1 - \frac{\lambda_1}{\lambda_1 + \mu_2} c_e + c_r^2 \\
 0 &\geq^? \frac{c_r^2}{c_r^1} - \left(\frac{\lambda_1}{\lambda_1 + \mu_2} \right) \frac{c_e}{c_r^1}
 \end{aligned} \tag{A.3}$$

When we check $d_{12} \leq c_r^1 - c_e$, we obtain an upper bound on $\frac{c_e}{c_r^1}$, i.e.

$$\begin{aligned}
 \frac{c_e}{c_r^1} &\leq \frac{\mu_1}{\lambda_1 + \mu_1}. \text{ We eliminate } \frac{c_e}{c_r^1} \text{ term in (A.3). Then, (A.3) is valid} \\
 \text{when } \frac{c_r^2}{c_r^1} &\leq \frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)}.
 \end{aligned}$$

Case 1.ii. $\frac{c_r^2}{c_r^1} \geq \frac{\lambda_1}{\lambda_1 + \mu_2}$

For all cases starting with ii., we will insert $\bar{d}_2 = \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2}$ when there exists an d_2 term in the equation.

Derivation of $\bar{d}_1$:

$$\begin{aligned}
 \bar{d}_1 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \mu_1} \text{ as in (4.19).} \\
 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} + \left(\frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \right) \frac{\mu_1}{\lambda_1 + \mu_1} \\
 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} \left(1 + \frac{\mu_1}{\lambda_1 + \lambda_2 + \mu_2} \right) + \frac{\lambda_2 \mu_1 c_r^2}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \\
 &= \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1 + \mu_2) c_r^1 + \lambda_2 \mu_1 c_r^2}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}
 \end{aligned}$$

We will use $\bar{d}_1$ value for all case started with case 1.ii.

Case 1.ii.a. $d_{12} \geq c_r^1 - c_e$

Check whether $d_1 \geq d_{12} + c_r^2$ holds:

We know that, we can re-rewrite d_{12} as $d_1 - \frac{\mu_1}{\mu_2}(d_1 - d_2)$. Then, inequality $d_1 \geq d_{12} + c_r^2$ becomes $\frac{\mu_1}{\mu_2}(d_1 - d_2) \geq c_r^2$, so that we need

to find $d_1 - d_2$.

$$\begin{aligned}
 d_1 - d_2 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \mu_1} - d_2 \\
 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} - d_2 \frac{\lambda_1}{\lambda_1 + \mu_1} \\
 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_1} - \left(\frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \right) \frac{\lambda_1}{\lambda_1 + \mu_1} \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_1} \left(c_r^1 \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu_2} \right) - \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu_2} c_r^2 \right) \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_1} \left(\frac{\lambda_2 + \mu_2}{\lambda_1 + \lambda_2 + \mu_2} c_r^1 - \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu_2} c_r^2 \right)
 \end{aligned}$$

Check whether $\frac{\mu_1}{\mu_2}(d_1 - d_2) \geq c_r^2$ holds:

$$\begin{aligned}
 c_r^2 &\stackrel{?}{\leq} \frac{\lambda_1 \mu_1}{(\lambda_1 + \mu_1) \mu_2} \left(\frac{(\lambda_2 + \mu_2) c_r^1}{\lambda_1 + \lambda_2 + \mu_2} - \frac{\lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \right) \\
 c_r^2 + \frac{\lambda_1 \lambda_2 \mu_1 c_r^2}{\mu_2 (\lambda_1 + \mu_1) (\lambda_1 + \lambda_2 + \mu_2)} &\stackrel{?}{\leq} \frac{\lambda_1 \mu_1 (\lambda_2 + \mu_2) c_r^1}{\mu_2 (\lambda_1 + \mu_1) (\lambda_1 + \lambda_2 + \mu_2)}
 \end{aligned}$$

which is only valid when $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1 \mu_1 (\lambda_2 + \mu_2)}{\lambda_1 \lambda_2 \mu_1 + \mu_2 (\lambda_1 + \mu_1) (\lambda_1 + \lambda_2 + \mu_2)}$. However, this upper bound is always smaller than $\frac{\lambda_1}{\lambda_1 + \mu_2}$, which makes this case invalid.

Case 2. $c_r^1 \geq d_1$ and $d_{12} + c_r^2 \geq d_1$

Case 2.i. $\frac{c_r^2}{c_r^1} \leq \frac{\lambda_1}{\lambda_1 + \mu_2}$

Case 2.i.a. $d_{12} \geq c_r^1 - c_e$

Derivation of $\bar{d}_1$

$$\begin{aligned}
 \bar{d}_1 &= \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \lambda_2 + \mu_1} + d_{12} \frac{\lambda_2}{\lambda_1 + \lambda_2 + \mu_1} \\
 &= \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_1} + d_2 \frac{\mu_1}{\lambda_1 + \lambda_2 + \mu_1} + \frac{\left(\frac{\mu_2 - \mu_1}{\mu_2} d_1 + \frac{\mu_1}{\mu_2} d_2 \right) \lambda_2}{\lambda_1 + \lambda_2 + \mu_1} \\
 \bar{d}_1 &= \frac{\mu_2 (\lambda_1 c_r^1 + \lambda_2 c_r^2) + d_2 \mu_1 (\lambda_2 + \mu_2)}{\mu_2 (\lambda_1 + \lambda_2 + \mu_1) - \lambda_2 (\mu_2 - \mu_1)} \\
 &= \frac{\mu_2 (\lambda_1 c_r^1 + \lambda_2 c_r^2) + \left(\frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2} \right) \mu_1 (\lambda_2 + \mu_2)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \\
 &= \frac{\lambda_1 c_r^1 (\mu_2 + \frac{\mu_1 (\lambda_2 + \mu_2)}{\lambda_1 + \mu_2}) + \lambda_2 \mu_2 c_r^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(1 + \frac{\mu_2^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) c_r^1 + \frac{\lambda_2 \mu_2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} c_r^2
 \end{aligned}$$

Derivation of $\bar{d}_{12}$

$$\begin{aligned}
 \bar{d}_{12} &= \frac{\mu_2 - \mu_1}{\mu_2} d_1 + \frac{\mu_1}{\mu_2} d_2, \text{ as in (4.17)} \\
 &= \frac{\lambda_1 c_r^1}{\mu_2 (\lambda_1 + \mu_2)} \left(\mu_1 + (\mu_2 - \mu_1) \left(1 + \frac{\mu_2^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) \right) \\
 &\quad + \frac{\lambda_2 (\mu_2 - \mu_1) c_r^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \\
 &= \frac{\lambda_1 c_r^1}{\mu_2 (\lambda_1 + \mu_2)} \left(\mu_2 + \frac{\mu_2^2 (\mu_2 - \mu_1)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) + \frac{\lambda_2 (\mu_2 - \mu_1) c_r^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \\
 &= \frac{\lambda_1}{\lambda_1 + \mu_2} \left(1 + \frac{\mu_2 (\mu_2 - \mu_1)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) c_r^1 + \frac{\lambda_2 (\mu_2 - \mu_1)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} c_r^2
 \end{aligned}$$

Check whether $c_r^1 \geq d_1$ holds:

$$\begin{aligned}
 c_r^1 &\stackrel{?}{\geq} d_1 \\
 c_r^1 &\stackrel{?}{\geq} \frac{\lambda_1}{\lambda_1 + \mu_2} \left(1 + \frac{\mu_2^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) c_r^1 \\
 &\quad + \frac{\lambda_2 \mu_2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} c_r^2 \\
 \frac{c_r^1}{\lambda_1 + \mu_2} \left(1 - \frac{\lambda_1 \mu_2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) &\stackrel{?}{\geq} \frac{\lambda_2 c_r^2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)},
 \end{aligned}$$

which is only valid when $\frac{c_r^2}{c_r^1} \leq \frac{\mu_1 (\lambda_2 + \mu_2)}{\lambda_2 (\lambda_1 + \mu_2)}$. Moreover, there exists another upper-bound on $\frac{c_r^2}{c_r^1}$ for the cases starting with list index

(i). Hence, in order to find an upper bound on $\frac{c_r^2}{c_r^1}$, we should take the minimum of those two upper bounds.

I. If $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$, then the upper bound is $\frac{\lambda_1}{\lambda_1 + \mu_2}$.

II. If $\mu_1(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$, then the upper bound is $\frac{\mu_1(\lambda_2 + \mu_2)}{\lambda_2(\lambda_1 + \mu_2)}$.

Check whether $d_1 \leq d_{12} + c_r^2$ holds:

We know that, we can re-rewrite d_{12} as $d_1 - \frac{\mu_1}{\mu_2}(d_1 - d_2)$. Then, inequality $d_1 \leq d_{12} + c_r^2$ becomes $\frac{\mu_1}{\mu_2}(d_1 - d_2) \leq c_r^2$.

$$\frac{\lambda_1\mu_1c_r^1}{\mu_2(\lambda_1 + \mu_2)} \left(1 + \frac{\mu_2^2}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}\right) + \frac{\lambda_2\mu_1c_r^2}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)} - \frac{\lambda_1\mu_1c_r^1}{\mu_2(\lambda_1 + \mu_2)} \stackrel{?}{\leq} c_r^2$$

$$\left(1 - \frac{\lambda_2\mu_1}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}\right)c_r^2 \stackrel{?}{\leq} \frac{\lambda_1\mu_1\mu_2}{(\lambda_1 + \mu_2)(\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2))}c_r^1$$

which is only valid when $\frac{\lambda_1\mu_1}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \leq \frac{c_r^2}{c_r^1}$.

Note that, the upper bounds we determined (I) and (II) are always greater than the lower bound.

Check whether $d_{12} \geq c_r^1 - c_e$ holds:

$$\frac{\lambda_2(\mu_2 - \mu_1)}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}c_r^2 + c_e \stackrel{?}{\geq} \left(1 - \frac{\lambda_1}{\lambda_1 + \mu_2} \left(1 + \frac{\mu_2(\mu_2 - \mu_1)}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}\right)\right)c_r^1$$

$$\frac{\lambda_2(\mu_2 - \mu_1)}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}c_r^2 + c_e \stackrel{?}{\geq} \frac{\mu_2}{\lambda_1 + \mu_2} \left(1 - \frac{\lambda_1(\mu_2 - \mu_1)}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}\right)c_r^1$$

$$\frac{\lambda_2(\mu_2 - \mu_1)}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}c_r^2 + c_e \stackrel{?}{\geq} \frac{\mu_2}{\lambda_1 + \mu_2} \left(\frac{\mu_1(\lambda_1 + \lambda_2 + \mu_2)}{\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)}\right)c_r^1$$

I. If $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$, then the lower bound on $\frac{c_e}{c_r^1}$ becomes:

$$\frac{\mu_1\mu_2(\lambda_1 + \lambda_2 + \mu_2) + \lambda_1\lambda_2\mu_1 - \lambda_1\lambda_2\mu_2}{(\lambda_1 + \mu_2)(\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2))},$$

which is always positive.

II. If $\mu_1(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$, then the lower bound on $\frac{c_e}{c_r^1}$ becomes:

$$\frac{\mu_1}{\lambda_1 + \mu_2}$$

Case 2.i.b. $d_{12} \leq c_r^1 - c_e$

Derivation of $\bar{d}_1$:

$$\begin{aligned}
 \bar{d}_1(\lambda_1 + \lambda_2 + \mu_1) &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 + d_{12} \lambda_2 \\
 &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 \\
 &\quad + \left(\frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2} d_1 + \frac{\mu_1}{\lambda_1 + \mu_2} d_2 \right) \lambda_2 \\
 \bar{d}_1 \left(\lambda_1 + \lambda_2 + \mu_1 - \frac{\lambda_2(\mu_2 - \mu_1)}{\lambda_1 + \mu_2} \right) &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 \left(1 + \frac{\lambda_2}{\lambda_1 + \mu_2} \right) \\
 &\quad + \frac{\lambda_1 \lambda_2 (c_r^1 - c_e)}{\lambda_1 + \mu_2} \\
 \bar{d}_1(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2) &= (\lambda_1 c_r^1 + \lambda_2 c_r^2)(\lambda_1 + \mu_2) \\
 &\quad + \frac{\lambda_1 \mu_1 (\lambda_1 + \lambda_2 + \mu_2)}{\lambda_1 + \mu_2} c_r^1 + \lambda_1 \lambda_2 (c_r^1 - c_e) \\
 \bar{d}_1 &= \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 + \lambda_2 \frac{(\lambda_1 + \mu_2) c_r^2 - \lambda_1 c_e}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}
 \end{aligned}$$

Derivation of $\bar{d}_{12}$:

$$\begin{aligned}
 \bar{d}_{12} &= \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2} d_1 + \frac{\mu_1}{\lambda_1 + \mu_2} d_2 \\
 &= \frac{\lambda_1 c_r^1}{\lambda_1 + \mu_2} \left(1 + \frac{(\mu_2 - \mu_1)(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} + \frac{\mu_1}{\lambda_1 + \mu_2} \right) \\
 &\quad + \frac{\lambda_2(\mu_2 - \mu_1) c_r^2}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} - \frac{\lambda_1 c_e}{\lambda_1 + \mu_2} \left(1 + \frac{\lambda_2(\mu_2 - \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \right) \\
 &= \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} c_r^1 + \frac{\lambda_2(\mu_2 - \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} c_r^2 \\
 &\quad - \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} c_e
 \end{aligned}$$

Check whether $d_1 \leq c_r^1$ holds:

We will insert d_1 value and bracket with the cost parameters.

$$\begin{aligned}
 \frac{\lambda_2(\lambda_1 + \mu_2) c_r^2 - \lambda_1 \lambda_2 c_e}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} &\stackrel{?}{\leq} \left(1 - \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)} \right) \\
 (\lambda_1 + \mu_2) c_r^2 - \lambda_1 c_e &\stackrel{?}{\leq} \frac{\mu_1 \mu_2 (\lambda_1 + \lambda_2 + \mu_2)}{\lambda_2 (\lambda_1 + \mu_2)} c_r^1
 \end{aligned} \tag{A.4}$$

Since we did not obtain an inequality which only depends on two cost parameters, we will continue to check other conditions.

Check whether $d_1 \leq d_{12} + c_r^2$ holds:

We will insert d_1 and d_{12} values and bracket with the cost parameters.

$$\begin{aligned} 0 &\stackrel{?}{\leq} \left(1 - \frac{\lambda_2((\lambda_1 + \mu_2) - (\mu_2 - \mu_1))}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}\right) c_r^2 - \frac{\lambda_1}{\lambda_1 + \lambda_2 + \mu_2} c_e \\ 0 &\stackrel{?}{\leq} (\lambda_1 + \mu_2) c_r^2 - \lambda_1 c_e \end{aligned} \quad (\text{A.5})$$

When we merge (A.4) and (A.5) and divide each side with c_r^1 , we obtain:

$$0 \stackrel{?}{\leq} (\lambda_1 + \mu_2) \frac{c_r^2}{c_r^1} - \lambda_1 \frac{c_e}{c_r^1} \stackrel{?}{\leq} \frac{\mu_1 \mu_2 (\lambda_1 + \lambda_2 + \mu_2)}{\lambda_2 (\lambda_1 + \mu_2)} \quad (\text{A.6})$$

Check whether $d_{12} \leq c_r^1 - c_e$ holds:

We will insert d_{12} value and bracket with the cost parameters.

$$\begin{aligned} &\frac{\lambda_2(\mu_2 - \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} c_r^2 + \left(1 - \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}\right) c_e \\ &\stackrel{?}{\leq} \left(1 - \frac{\lambda_1(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \mu_2)}\right) c_r^1 \end{aligned}$$

After cancellations, we divide each side with c_r^1 and obtain the following:

$$\left(\frac{\lambda_2(\mu_2 - \mu_1)}{\lambda_1 + \lambda_2 + \mu_2}\right) \frac{c_r^2}{c_r^1} + \left(\frac{\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)}{\lambda_1 + \lambda_2 + \mu_2}\right) \frac{c_e}{c_r^1} \stackrel{?}{\leq} \frac{\mu_1 \mu_2}{\lambda_1 + \mu_2} \quad (\text{A.7})$$

We let α and β represent the coefficients of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ for (A.6), respectively. A and B represent lower and upper bounds of (A.6), i.e. $A \leq \alpha \frac{c_r^2}{c_r^1} - \beta \frac{c_e}{c_r^1} \leq B$. Moreover, α' and β' represent the coefficients of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ for (A.7), respectively. C represent upper bound of (A.6), i.e. $\alpha' \frac{c_r^2}{c_r^1} + \beta' \frac{c_e}{c_r^1} \leq C$. Then, the upper bound on

$\frac{c_r^2}{c_r^1}$ becomes $\frac{B\beta' + C\beta}{\alpha\beta' + \alpha'\beta}$, where

$$\begin{aligned} B\beta' + C\beta &= \frac{\mu_1\mu_2(\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2))}{\lambda_2(\lambda_1 + \mu_2)} + \frac{\lambda_1\mu_1\mu_2}{\lambda_1 + \mu_2} \\ &= \frac{\mu_1\mu_2(\lambda_1 + \mu_1)(\lambda_2 + \mu_2)}{\lambda_2(\lambda_1 + \mu_2)}, \text{ and} \\ \alpha\beta' + \alpha'\beta &= \frac{(\lambda_1 + \mu_2)(\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)) + \lambda_1\lambda_2(\mu_2 - \mu_1)}{\lambda_1 + \lambda_2 + \mu_2} \\ &= \mu_2(\lambda_1 + \mu_1) \end{aligned}$$

Furthermore, the upper bound on $\frac{c_e}{c_r^1}$ becomes $\frac{C\alpha}{\alpha'\beta + \alpha\beta'}$, where

$$\begin{aligned} C\alpha &= \frac{\mu_1\mu_2(\lambda_1 + \mu_2)}{\lambda_1 + \mu_2} = \mu_1\mu_2, \text{ and} \\ \alpha'\beta + \alpha\beta' &= \frac{\lambda_1\lambda_2(\mu_2 - \mu_1) + (\lambda_1 + \mu_2)(\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2))}{\lambda_1 + \lambda_2 + \mu_2} \\ &= \mu_2(\lambda_1 + \mu_1) \end{aligned}$$

Note that there exists another upper-bound on $\frac{c_r^2}{c_r^1}$, which comes from the definition of Case 2.i. Hence, in order to find an upper bound on $\frac{c_r^2}{c_r^1}$, we should take the minimum of those two upper bounds.

I. If $\lambda_1\lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$, then the upper bound is $\frac{\lambda_1}{\lambda_1 + \mu_2}$.

II. If $\mu_1(\lambda_2 + \mu_2) \leq \lambda_1\lambda_2$, then the upper bound is $\frac{c_r^2}{c_r^1} \leq \frac{\mu_1(\lambda_2 + \mu_2)}{\lambda_2(\lambda_1 + \mu_2)}$.

Case 2.ii. $\frac{c_r^2}{c_r^1} \geq \frac{\lambda_1}{\lambda_1 + \mu_2}$

For this range of parameters $\bar{d}_2 = \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2}$.

Case 2.ii.a. $d_{12} \geq c_r^1 - c_e$

Derivation of $\bar{d}_1$:

$$\begin{aligned}
 \bar{d}_1(\lambda_1 + \lambda_2 + \mu_1) &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 + d_{12} \lambda_2 \\
 &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 + \lambda_2 \left(\frac{\mu_2 - \mu_1}{\mu_2} d_1 + \frac{\mu_1}{\mu_2} d_2 \right) \\
 \bar{d}_1(\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)) &= (\lambda_1 c_r^1 + \lambda_2 c_r^2) \mu_2 + d_2 \mu_1 (\lambda_2 + \mu_2) \\
 &= (\lambda_1 c_r^1 + \lambda_2 c_r^2) \mu_2 + \mu_1 (\lambda_2 + \mu_2) \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \\
 \bar{d}_1 &= (\lambda_1 c_r^1 + \lambda_2 c_r^2) \frac{\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2)}{(\lambda_1 + \lambda_2 + \mu_2)(\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2))}
 \end{aligned}$$

Derivation of $\bar{d}_{12}$:

$$\begin{aligned}
 \bar{d}_{12} &= \frac{\mu_2 - \mu_1}{\mu_2} d_1 + \frac{\mu_1}{\mu_2} d_2 \\
 &= \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\mu_2 (\lambda_1 + \lambda_2 + \mu_2)} \left((\mu_2 - \mu_1) \frac{\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} + \mu_1 \right) \\
 &= \frac{\mu_2 (\lambda_1 c_r^1 + \lambda_2 c_r^2)}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)}
 \end{aligned}$$

Check whether $d_1 \leq c_r^1$ holds:

We will insert $\bar{d}_1$ value and bracket with the cost parameters.

$$\begin{aligned}
 \lambda_2 (\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2)) c_r^2 &\stackrel{?}{\leq} ((\lambda_1 + \lambda_2 + \mu_2) (\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)) \\
 &\quad - \lambda_1 (\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2))) c_r^1,
 \end{aligned}$$

which is only valid when $\frac{c_r^2}{c_r^1} \leq \frac{\mu_1 (\lambda_2 + \mu_2)^2}{\lambda_2 (\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2))}$.

Check whether $d_1 \leq d_{12} + c_r^2$ holds:

We will insert $\bar{d}_1$ and $\bar{d}_{12}$ values and bracket with the cost parameters.

$$\begin{aligned}
 &\lambda_1 \left(\frac{\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2)}{(\lambda_1 + \lambda_2 + \mu_2) (\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2))} - \frac{\mu_2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} \right) c_r^1 \\
 &\stackrel{?}{\leq} \left(1 + \frac{\lambda_2 \mu_2}{\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)} - \lambda_2 \frac{\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2)}{(\lambda_1 + \lambda_2 + \mu_2) (\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2))} \right) c_r^2,
 \end{aligned} \tag{A.8}$$

which is valid when $\frac{\lambda_1 \mu_1 (\lambda_2 + \mu_2)}{\lambda_1 \lambda_2 \mu_1 + \mu_2 (\lambda_1 + \mu_2) (\lambda_1 + \lambda_2 + \mu_2)} \leq \frac{c_r^2}{c_r^1}$. It is obvious that, this lower bound is always smaller than $\frac{\lambda_1}{\lambda_1 + \mu_2}$. Hence,

$$\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1}.$$

Check whether $c_r^1 - c_e \leq d_{12}$ holds or not:

We will insert $\bar{d}_{12}$ value and bracket with the cost parameters.

$$\begin{aligned} \left(1 - \frac{\lambda_1 \mu_2}{\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)}\right) c_r^1 &\leq? \frac{\lambda_2 \mu_2}{\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)} c_r^2 + c_e \\ \frac{\mu_1(\lambda_2 + \mu_2)}{\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)} c_r^1 &\leq? \frac{\lambda_2 \mu_2}{\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)} c_r^2 + c_e, \end{aligned}$$

which is valid when $\frac{\mu_1(\lambda_2 + \mu_2)}{\mu_1(\lambda_2 + \mu_2) + \mu_2(\lambda_1 + \lambda_2 + \mu_2)} \leq \frac{c_e}{c_r^1}$. As a remark, the upper bound we found for $\frac{c_r^2}{c_r^1}$ must be greater than the lower bound on $\frac{c_r^2}{c_r^1}$, which comes from the definition of Case 2.ii. Hence, this case is only valid when $\lambda_1 \lambda_2 \leq \mu_1(\lambda_2 + \mu_2)$.

Case 2.ii.b. $d_{12} \leq c_r^1 - c_e$

Derivation of $\bar{d}_1$:

$$\begin{aligned} \bar{d}_1(\lambda_1 + \lambda_2 + \mu_1) &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 + d_{12} \lambda_2 \\ &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 \\ &\quad + \left(\frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2} d_1 + \frac{\mu_1}{\lambda_1 + \mu_2} d_2 \right) \lambda_2 \\ \bar{d}_1 \left(\lambda_1 + \lambda_2 + \mu_1 - \frac{\lambda_2(\mu_2 - \mu_1)}{\lambda_1 + \mu_2} \right) &= \lambda_1 c_r^1 + \lambda_2 c_r^2 + d_2 \mu_1 \left(1 + \frac{\lambda_2}{\lambda_1 + \mu_2} \right) \\ &\quad + \frac{\lambda_1 \lambda_2 (c_r^1 - c_e)}{\lambda_1 + \mu_2} \\ \bar{d}_1(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2) &= (\lambda_1 c_r^1 + \lambda_2 c_r^2)(\lambda_1 + \mu_2) \\ &\quad + \mu_1(\lambda_1 c_r^1 + \lambda_2 c_r^2) + \lambda_1 \lambda_2 (c_r^1 - c_e) \\ \bar{d}_1 &= \frac{(\lambda_1 + \mu_1 + \mu_2)(\lambda_1 c_r^1 + \lambda_2 c_r^2) + \lambda_1 \lambda_2 (c_r^1 - c_e)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \end{aligned}$$

Derivation of $\bar{d}_{12}$:

$$\begin{aligned}
 \bar{d}_{12} &= \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2}d_1 + \frac{\mu_1}{\lambda_1 + \mu_2}d_2 \\
 &= \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} + \frac{\mu_2 - \mu_1}{\lambda_1 + \mu_2} \left(\frac{(\lambda_1 + \mu_1 + \mu_2)(\lambda_1 c_r^1 + \lambda_2 c_r^2) + \lambda_1 \lambda_2 (c_r^1 - c_e)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \right) \\
 &\quad + \frac{\mu_1}{\lambda_1 + \mu_2} \left(\frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{\lambda_1 + \lambda_2 + \mu_2} \right) \\
 &= \frac{\lambda_1 c_r^1 + \lambda_2 c_r^2}{(\lambda_1 + \mu_2)(\lambda_1 + \lambda_2 + \mu_2)} \left(\mu_1 + \frac{(\mu_2 - \mu_1)(\lambda_1 + \mu_1 + \mu_2)}{\lambda_1 + \mu_1} \right) \\
 &\quad + \frac{\lambda_1(c_r^1 - c_e)}{\lambda_1 + \mu_2} \left(1 + \frac{\lambda_2(\mu_2 - \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \right) \\
 &= \frac{\mu_2(\lambda_1 c_r^1 + \lambda_2 c_r^2) + \lambda_1(\lambda_1 + \lambda_2 + \mu_1)(c_r^1 - c_e)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}
 \end{aligned}$$

Check whether $d_1 \leq c_r^1$ holds:

We will insert $\bar{d}_1$ value and bracket with the cost parameters.

$$\begin{aligned}
 &\frac{\lambda_2(\lambda_1 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}c_r^2 - \frac{\lambda_1 \lambda_2}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}c_e \\
 &\leq^? \left(1 - \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} \right)c_r^1
 \end{aligned}$$

After cancellations, we divide each side with c_r^1 and obtain the following:

$$\lambda_2(\lambda_1 + \mu_1 + \mu_2) \frac{c_r^2}{c_r^1} - \lambda_1 \lambda_2 c_r^2 \frac{c_e}{c_r^1} \leq \mu_1(\lambda_2 + \mu_2) \quad (\text{A.9})$$

Check whether $d_1 \leq d_{12} + c_r^2$ holds:

Realize that $\bar{d}_1 - \bar{d}_{12} = \frac{\lambda_2 c_r^2 + \lambda_1 c_e}{\lambda_1 + \lambda_2 + \mu_2}$. We insert this value into $d_1 - d_{12} \leq c_r^2$, we divide each side with c_r^1 and obtain the following:

$$(\lambda_1 + \mu_2) \frac{c_r^2}{c_r^1} - \lambda_1 c_e \geq 0. \quad (\text{A.10})$$

Check whether $d_{12} \leq c_r^1 - c_e$ holds:

We will insert $\bar{d}_{12}$ value and bracket with the cost parameters.

$$\begin{aligned} & \frac{\lambda_2 \mu_2}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)} c_r^2 + \left(1 - \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}\right) c_e \\ & \leq? \left(1 - \frac{\lambda_1(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)}{(\lambda_1 + \mu_1)(\lambda_1 + \lambda_2 + \mu_2)}\right) c_r^1 \end{aligned}$$

After cancellations, we divide each side with c_r^1 and obtain the following:

$$\lambda_2 \mu_2 \frac{c_r^2}{c_r^1} + (\lambda_1 \mu_2 + \mu_1(\lambda_2 + \mu_2)) \frac{c_r^2}{c_r^1} \leq \mu_1(\lambda_2 + \mu_2) \quad (\text{A.11})$$

We let α and β represent the coefficients of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ for (A.9), respectively. A represent the upper bound of (A.9). Moreover, α' and β' represent the coefficients of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ for (A.5), respectively. Furthermore, α'' and β'' represent the coefficients of $\frac{c_r^2}{c_r^1}$ and $\frac{c_e}{c_r^1}$ for (A.6), respectively. B represent the upper bound of (A.11).

- I. $\alpha \frac{c_r^2}{c_r^1} - \beta \frac{c_e}{c_r^1} \leq A$, (A.9).
- II. $0 \leq \alpha' \frac{c_r^2}{c_r^1} - \beta' \frac{c_e}{c_r^1}$, (A.10).
- III. $\alpha'' \frac{c_r^2}{c_r^1} + \beta'' \frac{c_e}{c_r^1} \leq B$, (A.11).

Combine I. and III. and obtain an upper bound on $\frac{c_r^2}{c_r^1}$, which is $\frac{A\beta'' + B\beta}{\alpha\beta'' + \alpha''\beta}$, where

$$\begin{aligned} A\beta'' + B\beta &= \mu_1 \mu_2 (\lambda_2 + \mu_2) (\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)) \\ &= \mu_1 (\lambda_1 + \mu_1) (\lambda_2 + \mu_2)^2, \text{ and} \\ \alpha\beta'' + \alpha''\beta &= \lambda_2 (\lambda_1 + \mu_1 + \mu_2) (\lambda_1 \mu_2 + \mu_1 (\lambda_2 + \mu_2)) + \lambda_1 \lambda_2^2 \mu_2 \\ &= \lambda_2 (\lambda_1 + \mu_1) (\mu_2 (\lambda_1 + \lambda_2 + \mu_2) + \mu_1 (\lambda_2 + \mu_2)). \end{aligned}$$

Then, $\frac{\lambda_1}{\lambda_1 + \mu_2} \leq \frac{c_r^2}{c_r^1} \leq \frac{\mu_1 (\lambda_2 + \mu_2)^2}{\lambda_2 (\mu_1 (\lambda_2 + \mu_2) + \mu_2 (\lambda_1 + \lambda_2 + \mu_2))}$, which is valid when $0 \leq \mu_2 (\lambda_1 + \lambda_2 + \mu_2) (\lambda_1 \lambda_2 - \mu_1 (\lambda_2 + \mu_2)) (\mu_1 - \mu_2)$. We know that $\mu_1 - \mu_2 < 0$. Hence, the condition is valid when $\lambda_1 \lambda_2 \leq \mu_1 (\lambda_2 + \mu_2)$.

Combine I. and II. and obtain an upper bound on $\frac{c_e}{c_r^1}$, which is $\frac{A\alpha'}{\alpha\beta' - \alpha'\beta}$, where

$$A\alpha' = \mu_1(\lambda_1 + \mu_2)(\lambda_2 + \mu_2), \text{ and}$$

$$\alpha\beta' - \alpha'\beta = \lambda_1\lambda_2(\lambda_1 + \mu_1 + \mu_2) - \lambda_1\lambda_2(\lambda_1 + \mu_2) = \lambda_1\lambda_2\mu_1.$$

This upper bound is always greater than 1.

Combine II. and III. and obtain an upper bound on $\frac{c_e}{c_r^1}$, which is $\frac{B\alpha'}{\alpha'\beta'' + \alpha''\beta'}$, where

$$B\alpha' = \mu_1(\lambda_1 + \mu_2)(\lambda_2 + \mu_2), \text{ and}$$

$$\alpha'\beta'' + \alpha''\beta' = (\lambda_1 + \mu_2)(\lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2)) + \lambda_1\lambda_2\mu_2$$

$$= \lambda_1\mu_2(\lambda_1 + \lambda_2 + \mu_2) + \mu_1(\lambda_1 + \mu_2)(\lambda_2 + \mu_2).$$

Combine I. and the lower bound on $\frac{c_r^2}{c_r^1}$ (L) and obtain an lower bound on $\frac{c_e}{c_r^1}$, which is $\frac{\alpha L - A}{\beta}$. We know that $\frac{c_r^2}{c_r^1} \leq \frac{c_e}{c_r^1}$. Then, $\max\{\frac{\alpha L - A}{\beta}, L\}$ is the lower bound on $\frac{c_e}{c_r^1}$. L is always bigger than $\frac{\alpha L - A}{\beta}$ because $\beta \leq \alpha$. Then, $L \leq \frac{c_e}{c_r^1}$.

Combine II. and the upper bound on $\frac{c_r^2}{c_r^1}$ (U) and obtain an upper bound on $\frac{c_e}{c_r^1}$, which is $\frac{\alpha' U}{\beta'}$, where

$$\alpha' U = (\lambda_1 + \mu_2) \frac{\mu_1(\lambda_2 + \mu_2)^2}{\lambda_2(\mu_2(\lambda_1 + \lambda_2 + \mu_2) + \mu_1(\lambda_2 + \mu_2))}, \text{ and}$$

$$\beta' = \lambda_1.$$

This upper bound is always bigger than the previous upper bound. Hence, it is not a strict upper bound.

Combine III. and the upper bound on $\frac{c_r^2}{c_r^1}$ (L) and obtain an upper

bound on $\frac{c_e}{c_r^1}$, which is $\frac{B-\alpha''L}{\beta''}$, where

$$\begin{aligned} B - \alpha''L &= \mu_1(\lambda_2 + \mu_2) - \lambda_2\mu_2 \frac{\lambda_1}{\lambda_1 + \mu_2} \\ &= \frac{\mu_1(\lambda_1 + \mu_2)(\lambda_2 + \mu_2) - \lambda_1\lambda_2\mu_2}{\lambda_1 + \mu_2}, \text{ and} \end{aligned}$$

$$\beta'' = \lambda_1\mu_2 + \mu_1(\lambda_2 + \mu_2).$$

This upper bound is the smallest of the upper bounds we found.

Appendix B

HIZMET HOSPITAL

Floor	Departments	Size	Restrictions
6	Transplantation & Urology	9	*Infected patients cannot be accepted although there are empty beds. *Donors also stay here. *Circumcision, kidney operations are accepted.
5	Cardiology & Cardiovascular Surgeries (KVC)	12	*Patients transferred from coroner ICU, medical patients of cardiology and KVCs surgical patients stay here. *Infected patients are not accepted. *Circumcision, cesarean etc. can be accepted.
4	Internal branches (Internal medicine, nephrology, thoracic medicine, gastroenterology etc.)	13	*No surgical stays. *Patients who stay here are usually too old, are dialysis patients from nephrology and generally suffer from multiple illnesses *These patients cannot stay in another department.
3	Reserved floor	12	*This floor is normally closed but used when there is a capacity problem. *If an infected patient is placed, the others cannot or used according to the predetermined rules by the infection committee.
2	Gynecology & Obstetrics	9	*Patients of this department usually placed in 5 th and 6 th floor when capacity problem occurs. *No other types of patients are accepted on this floor in any case because nurses are specialized.
1	Surgical branches	13	*If the floor is full, patients can be placed on 5 th and 6 th floor (depending on the risk of infection) floors.

Table B.1: Hizmet Hospital floor plans, their capacities and allowed patients in these floors

Appendix C

OCCUPANCY RATES OF HIZMET HOSPITAL

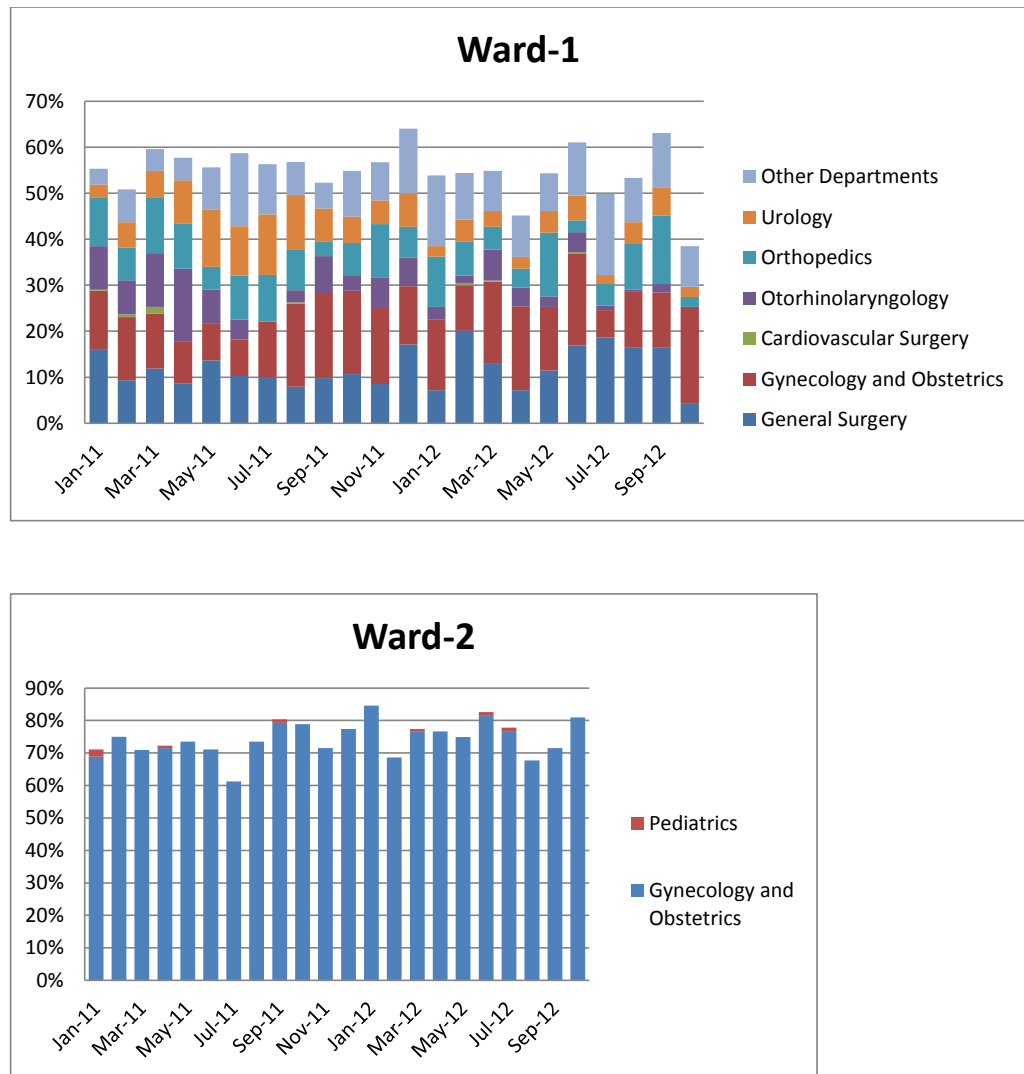


Figure C.1: Monthly average occupancy rates of surgical branches (Ward-1) and gynecology and obstetrics (Ward-2)

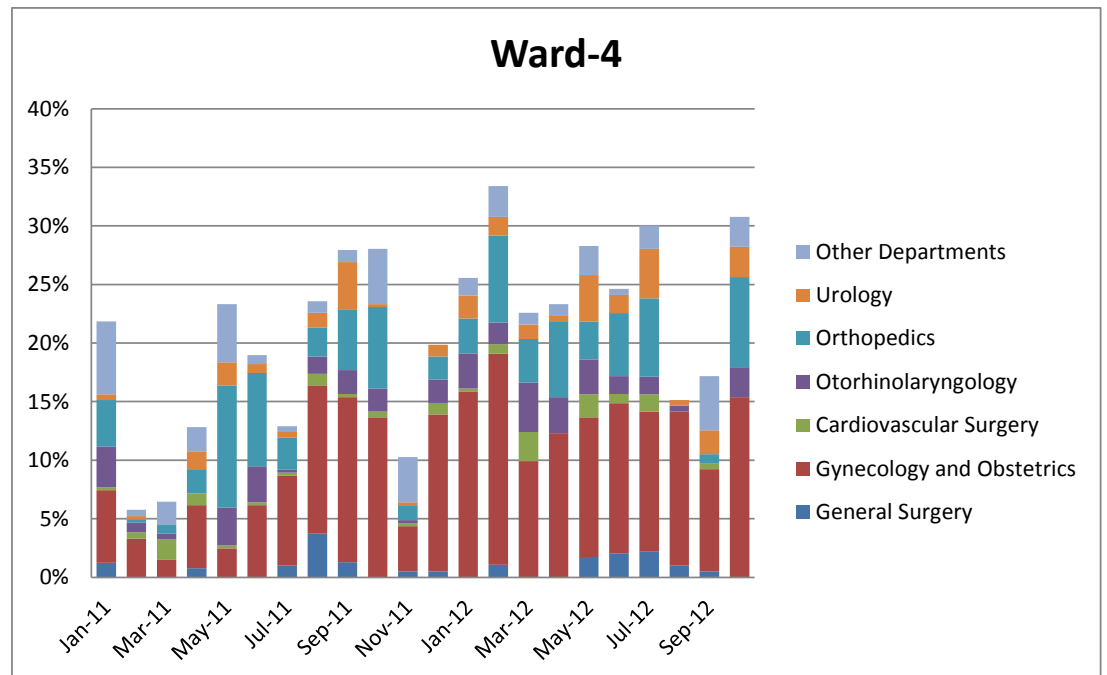
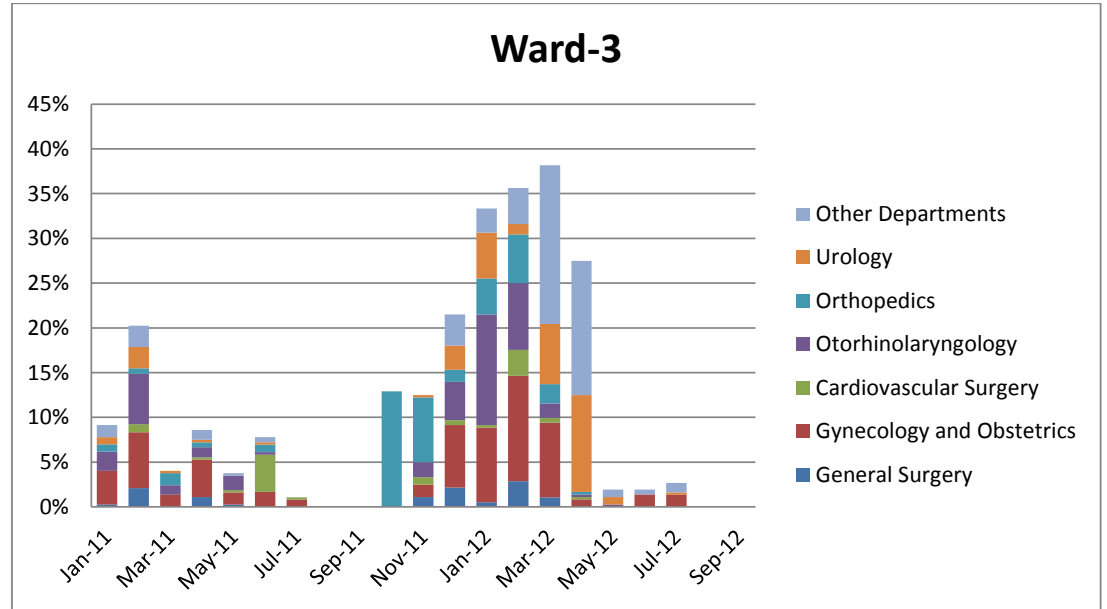


Figure C.2: Monthly average occupancy rates of the empty ward (Ward-3) and internal branches (Ward-4)

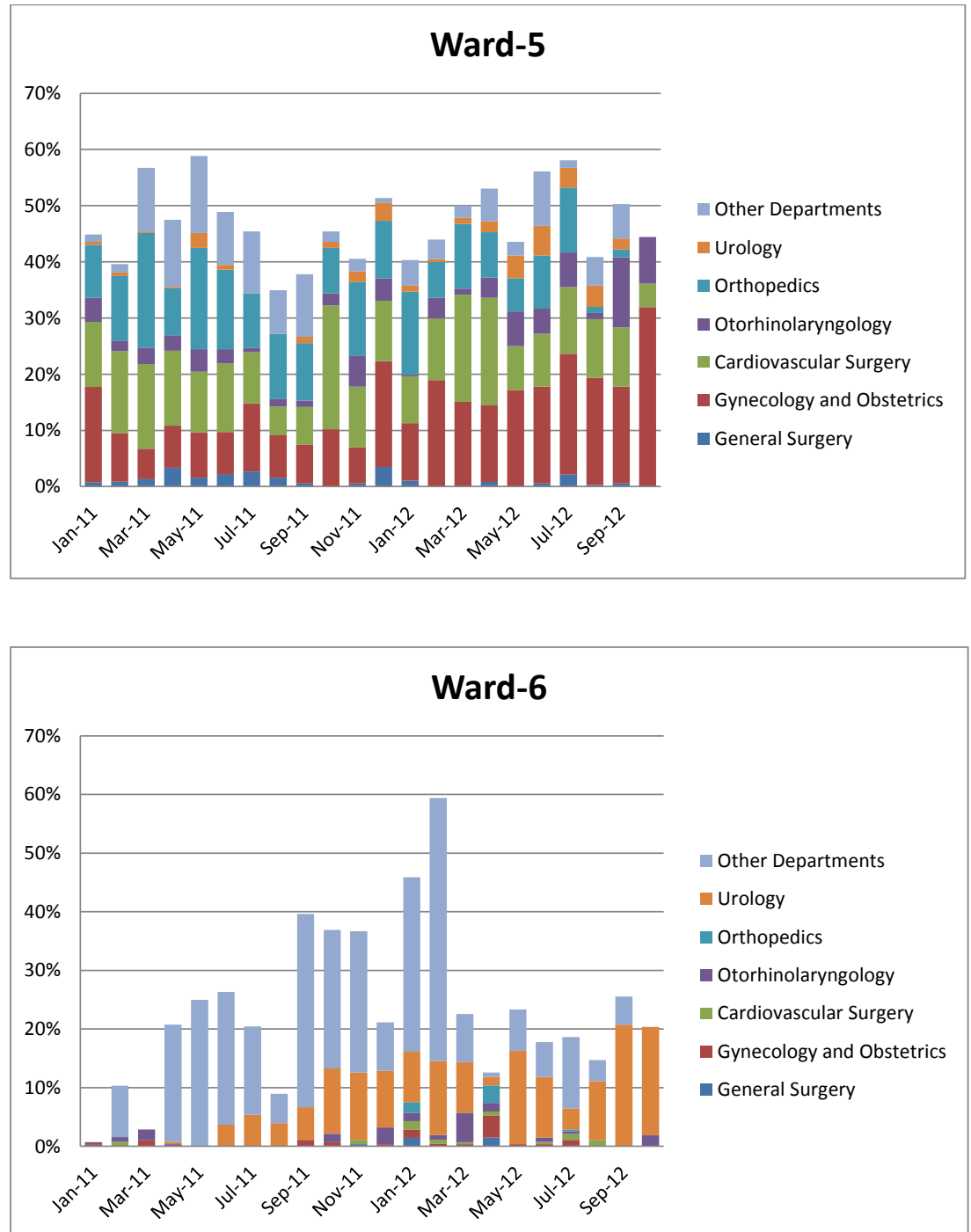


Figure C.3: Monthly average occupancy rates of cardiology and KVC (Ward-5) and urology and transplantation (Ward-6)

VITA

Nermin Elif Kurt was born in Istanbul, Turkey on March 7,1990. She received her B.Sc. degree in Electrical and Electronics Engineering from Bilkent University, Ankara in 2012. From September 2012 to August 2014, she worked as a teaching and research assistant in Koc University, Turkey and had studied for the “*Optimal patient admission policy under stochastic length of stay*” project which was sponsored by TÜBİTAK, since April 2013. She presented a poster in Operational Research Applied to Health Services Conference (2013). She also attended 2014 INFORMS Revenue Management and Pricing Section Conference.