Admission and Scheduling Control for Preventive Services

by

Derya Kunduzcu

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, 2009

Koç University

Graduate School of Sciences and Engineering

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

Derya Kunduzcu

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. Evrim Didem Güneş

_____

Assoc. Prof. O. Zeynep Akşin

_____

Assoc. Prof. Fikri Karaesmen

_____

Prof. Refik Güllü

Date: _____

"To my dearest family, Hatice, Mehmet, and Deniz, who believed and supported me in every step I have taken in my life"

ABSTRACT

In this study, we focus on a one server queue with exponential service times. We assume that the server provides colonoscopy service. We consider the capacity rationing of this facility within patients coming from different risk groups, the highest risk corresponding to colorectal-cancer patients. We assume that the facility providing the screening/diagnostic procedure operates in a dynamic random environment, which determines the demand for diagnostic services. The random environment represents the health of the population, where if the health of the population is better, the demand rate of symptomatic patients is lower. The system can exercise admission or scheduling or both controls. Scheduling lower risk patients is the screening process, improving the health of the whole population in the long run. We consider three different models, depending on which of the controls is used. The objective is to minimize total expected discounted costs over an infinite time horizon with a discount rate β as well as the long-run average costs. We establish the existence of optimal monotone policies under certain conditions. We study the effects of system parameters on the optimal policy and performance measures through a numerical study.

In the second part of the thesis, we use compartmental model to investigate the effects of operational controls on population dynamics. We use a four-stage health model (1) healthy, (2) polyps, (3) preclinical, or early stage, colorectal cancer (4) clinical, or late stage, colorectal cancer. We create four scenarios where the difference rises due to the allocation of the resources. We present a numerical analysis that compares the performances of the different scheduling strategies and explore the effect of system parameters in detail.

ÖZETÇE

Bu çalışmada, hizmet süreleri üssel olan tek işgörenli bir kuyruk modelini inceliyoruz. İşgörenin kolonoskopi hizmeti verdiğini kabul ediyoruz. Kolonoskopi aletinin farklı risk gruplarından gelen hastalar arasında nasıl paylaşılması gerektiği konusunu ele alıyoruz. Risk gruplarında, en yüksek risk grubu kolon kanseri olan hastaları temsil ediyor ki biz bunlara belirti gösteren hastalar anlamında belirtili hastalar diyeceğiz. Hem tarama hem de teşhis ve tedavi için kullanılan bu aletin dinamik ve rassal bir ortamda çalıştığını, ve ortamın durum değerlerinin teşhis ve tedavi isteğiyle gelen hasta hızını belirlediğini kabul ediyoruz. Rassal ortam halkın sağlık seviyesini temsil ediyor, öyle ki halk sağlığı iyi olduğunda belirtili hasta sayısı düşüyor. Sistemin uygulayabileceği iki çeşit kontrol var: belirtisiz hastaları kabul etme ve hastalar arasındaki hizmet önceliklerine karar verme. Hizmette belirtisiz hastalara öncelik vermek, tarama yaparak halk sağlığının uzun vadede iyileşmesini sağlar. Biz farklı kontroller kullanan üç farklı modeli düşüneceğiz: yalnızca kabul etme, yalnızca öncelik veya iki kararıda en iyileyen modeller. Amacımız sonsuz zaman içerisinde toplanacak beklenen indirilmiş maliyetleri veya uzun-vadede beklenen ortalama maliyetleri en azlamak. Bu modeller için, en iyi monoton politikaların var olduğunu gösterdik. Sistem parametrelerinin bu politikalar ve çeşitli başarım ölçütler üzerindeki etkileri sayısal örnekler üzerinde çalıştık.

Tezin ikinci kısmında, operasyonel kontrollerin halk sağlığı dinamiği üzerindeki etkilerini anlamak için 4-aşamalı bölmeli model kullandık: (1) sağlıklı, (2) polipli, (3) erken kanser teşhisi (4) geç kanser teşhisi. Kaynakların farklı kullanımlarını modelleyen 4 değişik senaryo yarattık. Farklı kaynak kullanımlarının başarım ölçütlerine olan etkilerini ve sistem parametrelerin etkilerini sayısal örnekler üzerinde detaylı olarak inceledik.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deepest gratitude to my advisors Lerzan Örmeci and Evrim Didem Güneş for their insightful discussions and never ending support during this work. I have learned and achieved more than I ever dreamed of in two-years by their precious guiding and help, and I feel myself very lucky and very proud of being one of their students.

I am grateful to Refik Güllü, Zeynep Akşin and Fikri Karaesmen for taking part in my thesis committee, for critical reading of this thesis and for their valuable suggestions and comments.

I also thank to TUBITAK (The Scientific and Technological Research Council of Turkey) for their generous financial support.

I would like to thank to all my friends at Koç University, Ayşegül Öztürk, Deniz Kubalı, Nihan Çömden, Özge Narin, Serra Caner, Turan Bulmuş, Çağrı Dedeoğlu and Bekir Yenilmez for their valuable friendship which I will not forget to all my life, for their encouragement when I needed and for all the fun and good times we shared together. Special thanks to Çağrı and Bekir for their help in numerous ways in my research.

I would like to thank my parents and my sister for believing me and trusting me at every step I have taken so far and for their endless and supporting love. I would also like to thank Melike Nacak for being the best of my friends and her fiancée Birkan Akbıyık for their kindness and cheerful support. And last but not least, I would like to thank Taştan Ersan Güresen not necessarily for coming along at the right time, but for the very special person he is. And for the incredible amount of patience he had with me in the last six months.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Treatment of a disease at early stages is the most effective form of prevention, therefore 'Screening saves lives' is the motto of the healthcare system. Screening increases life expectancy at the level of the individual and reduces mortality rate as well as treatment costs at the level of the population. Death rates have decreased over the past several years and increasing awareness of the importance of screening for the diseases is one of the major factors for this reduction.

The benefits of investment in screening services are realized as a reduction in disease prevalence and thus a reduction in treatment costs of the future. Since screening resources are limited and effective screening tests are expensive, a better control of screening services is crucial. If in addition, the resources have to be shared between screening and diagnostic services, a trade-off between the immediate need for diagnostic and the long-term benefit of screening will arise. This study focuses on this trade-off, by analyzing admission and scheduling control models for screening services in environments where screening and diagnostic services are provided by a shared resource. The main contribution of our study is introducing a modeling framework that can help better design and control screening services by providing insights for such systems. An application of the models is illustrated using colorectal cancer screening with colonoscopy as an example.

Colorectal cancer is the third most commonly diagnosed cancer and second leading cause of death in the United States (National Cancer Institute [43]). American Cancer Society [2] estimated that 146,970 men and women will be diagnosed with colorectal cancer this year, and 49,920 will die of the disease in the USA. Colorectal cancer is one of the most curable cancers if diagnosed at early stages. For this reason, colorectal cancer screening is critically

important. The demand for screening tests varies due to age, personal history of colorectal cancer or adenomatous polyps, and family history of colorectal cancer or adenomatous polyps. In particular, people of age 50-75 are the general target population for colonoscopy screening [89]. However, people with a colorectal cancer history in their family are considered to be at a higher risk, and therefore they are recommended to have screening starting from age 40 [63]. In general, the demand for screening is generated by individuals from different risk groups.

There are various screening tests available for colorectal cancer. ACG [77] recommends one of these screening tests for people at average risk for cancer and without any symptoms, beginning at age 50. The tests are flexible sigmoidoscopy, colonoscopy, double contrast barium enema, CT colonography, and fecal occult blood test (FOBT). They differ in cancer detection rates, and false-positive (or false-negative) results. There is no perfect screening test, however colonoscopy is widely viewed as the most accurate screening test among available screening tests for colorectal cancer. It has the highest level of sensitivity [1] and specificity [2] for detection of colorectal cancer [88]. Besides, colonoscopy has gained popularity in recent years as the most accurate test because it allows the doctor to see the entire colon and remove polyps [49]. Colonoscopy is an expensive procedure due to expensive equipments, and most countries have insufficient resources available to screen the entire population. Further, demand for colonoscopy exceeds the available capacity [13]. There are long waiting times (approximately six months) for colonoscopies and complaints about long waiting times have been in the news frequently [49, 25]. Limited capacity of colonoscopy raises the question of how to use the available colonoscopy capacity more effectively.

Typically colonoscopy is performed for two purposes: as a screening procedure for asymptomatic individuals, or as a diagnostic procedure for patients with a symptom (such as blood in the stool). If a polyp is found during screening colonoscopy it can be removed before it turns into cancer, thus colonoscopy may also *prevent* colorectal cancer. This eventually would mean a reduction in demand for diagnostic purposes, a reduction in mortality rate and a reduction

[1] Sensitivity is the probability of a positive test among patients with disease

[2] Specificity is the probability of a negative test among patients without disease

in treatment costs. The diagnostic procedure is required to start treatment of a person with cancer, for which a delay may have significant health consequences, therefore timely provision of service is crucial. For this reason, among the two types of demand, the diagnostic procedures are typically given higher priority. Although the procedure duration may be slightly longer when a polyp is removed, these two types of procedures (screening or diagnosis) are not significantly different in terms of costs and duration. If the diagnostic procedure confirms cancer, then further treatment is performed by other service providers. In this study we do not focus on the treatment services such as surgery, since they require different resources which we assume are available. Our focus is on a service such as colonoscopy, which is shared between screening demand and diagnosis demand, and therefore a rationing policy to use available resources effectively is needed.

We assume that the facility providing the screening/diagnostic procedure operates in a dynamic random environment, which determines the demand for diagnostic services. The random environment represents the health of the population, where we assume that if the health of the population is better, the demand rate of symptomatic patients is lower. In the literature, random environments are usually modeled as exogenous systems that cannot be influenced through the decisions of the system under consideration. However, as discussed above, providing screening services decreases the arrival rate of diagnostic services in the long-run. We model this effect by a possible improvement in the environment upon providing a screening service. We assume without loss of generality that there are E different environments $\{1, 2, \cdots, E\}$, where environments are ordered so that environment 1 represents the best and environment E the worst. The health of the population can change only gradually, so we assume that screening can improve an environment e to $e-1$. On the other hand, the worsening of the environment e to $e+1$ is an exogenous transition, which models the effect of deteriorating health conditions that decreases the effect of screening.

The demand arrivals to the facility providing the procedure occur according to a Poisson process. We model the arrival rate of symptomatic patients as a Markov- modulated Poisson process with rate $\lambda_H(e)$ where e is the current environment. We assume, without loss of gener-

ality, that asymptomatic patients form K different risk groups, and let λ_{L_i} be the arrival rate of asymptotic patients from risk group i. Throughout the thesis, we will refer to asymptomatic patients also as low risk patients.

The facility providing the screening/diagnostic services is modeled as a single server. Its service rate is taken as the overall service rate of all such facilities available to the population we are interested in. Such a representation reflects our aim to analyze the whole population, rather than optimizing a single facility. Our objective is to provide assistance in determining general screening policies, which considers the trade-offs between the future health of the population and the need of the current patients given the total capacity of such facilities.

The capacity of these facilities can be controlled through admission and scheduling policies. Admission control can be performed by declining some patients with screening request. This can be applied in practice by appropriate insurance coverage policies for different risk groups, or by dynamically refusing patients given the workload in the system. Scheduling control concerns the prioritization of screening vs. diagnostic demand, and a widely applied practice is strict prioritization of diagnostic procedures. In this thesis, we consider systems which can exercise only admission, only scheduling or both controls.

This study contributes to the understanding of better admission and scheduling control policies for screening services such as colonoscopy. Although the motivation of this work comes from colorectal cancer screening, the model developed here has potential applications in other service contexts, such as maintenance services, where maintenance or prevention type of works are performed along with a more costly repair services. The main characteristics of the systems in question are the following: First, different customer classes demand service from a single service provider where the delay and rejection costs are higher for one class, say class H. Second, customers of class H cannot be rejected. Finally, serving a low-risk class has a potential to reduce the demand from class H in the future. In this context, there is a trade-off between providing service to class H to avoid immediate costs, and providing service to a low-risk class to avoid future costs. We develop a Markov Decision Process (MDP) model for this system and analyze the optimal admission and scheduling policies to minimize system costs. Although

the model has applications in other services, we will use the colorectal cancer screening and colonoscopy service context throughout the thesis.

In the second part of the thesis, we use a dynamic compartmental model to explore the effects of operational controls on population dynamics. Dynamic compartmental models are useful to analyze the effects of policies in population dynamics. Although there is no study on screening of colorectal cancer which uses a compartmental model, this method is used to model screening of other diseases and effects of policies on population dynamics. We model a system with a single facility providing colonoscopy service and two types of risk groups. We develop a four-stage health model where we indicate precursor clinical state (polyps) which may cause inaccurate representation of disease progression if not considered. We consider various scheduling policies where we allocate the capacity between screening and diagnostic purposes. We present a numerical analysis that compares the performances of the scheduling strategies and explore the effect of system parameters (service rate, compliance rate and rate for seeking diagnosis)in detail. This chapter contributes to the understanding of better scheduling control policies for limited resources which are shared between diagnostic and screening purposes.

The remainder of the thesis is organized as follows. In Chapter 2, an overview of the previous studies about health care systems, admission and scheduling controls and environmental process is provided. The problem description and the structural properties for the problem are presented in Chapter 3. Basic structural results on optimal policies are stated in Chapter 4. Then, in Chapter 5, the effects of different parameters, such as service rate, probability of improving the environment on the performances of the models, and on the optimal policies are evaluated numerically. In Chapter 6, the framework is modeled by compartmental modeling and we present a numerical study that compares the performances of the different scheduling strategies . Finally, Chapter 7 concludes the thesis, and summarizes the main results.

Chapter 2

LITERATURE REVIEW

In this chapter, we provide details and references on advances in areas related to different aspects of this thesis.

2.1 Healthcare Systems Literature

Literature on healthcare systems relevant to this thesis can be classified into two categories: papers that develop simulation models to offer insights about the screening programs, and papers that focus on the allocation of the available resources to different demand classes in healthcare systems.

2.1.1 Simulation Models

We can review the papers in this stream in two groups. The first group investigates the cost-effectiveness of screening programs through modeling and the second group focuses on the disease process by considering compartmental modeling.

2.1.1.1 Cost-effectiveness of Screening Programs

Screening services research in the health care operations field typically focuses on screening of a certain disease and models different aspects of the screening programs. Here we briefly mention some examples from this literature on colorectal cancer. We refer the interested readers to [4] and [44] for extensive reviews on cancer screening.

The following studies target the whole population. Harper and Jones [38] develop a semi-Markov model for screening and treatment of colorectal cancer and evaluate alternative screening policies. These policies are no screening, annual Fecal Occult Blood Test (FOBT), annual

FOBT with sigmoidoscopy every 5 years, annual FOBT with sigmoidoscopy every 3 years, sigmoidoscopy at 50 years of age and sigmoidoscopy at 60 years of age. Frazier et al. [32] explore cost- effectiveness of alternative screening programs by developing an MDP model to represent progression of colorectal cancer. The alternative screning programs are constructed by various combination of screening tests and different time intervals. In [82], screening strategies (no screening, single colonoscopy and multiple colonoscopy) for colorectal cancer are modeled as Markov processes and the cost effectivenesses of these strategies are compared.

There are papers which concentrate on more specific groups. Eddy et al. [26] consider high risk patients having a first-degree relative with colorectal cancer and develop a mathematical model to estimate the cost-effectiveness of colorectal cancer screening strategies. Leshno et al. [57] compared the following strategies: no screening, one-time colonoscopic screening, colonoscopy, colonoscopy in a 10-year interval, annual FOBT, annual FOBT and sidmoidoscopy in a 5-year interval, and annual detection of altered human DNA in stool. They develop Partially Observed Markov Decision Process model to analyze the cost-effectiveness of these screening strategies. Further, Khandker et al. [47] examine the screening and surveillance methods for average-risk adults by using a decision model and cost-effectiveness framework. Wagner, Herdman and Wadhwa [90] analyze the cost effectiveness of the screening methods for people between the ages 65 and 85. In the studies evaluating cost effectiveness of screening strategies, the common conclusion is that: colonoscopy every 10 years, and sigmoidoscopy in a 5-years interval plus annual FOBT are the most cost-effective screening strategies.

Papers in this stream investigate the cost-effectiveness of screening programs but do not model any resource allocation issues. Different from these papers, in this thesis we consider resource allocation in screening.

2.1.1.2 Dynamic Compartmental Models

Dynamic compartmental models are useful to analyze the effects of policies in population dynamics. Although there is no study on screening of colorectal cancer which uses a compartmental model, this method is used to model screening of other diseases and effects of policies

on population dynamics. There is a wide literature but here we mention only a few examples. We can refer interested readers to [76], [22], [28], [62], [83], [55], [54], [41],and [46] for further information on compartmental modeling.

Brandeau et al. [10] develop a dynamic compartmental epidemic model to investigate costs and benefits of the programs for HIV screening of women of childbearing age. The model is a closed system with seven population classes (risk groups), each of which includes four disease stages. They explore the performance of different screening policies on targeted risk groups. In addition to policy evaluation, the authors study the sensitivity analyses of system parameters.

Zaric et al. [101] evaluate the cost-effectiveness of methadone maintanence treatment which is an effective way of decreasing the spread of HIV among injection drug users. Different than Brandeau et al. [10], Zaric et al. [101] design a compartmental model for HIV epidemic where the entry into and exit from the population are allowed. In this setting, the compartments are formed according to injection and risk levels. They study both health and economic outcomes of the model.

Güneş et al. [36] use a compartmental model to form a three-stage health model in breast cancer screening. The population is divided into twenty-one compartments according to health status and the state of progress through the health service system. Güneş et al. [36] analyze the effect of operational factors on breast cancer screening. Impacts of standards for minimum reading volume to quality, outreach with or without decentralization of service facilities, and the potential of queueing due to stochastic effects and limited capacity on health outcomes are explored.

In the second part of the thesis, we use a compartmental model to evaluate the effect of operational controls on population dynamics. We point out the limitations in OR models that are observed by Alagöz et al. [4]. In general, operations researchers model the diseases by ignoring precursor states which may cause inaccurate representation of disease progression. We overcome this limitation by developing a four-stage health model where we indicate precursor clinical state (polyps).

2.1.2 Resource Allocation

We focus on two streams in this context. While one stream focuses on outpatient scheduling (for a compherensive literature on outpatient scheduling see Çayırlı [18] and Mondschein [64]), the other stream considers the problem of dynamic allocation of the resources which is closest to our work.

There have been a number of studies on dynamic allocation of service capacity among several customer classes in healthcare systems. Typically, customer classes have different priority levels and costs associated with them.

Green, Savin and Wang [35] address the problem of managing patient demand for a diagnostic service. They consider various patient groups: emergency patients who must be served immediately, inpatients whose demands occur randomly during the day and outpatients who can be scheduled days or weeks in advance. The objective is to maximize expected profit for the diagnostic service. Serving outpatients is an essential source for revenue in health care systems since demands from inpatient and emergency patients are relatively low. Therefore, they investigate the design of outpatient appointment schedule. They establish structural properties of the optimal service policy as well as threshold appointment schedules and analyze performance of heuristic assignment policies numerically.

Patrick, Puterman and Queyranne [71] deal with a dynamic scheduling problem of multi-priority patients to a diagnostic facility. In this setting, emergency patients and inpatients are considered as patients with the highest priorities. Besides, there is incoming demand from multiple outpatient priority classes. Similar to Green et al. [35], emergency patients must be served as soon as possible. Inpatients can wait a maximum of one day and outpatients can often be booked in advance. The objective is to allocate available capacity to incoming demand so that waiting times are minimized. They formulate the booking problem as a Markov decision process and apply the method of approximate dynamic programming. They propose a booking policy and test it through simulation.

Kolisch and Sickinger [50] design an MDP model for resource scheduling of different demand groups (emergency, inpatient and outpatient) in radiology services. In this system, scheduled

outpatients arrive based on an appointment schedule and some do not show up. Inpatients and emergency patients arrive at random. The objective is to maximize the expected reward by allocating two parallel computer tomography scanners within a radiology department to the different demands of the patient classes.

As considered in the above studies, we focus on screening services in environments where screening and diagnostic services are provided by a shared resource, and model the control problem in such systems. Different from them, our model also includes the feedback effect of screening. Further, we study a queuing model with random arrivals whereas in these papers there are scheduled arrivals. Our stochastic system dynamics model differs from these works since two control mechanisms have not been considered all at once in the health care literature. We concentrate on joint admission and scheduling problem in a healthcare system. Therefore, we will now discuss the papers focused on admission control and scheduling.

2.2 Admission and Scheduling Control

The characterization of the admission control and the scheduling problems vary due to the types of servers, queues and the customers. In general, researchers address each of these controls separately in the literature. We provide a quick overview on the admission control policies.

Admission control is a well known approach to improve the performance of queuing systems. There have been many studies done on the control of arrivals to queueing systems. The earliest work on the control of admission to queuing system on $M/M/1$ is Naor's work[65]. Stidham [84] studies admission control to a single server queue with $GI/M/1$ system. For more information on admission control in single server systems, we refer the reader to [45], [58]. Further, there are papers on the optimal control of arrivals in networks of queues and multiple servers. In particular, control of admission to two queues in series [34], admission to the first queue of a series of queues [87], admission to parallel queues [24], and admission to a series of more than two queues [86], dynamic admission control in a two class loss system with c identical parallel servers [67]. Stidham [85] provides a summary of research papers on this area by focusing on various queueing systems where admission of jobs to servers are controlled. Waldmann and

Helm's [40] differs from the mentioned works since it considers the admission of customers to multiserver queues in a random environment. In general, it is shown that the optimal admission policies are of threshold type.

Like admission control, assignment of customers to resources is well studied in literature. Harrison [39] analyzes a single-server queuing system with several classes of customers. In Baras' [6] work, there are two types of customers demanding service from a single server with infinite buffer capacity. Liu and Nain [60] consider a multi-queue single server model. Stidham and Weber [92] review a number of models and results by classifying them into control groups such as service rates, admission, routing and scheduling. In addition, Crabill, Gross and Magazine [23] present a classified bibliography about, static (design) models, dynamic (control) models, and control of queue discipline. Earlier research papers on priority models, scheduling models and allocation of customers to multiple servers can be found in this bibliography. Scheduling in a system with multi-queues (finite or infinite capacity) where each queue has its own single server [42], in the systems with heterogeneous parallel servers [98] , in general two-node network [37], in the systems with two heterogeneous exponential servers [29], [59] are also studied. Optimality of shorter queue faster server policy [42], switch over policy [37], shortest queue policy [29], modified static policy [39], $c\mu$ rule [6], index rules for non preemptive scheduling of single server facility [60] and threshold policies [98], [59] are shown in these studies.

Our contribution is in modeling. The combined study of admission and scheduling control in dynamic programming has not been commonly addressed in the literature.We will now review the literature in which admission control and scheduling problems are addressed together.

Carr and Duenyas [14] discuss a joint order acceptance/rejection and sequencing problem in a production system with two classes of products. Type 1 products are made to stock, and type 2 products are made to order. Demands for type 1 product are met by inventory and decision maker decides whether or not to accept type 2 order. Markov Decision Process (MDP) is used as a tool to formulate the optimal admission and production control problem. It is shown that the optimal production policy and the optimal type 2 order acceptance policy is characterized by a switching curve. The complexity of the model leads the authors to explore the performance

of simpler policies.

Xu and Shantkihumar [99] consider a first-come, first-served service system with m parallel exponential servers. Customers arrive according to a Poisson process. The objective is to maximize the discounted and long run average profit through admission control. They show that the dual of the problem, the preemptive last come first serve system subject to expulsion control is equivalent to the original problem for the proposed cost/reward structure. Expulsion is a special type of admission control. In a system exercising expulsion control, the system does not reject the arriving customer. Instead of rejecting the new customers, it expels customers that are already in the system. They conclude that the optimal policy in the dual system is of a threshold type which implies the structure of the original problem as well.

Xu [97] applies Xu and Shantkihumar [99]'s approach in the case of a $M/M/2$ queueing system with heterogeneous processors and a single customer class. The system exercises both admission and scheduling controls under nonpreemptive first come first serve service discipline. The objective is to maximize the expected discounted and long run average profits. She converts the problem to its dual, a system which is subject to expulsion and scheduling control. She shows that the individually optimal policy in the dual is socially optimal in the original problem. She characterizes the optimal policy as a threshold type for a system subject to both controls.

Righter [80] extends Xu's [97] work to multiple customer classes. For class i customers, a reward R_i is gained for each service completion and a holding cost C_i per unit time is incurred. Priorities can be assigned for classes and this allows preemption. Righter aims to maximize the expected discounted profit for the system. She extends Xu's results for the single class system, and then studies the multiclass system by focusing on two cases where in the first one expulsion is not permitted and in the second one it is permitted. She derives that in both cases the optimal policy is determined by thresholds on the number of customers of each type in the system. She also shows that her results hold for the case of a finite buffer.

Systems subject to admission and scheduling control are complex. Therefore it is difficult to obtain exact results. The above studies deal with admission/scheduling problems by converting them to dual problems. These types of systems can be also studied by using approximations.

The following studies approximate the optimal policies.

Plambeck, Kumar, and Harrison [72] study a multi-class, single server queue with upper bound constraints on the throughput time of jobs. The queue is assumed to be in heavy traffic. They analyze an admission and sequencing policy under heavy-traffic conditions. The objective is to minimize the rejection penalties. They propose an asymptotically optimal scheduling policy and study fluid-scale analysis of this proposed policy.

As a complementary study of [72], Maglaras and Mieghem [61] consider a multi-class queuing network setting with certain lead time constraints. The jobs vary in arrival rates, and processing requirements. They take the lead-times as given and focus on controling the system subject to these lead-time constraints via admission and sequencing controls. Instead of a stochastic system, they study the fluid analysis of joint admission and sequencing control under delay constraints, since lead-time constraints can be guaranteed in the fluid model. They demonstrate the performance of the policies through a simulation experiment and provide preliminary results on admission and sequencing control.

Building up on [72], Ata [5] focuses on a queuing model of a make-to-order production system with a single server. In the model, there are different classes of demands with a rigid due date lead time. The system should decide on whether or not to accept orders and sequencing the orders by assuring that the due date constraints are satisfied. The objective is to minimize lost revenues in the long run. He employs a Brownian approximation method under heavy traffic conditions. The corresponding Brownian control problem of the original system is solved explicitly, and an effective scheduling policy is provided.

Bassamboo, Harrison, Zeevi [7] consider a service system where there are multiple server pools and multiple customer classes. The system exercises two controls. The customer can be blocked or admitted upon her/his arrival. Customers who wait longer can abandon the system. The objective of this paper is to minimize the expected operating costs (sum of blocking, holding and abandonment costs) over a fixed and finite planning horizon. They implement the stochastic fluid approximation method to derive admission and routing controls, and show that their implementation is asymptomatically optimal.

2.3 Environmental Processes

Most of the literature on stochastic models in operations research deal with models in which the system parameters are constant. However, this setting is not applicable to many real life problems. To overcome this limitation, researchers let parameters change by the randomly changing environment, which affects the model as a whole. A widely-used approach to insert an effect of fluctuating environment is defining a secondary Markov process. A process defined as such is called a Markov-modulated process. The Markov-modulated processes can affect the system either internally or externally. The external effect can be via exogenous environmental variations, and the internal effect can be via the decisions of the system under consideration. Most of the research consider extraneous influence of the Markov-modulated processes. Fluctuating environment can be applied in queuing, reliability, inventory models and finance. Since fluctuating environment in queuing models is closest to our work, we explain them in detail. First, we present the papers on exogenous Markov-modulated processes, then we state the papers on endogenous Markov-modulated processes.

The general set up for environmental processes in queueing context is as follows. There is a queue which resides in an environment fluctuating within m environments. Sojourn time in any environment is random, and the arrival and service rates depend on the environment. The earliest work that considered the effect of a fluctuating environment in $M/M/1$ queuing model was studied by Eisen and Tainiter [27]. They consider a system which fluctuates between two environments and obtained some steady state results. Yechiali and Naor [100] study $M/M/1$ queuing system in which the system can be in either of two feasible levels. Neuts [66] generalizes Yechiali and Naor's [100] research to the $M/G/1$ case with m environments. Purdue [74] studies the $M/M/1$ queue which is subject to extraneous phase changes. The environment process is modeled as an m-state irreducible Markov chain in continuous time. He derives the busy period, equilibrium conditions, and probabilities of an empty system. Prabhu and Zhu [73] analyze a model where customer arrival and service rates are modulated by a Markov process. They explore the properties of waiting time, idle time and busy periods. Righter [79] considers a

system where there are incoming resources and they need to be assigned to available activities. She investigates the system with random changes in the arrival rate of resources, random changes in the activity values and deadline rates. The objective is to assign arriving resources to available activities so that total expected return is maximized. She establishes the structure of the optimal policy. Çil, Örmeci and Karaesmen [19] investigates the effect of parameters on optimal policies for a number of queuing and inventory control models where the exogeneous environment is modeled by a Markov chain.

Özekici and Soyer [70] introduce a periodic-review reliability model where each component's survival probability depends on the state of the environment in which the components operate. They build a general set up for any network which is affected by a fluctuating environment, and present results for network reliability assessments. Özekici [68] analyzes different inventory, queueing and reliability models where demands, arrival and service processes and component lifetimes vary due to randomly changing environments. He investigates the implications of environmental processes.

Özekici and Parlar [69] consider periodic-review inventory model with unreliable suppliers. Demand, supply and cost parameters vary stochastically in a random environment. The environmental process follows a time-homogeneous Markov Chain. They analyze the optimality of environment-dependent base-stock and (S, s) policies. Feldman [31] considers (s, S) continuous review inventory policy. He derives the steady-state distribution of the inventory position in a continuous review inventory model where the demand is a Markov modulated compound Poisson process. Song and Zipkin [81] study an inventory model where the demand process is a Markov-modulated Poisson process. The demand fluctuates with a random environment due to economic fluctuations, or stages in the product life-cycle. They discuss the implications of optimal policies. Erdem and Özekici [30] design a single item inventory model that incorporates both random yield and random environment. The supply and the demand processes are modulated by a Markov chain. Also the cost parameters change randomly by the state of the environment. They show that if the levels are environment dependent, the well-known base-stock structure is optimal. Gayon et al. [33] investigate the effects of different pricing strategies avail-

able to a production inventory system with capacitated supply where the demand is generated by a Markov-modulated Poisson process. Different than other studies, the demand depends on the state of the environment and the offered price. They focus on static pricing, environment-dependent pricing, and dynamic pricing policies. They obtain structural results for the optimal pricing and inventory ordering policies.

Çakmak and Özekici [15] study a portfolio selection problem in a stochastic market through optimization. In the model, there are one riskless asset with known return and m risky assets with random returns where the returns in both type of asset depends on the state of the environment. They determine the mean variance efficient frontier which shows the best possible return for a given amount of risk.

Çanakoğlu and Özekici [16] considers multiperiod portfolio selection problem in a stochastic market with exponential utility functions. In this setting, the objective of the investor is to maximize the expected value of a utility function of the terminal wealth. The market states are modulated by a Markov process. The states of the market describe the prevailing economic, financial, social conditions that affect the returns of the assets and utility function. They study the structure of the optimal policy.

Çanakoğlu and Özekici [17] extend the discussion in [16] to multiperiod portfolio optimization by considering investors with logarithmic and power utility. Stochastic market is represented by an external process which affects returns on risky assets and utility function. They assume that the random changes in the market states are depicted by a Markov chain. They characterize the optimal policy.

The above studies investigate the models in a randomly changing environment. However, they do not consider the environment change in the context of an internal influence. Next, we will present the researches with endogenous Markov-modulated processes.

Burman and Smith [12] consider a single server queueing system. In this model, the server represents an access switch and the customers represent packets of data that are generated by data terminals or computers. The customers generate the packets according to a nonhomogeneous Poisson process whose rate is proportional to the number of customers actively generating

data. Therefore, arrivals are modeled by a Markov-modulated process. They characterize the Markov process by its infinitesimal generator. They derive closed-form expressions for the mean delay and mean number in the queue in light and heavy traffic. Gaver and Lehoczky [56] consider a system where the service rate was modulated by an M/M/c/c queueing system and extends the methodology of [12]. Cheng and Sethi [21] study the joint inventory promotion decision problem. The demand process is driven by a Markov chain, which is influenced by uncertain environmental factors and promotion decisions. They characterize the general structure of the optimal policies.

In the studies reviewed so far [except [21]], the demand and service mechanism are modeled either as exogenous or endogenous Markovian processes. Further, there is a limited literature on the systems that are influenced by endogenous Markov-modulated processes. The most important contribution of this thesis is the formulation of environment. We provide a model where these two factors (endogenous and exogenous) are considered together and influence the environment in which the system resides in. The system is under the external influence caused by the deterioration of health conditions. In addition, the effect of screening is modeled as an internal factor that improves the environment which leads to a change in the state of the environment. The arrival rates vary due to the state of the environment. We employ this approach to construct a complex queuing model, where we explore the trade-off between decreasing future risk level of the population and the emergency of incoming patients with a limited capacity of service.

Chapter 3

PROBLEM DEFINITION

In this study, we consider a single server system with different demand groups. The server provides screening and diagnostic services. This chapter describes the process in the facility providing the screening/diagnostic services, how it can be modeled as an Markov Decision Process (MDP) and presents necessary information which enable us to establish structural properties. Section 3.1 describes the model in detail. Section 3.2 presents performance criteria for optimality. In section 3.3, we introduce event operators to represent three possible systems (identified by the type of control(s) they use), as defined in Koole [51]. These are the building blocks of the MDP models to be analyzed later. Finally, section 3.4 introduces certain structures for the value functions, and summarizes which operators preserve these properties, by specifying certain conditions when necessary.

3.1 Description of the system

This section describes the system and how we model it as an MDP. The notation is summarized in Table 3.1.

Demand and Service: There is a single server with an exponential service rate μ. We also assume non-idling server. By non-idling, we prevent such an expensive server to stay idle which will not cost effective. We also assume preemptive scheduling. The demand for the service comes from patients with different risk profiles. We consider K types of asymptomatic patients, representing population groups with different risk factors demanding screening service, and one type of symptomatic patients, representing patients demanding diagnostic service. λ_{L_i} represents the arrival rate of the patient without symptoms who belongs to risk group L_i,$1 \leq i \leq K$. Risk groups are ordered so that a patient in group L_i has a higher risk than one in group L_{i+1}. Patients

without symptoms demand screening service and they may or may not have colorectal cancer. We will assume that they do not have a cancer which would require further treatment. On the other hand, we assume all symptomatic patients have cancer or serious illness. Symptomatic patients arrive according to an environment dependent Poisson process with rate $\lambda_H(e)$, where e denotes the environment, and asymptomatic patients belonging to group L_i arrive according to a Poisson process with rate λ_{L_i}.

Environment: The environment represents the health level of the population with respect to the sickness being considered. We denote the state of the environment process by e, with $e \in \{1, 2, \cdots, E\}$. We assume that if no screening is provided, the environment process will always be in state E. Hence, we can label this state as the *worst* state. We assume, without loss of generality, that the arrival rates of high risk group patients are ordered in the environment e, so that $\lambda_H(e) < \lambda_H(e+1)$, for all e. In words, we can say that environment $e+1$ is worse than environment e, in terms of the high risk arrival rates. We let $\bar{\lambda}_H$ be the maximum arrival rate of the high risk patients, i.e. $\bar{\lambda}_H = \lambda_H(E)$. We note that the environment improves or worsens gradually. Hence, we will allow the environment process to move only to $e-1$ or $e+1$ from environment e.

Screening decreases the proportion of symptomatic patients in the population in the future, which, in turn, decreases the disease prevalence and the arrival rate of symptomatic patients. To model the effect of screening, we assume that if an asymptomatic patient is served in environment e, then the environment improves to $(e-1)$ with probability $p_{e,e-1}$, and remains in e with probability $1 - p_{e,e-1}$. Thus, $p_{e,e-1}$ represents the impact of a single screening service on the number of symptomatic patients in the population, introduced as a proxy for the long-term health effects of screening. This formulation allows us to incorporate the effect of screening into the Markov decision model. Admittedly, improving the environment by screening only one patient is not very realistic. This problem can be overcome by setting low values for $p_{e,e-1}$ and by defining a large number of environments to represent the gradual improvement of the population via screening.

We assume that the environment deteriorates (i.e., shifts from e to $(e+1)$) with rate $\gamma_{e,e+1}$

due to uncontrollable factors such as aging and death. This models the diminishing effects of screening through time. We consider the indirect effect of worsening environment if asymptomatic patients were not screened. Asymptomatic patients will develop colorectal cancer if they wait too much. However, we do not model this fact. We let $\bar{\gamma}$ be the maximum deterioration rate of the environment, such that $\bar{\gamma} = \max_{1 \leq e \leq E-1}\{\gamma_{e,e+1}\}$.

Costs and Decision Structure: Holding costs $c_H \geq 0$ and $c_L \geq 0$ are incurred for each symptomatic or asymptomatic patient in the system, respectively. Since the burden of a symptomatic patient on the system is more than the burden of an asymptomatic patient, we assume that $c_H \geq c_L$. In addition, a fixed cost $c > 0$ is incurred for each symptomatic patient. In reality, a symptomatic patient may or may not have colorectal cancer, and so the treatment costs differ for each patient. The fixed cost c in our model represents the average treatment cost of a symptomatic patient to the system.

In such a setting, the inflow of patients can be controlled by an admission decision to minimize the total system costs. Symptomatic patients are always admitted, while asymptomatic patients may be either accepted or rejected. We incur an immediate cost of denying screening service to an asymptomatic patient, to represent the risk taken by the patient and the system. We assume that rejection costs are ordered as $r_{L_1} > r_{L_2} > r_{L_3} > \ldots > r_{L_K}$, in accordance with the ordering of the risk groups where $r_{L_i} \geq 0$ for all i. The second possible decision is scheduling of the admitted patients. The service provider can give priority to either asymptomatic or symptomatic patients depending on the state of the system. When a patient is served, a screening cost, s, is incurred. A schematic representation of the model is given in Figure 3.1.

We will model three different systems depending on which controls they use: one which exercises both admission and scheduling controls, one exercising only scheduling control while admitting all customers, and finally one which controls the admission of asymptomatic patients while always prioritizing the symptomatic patients in scheduling. In all these systems, the state of the system can be defined as (e, x_H, x_L), where e represents the state of the environment and x_H and x_L denote the number of symptomatic and asymptomatic patients in the system, respectively. We let S be the state space of the system and define it as follows: $S = \{(e, x_H, x_L) :$

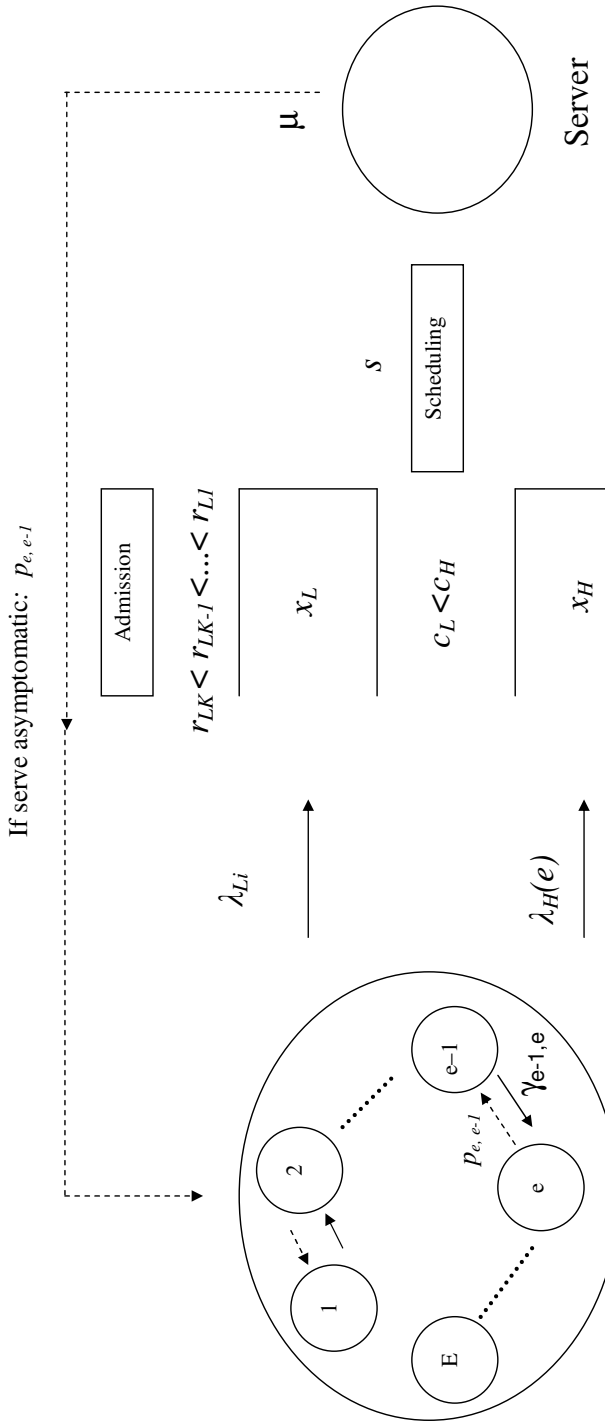


Figure 3.1: Schematic representation of the model.

Parameters	**Definition**
e	Current environment state
$\lambda_H(e)$	The arrival rate of symptomatic patient in environment e
λ_{L_i}	The arrival rate of asymptomatic patient belonging to group L_i
c_H	Holding cost of a symptomatic patient
c_L	Holding cost of an asymptomatic patient
μ	Service rate
$p_{e,e-1}$	Probability of moving from e to $e-1$ upon screening a asymptomatic patient
s	Screening cost
r_{L_i}	Rejection cost for an asymptomatic patient belonging to group L_i
c	Cost of a symptomatic patient
x_H	Number of symptomatic patients in the system
x_L	Number of asymptomatic patients in the system
$\gamma_{e,e+1}$	Deteriorating rate of environment from e to $e+1$
β	Discount (exponential failure) rate
$v_n(e, x_H, x_L)$	Optimal total cost of n stage problem with patients (x_H, x_L) and environment e

Table 3.1: Summary of the notation used in the thesis

$(x_H, x_L) \geq 0, e \in \{1, \ldots, E\}\}$.

All these models can be formulated as a discrete time Markov decision process with the objective of minimizing total expected discounted costs over a finite time horizon with a discount rate β. The discount rate β can be considered as an exponential failure rate such that the system shuts down in an exponentially distributed time with rate β. The maximum possible rate out of any state (e, x_H, x_L) is $R = \sum_{i=1}^{K} \lambda_{L_i} + \bar{\lambda}_H + \mu + \bar{\gamma} + \beta$. Since exponential transition rate out of any state is finite, we use uniformization and normalization to analyze the system in discrete time. We rescale the time by setting $R = 1$, so that the system will be observed at exponentially distributed intervals with mean 1. There will be a potential arrival with probability $(\sum_{i=1}^{K} \lambda_{L_i} + \bar{\lambda}_H)$, a potential service completion with probability μ, a change in the environment with probability $\gamma_{e,e+1}$, a system failure with probability β, and a fictitious transition with probability $\bar{\lambda}_H - \lambda_H(e) + \bar{\gamma} - \gamma_{e,e+1}$.

3.2 Performance Criteria

The performance criterion for optimality can be either the discounted cost criterion over an infinite horizon or long run average cost criterion. Throughout this section, we will pay

attention to these criteria. Meanwhile, we will provide information about algorithms that is used for obtaining optimal policies.

Below, we introduce the notations we will use throughout this section.

Notations	
a	An action
A_s	Set of available actions in state s
d	Decision rule
$d(s)$	Action chosen by decision rule d in state s
d_t	Decision rule at decision epoch t
e	Vector in which all components equal 1
$p(j\|s,a)$	Stationary version of $p_t(j\|s,a)$
$p_t(j\|s,a)$	Probability system occupies state j at decision epoch $t+1$ when action a is chosen in state s at each decision epoch t
$r(s,a)$	Stationary version of $r_t(s,a)$
$r_t(s,a)$	(Expected) present value of one period reward if system is in state s at decision epoch t and action a is chosen
s	A state
S	Set of states
v	An element of a normed linear space of functions on S
V	Space of bounded functions on S

Table 3.2: Summary of the notation used in this section

3.2.1 Discounted Criterion

The objective of this criterion is to minimize (maximize) expected total discounted cost (profit) over an infinite horizon. The following theorem provides the conditions to ensure the existence of the optimal policies.

Theorem 1 ***(Theorem 6.2.10 of [75])***

Assume S is discrete, and either

(a) *A_s is finite for each $s \in S$, or*

(b) *A_s is compact, $r(s,a)$ is continuous in a for each $s \in S$, and for each $j \in S$ and $s \in S$, $p(j|s,a)$ is continuous in a, or*

(c) *A_s is compact, $r(s,a)$ is upper semicontinuous in a for each $s \in S$, and for each $j \in S$ and $s \in S$, $p(j|s,a)$ is lower semicontinuous in a.*
Then there exists an optimal deterministic stationary policy.

In our model, states are constructed by environment and number of patients in the system. Thus, S is countable. Further, there are four actions available for a fixed state (e, x_H, x_L), these are admission (reject or admit asymptomatic patients) and scheduling (serve asymptomatic or symptomatic patients) decisions. Hence, A_s is finite for each $s \in S$. The model satisfies the conditions of Theorem 1, so that for our problem, there is always an optimal deterministic stationary policy.

As it is known that S is a countable set, we will introduce the assumptions which should be satisfied in order to use value iteration algoritm to find the structure of the optimal policies [75];

Assumption 1 *i)There exists a constant $\mu < \infty$ such that*

$$\sup_{a \in A_s} |r(i,a)| \leq \mu w(i). \tag{3.1}$$

ii)There exists a constant $0 \leq \kappa < \infty$, for which

$$\sum_{j \in S} p(j|i,a) w(j) \leq \kappa w(i), \tag{3.2}$$

iii)For each λ, $0 \leq \lambda < 1$, there exists an α, $0 \leq \alpha < 1$ and an integer J such that

$$\lambda^J \sum_{j \in S} P_\pi^J(j|i) w(j) \leq \alpha w(i), \tag{3.3}$$

for all $\pi = (d_1, \ldots, d_J)$ where $d_k \in D_{MD}$; $1 \leq k \leq J$.

We let $w(e, x_H, x_L) = c_H x_H + c_L x_L + c + r + s$. Then for $\mu = 1$,

$$\sup_{a \in A_s} |r(i,a)| \leq \mu w(i) \tag{3.4}$$

holds. Therefore the model satisfies the condition (3.1).

$$\begin{aligned} \sum_{j\in S} p(j|i,a)w(j) &\leq w(e, x_H+1, x_L) && (3.5)\\ &= c_H x_H + c_H + c_L x_L + c + r + s && (3.6)\\ &\leq \kappa(c_H x_H + c_L x_L + c + r + s) && (3.7) \end{aligned}$$

holds with $\kappa = 1 + \frac{c_H}{c}$. Note that inequality (3.5) is true due to $c_H \geq c_L$. Therefore, our system satisfies (3.2).

$$\lambda^J \sum_{j\in S} P_\pi^J(j|i)w(j) \leq \lambda^J w(e, x_H+J, x_L). \tag{3.8}$$

Therefore, it is sufficient to show that

$$\lambda^J w(e, x_H+J, x_L) \leq \alpha w(e, x_H, x_L). \tag{3.9}$$

which is equal to

$$\lambda^J(c_H(x_H+J) + c_L x_L + c + r + s) \leq \alpha(c_H x_H + c_L x_L + c + r + s) \tag{3.10}$$

We let $\lambda = \alpha$. Since $\lambda^J < \lambda$, inequality (3.10) is simplified into:

$$\lambda^{J-1}(x_H + J) \leq x_H. \tag{3.11}$$

Consequently, for J sufficiently large, (3.11) holds. Therefore (3.3) holds. Hence, for this model we can use value iteration algorithm.

We let $v_n(e, x_H, x_L)$ be the total β-discounted minimal cost of the system when it is currently in state (e, x_H, x_L) and n transitions remain in the horizon. We define value iteration algorithm as follows:

3.2.2 Value Iteration Algorithm

1. Select $v_0 \in V$ specify $\epsilon > 0$ and set $n = 0$.

2. For each $i \in S$, compute $v_{n+1}(i)$ by

$$v_{n+1}(i) = \max_{a \in A_s} \left\{ r(i,a) + \sum_{j \in S} \theta p(j|i,a) v_n(j) \right\}. \tag{3.12}$$

3. If

$$||v_{n+1} - v_n|| = \epsilon(1-\theta)/2\theta,$$

go to step 4. Otherwise increment n by 1 and return to step 2.

4. For each $i \in S$, choose

$$d_\epsilon(i) \in \arg\max_{a \in A_s} \left\{ r(i,a) + \sum_{j \in S} \theta p(j|i,a) v_n(j) \right\}. \tag{3.13}$$

We let $v(e, x_H, x_L)$ be the total expected β-discounted cost over an infinite horizon. Then, for $\beta > 0$,

$$v(e, x_H, x_L) = \lim_{n \to \infty} v_n(e, x_H, x_L).$$

All our results are shown under the objective of minimizing total expected β-discounted cost for a finite number of transitions, n. The value iteration algorithm for discounting criterion ensures that the results for v_n extend to v.

3.2.3 Long Run Average Criterion

The objective of the Markov Decision problem is to minimize the expected long run average cost. The conditions in [75] are not met by our model for the stationary average optimal policy. In our model, the state space S is infinite. Weber and Stidham [91] show that even if the state space is not finite and one-stage costs are unbounded, average cost optimal policy can be

determined by taking the limit of discounted cost optimal policies. The conditions for limiting scheme with the corresponding explanations of our model are stated below;

Assumption 2 *i) The state space X is countable.*

ii) The set of actions $A(i)$ which is available in state i is a compact metric space.

iii) The probability $P_a(i,j)$, of transition to state j when action a is taken in state i, is continuous in $a \in A(i)$.

iv) The one-stage cost $C_a(i)$, of taking action a in state i, is non-negative and continuous in $a \in A(i)$.

v)It is possible to go from any state i to any other state j with finite expected cost.

vi)For each i there are only finitely many j for which $P_a(i,j) > 0$ for some $a \in A(i)$.

vii) If there is some policy which achieves a finite average cost, say y^*, then the number of states in which the one-stage cost can be no more than y^* is finite.

Now, we will investigate whether or not these assumptions are met by our model.

In our model, $S = \{(e, x_H, x_L) : (x_H, x_L) \geq 0, e \in \{1, \ldots, E\}\}$ is countable. In the model, there are four actions available for a fixed state (e, x_H, x_L), these are admission (reject or admit asymptomatic patients) and scheduling (serve asymptomatic or symptomatic patients) decisions. Since any finite set is compact, the set of actions is compact. Since the action space is discrete, if we take two points $|a - a'| < \epsilon$, then these are exactly same points. Hence $|P_a(i,j) - P_{a'}(i,j)| = |P_a(i,j) - P_a(i,j)| = 0$. Therefore, the continuity holds by definition, which shows that our model satisfies (iii). Since the costs in the system are nonnegative, one stage costs, $C_a(i)$, are non negative. Similarly, taking two points $|a - a'| < \epsilon$ implies that $|C_a(i) - C_{a'}(i)| = |C_a(i) - C_a(i)| = 0$. Therefore, condition (iv) is satisfied, too. We have a recurrent Markov process. It is possible go from any state i to the state $(2, 0, 0)$ since the system can reduce the number of patients in the system by screening patients, and the environment move to $e = 2$ with a rate γ by deterioration of health conditions. We initialize $v_0(z) = 0$ for all z, therefore being a sum of finite numbers $v_n(i)$ and $v_n(j)$ are finite, which implies that their difference is finite. From a state (e, x_H, x_L),the system can move to states $(e, x_H, x_L + 1)$, $(e, x_H - 1, x_L)$, $(e, x_H, x_L - 1)$ and $(e-1, x_H, x_L - 1)$ with probability greater than zero. Since costs are increasing

in all components, for a finite average cost, there are finitely many states with less finite average cost.

We show that all the conditions are met, so that there exists an average cost optimal policy.

For average cost criterion, value iteration algorithm may not converge or it may cause numerical instability. Therefore, we use relative value iteration for finding optimal policies. The relative value algorithm is given below.

3.2.4 Relative Value Iteration Algorithm

1. Select $v_0 \in V$, choose $s^* \in S$, specify $\epsilon > 0$, set

$$w_0 = v_0 - v_0(s^*)e \quad \text{and} \quad n = 0. \tag{3.14}$$

2. Set

$$v_{n+1} = Lw_n \quad \text{and} \quad w_{n+1} = v_{n+1} - v_{n+1}(s^*)e. \tag{3.15}$$

3. If

$$sp(v_{n+1} - v_n) < \epsilon,$$

go to step 4. Otherwise, increment n by 1 and return to step 2.

4. Choose

$$d_\epsilon \in \arg\max_{d \in D}\{r_d + P_d v_n\}. \tag{3.16}$$

We choose $s^* = (2, 0, 0)$. In our model, L is the combination of operators which we will mention in the next section.

We will define the operators which specify the consequences of each of the events described above in the next section. Different combinations of these operators represent the three models

we are interested in. Section 3.4 will prove that these operators preserve certain properties of a function $f(e, x_H, x_L)$. This implies, by induction, that $v_n(e, x_H, x_L)$ has this property for all n if $v_0(e, x_H, x_L)$ has them.

3.3 Operators

We use event based dynamic programming method introduced by Koole [51] to formulate our models. Koole constructed the value functions as a combination of event operators and investigated the properties of the event operators. A value function preserves the properties if all event operators that form it has these properties. Therefore, we use the preserved properties of event operators to characterize the structure of optimal policy. At each transition epoch, there can be an arrival of a symptomatic patient, an arrival of an asymptomatic patient, a service completion, and a fictitious transition. As in Koole [51] and Çil et al. [19], we will introduce event operators for each event and study their properties. The event operators used to model system transitions in these cases are explained below in detail. Let f be a generic function defined on the state space S, $f : S \to \mathbb{R}$.

Arrival process of symptomatic patients: The system admits all symptomatic patients by incurring a treatment cost of $c > 0$. We define an arrival operator, T_{ARR_H}, to represent the arrival process of queueing system:

$$T_{ARR_H} f(e, x_H, x_L) = \alpha[f(e, x_H + 1, x_L) + c] + (1 - \alpha) f(e, x_H, x_L), \tag{3.17}$$

where $\alpha = \lambda_H(e)/\bar{\lambda}_H$. $\bar{\lambda}_H$ is the maximal arrival rate of symptomatic patients. Hence, α is the probability that a symptomatic patient arrives at the system in environment e incurring a fixed cost of c, and $1 - \alpha$ is the probability of a fictitious arrival.

Arrival process of asymptomatic patients: The arrival of asymptomatic patients can be modeled in two different ways, depending on whether the arrivals are controlled or not.

The operator, T_{ADM_i}, models the admission control of asymptomatic patients. If there is an arrival of an asymptomatic patient, we compare the costs of the actions. Asymptomatic patient

from the group L_i would be accepted for screening only if

$$f(e, x_H, x_L) + r_{L_i} \geq f(e, x_H, x_L + 1).$$

Hence, for $1 \leq i \leq K$:

$$T_{ADM_i} f(e, x_H, x_L) = \min\{f(e, x_H, x_L) + r_{L_i}, f(e, x_H, x_L + 1)\}. \tag{3.18}$$

The operator $T_{ARR_{L_i}}$, on the other hand, models the arrival of asymptomatic patients when they are always admitted to the system. When a patient without symptoms arrives with rate λ_{L_i} in state (e, x_H, x_L), the system moves to state $(e, x_H, x_L + 1)$:

$$T_{ARR_{L_i}} f(e, x_H, x_L) = f(e, x_H, x_L + 1). \tag{3.19}$$

for $1 \leq i \leq K$.

Departure Process: The departure process can also be modeled in two ways, depending whether the system is exercising scheduling control or not. In this framework, we assume a non-idling policy.

The scheduling operator T_{SCH} models the choice of whether to serve an asymptomatic patient or a symptomatic patient, incurring the screening cost, s. Asymptomatic patients would be served only if

$$g(e, x_H, x_L - 1) \leq f(e, x_H - 1, x_L).$$

g function depends on the environment. Since we label environment 1 as the best environment, environment can not improve for $e = 1$. In this case, the system serves an asymptomatic patient if

$$f(1, x_H, x_L - 1) \leq f(1, x_H - 1, x_L)$$

for any state $(1, x_H, x_L)$.

However, for $e \in \{2, \ldots, E\}$, if an asymptomatic patient is served, then the environment will

improve with probability $p_{e,e-1}$, and remain the same with probability $(1-p_{e,e-1})$. So for any state (e, x_H, x_L) we compare the functions $p_{e,e-1}f(e-1, x_H, x_L-1)+(1-p_{e,e-1})f(e, x_H, x_L-1)$ and $f(e, x_H-1, x_L)$. Moreover, in both cases a screening cost, $s \geq 0$, is incurred. Thus, the scheduling operator T_{SCH} is defined as:

$$T_{SCH}f(e, x_H, x_L) = \min\{f(e, x_H-1, x_L), g(e, x_H, x_L-1)\} + s, \tag{3.20}$$

where

$$g(e,x,y) = \begin{cases} f(1,x,y) & \text{if} \quad e=1 \\ p_{e,e-1}f(e-1,x_H,x_L-1)+(1-p_{e,e-1})f(e,x_H,x_L-1) & \text{otherwise.} \end{cases}$$

The operator T_{DEP} models the system which always serves the symptomatic patients first:

$$T_{DEP}f(e, x_H, x_L) = \Psi(e, x_H, x_L) + s,$$

where

$$\Psi(e,x,y) = \begin{cases} f(e,x-1,y) & x \neq 0, \text{and } e \in \{1,\ldots,E\} \\ f(1,x,y-1) & x=0, \text{and } e=1 \\ p_{e,e-1}f(e-1,x,y-1)+(1-p_{e,e-1})f(e,x,y-1) & x=0, \text{and } e \in \{2,\ldots,E\}. \end{cases}$$

Environment Shift: In the deterioration operator, T_{DET}, either a transition from an environment e to $e+1$ occurs with probability $\tau = \gamma_{e,e+1}/\bar{\gamma}$ or a fictitious environment change occurs with probability $1-\tau$:

$$T_{DET}f(e, x_H, x_L) = \tau f(e+1, x_H, x_L) + (1-\tau)f(e, x_H, x_L). \tag{3.21}$$

Cost: The cost operator, T_{COST}, indicates a non-negative holding cost, $h(e, x_H, x_L) = c_H x_H + c_L x_L$, to the system:

$$T_{COST}f(e, x_H, x_L) = f(e, x_H, x_L) + h(e, x_H, x_L). \tag{3.22}$$

Uniformization: The operator, T_{UNIF}, forms the convex combination of the functions f_j with propabilities $p_j > 0$ for all j. For any function $f_j : S \to \mathbb{R}$,

$$T_{UNIF}(\{f_j(e, x_H, x_L)\}; \{p_j\}) = \sum_j p_j f_j(e, x_H, x_L). \tag{3.23}$$

We can model discounting criterion by satisfying the condition $\sum_{j=1}^{m} p_j < 1$, and long-run average criterion by $\sum_{j=1}^{m} p_j = 1$.

3.4 Structural Properties Preserved by the Operators

In this section, we define certain structural properties of the operators.

We define the following notations to represent the effects of additional patients, deterioration of environment and interchanging patients respectively.

$$\Delta_H(e, x_H, x_L) = v(e, x_H + 1, x_L) - v(e, x_H, x_L), \tag{3.24}$$

$$\Delta_L(e, x_H, x_L) = v(e, x_H, x_L + 1) - v(e, x_H, x_L), \tag{3.25}$$

$$\Delta_e(e, x_H, x_L) = v(e + 1, x_H, x_L) - v(e, x_H, x_L), \tag{3.26}$$

$$\Delta_{HL}(e, x_H, x_L) = v(e, x_H + 1, x_L) - v(e, x_H, x_L + 1). \tag{3.27}$$

Equations 3.24 and 3.25 represents the burden for an additional symptomatic or asymptomatic patients respectively in state (e, x_H, x_L), and Equation 3.26 indicates the burden of worsening environment e to $e+1$ in state (e, x_H, x_L). Equation 3.27 represents the burden of switching an asymptomatic patient with a symptomatic patient.

First we will consider the monotonicity properties. We call a value function increasing in x if $f(x) \leq f(x+1)$. Now, we give the modified definitions of the monotonicity properties for our

model:

$$Inc(x_H) : \Delta_H(e, x_H, x_L) \geq 0, \tag{3.28}$$

$$Inc(x_L) : \Delta_L(e, x_H, x_L) \geq 0, \tag{3.29}$$

$$Inc(e) : \Delta_e(e, x_H, x_L) \geq 0. \tag{3.30}$$

Equation 3.28 implies that when a symptomatic patient enters the system, the expected total cost increases. In other words, a positive burden is incurred for an additional symptomatic patient. Similarly, equation 3.29 implies a burden for an additional low risk patient on the system. Lastly, Equation 3.30 implies that deterioration of an environment has a burden on the system. In other words, when environment worsens, proportion of high risk patients in the population increases, so that expected total cost increases, too.

We also call a function decreasing in x if $f(x) \geq f(x+1)$. We let $v^p(e, x_H, x_L)$ be the value function in state (e, x_H, x_L) where p is the probability of improving the environment upon screening. then we define $Dec(p)$ property as follows:

$$Dec(p) : v^{p+\epsilon}(e, x_H, x_L) \leq v^p(e, x_H, x_L), \forall\, \epsilon > 0. \tag{3.31}$$

Next, we let $Diag(x_H, x_L)$ represent the monotonicity on the diagonal and define it:

$$Diag(x_H, x_L) : \Delta_{HL}(e, x_H, x_L) \geq 0. \tag{3.32}$$

Due to the complex structure of scheduling, this property does not have a general implication on scheduling. However, for $e = 1$, it characterizes the optimal scheduling policy, as prioritizing symptomatic patients.

Further, we consider the inverse diagonality property by taking account of the environment factor. We represent it by $IDiag_e(x_H, x_L)$ and describe it as follows:

$$IDiag_e(x_H, x_L) : v(e-1, x_H+1, x_L) \leq v(e, x_H, x_L+1). \tag{3.33}$$

First, we assume that the state is $(e-1, x_H, x_L+1)$. We consider short run and long run effects. The short-run burden can be represented by switching one asymptomatic patient to a symptomatic patient. Mathematically, it is represented as:

$$v(e-1, x_H+1, x_L) - v(e-1, x_H, x_L+1).$$

Further,

$$v(e, x_H, x_L+1) - v(e-1, x_H, x_L+1)$$

expresses the burden of worsening environment which can be considered as a long-run effect. The difference is equal to $IDiag_e(x_H, x_L)$:

$$\begin{aligned} & v(e, x_H, x_L+1) - v(e-1, x_H, x_L+1) - v(e-1, x_H+1, x_L) + v(e-1, x_H, x_L+1) \\ = \;& v(e, x_H, x_L+1) - v(e-1, x_H+1, x_L). \end{aligned}$$

Therefore, $IDiag_e(x_H, x_L) \geq 0$ implies that the burden of switching one asymptomatic patient to a symptomatic patient is less than the burden of worsening the environment. In other words, long run effects are stronger than short run effects. This property is important to understand the structure of the optimal scheduling policy in environment $e \geq 2$.

We also consider the convexity in x_L:

$$Conv(x_L) : \Delta_L(e, x_H, x_L) \leq \Delta_L(e, x_H, x_L+1), \tag{3.34}$$

which specifies the existence of an optimal admission policy as a threshold type.

Finally, we define supermodularity properties. First one is the supermodularity of the value functions in (x_H, x_L) and the second one is the supermodularity of the value functions in (e, x_L).

$$Sup(x_H, x_L) \;:\; \Delta_H(e, x_H, x_L) \leq \Delta_H(e, x_H, x_L+1), \tag{3.35}$$

$$Sup(e, x_L) \;:\; \Delta_e(e, x_H, x_L) \leq \Delta_e(e, x_H, x_L+1). \tag{3.36}$$

Equation 3.35 implies that the optimal admission policy can be characterized with a threshold on x_H, and equation 3.36 implies that it can be characterized with a threshold on e.

Table 3.3 summarizes the results. Ticks in the table imply that the event operator preserves the corresponding property. For some operators, additional conditions are required. These conditions are remarked as superscripts. Moreover, the operator T_{UNIF} preserves a property if all functions constituting the operator preserves this property. All proofs are stated in Appendix.

If a certain operator T is defined as $Tv_n = v_{n+1}$, where T is a combination of the operators T_{ARR_H}, T_{ADM_i}, $T_{ARR_{L_i}}$, $T_{SCH}(T_{DEP})$, T_{COST}, and T_{DET} then the properties $Inc(x_H)$, $Inc(x_L)$, $Inc(e)$, $Dec(p)$ and $Diag(x_H, x_L)$ are always preserved by T. $IDiag_e(x_H, x_L)$ is preserved under some conditions. For the operator, T_{COST}, the condition $\frac{c_H - c_L}{c} \leq \lambda_H(e+1) - \lambda_H(e)$, and for the operator T_{DET}, the condition $\lambda_H(e+1) - \lambda_H(e) \geq \gamma_{e,e+1}$ should be met. Moreover, those operators, if alone, are not enough for $IDiag_e(x_H, x_L)$ property. They should form the model with the operator T_{ARR_H}. Then we define an operator U as $Uv_n = v_{n+1}$, where U is a combination of the operators T_{ARR_H}, T_{ADM_i}, $T_{ARR_{L_i}}$, T_{DEP}, T_{COST}, and T_{DET}. We can say that U preserve $Conv(x_L)$ property anyway. However, it preserves $Sup(x_H, x_L)$ property if T_{DEP} preserves $Sup(e, x_L)$ property. Also, it preserves $Sup(e, x_L)$ property if T_{ARR_H} preserves $Sup(x_H, x_L)$ property.

The first three properties show that all operators preserve monotonicity with respect to x_H, x_L and e respectively. Hence, the value functions of all the models we consider will be increasing in x_H, x_L and e. This is very intuitive, since increasing the number of symptomatic patients or low risk patients in the system can only increase the total cost. Moreover, the assumption $\lambda_H(e) < \lambda_H(e+1)$ implies that environment e is better than environment $e+1$. In other words, the total minimal costs for each environment state e have the same ordering with the arrival rates of symptomatic patients of corresponding environment. Further, the value functions are decreasing in p. This is expected since as the probability of improvement the environment increases, the costs should decrease. An observation from Table 3.3 is on the operator T_{SCH}, as it does not preserve a number of properties. Accordingly, we will see below that systems exercising scheduling control will not have certain monotonicity properties.

	Preserved Properties								
Operator	$Inc(x_H)$	$Inc(x_L)$	$Inc(e)$	$Dec(p)$	$Diag(x_H, x_L)$	$IDiag_e(x_H, x_L)$	$Conv(x_L)$	$Sup(x_H, x_L)$	$Sup(e, x_L)$
T_{ARR_H}	√	√	√	√	√	√	√	√	√$^{\circ}$
T_{ADM_i}	√	√	√	√	√	√	√	√	√
$T_{ARR_{L_i}}$	√	√	√	√	√	√	√	√	√
T_{SCH}	√	√	√	√	√	√	–	–	–
T_{DEP}	√	√	√	√	√	√	√	√*	√
T_{COST}	√	√	√	√	√	√$^{\bullet}$	√	√	√
T_{DET}	√	√	√	√	√	√†	√	√	√

† $\lambda_H(e+1) - \lambda_H(e) \geq \gamma_{e,e+1}$, and it should enter the model with T_{ARR_H}
$^{\bullet}$ $\frac{c_H - c_L}{c} \leq \lambda_H(e+1) - \lambda_H(e)$, and it should enter the model with T_{ARR_H}
* $f(e, x_H, x_L) : Sup(e, x_L)$
$^{\circ}$ $f(e, x_H, x_L) : Sup(x_H, x_L)$

Table 3.3: Properties preserved by the operators when the function $f(e, x_H, x_L)$ has the corresponding property

Chapter 4

MODEL FORMULATION AND STRUCTURAL PROPERTIES OF OPTIMAL POLICIES

In this chapter, we will introduce three models differentiating on controls, and establish structural properties of these models.

4.1 Systems Exercising Both Scheduling and Admission Control

In this section, we model systems which exercise both scheduling and admission control. We study the implications of the structural properties preserved by the operators on the structure of optimal policies. As before, we let $v_n(e, x_H, x_L)$ be the total expected β-discounted cost of such a system with n transitions remaining in the horizon. Using the event based dynamic programming technique, the optimality equations are given by:

$$\begin{aligned} v_{n+1}(e, x_H, x_L) = T_{COST}(T_{UNIF}(\{T_{ARR_H} v_n(e, x_H, x_L), \{T_{ADM_i} v_n(e, x_H, x_L)\}_i, \\ T_{SCH} v_n(e, x_H, x_L), T_{DET} v_n(e, x_H, x_L)\}; \{\bar{\lambda}_H, \{\lambda_{L_i}\}_i, \mu, \bar{\gamma}\})), \end{aligned} \tag{4.1}$$

with boundary conditions

$$\begin{aligned} v_n(e, -1, x_L) &= v_n(e, 0, x_L), \\ v_n(e, x_H, -1) &= v_n(e, x_H, 0), \end{aligned}$$

and,

$$v_0(e, x_H, x_L) = 0, \quad \forall\ (e, x_H, x_L).$$

Remark: From now on, we will use v_n instead of $v_n(e, x_H, x_L)$.

This model preserves $Inc(x_H)$, $Inc(x_L)$, $Inc(e)$, and $Dec(p)$ and $Diag(x_H, x_L)$ properties for any parameter set. Further, it preserves $Diag_e(x_H, x_L)$ property if the necessary conditions are satisfied.

4.1.1 Scheduling Control

Intuitively, we expect the system to always give priority to symptomatic patients. However, this is not valid for the whole model since we assume that screening asymptomatic patients will decrease the arrival rate of the symptomatic patients in the long run. We aim to observe the effect of p on the optimal policy. Since screening asymptomatic patients could not improve the environment, scheduling policy is characterized for the best environment as giving priority to symptomatic patients. In this context, the operator $Diag(x_H, x_L)$ will have direct implications on the optimal scheduling policy. The following proposition characterizes the scheduling policy for the systems which exercises both admission and scheduling control.

Proposition 1

i*) If $e = 1$ and $x_H > 0$, then symptomatic patients are always scheduled first.*

ii*) If $p = 1$, and the following conditions are satisfied, then asymptomatic patients gain priority in $e \in \{2, \dots, E\}$:*

$$(a) \qquad \frac{c_H - c_L}{c} \leq \lambda_H(e) - \lambda_H(e-1), \tag{4.2}$$

$$(b) \qquad \lambda_H(e) - \lambda_H(e-1) \geq \gamma_{e-1,e}. \tag{4.3}$$

iii*) Otherwise, the optimal scheduling policy is dynamic.*

In particular, i) does not hold for $e \in \{2, \dots, E\}$. In order to serve symptomatic patients first in $e \in \{2, \dots, E\}$, the following criterion should be satisfied;

$$v(e, x_H - 1, x_L) \leq p_{e,e-1} v(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) v(e, x_H, x_L - 1). \tag{4.4}$$

However, inequality (4.4) is not satisfied for all parameter values. Actually we can show that for

$e > 1$, under certain conditions asymptomatic patients obtain priority. We consider the extreme case where $p_{e,e-1} = 1$. Then our scheduling decision is characterized by

$$v(e, x_H - 1, x_L) - v(e - 1, x_H, x_L - 1). \tag{4.5}$$

The expression (4.5) is greater than zero if the conditions stated in (4.2) and (4.3) are satisfied. We check the conditions in order to see when the system satisfies these conditions. Since, $c_H \geq c_L$, $c_H - c_L \geq 0$. c is a huge number and as c goes to infinity, $\frac{c_H - c_L}{c}$ goes to zero. It implies $\lambda_H(e) \geq \lambda_H(e-1)$, which is one of the model's assumption. Therefore, (a) is valid for our system. In (b), the change in the arrival rates of symptomatic patients which can be considered as an effect in the medium term has a higher impact than the deterioriation rate of long term other effects. Hence the equations (4.2) and (4.3) will generaly hold. Therefore, asymptomatic patients obtain priority if conditions in Proposition 1 are satisfied. More generally, we can conclude that asymptomatic patients may obtain priority. Therefore (i) cannot be extended for other environments. Figure 4.1 is the illustration of Proposition 1. The shaded regions imply that asymptomatic patients obtain priority. The conditions (a) and (b) are satisfied for the parameter set. First graph represents part (i), second graph part (ii), and third graph part (iii).

Further, we can make the following conjecture.

Conjecture 3 *Let $\overline{p}$ be the largest value such that symptomatic patients always obtain priority and $\underline{p}$ be the smallest value such that asymptomatic patients always obtain priority. Then for any number $p < \overline{p}$ symptomatic patients always obtain priority and for any number $p > \underline{p}$ asymptomatic patients always obtain priority. For other p values we have dynamic scheduling.*

We illustrate this observation with the following example.

Example 1 We consider long-run average cost criterion with the following parameter values. $\lambda_L = 11$, $\lambda_H = [0.36\ 1.2]$, $\mu = 12$, $c_H = 1$, $c_L = 0.02$, $r = 0.5$, $s = 0$, $c = 35100$, and $\gamma = 1/3600$. In $e = 1$, the system serves symptomatic patients as proved. We change p values to observe the optimal scheduling policy in $e = 2$. We observe that if $p \in [0, 0.0008278]$, we give priority to symptomatic patients ($\overline{p} = 0.0008278$). In other words, screening asymptomatic patients is

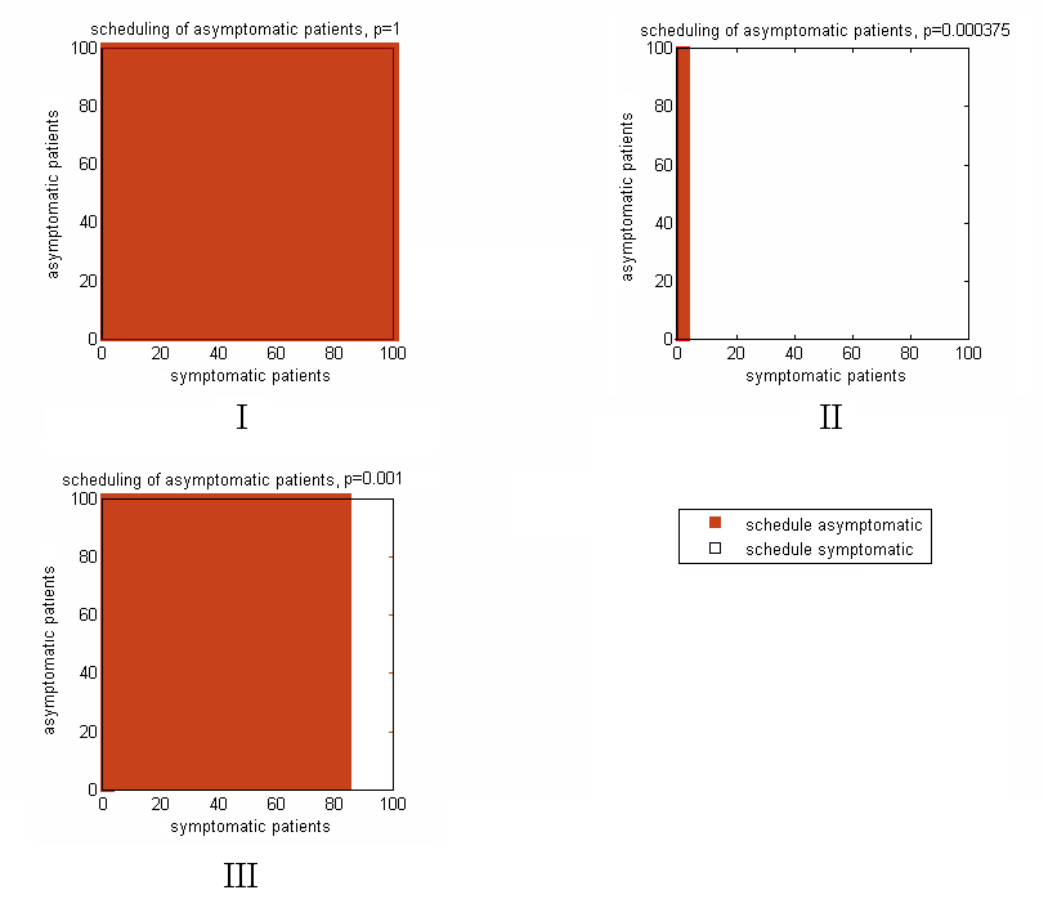


Figure 4.1: Illustration of optimal scheduling policies for different p values (Remaining parameters are as follows; c_H=1, c_L=0.02, λ_L=1.6, $\lambda_H(2)$=1.82, $\lambda_H(1)$=0.55, μ=1.73, γ=0.0004, β=0, r=0.5, s=0, and c=35100).

not enough to improve the environment. So the system chooses to serve symptomatic patients. When $p \in [0.76, 1]$, the system gives priority to asymptomatic patients since in the long run the number of symptomatic patients decreases ($\underline{p} = 0.76$). Otherwise, there is a dynamic scheduling.

4.1.2 Admission Control

Intuitively, we expect rejection costs to be inversely proportional with rejection regions for all environments, since the rejection costs are ordered but screening costs are the same. The implication of this result on the optimal policy is given in Proposition 2:

Proposition 2 *Given (e, x_H, x_L), let k be the smallest integer such that; if it is optimal to accept a patient from the group L_k, then it is optimal to accept a patient from the group L_i whenever $k \geq i$, where $r_{L_K} < r_{L_{K-1}} < \ldots < r_{L_2} < r_{L_1}$.*

Proof. There are K different risk groups with rejection costs: $r_{L_K} < r_{L_{K-1}} < \ldots < r_{L_2} < r_{L_1}$. Let A_i denote the set where the asymptomatic patient in the i^{th} group is admitted, and R_i denote the set where the asymptomatic patient in the i^{th} group is rejected. Suppose that the state is (e, x_H, x_L) and k is the smallest integer such that asymptomatic patient is admitted. Let an asymptomatic patient be in A_k (by the definition of k, this patient is in R_{k+1}). In other words:

$$v(e, x_H, x_L) + r_{L_k} \geq v(e, x_H, x_L + 1),$$

and

$$v(e, x_H, x_L) + r_{L_{k+1}} < v(e, x_H, x_L + 1).$$

By assumption,

$$v(e, x_H, x_L) + r_{L_1} > v(e, x_H, x_L) + r_{L_2} > \ldots > v(e, x_H, x_L) + r_{L_k} > v(e, x_H, x_L + 1).$$

This implies that in state (e, x_H, x_L), if the patient in the k^{th} group is admitted then the patients who are in the groups with greater rejection costs than k^{th} group in that state is also admitted.

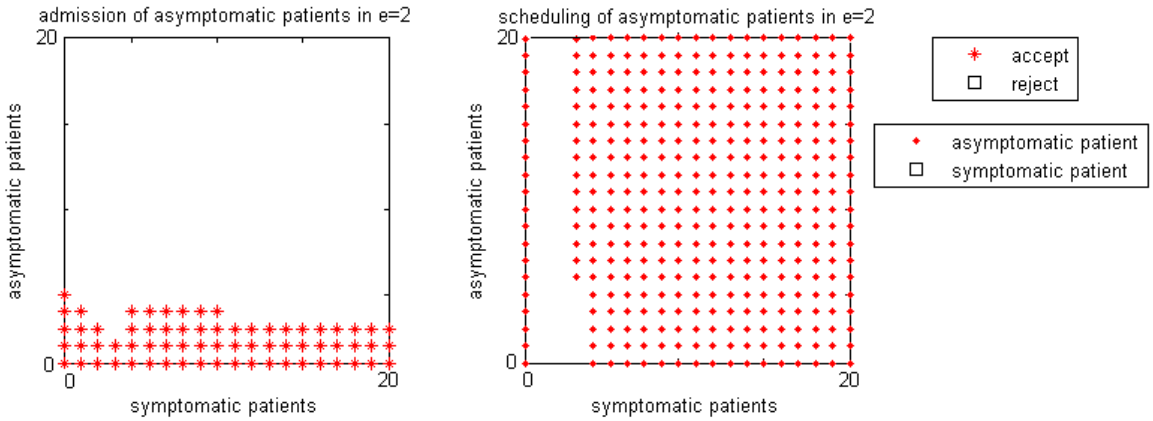


Figure 4.2: Optimal Policy with Discounting Model

Hence,

$$v(e, x_H, x_L) + r_{L_j} \geq v(e, x_H, x_L + 1)$$

means the patient is in A_j in state (e, x_H, x_L) for all $j \leq k$. Therefore if we order the sets, we obtain

$$A_K \subset A_{K-1} \subset \ldots \subset A_2 \subset A_1,$$

and

$$R_1 \subset R_2 \subset \ldots \subset R_{K-1} \subset R_K.$$

□

Further, we observe that if symptomatic patients are not always scheduled first, we can not characterize the structure of the optimal admission policy due to the effect of event operator T_{SCH}. We illustrate this with an example.

Example 2 We consider the discounting model with parameters c_H=100, c_L=0.02, λ_L=11, $\lambda_H(2)$=1.2, $\lambda_H(1)$=0.36, μ=12, γ=0.01, β=0.35, p=0.096, r=0.5, s=0, c=30000. Figure 4.2

shows the optimal policy of the system. We observed that rejection region is not increasing in symptomatic patients. Thus, we cannot show the existence of the threshold on symptomatic patients.

As observed, this model is difficult to analyze in terms of optimal policies due to scheduling. So we simplify the model by considering dynamic admission model given symptomatic patients are always scheduled first, irrespective of the environment, and analyze this special case in the next section.

Before moving to dynamic admission model, we analyze the case where there is a dynamic scheduling with a fixed admission policy as admit all. We rewrite the model.
For $e \in \{1, \ldots, E\}$,

$$v_{n+1} = T_{COST}(T_{UNIF}(\{T_{ARR_H} v_n, \{T_{ARR_{L_i}} v_n\}_i, T_{SCH} v_n, T_{DET} v_n\}; \{\bar{\lambda}_H, \{\lambda_{L_i}\}_i, \mu, \bar{\gamma}\})). \quad (4.6)$$

The new operator $T_{ARR_{L_i}}$ preserves the monotonicity properties in x_H, x_L, and e. Further, it maintains $Dec(p)$, $Diag(x_H, x_L)$ and $IDiag_e(x_H, x_L)$ properties. All the results of the previous model apply to this case. However, due to the complex behaviour of the operator T_{SCH} we can not characterize the structure of optimal scheduling policy in this model as well.

4.2 System Exercising Only Admission Control

In this model, we assume that symptomatic patients have priority, so that if there is a symptomatic patient in the system, he/she will be served. Therefore, an asymptomatic patient will be served only when there is no symptomatic patients in the system. The model is presented below;
For $e \in \{1, \ldots, E\}$,

$$v_{n+1} = T_{COST}(T_{UNIF}(\{T_{ARR_H} v_n, \{T_{ADM_i} v_n\}_i, T_{DEP} v_n, T_{DET} v_n\}; \{\bar{\lambda}_H, \{\lambda_{L_i}\}_i, \mu, \bar{\gamma}\})). \quad (4.7)$$

The operator T_{DEP}, which replaces T_{SCH}, preserves the monotonicity properties in x_H, x_L, e, p and other diagonality properties. Thus, the results derived in Section 4.1 are still valid for this

model.

$Sup(x_H, x_L)$ property implies that if the system rejects an asymptomatic patient for any pair (x_H, x_L), then it will still reject him/her when there are more symptomatic patients in the system. In other words, the cost of an additional asymptomatic patient to the system should increase in the number of symptomatic patients in the system. The operators, T_{ARR_H}, T_{ADM_i}, T_{DEP}, T_{DET}, T_{COST} and T_{UNIF} preserve the $Sup(x_H, x_L)$ property if they also preserve $Sup(e, x_L)$ property. Supermodularity guarantees the existence of optimal threshold policies, so we conclude that the optimal policy is of threshold type for x_H. The following theorem states the supermodularity of value functions in (x_H, x_L). The proofs for these results are provided in the Appendix.

Theorem 1 *Given (e, x_L), there exists an optimal threshold $l_H(e, x_L)$ on x_H such that if $x_H \geq l_H(e, x_L)$ it is optimal to reject the incoming asymptomatic patient, otherwise it is optimal to admit her.*

Proof. We would like to show that an optimal threshold policy exists. We define

$$l_H(e, x_L) = \arg\min\{x_H : v_n(e, x_H, x_L + 1) - v_n(e, x_H, x_L) - r \geq 0\}.$$

Assume that $l_H(e, x_L) = x_H^*$, so that the following inequality is satisfied,

$$v_n(e, x_H^*, x_L + 1) - v_n(e, x_H^*, x_L) - r \geq 0.$$

By $Sup(x_H, x_L)$ property,

$$v_n(e, x_H^* + 1, x_L + 1) - v_n(e, x_H^* + 1, x_L) - r \geq v_n(e, x_H^*, x_L + 1) - v_n(e, x_H^*, x_L) - r \geq 0.$$

Hence, given (e, x_L), if the system rejects an asymptomatic patient when there are x_H^* symptomatic patients in the system, then it rejects an asymptomatic patient when there are more

symptomatic patients in the system. Since x_H^* is the smallest value that satisfies;

$$v_n(e, x_H^*, x_L + 1) - v_n(e, x_H^*, x_L) - r \geq 0,$$

Asymptomatic patients will be admitted for all values less than x_H^*. So we proved the existence of optimal threshold $l_H(e, x_L)$.

□

Now the following example illustrates the thresholds.

Example 3 We consider a system with two environments over an infinite horizon, where we set parameters as c_H=1, c_L=0.008, λ_L=11, $\lambda_H(2)$=1.2, $\lambda_H(1)$=0.36, μ=12, γ=1/3650, β=0, p=0.0005, r=0.5, s=0, and c=35100. The optimal policy is given in Figure 4.3. The thresholds for symptomatic patients can be easily derived from Figure 4.3: $l_H(1,0) = 6$, $l_H(1,1) = 5$, $l_H(1,2) = 4$, $l_H(1,3) = 3$, $l_H(1,4) = 2$, $l_H(1,5) = 1$, and $l_H(1,6) = 0$.

$Sup(e, x_L)$ property implies that if the system admits an asymptomatic patient in an environment e^* with a pair (x_H, x_L) where the proportion of symptomatic patients is high, then the system admits an asymptomatic patient in a better environment $e \leq e^*$. The operators, T_{ARR_H}, T_{ADM_i}, T_{DEP}, T_{DET}, T_{COST} and T_{UNIF} also preserve $Sup(e, x_L)$ property if $Sup(x_H, x_L)$ is preserved. Therefore, optimal policy is of threshold type for e as stated in the following theorem.

Theorem 2 *Given (x_H, x_L), there exists a threshold policy $l_e(x_H, x_L)$ on e such that if $e \geq l_e(x_H, x_L)$ it is optimal to reject the incoming asymptomatic patient, otherwise it is optimal to admit her.*

Proof. We define

$$l_e(x_H, x_L) = \arg\min\{e : v_n(e, x_H, x_L + 1) - v_n(e, x_H, x_L) - r \geq 0\},$$

and refer to $Sup(e, x_L)$ property. The rest is similar to the proof of Theorem 1.

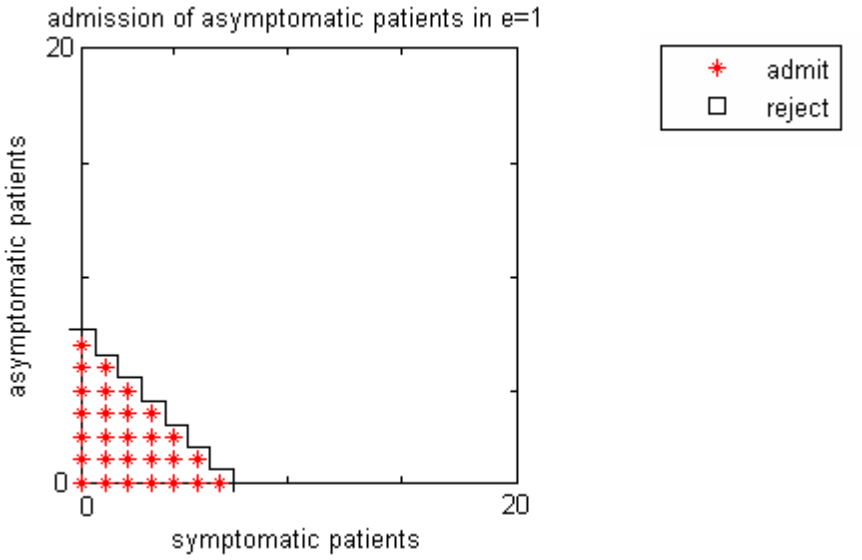


Figure 4.3: Illustration of optimal policies for the system in Example 3.

□

Let us give a representative example of the result.

Example 4 We consider a system with two environments over an infinite horizon, for which we set c_H=1, c_L=0.005, λ_L=11, $\lambda_H(2)$=1.2, $\lambda_H(1)$=0.36, μ=12, γ=1/3650, β=0, p=0.0005, r=0.5, s=0, and c=35100. Figure 4.4 shows the optimal policies. The triangle in the right figure covers the states in which asymptomatic patients are admitted to the system in $e = 2$. In a better environment namely $e = 1$, we observe that asymptomatic patients are also admitted in those states. More explicitly, if an asymptomatic patient is accepted in $e = 2$, then s/he will be accepted in $e = 1$.

Moreover, the operators, T_{ARR_H}, T_{ADM_i}, T_{DEP}, T_{DET}, T_{COST} and T_{UNIF} preserves convexity in x_L which guarantees the existence of an optimal threshold policy on x_L. An immediate

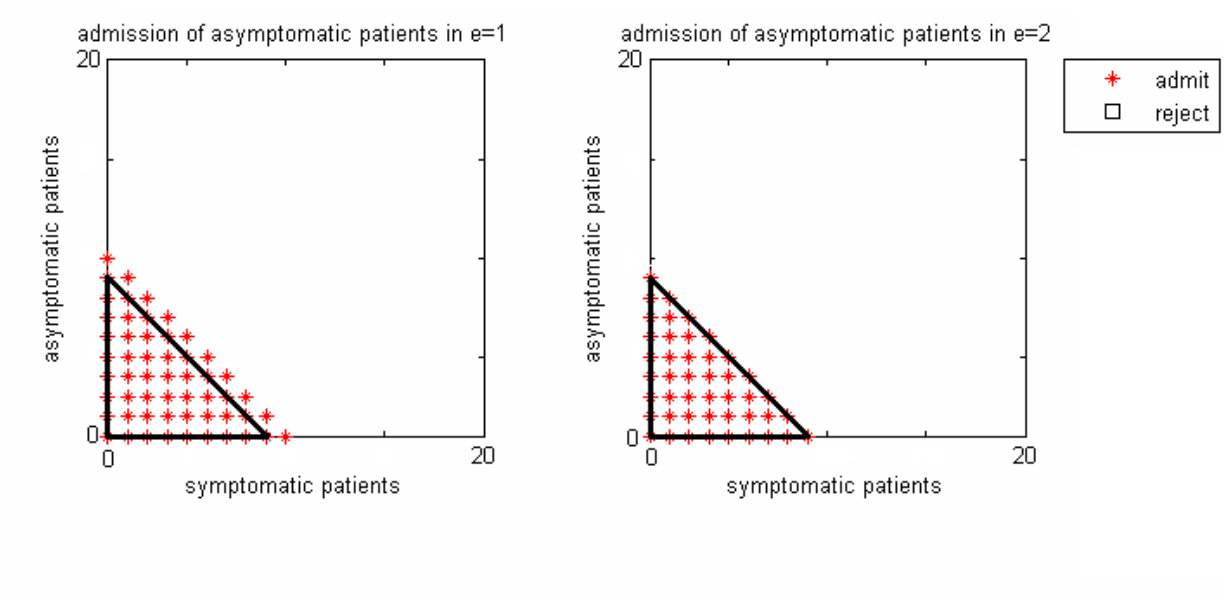


Figure 4.4: Illustration of optimal policies for the system in Example 4.

consequence of convexity can be expressed via thresholds on the number of asymptomatic patients in the system.

Theorem 3 *Given (e, x_H), there exists an optimal threshold $l_L(e, x_H)$ on x_L such that if $x_L \geq l_L(e, x_H)$ it is optimal to reject the incoming asymptomatic patient, otherwise it is optimal to admit her.*

Proof. We define

$$l_L(e, x_H) = \arg\min\{x_L : v_n(e, x_H, x_L + 1) - v_n(e, x_H, x_L) - r \geq 0\},$$

and refer to $Conv(x_H, x_L)$ property. The rest is similar to the proof of Theorem 1.

□

Let us visualize the result with an example.

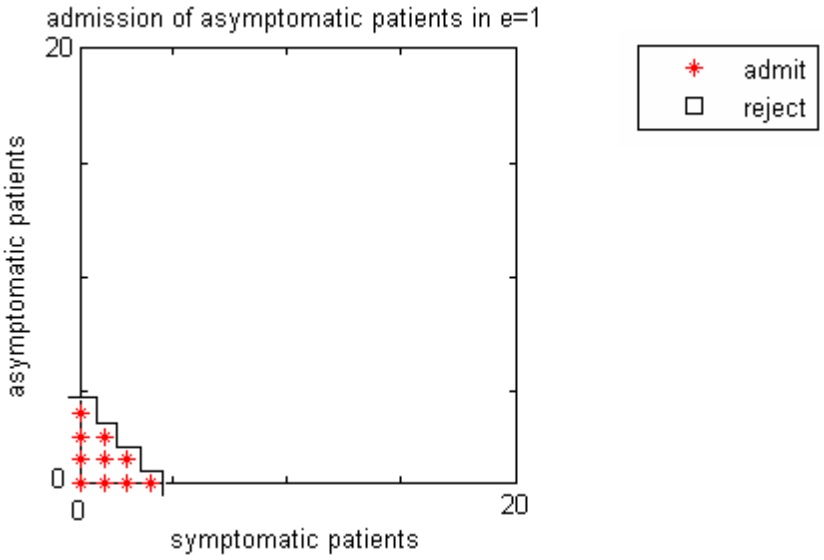


Figure 4.5: Illustration of optimal policies for the system in Example 5.

Example 5 We consider a system with two environments over an infinite horizon, where we set parameters as c_H=1, c_L=0.02, λ_L=11, $\lambda_H(2)$=1.2, $\lambda_H(1)$=0.36, μ=12, γ=1/3650, β=0, p=0.0005, r=0.5, s=0, and c=35100. The optimal policy is given in Figure 4.5. The thresholds for asymptomatic patients are as follows: $l_L(1,0) = 3$, $l_L(1,1) = 2$, $l_L(1,2) = 1$, and $l_L(1,3) = 0$.

4.2.1 Conclusion

In this chapter, we establish properties for a system which exercises both admission and scheduling controls. In particular, we show that if the prevalence of the disease is low, then it is optimal to always schedule symptomatic class first. However, if the demand from symptomatic class is high, i.e. the disease prevalence is high, and if screening is effective to reduce disease prevalence, then asymptomatic class can be given priority. Moreover, we show that in each state there exists a threshold on the risk groups, such that asymptomatic patients with higher

Both Controls	Admission Control	Scheduling Control
Increasing in x_H, x_L, e Monotonicity on the diagonal	Increasing in x_H, x_L, e Monotonicity on the diagonal	Increasing in x_H, x_L, e Monotonicity on the diagonal Threshold policies on x_H, x_L, e

Table 4.1: Summary of the results

risk than the threshold are admitted to the system, while the others are rejected. Then, we consider a system which always gives priority to the symptomatic patients for service. The system exercises its control only through the admission of asymptomatic patients. We show that optimal admission policies are of threshold type, where the thresholds are on the number of symptomatic patients, the number of asymptomatic patients, and on the environments. Table 4.1 summarizes the results.

Chapter 5

NUMERICAL RESULTS

In this chapter, we will discuss the sensitivity of optimal policies and performance measures to model parameters through numerical examples. Our objective is to present insights into how the system responds to the changes in the system parameters.

This chapter is structured as follows. Section 5.1 provides the set-up for numerical study. Section 5.2 explores the effect of system parameters on the optimal policies for systems exercising different controls. Section 5.3 investigates the impact of system parameters on the performance measures.

5.1 Set-up of the Numerical Study

The first step in analyzing the model described in the previous chapters is to create a base case scenario using realistic values for the parameters. We consider a case where we estimate the parameters of the system according to published data and statistical information from national agencies. In general, the model was implemented with data for the State of New Hampshire, obtained from [13]. We conduct a numerical analysis with two risk groups and two environments. Since colonoscopy cannot be performed on weekends, we assume that there are approximately 250 days in a year excluding weekends, so that μ and other parameters will be consistent.

We briefly explain how we set the parameters. On the average 1.73 colonoscopies can be performed in a day [13]. So, we take the service rate equal to 1.73. Therefore, $\mu = 1.73$ per day. In [13], the total population for a single server is approximately 4000, therefore we let population size be $N = 4000$. For the average risk population, colonoscopy is recommended for every ten years (World Health Organization [93]). Therefore, we assume that one-tenth of the population demand colonoscopy each year. With this information, we estimate the arrival rate

of asymtomatic patients to be 1.6 per day. So, we set $\lambda_L = 1.6$. From SEER (Surveillance, Epidemiology and End Results) data [53], we obtain that the incidence rate (new incidences in a year) by 2005 is 45.5 (per 100,000 person). Population size times incidence rate gives us the arrival rate of the symptomatic patient in the bad environment. Therefore, we estimate that the arrival rate of symptomatic patients in bad environment is 1.82 per day. Thus, $\lambda_H(2) = 1.82$. National Polyp Study [95], suggests that a periodic colonoscopy could prevent 76% to 90% of colon cancers. Being conservative, we assume that 70% of colon cancers will be prevented by periodic colonoscopy. Therefore, the arrival rate of the symptomatic patient in the good environment is equal to 30% of the arrival rate of the symptomatic patient in the bad environment. With this information, we estimate arrival rate of symptomatic patients in good environment to be 0.55 per day, $\lambda_H(1) = 0.55$. According to World Health Organization [93], for an effective screening test, 70% of the population at risk needs to be screened. With this information, we further approximate the probability of improving the environment upon screening an asymptomatic patient to be 0.000357 ($p = 0.000357$) from the expression $1/(N(1 - \text{incidence rate})0.7)$. Since colonoscopy should be repeated every ten years, the effectiveness of screening lasts for ten years. Therefore, we assume that the expected duration in good environment is 10 years. Thus, deterioriation rate of environment is 0.0004 per day, $\gamma = 0.0004$.

Brown [11] estimates that the average treatment cost for colorectal cancer is approximately \$35,100. So we let treatment cost of symptomatic patient to be 35,100 ($c = 35,100$). Further, we let rejection cost be $r = 0.5 > 0$ since it is important for characterizing optimal admission policy. However, we let screening costs be $s = 0$ because it is the same for both patient types. Since delaying symptomatic patients is more costly than asymptomatic patients, we let $c_H = 1$, and $c_L = 0.02$. The base case parameter values are listed in Table 5.1.

In this chapter, we focus on four systems with different control mechanisms defined in Table 5.2.

The solution algorithm for the presented models were programmed in Matlab Version 7.0. We compute the optimal policies numerically by the relative value iteration algorithm by allowing a maximum of 100 patients in the system.

Parameters	**Definitions**
$\lambda_L = 1.6$	The arrival rate of asymptomatic patient
$\lambda_H(1) = 0.55$	The arrival rate of symptomatic patient in environment 1
$\lambda_H(2) = 1.82$	The arrival rate of symptomatic patient in environment 2
$\mu = 1.73$	Service rate
$p = 0.000357$	Probability of moving from worse environment to best environment upon screening a asymptomatic patient
$\gamma = 0.0004$	Deteriorating rate of environment
$\beta = 0$	Discount (exponential failure) rate
$c = 35100$	Cost of a symptomatic patient
$c_H = 1$	Holding cost of a symptomatic patient
$c_L = 0.02$	Holding cost of an asymptomatic patient
$r = 0.5$	Rejection cost for an asymptomatic patient
$s = 0$	Screening cost

Table 5.1: Set up values of the numerical problem

Policy	Description
Model B	System exercising both controls
Model S	System exercising only scheduling control
Model A	System exercising only admission control
Model N	System does not exercise any control

Table 5.2: Models with description

For sensitivity analysis, we focus on the system parameters μ and p. The service rate, μ, for the base case is 1.73, which is insufficient for the total demand in both environments ($\lambda_L + \lambda_H(2) > \lambda_L + \lambda_H(1) > \mu$). We set $\mu = 2.151$ which represents the case when the service capacity is barely sufficient for total demand in environment 1. Setting μ=2.5 and 3, we analyze the system when the service capacity is sufficient in environment 1, but not in environment 2. Finally, we take $\mu = 3.5$ which represents the sufficient capacity in both environments. For the p values we have 0.001 and 0.01 in addition to the base case $p = 0.000357$, in order to capture the effect of screening.

We present varying optimal policies in a two dimensional graph. The lines on the graphs represent thresholds. The system admits asymptomatic patients if the number of the patients (asymptomatic and symptomatic) in the system is less than threshold value for admission policy. For scheduling policy, asymptomatic patients obtain priority in any state in which the number

of patients is less than the one that are determined by thresholds.

In this chapter, we scale down the graphs for visual simplicity. The dimensions of the graphs may not be equal. For example, in Figure 5.2, $x_H \leq 10$, and $x_L \leq 10$, in the admission for $e = 1$, $x_H \leq 100$, and $x_L \leq 10$, for $e = 2$, and finally for scheduling $x_H \leq 100$, and $x_L \leq 100$. Moreover, we exclude the graphs of scheduling policy in $e = 1$ since we have shown that the system schedules symptomatic patients first.

5.1.1 Optimal Policies of the Base Case

We find the optimal admission and scheduling policy for the Model B. In this case, asymptomatic patients are never admitted. Symptomatic patients obtain priority for scheduling in both environments. In the base case, the holding cost ratio (c_H/c_L) is 50. If we increase this ratio by decreasing the holding cost of asymptomatic patients, then the opimal admission policy changes. Figure 5.1 summarizes the results for this case. The two graphs at the top of Figure 5.1 present optimal admission policy. We observe that when number of patients in the system is low, asymptomatic patients are admitted in $e = 1$. When the holding cost of an asymptomatic patients is lower, the system decides to admit asymptomatic patients since they bring less burden on the system. The two graphs at the bottom of Figure 5.1 provides the optimal scheduling policy which shows that symptomatic patients obtain priority for both environments. In Model S, symptomatic patients obtain priority in $e = 2$, and in Model A, the asymptomatic patients are never admitted in both environments.

The reason why symptomatic patients are scheduled first is; there is an insufficient capacity of colonoscopy service and estimated p value is not high enough to improve the environment upon screening an asymptomatic patient, so the system chooses to screen the patients with colorectal cancer. In health care systems, there is a prioritization in the urgent cases considering patients with colorectal cancer. Further, The New York City Department of Health and Mental Hygiene [8] recommends to prioritize symptomatic patients in order to allocate limited colonoscopy based on need and use time efficiently. Thus, the current practice of allocating resources is consistent with our results depending on the estimated parameters.

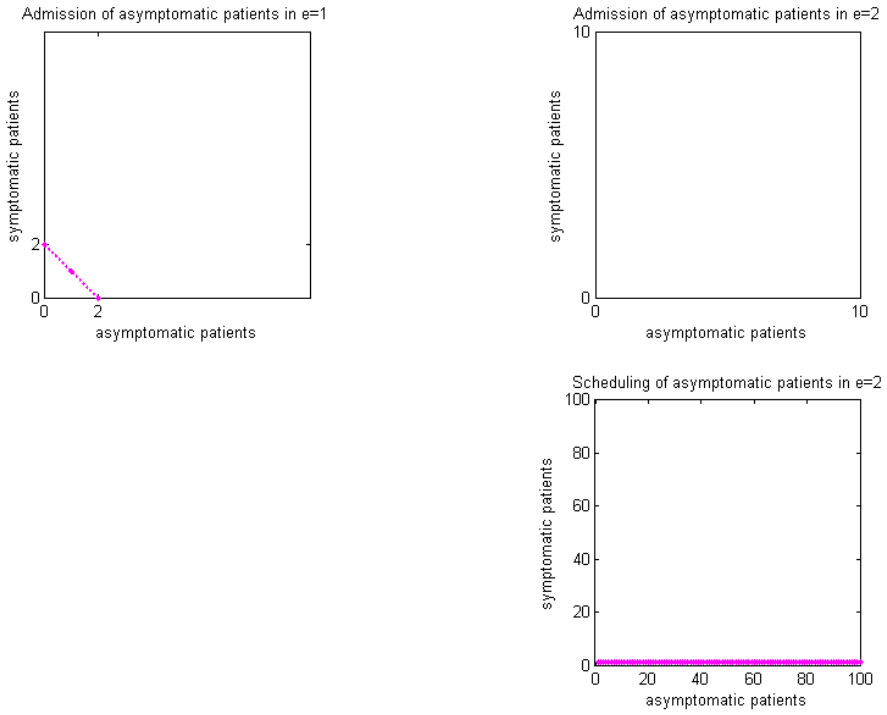


Figure 5.1: Optimal policies for $c_L = 0.005$ for Model B.

5.2 Policies

In this section, we analyze how optimal policies change as the model parameters vary.

5.2.1 Model B

First we study the effects of system parameters on the optimal policy for the model in which both admission and scheduling decisions are available.

5.2.1.1 Sensitivity of p on Model B

We consider the effect of p on the optimal policy. Figure 5.2 summarizes the optimal admission and scheduling policies for the p values. Note that the admission regions in $e = 1$ are the

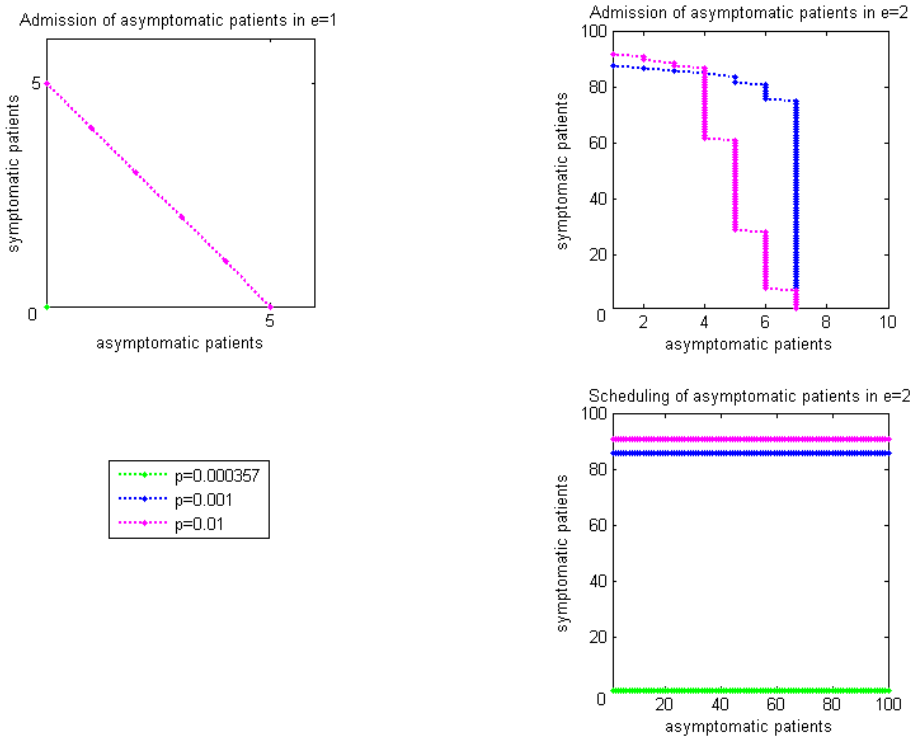


Figure 5.2: Optimal policies with corresponding p values for Model B.

same for the cases $p = 0.01$ and $p = 0.001$. The optimal admission policies are of threshold type, where the thresholds are on the number of symptomatic and asymptomatic patients. We observe that when the number of patients in the system is low, asymptomatic patients are admitted. For $e = 2$, we observe that if there is a high probability of improving the environment by giving priority to asymptomatic patients, then the system serves asymptomatic patients in order to decrease the arrival rate of symptomatic patients in the long run. In particular, asymptomatic patients are admitted in $e = 2$, if the number of asymptomatic patients is relatively low in the system. However, we do not observe any monotonic behaviour in optimal admission policy for $e = 2$.

5.2.1.2 Sensitivity of μ on Model B

Next, we analyze the effect of μ, which is significant for both controls (See Figure 5.3). Recall that in the good environment, symptomatic patients are scheduled first. As service rate increases, the system is more likely to reach any state $(1, 0, x_L) \in S$. Therefore, the system wants to schedule asymptomatic patients as well. Hence it admits more asymptomatic patients as μ increases. As can be observed in Figure 5.3, thresholds for admission region in $e = 1$ increases with μ.

For the bad environment, the increase in service rate results in an increase in the number of states, where asymptomatic patients obtain priority. As we schedule asymptomatic patients in more states, we also admit more asymptomatic patients especially when the number of asymptomatic patients is low in the system.

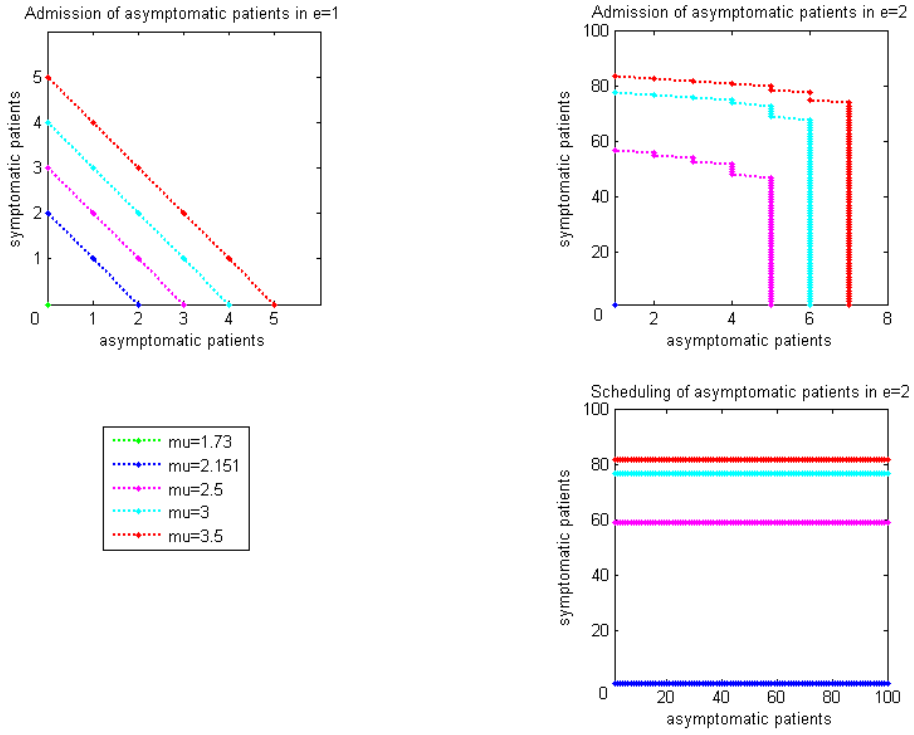


Figure 5.3: Optimal policies with corresponding service rates for Model B.

5.2.2 Model S

We consider the system exercising scheduling control. In this system, asymptomatic patients are always admitted. Thus, we provide only optimal scheduling policy in $e = 2$.

5.2.2.1 Sensitivity of p on Model S

We will examine the effect of p on the optimal scheduling policy. Figure 5.4 reports the

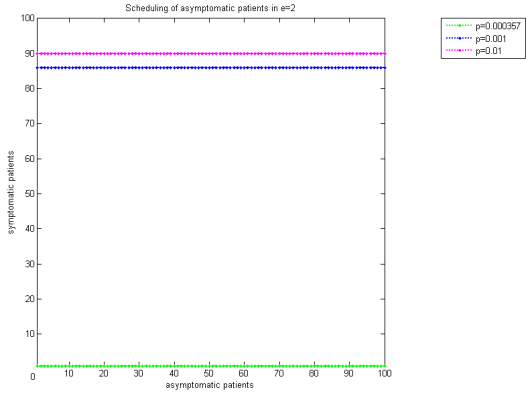


Figure 5.4: Optimal policies corresponding to p values for Model S.

results on scheduling control. We can conclude that, as p increases, the system serves more asymptomatic patients. Figure 5.2 and Figure 5.4 are similar in the scheduling decision in $e = 2$. In particular, we can say that the admission decision does not affect the scheduling policy significantly.

5.2.2.2 Sensitivity of μ on Model S

Now, we analyze the effect of μ on scheduling decision. As indicated in Figure 5.5, if the number of symptomatic patients are high, the system serves them first. Otherwise, the system gives priority to asymptomatic patients. As μ increases, the system serves asymptomatic patients

in more states. We can also conclude that the effect of μ on the optimal policies is similar in both models considered so far.

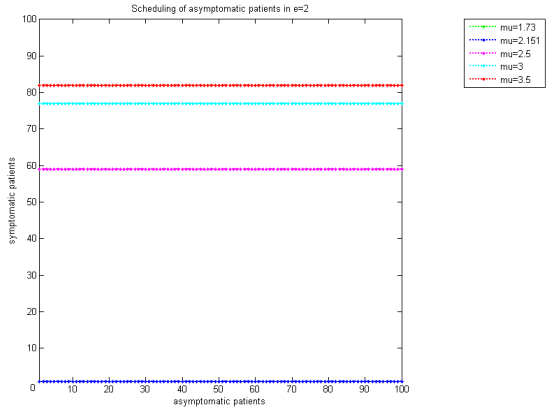


Figure 5.5: Optimal policies corresponding to service rates for Model S.

5.2.3 Model A

Finally, we consider the model where symptomatic patients are always scheduled first in both environments. Thus, we do not provide graphs representing scheduling policies.

5.2.3.1 Sensitivity of p on Model A

We investigate the impact of p on the optimal admission policy. Figure 5.6 shows the optimal admission policies. Note that the admission regions in both environments are the same for the cases $p = 0.01$ and $p = 0.001$. This figure shows the significant effect of scheduling policies on admission decisions, as opposed to the insignificant effect of admission on scheduling. If the system prioritizes symptomatic patients, then only very few asymptomatic patients are admitted in $e = 2$. Figure 5.6 differs significantly from Figure 5.2 which shows the sensitivity of admission policies to scheduling policies with varying p values.

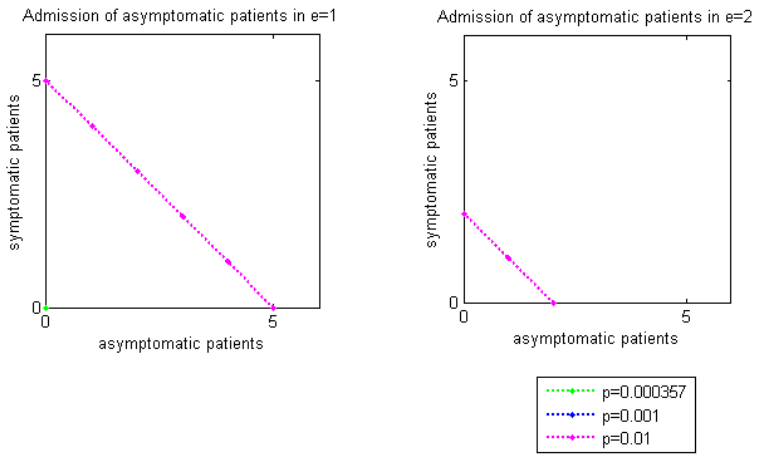


Figure 5.6: Optimal policies corresponding to p values for Model A.

5.2.3.2 Sensitivity of μ on Model A

We compare the effects of μ on Model A (see Figure 5.7) and Model B (see Figure 5.3). While the admission region of asymptomatic patients in $e = 1$ remains unchanged, this region is significantly smaller in $e = 2$ in Model A when compared to that in Model B. This shows that the interaction of two control mechanisms can be quite high. When symptomatic patients have priority, then the system does not choose to admit asymptomatic patients.

5.3 Performance Measures

In this section, we will conduct a computational study to discuss the sensitivity of performance measures to model parameters. Typical performance measures are the number of patients in the system and the proportion of time the system stays in both environments, which affects long run arrival rate of symptomatic patients in the long run.

Firstly, we will explore the effect of μ on all performance measures. We begin with the

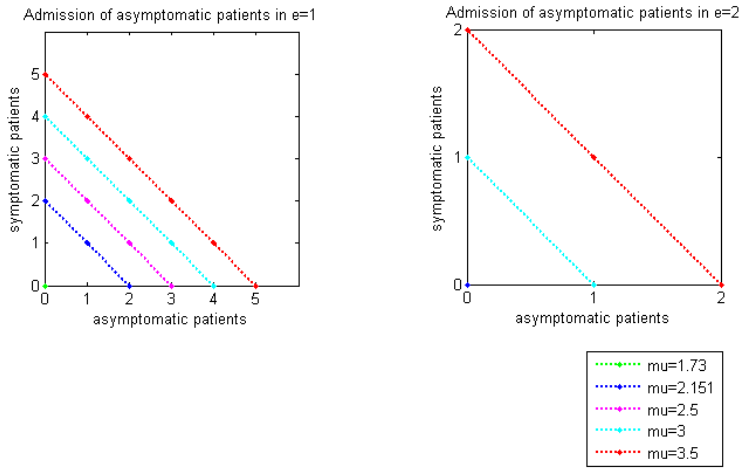


Figure 5.7: Optimal policies corresponding to μ values for Model A.

effective arrival rate of symptomatic patients. We define it as follows:

$$\lambda_{\text{eff}} = \pi_1 \lambda_H(1) + \pi_2 \lambda_H(2),$$

where π_e is the long-run proportion of time the system stays in e. The results are stated in Figure 5.8.

When the capacity is insufficient and p is relatively low, we observe that all models offer same λ_{eff} (for $\mu \leq 2.1$). For $\mu \leq 2.1$, symptomatic patients obtain priority and $\pi_e, 1 \leq e \leq 2$ does not differ among the models. For $3.4 \geq \mu \geq 2.1$, we have $\lambda_L + \lambda_H(2) \geq \mu \geq \lambda_L + \lambda_H(1)$ so that the capacity can be sufficient in some cases. Therefore, optimal scheduling policy changes and the asymptomatic patients begin to obtain priority (refer to Figure 5.3). Model B and Model S provide slightly lower λ_{eff} than Model A and Model N do, since they screen asymptomatic patients which will improve the health of the population and increase the amount of time spent in $e = 1$, π_1. For $3.4 \leq \mu$, the difference in λ_{eff} increases between the models B, S and A, N, since the capacity is sufficient in all cases. We note that Model B and S perform better in

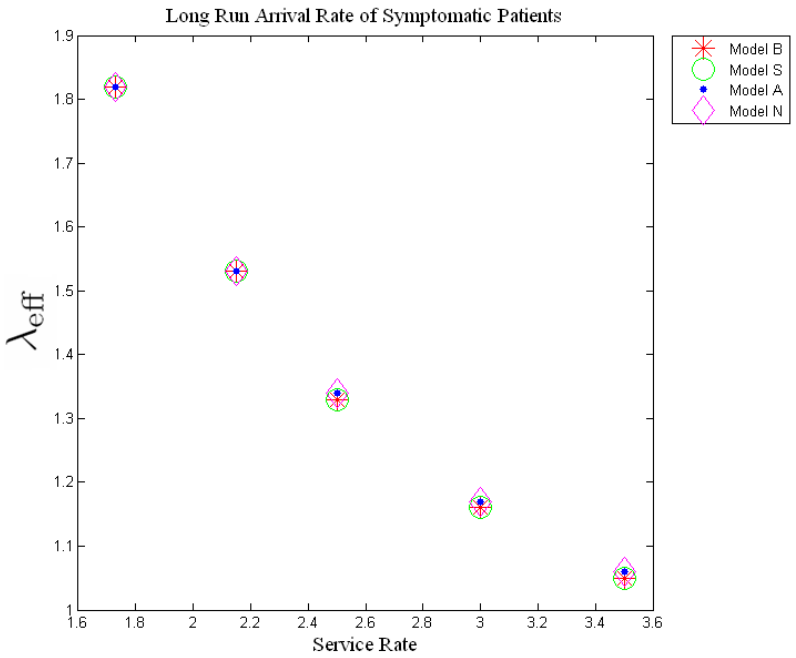


Figure 5.8: The effective arrival rate of symptomatic patients, λ_{eff} for μ values for p=0.000357 .

improving the health of poulation in the long run. Hence we can conclude that scheduling is a more effective control.

We continue with the long run average number of symptomatic patients in the system. Figure 5.9 presents the results for varying service rates.

Since the scheduling policy is to give priority to symptomatic patients, Model A and N always give priority to symptomatic patients. As μ increases, more symptomatic patients receive service. Once they are served, they leave the system. Therefore, the number of symptomatic patients decreases. The behaviour of Model S and B are similar to Model A and N for $\mu \leq 2.151$, (Recall that for $\mu = 1.73$ and $\mu = 2.1$, symptomatic patients obtain priority). For $\mu \geq 2.1$, since the asymptomatic patiens starts to obtain priority in Model S and B (refer to Figure 5.3), the number of symptomatic patients in the system increases. When $\mu \geq \lambda_L + \lambda_H(2)$ ($\mu \geq 3.4$), the capacity is sufficient for all types of patients and in both environments which decreases the

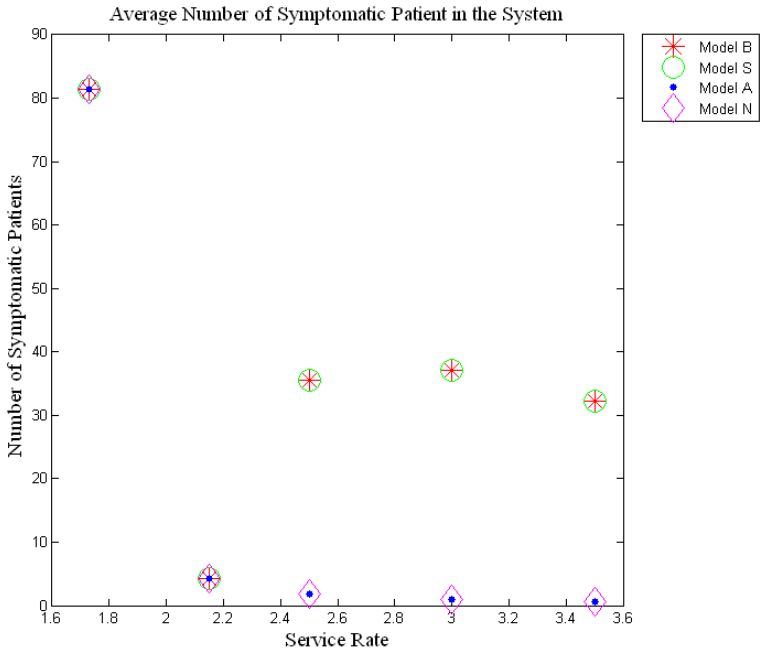


Figure 5.9: Long-run average number of symptomatic patients for μ values for p=0.000357.

number of symptomatic patients in the system decreases.

Figure 5.10 shows the long-run average number of asymptomatic patients. Model N and Model S admit all asymptomatic patients. Therefore the number of asymptomatic patients are high in these models. As μ increases from 1.73 to 2.1, the system is more likely to reach any state $(e, 0, x_L) \in S$ where asymptomatic patients will be served. Hence the number of asymptomatic patients decreases in μ. For $\mu \geq 2.1$, the Model N and Model S begin to differ, since asymptomatic patients begin to obtain priority in optimal scheduling. Therefore, the average number of asymptomatic patients in Model S is less than that in Model N. Additionally, the number of asymptomatic patients increases slightly for the remaining models (Model A and B) since as μ increases, the systems admit more asymptomatic patients.

Now we will focus on the effect of p on performance measures. Figure 5.11 presents the results.

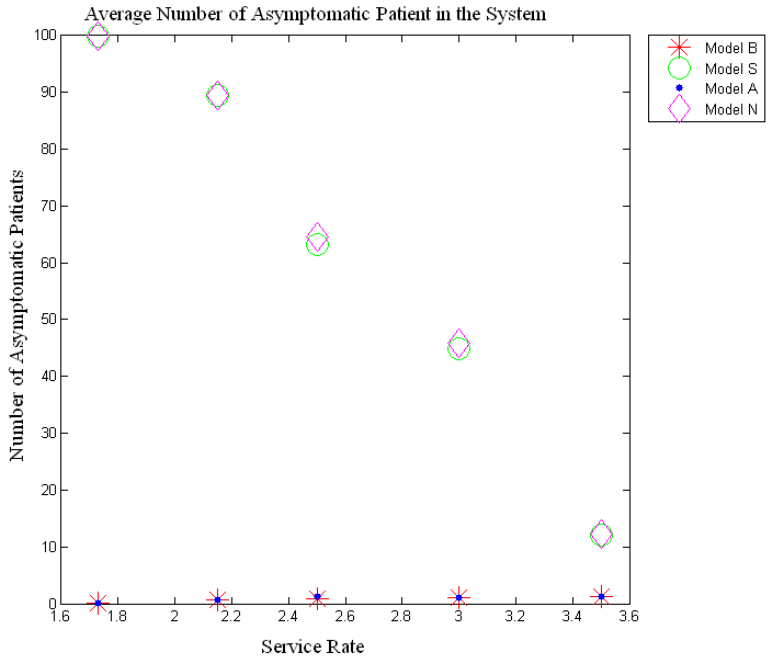


Figure 5.10: Long-run average number of asymptomatic patients for μ values for p=0.000357.

Recall that in this analysis, $\mu = 1.73$ which means there is insufficient capacity. We observe that in addition to insufficient capacity, if p values are low, all models perform almost the same with a high λ_{eff}. As p increases to 0.01, the effectiveness of screening increases. λ_{eff} decreases only a little for the Model A and N because these systems give priority to symptomatic patients. However, for the Models B and S, λ_{eff} decreases significantly, since these systems give priority to asymptomatic patients and we observe the effect of screening which improves the health of the population. Screening high risk groups will have a higher p value; hence it is important to identify the risk groups and screen them rather than screening the whole population.

Figure 5.12 shows the effect of p on the long-run average number of symptomatic patients in the system. With small p and insufficient capacity, all models perform the same. When p increases, we observe the effect of screening in Models B and S. High p improves the environment faster, and decreases the number of symptomatic patients in the population. Since the effective

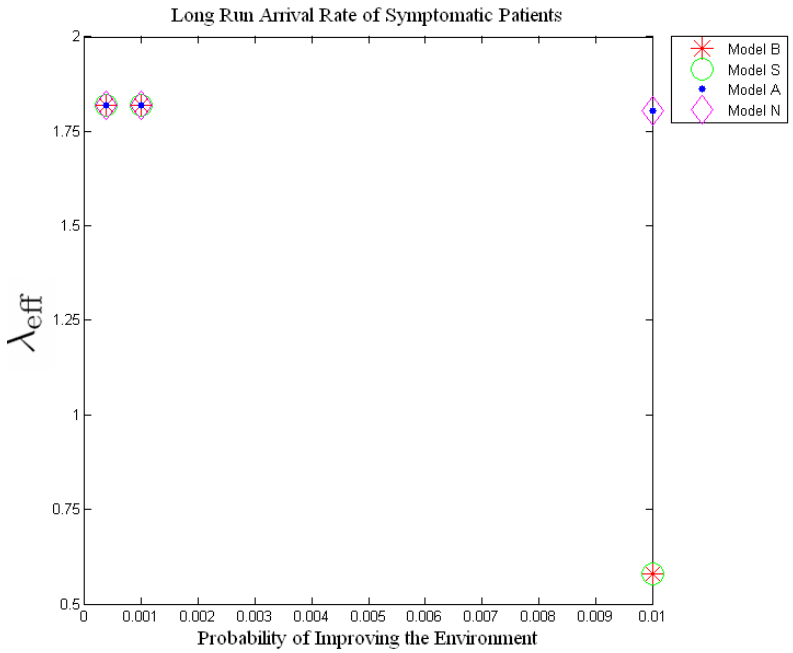


Figure 5.11: The effective arrival rate of symptomatic patients, λ_{eff}, for p values for $\mu = 1.73$.

arrival rate of symptomatic patients decreases, less symptomatic patients demand colonoscopy which implies a reduction in the number of symptomatic patients in the system. Even though Model A and N prioritizes symptomatic patients, the number of symptomatic patients is much higher when compared to Model B and S showing the effectiveness of screening.

Next, we analyze the effect of p on the long run average number of asymptomatic patients in the system. As seen in Figure 5.13, the number of asymptomatic patients are high in Model N and S for low p values. As p increases, Model N has the same number of asymptomatic patients in the system, since it is not subject to any control. In Model S, the number of asymptomatic patients decreases, because as p increases asymptomatic patients gain priority. For Model B and A, the asymptomatic patients are admitted to the system in a fewer number of states. Therefore, the values do not vary as much.

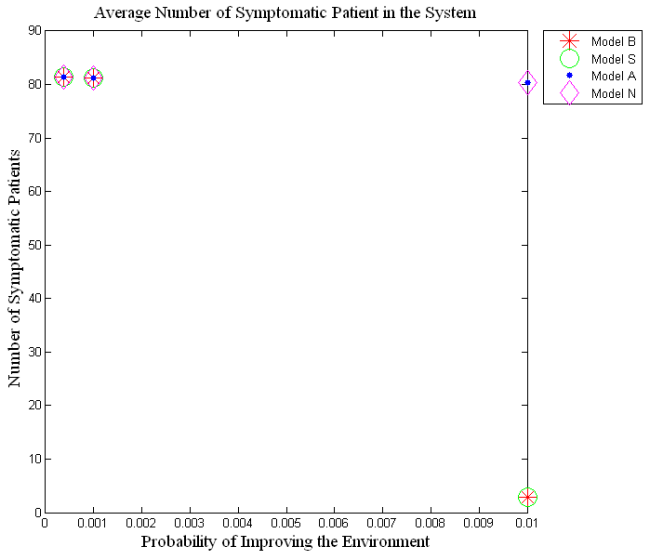


Figure 5.12: Long-run average number of symptomatic patients for p values for $\mu = 1.73$.

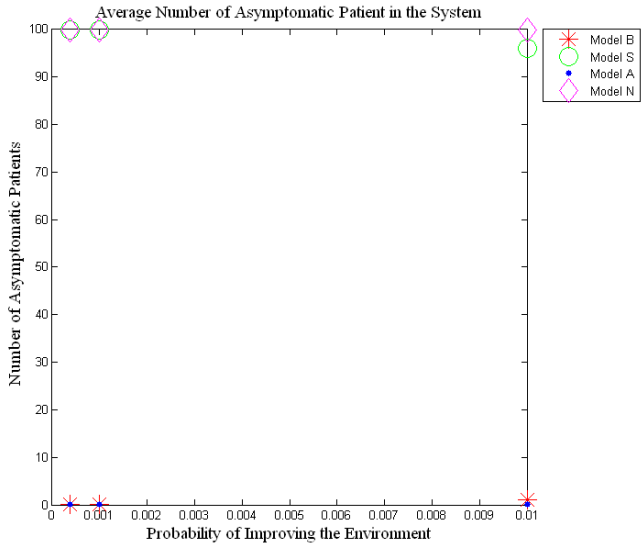


Figure 5.13: Long-run average number of asymptomatic patients for p values for $\mu = 1.73$.

5.3.1 Conclusion

In this chapter, we aim to determine the effects of factors on the optimal policies and performance measures. In general, increasing μ and p increase the number of states in which asymptomatic patients are admitted and the number of states in which asymptomatic patients are scheduled. However, admission region and scheduling policy may not be monotone. We observe that admission policy does not affect the optimal scheduling policy, but scheduling policy has a significant influence on admission. Indeed, focusing on scheduling control and exploring the effective ways of capacity allocation rather than admission control is advantageous for preventive services. Hence, in practice only scheduling can be considered.

For low μ and p, the performance of the models do not deviate from each other significantly. For low p and high μ, Model A and N have lower symptomatic patients in the system than Models B and S due to prioritizing of symptomatic patients and the low long-run effect of p. Model A and B have lower asymptomatic patient in the system than Model N and S due to admission. For low p, the models induce rather close effective arrival rates. The models start differing when p increases. High p promises an improvement in the health level of the population. Therefore we would like to increase p. One way of this is to screen proper risk groups. If there is a patient who is a potential candidate for colorectal cancer in future, screening him is more valuable than screening a patient with relatively low risk. Therefore, classifying risk groups accurately is crucial and will reduce the health costs of the population.

Chapter 6

DYNAMIC COMPARTMENTAL MODEL FOR COLORECTAL CANCER SCREENING

6.1 Introduction

In this chapter, we will investigate the effects of operational controls on population dynamics. Our analysis is based on a dynamic compartmental model of the colorectal cancer. Dynamic compartmental models are widely used to model the progression of the diseases. Compartments are the homogenous, disjoint groups of the population, which are defined by stage of the cancer and screening process in that case. For screening programs, most of the research focuses on three-stage health models (healthy, early stage, late stage) where they ignore the stages which will influence the disease progress [4]. A good example of these stages is polyps for colorectal cancer. During colonoscopy, polyps if there are any, are removed, which shows the effectiveness of the screening programs. If we do not include that process, then we will not be able to model the interaction between the colorectal cancer development and screening processes. Therefore, in this chapter, we consider a four-stage health model. We create four scenarios where the difference arises due to the allocation of the resources. We present a numerical analysis that compares the performances of the different scheduling strategies and explore the effect of system parameters in detail.

6.2 Model Description

Our analysis is based on a fluid model of colorectal cancer in a population of individuals aged 50 to 80. A schematic of the model is presented in Figure 6.1.

Colorectal cancer has four stages depending on the progress of the disease. Stage I is the least advanced and stage IV is the most advanced stage. We assume that preclinical (early

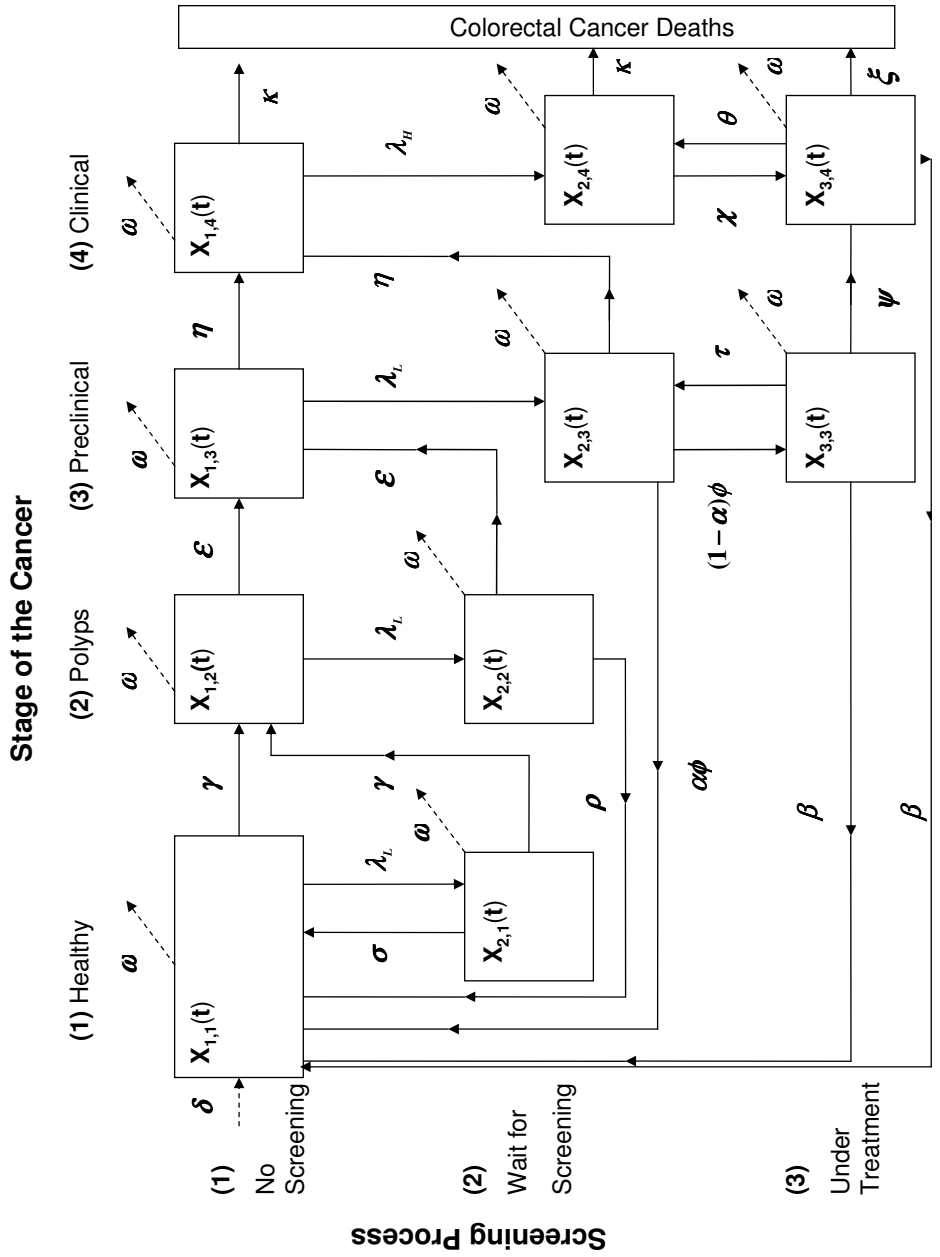


Figure 6.1: Fluid Model

stage) cancer includes stages I, II, and clinical (late stage) cancer includes stages III, IV in this model. We assume a four-stage health model: (1) healthy, (2) with polyps, (3) preclinical (early stage), colorectal cancer (4) clinical (late stage), colorectal cancer. Moreover, we have three groups in the screening process: (1) no screening, (2) waiting for screening, (3) under treatment The number of individuals in the stage of cancer $j = 1, 2, 3, 4$ and screening process state $i = 1, 2, 3$ is denoted $X_{i,j}$. This results in a model with 10 compartments. Healthy people, individuals with polyps, individuals having preclinical cancer and individuals having clinical cancer are represented by $\sum_i X_{i,1}$,$\sum_i X_{i,2}$,$\sum_i X_{i,3}$ and $\sum_i X_{i,4}$ respectively. We assume 100% sensitivity of the colonoscopy, and the polyps are removed during screening service. Therefore, healthy people and individuals having polyps do not go under treatment after being screened. Thus, we do not define the compartments $X_{3,1}$ and $X_{3,2}$. Individual flows through the service process are represented vertically in Figure 6.1, and changes in cancer stages are illustrated with horizontal flows from compartment to compartment.

In our system, we consider healthy people, individuals with colon polyps and preclinical cancer as low risk patients. Low risk patients demand screening service with rate λ_L. Healthy people will develop colon polyps with rate γ. Colon polyps can transform to early stage colorectal cancer with a rate ϵ. Further, rate of cancer advancing from early stage to late stage is η. Patients having colorectal cancer demand service with rate λ_H. Healthy people, individuals with colon polyps, preclinical and clinical cancer will complete the service with rate σ, ρ, ϕ and χ respectively. These rates allow us to model different scheduling policies. We assume 100% sensitivity of colonoscopy. α % of the preclinical patients are cured and become healthy after the screening process, and $(1 - \alpha)$ % of them continue to treatment. After being screened, an individual with preclinical cancer who continues treatment demands service again with rate τ. Also, the disease may progress in spite of the treatment and the patient will undergo the same treatment with clinical patients with rate ψ. Further, we assume that individuals with preclinical and clinical cancer become healthy if they survive ten years during treatment. If there is no recurrence in a ten year time interval, then the patient is considered as healthy person [20] and this occurs with rate β. Patients with clinical cancer demand service again with rate θ after the

first service.

Exit out of the population from any compartment occurs via maturation of 80-year-olds with rate ω. Colorectal cancer related deaths are assumed to affect only the individuals with clinical cancer $X_{1,4}$, $X_{2,4}$, and $X_{3,4}$. Individuals, who are not under treatment, can leave the system with rate κ. If they are under treatment, then they can die and leave the population with rate ξ. We ignore the non-cancer related deaths. Entry into the population occurs with rate δ due to the maturation of 49-year-olds. We assume a constant population so that a new individual enters the target population as an exit occurs. Therefore we assume that the rate of entering to the population is equal to the rate of leaving the population due to aging plus mortality rate, $X_{mortality}$, where we will define mortality rate in the next section. The population dynamics with these assumptions is represented by the following differential equations:

$$\begin{aligned}
\frac{dX_{1,1}(t)}{dt} &= \omega \sum_{j=1}^{4}\sum_{i=1}^{3} X_{i,j}(t) + X_{mortality} + \sigma X_{2,1}(t) + \rho X_{2,2}(t) \\
&\quad + \alpha\phi X_{2,3}(t) + \beta(X_{3,3}(t) + X_{3,4}(t)) - (\gamma + \lambda_L + \omega)X_{1,1}(t) = 0 \\
\frac{dX_{1,2}(t)}{dt} &= \gamma(X_{1,1}(t) + X_{2,1}(t)) - (\lambda_L + \epsilon + \omega)X_{1,2}(t) = 0 \\
\frac{dX_{1,3}(t)}{dt} &= \epsilon(X_{1,2}(t) + X_{2,2}(t)) - (\lambda_L + \eta + \omega)X_{1,3}(t) = 0 \\
\frac{dX_{1,4}(t)}{dt} &= \eta(X_{1,3}(t) + X_{2,3}(t)) - (\lambda_H + \omega + \kappa)X_{1,4}(t) = 0 \\
\frac{dX_{2,1}(t)}{dt} &= \lambda_L X_{1,1}(t) - (\sigma + \gamma + \omega)X_{2,1}(t) = 0 \\
\frac{dX_{2,2}(t)}{dt} &= \lambda_L X_{1,2}(t) - (\rho + \epsilon + \omega)X_{2,2}(t) = 0
\end{aligned}$$

$$\frac{dX_{2,3}(t)}{dt} = \lambda_L X_{1,3}(t) + \tau X_{3,3}(t) - (\phi + \eta + \omega)X_{2,3}(t) = 0$$

$$\frac{dX_{2,4}(t)}{dt} = \lambda_H X_{1,4}(t) + \theta X_{3,4}(t) - (\chi + \omega + \kappa)X_8(t) = 0$$

$$\frac{dX_{3,3}(t)}{dt} = (1-\alpha)\phi X_{2,3}(t) - (\beta + \tau + \psi + \omega)X_9(t) = 0$$

$$\frac{dX_{3,4}(t)}{dt} = \chi X_{2,4}(t) + \psi X_{3,3}(t) - (\theta + \beta + \omega + \xi)X_{3,4}(t) = 0$$

We focus on different policies and compare the effects of these policies. We create four cases where the difference arises due to the allocation of the resources. The policies includes (a) priority for high risk patients, (b) priority for low risk patients, (c) no screening for low risk patients and, (d) rationing of resources among high and low risk patients. Before moving to policies, we will estimate the values of parameters based on recent available data in order to make accurate analyses for real life problems.

6.3 Parameter Estimation

The parameters are estimated from medical journal articles or national statistical publications. Where that was not possible, we made reasonable assumptions so that the simulation output was of the same magnitude as real life cases. The parameter values corresponding to our base case described by Table 6.1.

For people over 50, it has been recommended to have a colonoscopy every ten years [77]. We assume 100% compliance for colonoscopy. Hence, the arrival rate of low risk patients is 0.1 per year, λ_L=0.1. We assume that high risk patient demands service within a month on the average, therefore λ_H=12 per year. In previous chapter, service rate is estimated as 1.73 per day where the population size is 4,000. For a simulation model, this population size is not enough to make accurate estimations. Therefore, service rate is scaled up to 21 per day for a targeted population of 50,000 people.

Parameters	**Description**
$\lambda_L = 0.1$	arrival rate/person/year for screening from low risk group
$\lambda_H = 12$	arrival rate/person/year for screening from high risk group
$\mu = 5,412$ (21 per day)	service rate/person/year
$\omega = 0.033$	maturation rate/person/year out of the population
$\gamma = 0.004$	rate/person/year of developing colon cancer
$\beta = 0.058$	treatment completion rate/person/year after early diagnosis
$\kappa = 0.29$	death rate/person/year from late stage cancer without treatment
$\alpha = 0.45$	percentage of to be cured in preclinical stage
$\epsilon = 0.1$	rate/person/year of acquiring preclinical cancer
$\eta = 0.2$	rate/person/year of cancer advancing from preclinical to clinical stage
$\psi = 0.04$	rate/person/year of cancer advancing from preclinical to clinical stage during treatment
$\theta = 0.5$	rate/person/year for re-screening, from preclinical stage
$\tau = 0.24$	rate/person/year for re-screening, from clinical stage
$\xi = 0.1$	death rate/person/year from late stage cancer during treatment

Table 6.1: Base case problem parameters

Since the targeted age interval is 30 years and after 30 years a person leaves the population, we assume $\omega = 1/30$ per year. We estimate ψ to be 0.04 per year by fitting the relative survival rates for early stage cancer, assuming an exponential distribution for the time to be cancer. We assume that expected remaining lifetime of a patient with clinical cancer who undergoes treatment is 10 years, therefore we let $\xi = 0.1$ [9]. $X_{1,4}$ and $X_{2,4}$ includes stage III type cancer patients as well as stage IV type (distal) cancer patients. Mean survival rate for distal cancer is 1.9 years [52]. Therefore, the expected remaining lifetime for people in compartments $X_{1,4}$ and $X_{2,4}$ should be greater than 1.9. Also it should be less than $1/\xi$ since the probability of dying of cancer without treatment is relatively high. With respect to this information, we estimate κ to be 0.29 per year (3.5 years of survival). We estimate $\beta = 0.058$ per year assuming an exponential survival process [48]. Colorectal cancer is curable more than 90 % of the time if it is detected early (at stage I) according to American Gastroenterological Association [3]. Since the compartment $X_{2,4}$ includes two stages (I, II), we take α=0.45. On the average, it takes 10 years for a colon polyp to transform into a cancer [94]. So ϵ=0.1 per year. Further, an early stage

cancer turns into a late stage cancer in approximately 3 years in a high risk group [96]. Since we assume an average risk group population, we take $\eta = 1/5$. According to American Cancer Society [1], colonoscopy is recommended within a year after surgery and it should be repeated in three years. Then, it will be followed by repeat examinations at 5-year intervals. Since the expected remaining lifetime is 25 years for early stage cancer, after being screened, patients with colon cancer demand 6/25 service in a year. Hence, we take θ=6/25. We set τ=1/2 per year since colonoscopy is recommended once every two years for high risk population [78]. We can not reach any data that give an insight about the rate of developing colon polyps. Therefore we estimate it in order to obtain model output close to the real data. We assume that polyp developing rate is greater than incidence rate (which is 157.02 per 100,000 person), and let it be 400 per 100,000 person. Assuming an exponential distribution for developing polyps, we set $\gamma = 0.004$.

The model was validated against the incidence and mortality data given by National Cancer Institute [53]. Table 6.2 presents the 5-year age specific incidence and mortality rates. We calculated the incidence rate for people between 50-80 ages as 157.02 per 100,000 person by taking the average of 5-year age specific incidence rates. In our model, we assume that incidence rate is determined by two factors. Firstly, it includes the number of patients who die from colorectal cancer without any diagnosis. It also includes the number of patients who are diagnosed after service. However, we do not count the number of patients who are under treatment and demand colonoscopy in order to prevent double counting. We define incidence rate as

$$\frac{dX_{incidence}(t)}{dt} = (\phi - \tau)X_{2,3} + (\chi - \theta)X_{2,4} + \kappa(X_{1,4} + X_{2,4}).$$

Moreover, we use this parameter as one of our performance measures in the following sections. Death rate estimations are done taking the average of 5-year mortality rates among the people between 50-80 ages (in Table 6.2). We estimate the mortality rate to be 53.08 per 100,000

Age Interval	Incidence Rates per 100,000 person	Mortality Rates per 100,000 person
50-54	53.9	13.7
55-59	76.7	22.3
60-64	114.2	36.5
65-69	173.6	55.5
70-74	230.2	79.3
75-79	293.5	112.2

Table 6.2: Age specific incidence and mortality rates per 100,000 ([53])

person. In this part, we introduce mortality rate, $X_{mortality}$, and define it as follows:

$$\frac{dX_{mortality}(t)}{dt} = \kappa(X_{1,4}(t) + X_{2,4}(t)) + \xi X_{3,4}(t). \tag{6.1}$$

This will also be the performance measure of the model and we will analyze the effects of system parameters on performance measures in detail later. We verify our simulation parameters by using the model where high risk patients always obtain priority since it is the current practice of health care systems. As indicated in Table 6.3, the parameter values give a good fit with incidence and mortality data from National Cancer Institute [53].

Source of estimate	Incidence	Colorectal Cancer Deaths
[53]	157.02	53.08
Model Estimate	149.345	54.880
|Error|	4.8 %	3.4%

Table 6.3: Comparison of statistics with simulation results (per 100000)

Moreover, we use the number of people with cancer in the population as one of the performance measures. It will be a good measure for determining the effectiveness of screening programs. We express it as follows;

$$\#\,\text{Cancer} = \sum_{j=3}^{4}(X_{1,j}(t) + X_{2,j}(t) + X_{3,j}(t)). \tag{6.2}$$

For our calculations, we set time division to be one day. Data were collected after the model had reached steady state. Our simulation model was run on a PC platform in Vensim® PLE for Windows Version 5.9 with Windows XP SP2, 1.73 GHz Pentium Processor.

6.4 Policies

Now we will continue with the models to explore their performances. As stated before, we have four policies which differs due to resource allocation. For simplicity, we enumerate the policies. Policies are summmarized in Table 6.4.

Policy	Description
Policy H	High risk patients always obtain priority
Policy L	Low risk patients always obtain priority
Policy NS	No screening demand from low risk patients
Policy R	Rationing the capacity between the two groups

Table 6.4: Policies with description

6.4.1 Policy H: Priority for High Risk Patients

In this policy, high risk patients have the priority. If a high risk patient enters the system, she/he will be served as soon as possible. Resource capacity should be allocated for low risk patients only if there is no high risk patients in the system. The flow going out of $X_{2,4}$ is,

$$\chi X_{2,4} = \mu \min\{1, X_{2,4}\}. \tag{6.3}$$

The remaining service capacity is given to low risk patients. We allocate the remaining service between healthy people, individuals with polyps and individuals with preclinical cancer randomly since polyps, and early stage cancer may not be observed. The flow going out of $X_{2,j}$, $1 \leq j \leq 3$ when the service is given, is such that

$$\sigma X_{2,1} + \rho X_{2,2} + \phi X_{2,3} = \mu \min\{\max\{1 - X_{2,4}, 0\}, X_{2,1} + X_{2,2} + X_{2,3}\}. \tag{6.4}$$

More specifically, we can represent the flow going out of $X_{2,j}$, $(1 \leq j \leq 3)$

$$\begin{cases} \frac{X_{2,j}}{X_{2,1}+X_{2,2}+X_{2,3}} \mu \min\{\max\{1 - X_{2,4}, 0\}, X_{2,1} + X_{2,2} + X_{2,3}\} & \text{if} \quad X_{2,1} + X_{2,2} + X_{2,3} \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

For this policy, we explore the impact of μ on its performance. Table 6.5 shows the results.

	Service rate							
	5	10	15	21*	25	30	40	50
Incidence	212.852	181.629	149.345	149.345	149.345	149.345	149.345	149.345
Mortality	107.39	83.0392	54.8835	54.8806	54.8794	54.8784	54.8771	54.8764
# Cancer	1040.67	821.786	573.15	573.15	573.151	573.151	573.151	573.151

Table 6.5: The Effect of μ on the Performance Measures for Policy H

* Estimated parameter

As capacity increases, the opportunity of satisfying demand increases. The server can serve more individuals, which increases the probability of screening low risk patients in this policy. Once low risk patients receive service, the polyps, if there are any, are removed and the patients become healthy. Therefore, the risk of developing colorectal cancer decreases. This causes a reduction in the incidence rate. Increased capacity enables the system to screen more individuals. After being screened by colonoscopy, if necessary the patients undergo surgery, or they continue treatment (i.e. periodical colonoscopy). The actions that are taken to prevent the progression of the cancer decreases the mortality rate. Accordingly, as service capacity increases, the number of people with cancer decreases since early detection reduces the probability of developing colorectal cancer.

Compliance is an important variable which affects the demand of colonoscopy. Compliance rate indicates the proportion of people who demand colonoscopy in low risk group. Thus, as compliance rate increases, demand for colonoscopy of screening purposes increases. Now, we will analyze different compliance rates of low risk patients. Table 6.6 summarizes the results.

As stated above, an increase in compliance rate implies more demand for colonoscopy by

	Compliance(%)				
	0	30	50	70	100*
Incidence	240.957	203.113	183.729	168.392	149.345
Mortality	127.356	94.753	79.474	67.9716	54.8806
# Cancer	1222.56	933.285	796.474	692.782	573.15

Table 6.6: Impact of Compliance for Policy H

* Estimated parameter

low risk patients. If these individuals have colon polyps, then the polyps are removed since we assume 100% sensitivity for colonoscopy. Therefore, transformation of polyps into cancer is prevented. This decreases the incidence rate. Mortality rate also decreases because early detection enables the patients to overcome the disease. As observed in Table 6.6, compliance rate also increases the health level of the population by decreasing the number of people with cancer, which shows the effectiveness of screening.

	Rate for Seeking Diagnosis					
	1	2	3	4	6	12*
Incidence	147.079	148.181	148.623	148.848	149.096	149.345
Mortality	63.1242	59.2377	57.695	56.8599	55.8032	54.8806
# Cancer	551.832	562.592	566.88	569.152	570.639	573.15

Table 6.7: The Effect of λ_H for Policy H

* Estimated parameter

In addition to compliance rate, one factor that affects the demand is the average time to seek diagnosis. As seen in Table 6.7, as arrival rate of high risk group increases, incidence rate will increase. This is expected since every high risk patient will be diagnosed with cancer due to our assumption of %100 specificity for the colonoscopy. At the same time, the chance of serving low risk patients decreases, because the capacity is allocated to clinical patients. Therefore, we do not observe the beneficial effects of screening. Mortality rate decreases because patients with clinical cancer go under treatment. Moreover, a reduction in the average time to seek diagnosis decreases the deaths among patients without diagnosis. More high risk patients begin treatment since they are always scheduled first. This decreases the capacity allocated to low risk patients

and healthy people continue to develop symptoms. Therefore, in the overall population, the number of people with cancer increases.

6.4.2 Policy L: Priority for Low Risk Patients

In this policy, low risk patients obtain priority. The flow going out of $X_{2,j}$, $1 \leq j \leq 3$ when the service is given, is

$$+\sigma X_{2,1} + \rho X_{2,2} + \phi X_{2,3} = \mu \min\{1, X_{2,1} + X_{2,2} + X_{2,3}\}. \tag{6.5}$$

More specifically, we can represent the flow going out of $X_{2,j}$, $(1 \leq j \leq 3)$

$$\begin{cases} \frac{X_{2,j}}{X_{2,1}+X_{2,2}+X_{2,3}} \mu \min\{1, X_{2,1} + X_{2,2} + X_{2,3}\} & \text{if} \quad X_{2,1} + X_{2,2} + X_{2,3} \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The remaining service is given to high risk patients. So the flow going out of $X_{2,4}$ is

$$\chi X_{2,4} = \mu \min\{\max\{1 - X_{2,1} - X_{2,2} - X_{2,3}, 0\}, X_{2,4}\}. \tag{6.6}$$

	Service rate							
	5	10	15	21	25	30	40	50
Incidence	186.126	160.474	149.345	149.345	149.345	149.345	149.345	149.345
Mortality	173.936	132.676	54.8835	54.8806	54.8794	54.8784	54.8771	54.8764
# Cancer	804.778	633.276	573.15	573.15	573.151	573.151	573.151	573.151

Table 6.8: The Effect of μ for Policy L

Table 6.8 summarizes the results for service rate. As capacity increases, more low risk patients are screened, which implies more polyps are removed. Therefore the probability of developing cancer reduces. This decreases the incidence rate. By the same logic, early detection reduces mortality rate. Removing polyps and early detection also increases the health level of the population by preventing development of cancer. Therefore as service capacity increases,

the number of people with cancer decreases. Note that we do not observe the non monotonic property between the service rate and the number of cancerous people. This is expected because the system do not allocate capacity to high risk patients when $\mu = 0$ and $\mu = 1$ (total demand from low risk patients is 4.61) for this system. Deaths can occur via high risk patients and since they do not receive service the number of deaths are close to each other in both cases. So deaths can not manipulate the monotonicity property.

Table 6.9 and Table 6.10 present the effects of compliance rates and arrival rate of high risk patients on the performance measures respectively. Similar effects with Policy H are observed for compliance rates and arrival rate of high risk patients.

	Compliance(%)				
	0	30	50	70	100
Incidence	240.957	203.113	183.729	168.392	149.345
Mortality	127.356	94.753	79.474	67.9716	54.8806
# Cancer	1222.56	933.285	796.474	692.783	573.15

Table 6.9: Impact of Compliance on Performance Measures for Policy L

	Rate for Seeking Diagnosis					
	1	2	3	4	6	12
Incidence	147.079	148.181	148.623	148.848	149.096	149.345
Mortality	63.1242	59.2377	57.695	56.8599	55.8032	54.8806
# Cancer	551.832	562.592	566.88	569.152	570.639	573.15

Table 6.10: The Effect of λ_H for Policy L

6.4.3 Policy NS: No Screening For Low Risk Patients

In this case, we never let low risk patients be screened. The colonoscopy service is only used for diagnostic purposes. Therefore, we set $\lambda_L = 0$. The flow going out of $X_{2,4}$ is

$$\chi X_{2,4} = \mu \min\{1, X_{2,4}\}. \tag{6.7}$$

As observed in Table 6.11 in the case of no screening, as service capacity increases, we observe an insignificant increase in incidence rate, and an insignificant decrease in mortality rate. This is due to the absence of early detection. Also, there is an insignificant increase in the number of people with cancer. The values are relatively higher compared to previous two policies. This shows the effectiveness of screening programs.

	Service rate							
	5	10	15	21	25	30	40	50
Incidence	240.945	240.953	240.955	240.957	240.957	240.958	240.958	240.959
Mortality	127.399	127.371	127.362	127.356	127.354	127.353	127.35	127.349
# Cancer	1222.45	1222.52	1222.54	1222.56	1222.56	1222.57	1222.58	1222.58

Table 6.11: The Effect of μ for Policy NS

Table 6.12 shows the response of incidence rate, mortality rate and number of people with cancer for the rate of seeking policy when there is no screening program for low risk patients. We observe that incidence rate increases and mortality rate decreases as the average time for seeking diagnosis decreases. No screening worsens the health of the population, therefore the number of people with cancer increases.

	Rate for Seeking Diagnosis					
	1	2	3	4	6	12
Incidence	235.62	238.212	239.207	239.785	240.346	240.957
Mortality	147.371	137.519	133.733	131.725	129.59	127.356
# Cancer	1171.38	1195.18	1205.29	1210.82	1216.52	1222.56

Table 6.12: The Effect of λ_H for Policy NS

6.4.4 Policy R: Rationing Capacity

This policy allows high and low risk patients to share the resource. Let c be the percentage of the server which is reserved for low risk patients, and $1 - c$ be the percentage of the server which is reserved for high risk patients. The flow going out of $X_{2,j}$, $1 \leq j \leq 3$ when the service

is given, is

$$\sigma X_{2,1} + \rho X_{2,2} + \phi X_{2,3} = c\mu \min\{1, X_{2,1} + X_{2,2} + X_{2,3}\}. \tag{6.8}$$

More specifically, we can represent the flow going out of $X_{2,j}$, $(1 \leq j \leq 3)$

$$\begin{cases} \frac{X_{2,j}}{X_{2,1}+X_{2,2}+X_{2,3}} c\mu \min\{1, X_{2,1} + X_{2,2} + X_{2,3}\} & \text{if} \quad X_{2,1} + X_{2,2} + X_{2,3} \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We also define the flow going out of $X_{2,4}$ as

$$\chi X_{2,4} = (1 - c)\mu \min\{1, X_{2,4}\}. \tag{6.9}$$

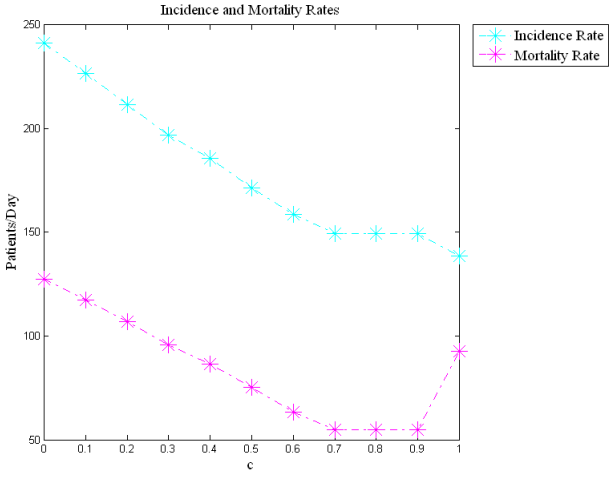
	c								
	0	0.1	0.2	0.5	0.6	0.7	0.8	0.9	1
Incidence	241.28	226.71	185.641	171.545	158.558	149.342	149.339	149.331	138.745
Mortality	127.528	117.232	86.6744	75.1386	63.2957	54.8944	54.9034	54.9312	92.7442
# Cancer	1224.2	1130.49	853.696	750.866	647.063	573.12	573.092	573.013	471.219

Table 6.13: The Impact of Capacity Rationing on Performance Measures

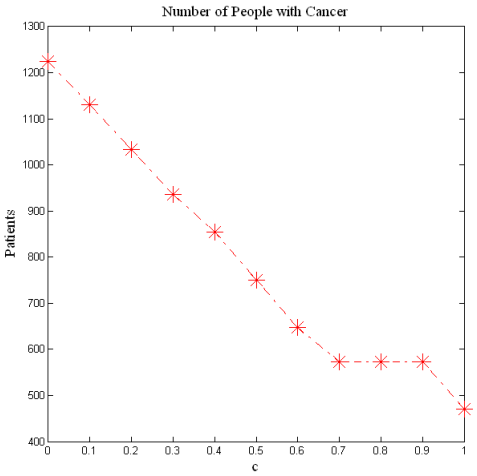
As indicated in Table 6.13, the effect of screening becomes more significant when the system allocates more capacity to low risk patients. As c increases, incidence rate and the number of people with cancer decrease (See Figure 6.2). Hence, allocating more capacity to low risk patients improves these performance measures. However, for high values of c, the system allcoates less capacity to high risk demand which increases mortality rates. Moreover allocating all capacity to one type of patients will not be effective since the remaining service capacity, if any, can not be used. Therefore to avoid ineffective use of service capacity, we analyze only the cases where $c = 20\%, 50\%$ and 80%.

6.4.5 Model Comparison

In this section we compare the policies in detail. Figure 6.3 includes the incidence rates of



(a) Incidence and Mortality rates with c values



(b) Number of People with Cancer with c values

Figure 6.2: Performance measures for c values

different scheduling policies respectively. We observe that the highest incidence rate belongs to Policy NS. This is intuitive, since if the system does not perform colonoscopy for screening purposes, then the system can not detect the abnormalities and remove them when they are less harmful. So incidence rate will be higher compared to other policies. We analyze the system and observe that whenever service capacity is more than 14 ($\mu \geq 14$), the capacity exceeds the total demand. As observed in Figure 6.3 , when the demand is less than the capacity, the performance measures differ in Policy H and Policy L, otherwise they behave similarly. Incidence rates are the same for the capacity values greater than $\mu = 14$. This result is expected since in the case of sufficient capacity, the system meets all the demand. Therefore, policies converge to each other. However, when capacity is not enough then the allocation of the capacity becomes significant on performance measures. In Policy H, since high risk patients are diagnosed and considered in the incidence rate, the incidence rate is higher than in Policy L. For the rationing policies, we observed that as the system allocate more capacity to low risk patients, incidence rate decreases. Moreover, the incidence rates for rationing policies are bounded by the incidence rate of Policy

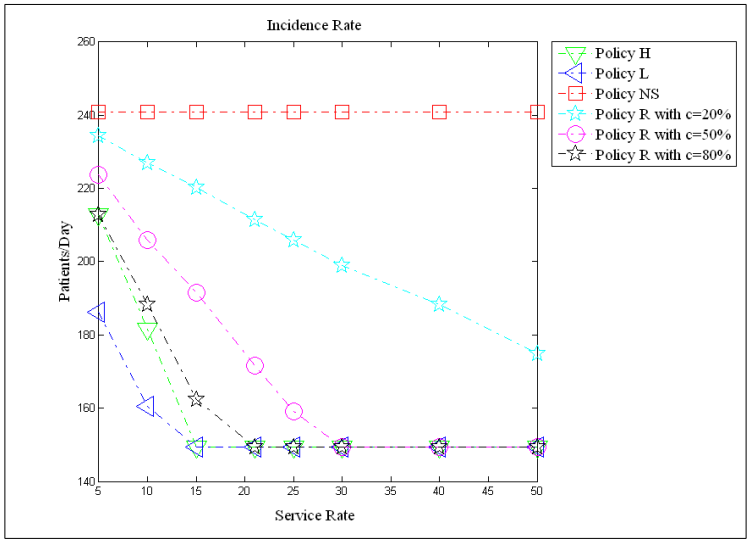


Figure 6.3: Comparison of incidence rates with varying μ

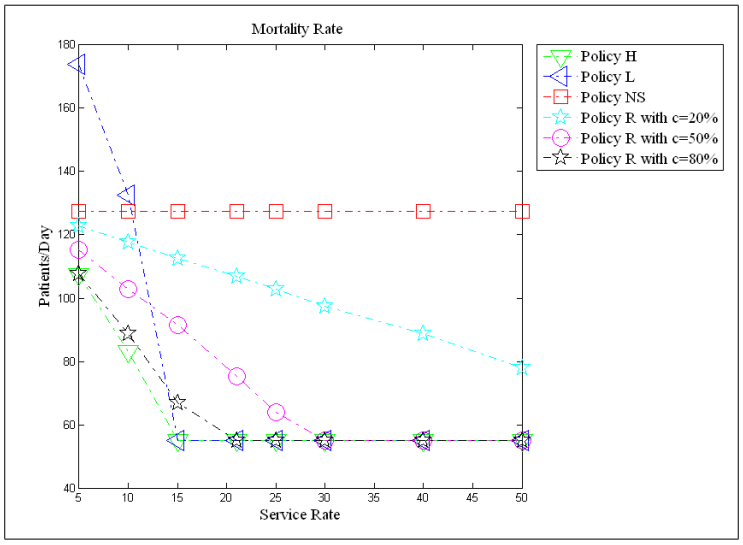


Figure 6.4: Comparison of mortality rates with varying μ

NS from above and Policy H from below.

Figure 6.4 indicates the sensitivity of mortality rates to service capacity in various policies. Policy H has the lowest mortality rate among all the policies, since the deaths can occur via high risk patients, and giving priority to them decreases mortality rate more than it does by giving priority to low risk patients. However, we also observe that if we serve only high risk patients and do not screen low risk patients, then this leads ineffective use of capacity, and results with high level of mortality rate (Policy NS). For $\mu < 14$, in Policy L, the service capacity does not satisfy the low risk demand which prevents the system to allocate capacity for high risk patients, therefore the mortality rate for Policy L is the highest among other policies for $\mu < 14$. In other words, we can state that the capacity is neither enough for low risk patients nor high risk patients. As μ increases, Policy L begins to satisfy certain level of low risk demand. Satisfying such demand returns similar effects with allocating $c\%$ of capacity to low risk demand and remaining capacity to high risk demand. Therefore, Policy L intersects with Policy R. As capacity becomes enough to serve all type of patients in Policy L and Policy H, both policies give the same mortality values, and these values are almost constant. In rationing policies, we observed that as c increases, mortality rate decreases. (Note that we are analyzing effective c values.)

As observed in Figure 6.5, no screening policy results with a highest level of cancerous population. Policy H and Policy R follow non monotonic trends in service rate. As stated in the analysis of Policy H, at the low values of μ, deaths when $\mu = 0$ can be high enough to supress the monotone behaviour of service capacity. As c increases, the number of people with cancer decreases since the system improves the health level of low risk patients by removing polyps in precursor stage, or apply effective treatment methods as a result of early detection. When low risk patients have the priority, the number of people with cancer is less than other policies, because the system can minimize the flow from precursor stage to preclinical stage and maximize the flow from preclinical stage to healthy stage.

Figure 6.6, Figure 6.7 and Figure 6.8 show the incidence rates, mortality rates and the number of people with cancer for varying compliance rates, respectively. We observed that compliance

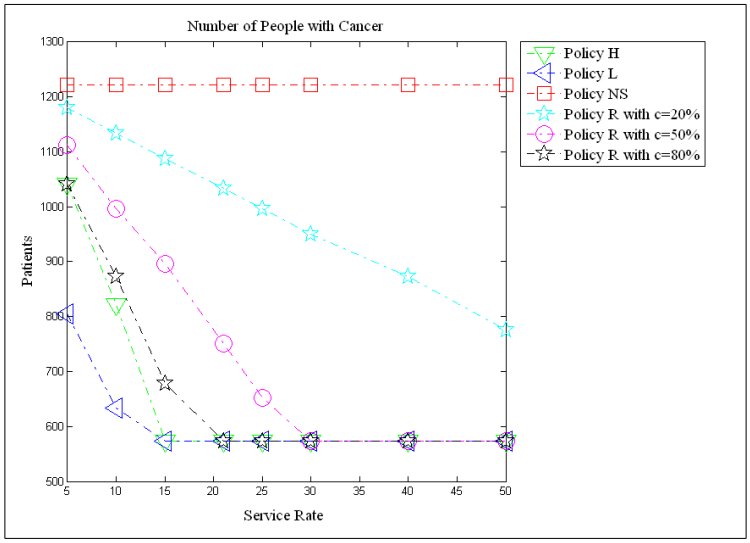


Figure 6.5: Comparison of the number of cancerous people with varying μ

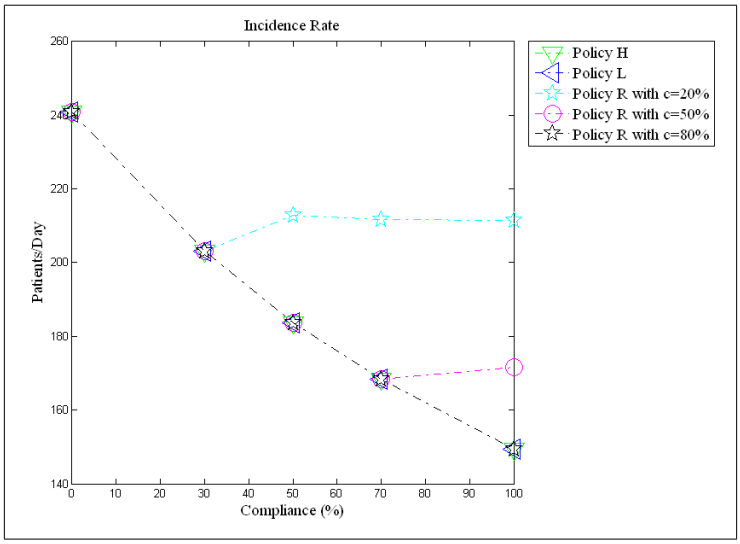


Figure 6.6: Comparison of incidence rates with varying compliance rates

rates influence the performance measures similarly. Note that Policy H and Policy L behave similarly, since we ran the simulation for $\mu = 21$, which is a sufficient capacity. Moreover, when $c = 80\%$ we have similar results with Policy H and L. Recall that this is true for the cases where there is enough capacity (See Figure 6.3, Figure 6.4 and Figure 6.5). Incidence rate, mortality rate and the number of people with cancer decreases with compliance rates for these policies. Policy R with $c = 20\%$ differs from other policies if compliance rate is greater than 30 %, the performance measures start increasing beginning from that level, since the capacity allocated to low risk patients do not meet the demand of low risk patients. This advances the flow from non-cancer stages to cancer stages. Hence incidence rate, mortality rate and the number of people with cancer increases. The same reason can be concluded for Policy R with $c = 50\%$, too. For the compliance rates greater than 70 %, Policy R with $c = 50\%$ differentiates from Policy H and L.

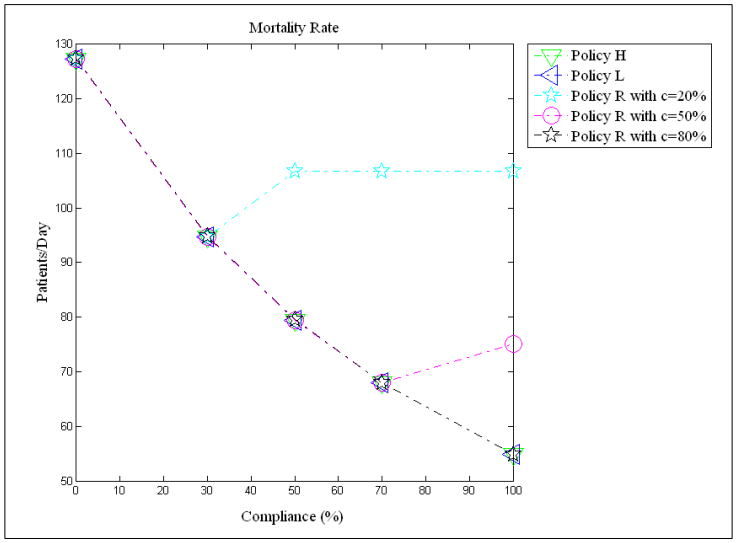


Figure 6.7: Comparison of mortality rates with varying compliance rates

The results for varying λ_H values can be found in Figure 6.9, Figure 6.10 and Figure 6.11. We conclude that the most insignificant parameter is λ_H. The change in λ_H, does not influence

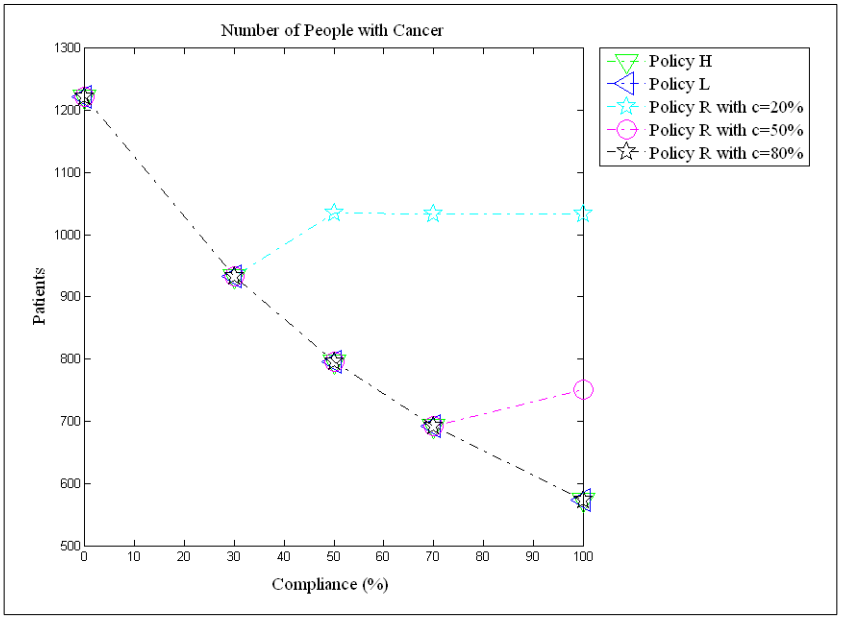


Figure 6.8: Comparison of the number of cancerous people with varying compliance rates

the performance measures at all. There are slight differences between the data points. If average time for seeking diagnosis decreases, then more high risk patients can receive service. Screening high risk patients will increase the incidence rate since they will be diagnosed with cancer. As average time for seeking diagnosis decreases, the deaths decreases. Moreover, the number of people with cancer increase with the arrival rate of high risk patients, since there is a decrease in mortality rate meanwhile there is a flow from precursor stage to cancer stages. As verified in figures (Figure 6.9, Figure 6.10 and Figure 6.11), incidence rate and the number of people with cancer are increasing in λ_H, whereas mortality rate is decreasing in λ_H. We observe that the effect of not screening low risk patients is drastic. Since capacity is enough, we observe that the performance measures for Policy H and Policy L are same. Moreover, different practices of Policy R vary between these values.

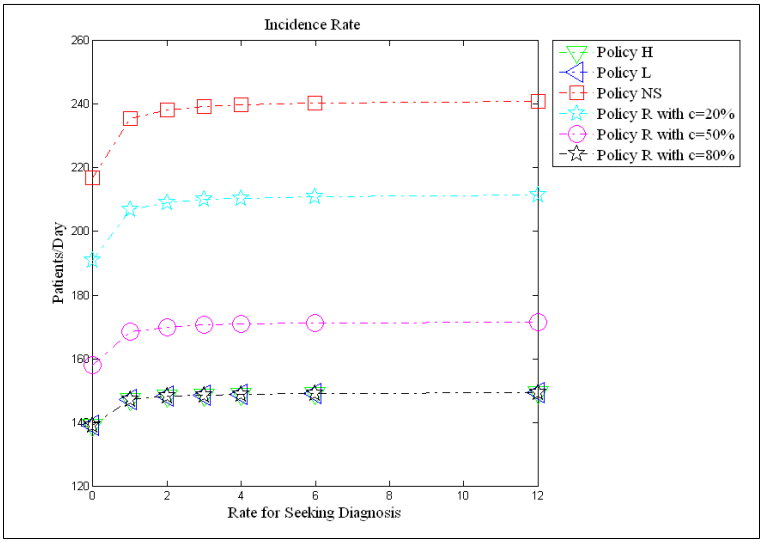


Figure 6.9: Comparison of incidence rates with varying λ_H

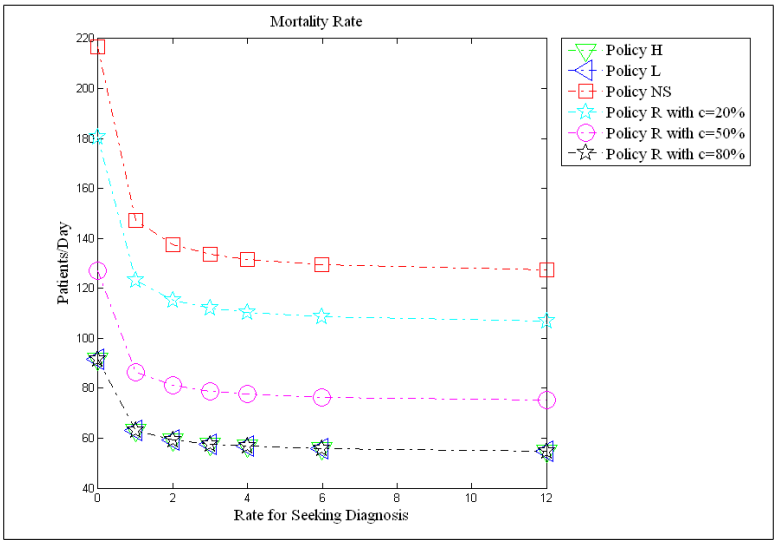


Figure 6.10: Comparison of mortality rates with varying λ_H

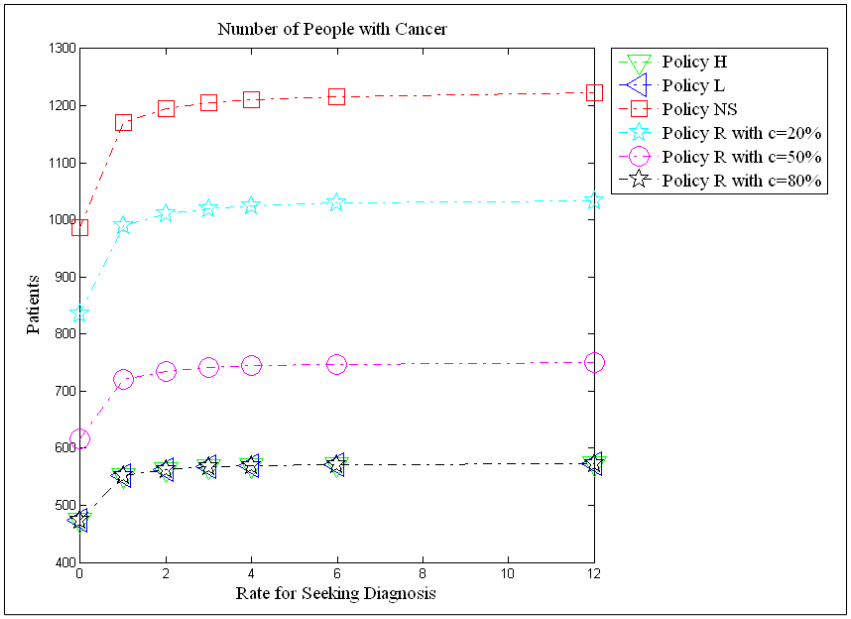


Figure 6.11: Comparison of the number of cancerous people with varying λ_H

6.4.6 Conclusion

As mentioned before, emphasising on scheduling alone is more cost-effective than admission control. Therefore in this chapter, we consider possible scheduling policies and analyze performance of the policies. In this study, we consider four stage health model for colorectal cancer screening. By doing this, we aim to obtain highest efficiency from screening methods as a result of classifying the risk groups accurately.

There have been decreases in incidence and mortality rates year to year according to Seer data [53]. One reason for that trend can be an increasing awareness of the importance of screening in the population. This can increase the demand for screening. In general, we observe that as policies improve, incidence and mortality rates decrease. The results in this chapter confirm these findings.

When service capacity is enough, incidence rate, mortality rate and the number of patients with cancer behave similarly to compliance rates for all policies. Also in the case of enough

capacity, an increase in the rate for seeking diagnosis leads an increase in incidence rate and the number of cancerous patients whereas a decrease in mortality rates for all policies. If the capacity is not enough, the behaviour of the performance measures can vary. We can observe non monotonic behaviours as well as reordering in the rankings. Therefore, capacity allocation should receive attention in this case.

In Model A (systems exercising only admission decision and give priority to high risk patients) in the previous chapter, we observe that as μ increases, the system can decrease the number of high risk patients more easily and allocate the remaining service to low risk patients. Serving low risk patients reduces the long- run effective arrival rate for high risk patients in the future. This verifies the importance of screening. The system saves lives by screening low risk patients as a result of early detection. Policy H, a duplicate of Model A in this chapter, states the same results. In this policy as μ increases, incidence and mortality rates decrease. This shows the beneficial effect of screening programs.

Policy NS returns the highest value for the number of cancerous people, incidence and mortality rates. No screening worsens the health level of the population and deterioriates the environment. We can also say that the models which do not screen asymptomatic patients, namely Model A and N, also has the highest values for long- run effective arrival rate for symptomatic patients. This is unfavorable, since it shows that the ratio of symptomatic patients are high. That is, these models worsen the population health compared to Model S and B. Hence, we conclude that in both chapters, the systems that do not use capacity for screening purposes gives the unfavorable results for the population health, hence the results are consistent.

Chapter 7

CONCLUSION

In this thesis, we consider a single server where the capacity is shared between screening and diagnostic services. There are different risk groups who demand service for screening purposes. There are high risk patients who demand service for diagnostic purposes. The feedback effect of screening can be observed after service completion. For such a system, we aim to provide insights of better admission and scheduling control policies for screening services.

Firstly, we study the micro level analysis of the system. We assume that the facility providing the screening/diagnostic procedure operates in a dynamic random environment, which determines the demand for diagnostic services. The random environment represents the health of the population, where if the health of the population is better, the demand rate of symptomatic patients is lower. We provide a model where the system is under the external influence which is caused by deterioration of health conditions. Moreover, scheduling lower risk patients is the screening process, improving the health of the whole population in the long run. We employ this approach to model complex queuing model where we explore the trade off between decreasing future risk level of the population and the emergency of incoming patients with a limited capacity of service. The system can exercise admission or scheduling or both controls. We consider three different models, depending on which of the controls is used.

Our main contribution in this framework is the formulation of the environment. We provide a model where endogenous and exogenous facrors are considered together and influence the environment in which the system resides in. The system is under the external influence caused by the deterioration of health conditions. In addition, the effect of screening is modeled as an internal factor that improves the environment which leads to a change in the state of the environment. The arrival rates vary due to the state of the environment.

In the models where two decisions are available, interaction can be observed between the decisions. In this study, we have shown that admission control does not influence the scheduling control at all. However, scheduling policy has an significant effect on admission policy. Hence, instead of two controls, considering only scheduling control can be an effective practice for health care systems. Allocating more time and fund for determining the scheduling strategy increases the effectiveness of screening programs, which results with a reduction in disease prevalence and thus a reduction in treatment costs of the future. The current practice in the health care system is to prioritize high risk patients. However, the numerical results show that this practice is not optimal. By allocating capacity to low risk patients, we can improve the health conditions. Therefore, we stress the importance of capacity allocation. Moreover, p is the most deterministic parameter for the performance of the policies. High p improves the future health of the population. High level of p can be obtained by screening the right risk group. Therefore classifying the risk groups accurately is crucial. In order to decrease health costs in future, insurance companies may cover the costs of screening procedures for the individuals who belong to risk group with higher p value.

The model we propose represents a step in further understanding in the effective allocation of capacity for diagnostic and screening services via admission and scheduling control. It can ve extended in many aspects. For an extension, we can introduce a new operator, T_{WORSE}, and let it represent the process of developing cancer while waiting for the service in the queue where M is the fixed cost of developing cancer. We define it as follows:

$$T_{WORSE}f(e, x_H, x_L) = f(e, x_H + 1, x_L - 1) + M.$$

We show that this operator preserves all the properties defined in Chapter 3. The proofs are in the Appendix B.1.

One of the shortcomings of the model is we assume same $p_{e,e-1}$ values for different asymptomatic groups because estimating specific $p_{e,e-1}$ values for risk groups is difficult.

Another limitation of this study is we assume preemptive scheduling. For non-preemptive

scheduling, we can define a new state where we also record which type of patient the system serves. By doing this, we formulate a new state space, $S_{new} = \{(A, e, x_H, x_L) : A \in \{H, L\}\ (x_H, x_L) \geq 0, e \in \{1, \ldots, E\}\}$ where H indicates that symptomatic patients receives service whereas L stands for asymptomatic patient. Then

$$v_{n+1}(A, e, x_H, x_L) = T_{COST}(T_{UNIF}(\{T_{ARR_H} v_n(A, e, x_H, x_L), \{T_{ADM_i} v_n(A, e, x_H, x_L)\}_i,$$
$$T_{SCH} v_n(A, e, x_H, x_L), T_{DET} v_n(A, e, x_H, x_L)\}; \{\bar{\lambda}_H, \{\lambda_{L_i}\}_i, \mu, \bar{\gamma}\})), \tag{7.1}$$

and

$$T_{SCH} f(A, e, x_H, x_L) = \min\{f(H, e, x_H - 1, x_L), g(L, e, x_H, x_L - 1)\} + s, \tag{7.2}$$

where

$$g(L, e, x, y) = \begin{cases} f(L, 1, x, y) & \text{if} \quad e = 1 \\ p_{e,e-1} f(L, e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(L, e, x_H, x_L - 1) & \text{otherwise.} \end{cases}$$

We show that this new model preserves $Inc(x_H)$, $Inc(x_L)$, $Inc(e)$, $Diag(x_H, x_L)$, $IDiag_e(x_H, x_L)$ and $Conv(x_L)$ properties (See Appendix B.2). We continue with the following conjecture.

Conjecture 4 *Value functions preserve $Sup(x_H, x_L)$ and $Sup(e, x_L)$ properties for state space S_{new}.*

In this thesis, we have considered linear holding costs for risk group. We have shown various properties of value functions. As the number of patients increase in the system, the burden of one additional patient will increase, therefore we can also assume non-decreasing and convex costs in x_H and x_L. Under non-linear holding costs, we lose some of the properties of value functions. The following conjecture includes these properties:

Conjecture 5 *The properties, $Diag(x_H, x_L)$, $IDiag_e(x_H, x_L)$, $Sup(e, x_L)$ and $Sup(x_H, x_L)$ are preserved by convex holding costs.*

The remaining properties are preserved by convex holding costs (See Appendix B.3).

Secondly, we use compartmental model to analyze the effects of operational controls on population dynamics. While doing this, we pay attention to determine the risk groups. We develop a four-stage health model where we indicate precursor clinical state (polyps) which may cause inaccurate representation of disease progression if not considered. We consider various scheduling policies where we allocate the capacity between screening and diagnostic purposes. Numerical results suggest that screening low risk patients decrease the incidence rate of colorectal cancer as well as mortality rate. Moreover, screening improves the health of the population. When capacity is enough to meet the demand, then the scheduling strategies may converge to each other. If this is not the case, a trade-off between the immediate need for diagnostic and the long-term benefit of screening will arise. Therefore service providers should pay attention to the allocation of the resource for screening or diagnostic purposes. Applying effective screening programs will decrease the burden on the health system and improves the health level of the population.

BIBLIOGRAPHY

[1] ACS. Detailed guide: Colon and rectum cancer, what happens after treatment?, 2009. http://www.cancer.org/docroot/CRI/content/CRI_2_4_5x_What_happens_after_treatment_10.asp.

[2] ACS. Estimated new cancer cases and deaths for 2009, 2009. http://seer.cancer.gov/csr/1975_2006/results_single/sect_01_table.01.pdf.

[3] AGA. The facts about colorectal cancer, 2009. http://www.gastro.org/wmspage.cfm?parm1=682.

[4] O. Alagöz, T. Ayer, and F. S. Erenay. Operations research models for cancer screening. In *Encyclopedia for Operations Research and Management Science*. John Wiley and Sons, 2009.

[5] B. Ata. Dynamic control of a multiclass queue with thin arrival streams. *Operations Research*, 54(5):876–892, 2006.

[6] J. S. Baras, A. J. Dorsey, and A. M. Makowski. Two competing queues with linear costs and geometric service requirments: The $\mu c-$rule is often optimal. *Advanced Applied Probability*, 17:186–209, 1985.

[7] A. Bassamboo, J. M. Harrison, and A. Zeevi. Dynamic routing and admission control in high-volume service systems: Asymptotic analysis via multi-scale fluid limits. *Queueing Systems*, 51:249–285, 2005.

[8] M. Bassett and et al. *A Practical Guide to Increasing Screening Colonoscopy*. Cancer Prevention and Control Program, Bureau of Chronic Disease Prevention and Control,

The New York City Department of Health and Mental Hygiene and New York Citywide Colon Cancer Control Coalition, 2006.

[9] M. D. Bethesda. Prevalence and cost of esrd therapy. In *US Renal Data System: USRDS 1994 Annual Data Report*, chapter 3, pages 23–43. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 1994.

[10] M. L. Brandeau, D. K. Owens, C. H. Sox, and R. M. Wachter. Screening women of childbearing age for human immunodeficiency virus: A model-based policy analysis. *Management Science*, 39(1):72–92, 1993.

[11] M. L. Brown, G. F. Riley, A. L. Potosky, and R. Etzioni. Obtaining long-term disease specific costs of care: application to medicare enrollees diagnosed with colorectal cancer. *Med Care*, 37:1249–1259, 1999.

[12] D. Y. Burman and D. R. Smith. An asymptotic analysis of a queueing system with markov-modulated arrivals. *Operations Research*, 34(1):105–119, 1986.

[13] L. Butterly, C. Olenec, M. Goodrich, P. Carney, and A. Dietrich. Colonoscopy demand and capacity in new hampshire. *American Journal of Preventive Medicine*, 32(1):25–31, 2007.

[14] S. Carr and I. Duenyas. Optimal admission control and sequencing in a make-to-stock make-to-order production system. *Operations Research*, 48(5):709–720, 2000.

[15] U. Çakmak and S. Özekici. Portfolio optimization in stochastic markets. *Mathematical Methods of Operations Research*, 63:151–168, 2006.

[16] E. Çanakoğlu and S. Özekici. Portfolio selection in stochastic markets with exponential utility functions. *Ann Oper Res*, 166:281–297, 2009.

[17] E. Çanakoğlu and S. Özekici. Portfolio selection in stochastic markets with hara utility functions. *European Journal of Operational Research*, 2009.

[18] T. Çayırlı and E. Veral. Outpatient scheduling in health care: a review of literature. *Production and Operations Management*, 12(4):519–549, 2003.

[19] E. B. Çil, L. Örmeci, and F. Karaesmen. Effects of system parameters on the optimal policy structure in a class of queueing control problems. *Queueing Systems*, 61:273–304, 2009.

[20] J. A. Chan, J. A. Meyerhardt, A. T. Chan, E. L. Giovannucci, G. A. Colditz, and C. S. Fuchs. Hormone replacement therapy and survival after colorectal cancer diagnosis. *Journal of Clinical Oncology*, 24(36):5680–5686, 2006.

[21] F. Cheng and S. P. Sethi. A periodic review inventory model with demand influenced by promotion decisions. *Management Science*, 45(11):1510–1523, 1999.

[22] C. R. Chien and T. H. H. Chen. Mean sojourn time and effectiveness of mortality reduction for lung cancer screening with computed tomography. *Int. J. Cancer*, 122:2594–2599, 2008.

[23] T. B. Crabill, D. Gross, and M. J. Magazine. A classified bibliography of research on optimal design and control of queues. *Operations Research*, 25(2):219–232, 1997.

[24] E. Davis. Optimal control of arrivals to a two-server queuing system with seperate queues. Ph. D. Dissertation Program in Operations Research, N.C. State University, Raleigh, 1977.

[25] E. Donnellan. Nine-month wait for bowel cancer test, 2008. http://www.irishtimes.com/newspaper/ireland/2008/1115/1226700611205.html.

[26] D. M. Eddy, F. W. Nugent, J. F. Eddy, J. Coller, V. Gilbertsen, L. S. Gottlieb, R. Rice, P. Sherlock, and S. Winawer. Screening for colorectal cancer in a high-risk population. results of a mathematical model. *Gastroenterology*, 92(3):682–692, 1987.

[27] M. Eisen and M. Tainiter. Stochastic variations in queuing processes. *Operations Research*, 11(6):922–927, 1963.

[28] E. El-Darzi, C. V. b, T. Chaussalet, and P. Millard. A simulation modelling approach to evaluating length of stay, occupancy, emptiness and bed blocking in a hospital geriatric department. *Health Care Management Science*, 1:143–149, 1998.

[29] A. Ephremides, P. Varaiya, and J. Walrand. A simple dynamic routing problem. *IEEE Transactions on Automatic Control*, 25(4):690–693, 1980.

[30] A. S. Erdem and S. Özekici. Inventory models with random yield in a random environment. *Int. J. Production Economics*, 78:239–253, 2002.

[31] R. M. Feldman. A continuous review (s, s) inventory system in a random environment. *Journal of Applied Probability*, 15(3):654–659, 1978.

[32] A. Frazier, G. Colditz, C. Fuchs, and K. Kuntz. Cost-effectiveness of screening for colorectal cancer in the general population. *Journal of American Medical Association*, 284(15):1954–1961, 2000.

[33] J. P. Gayon, I. Talay, F. Karaesmen, and E. L. Örmeci. Optimal pricing and production policies of a make-to-stock system with fluctuating demand. *Probability in the Engineering and Informational Sciences*, 23:205–230, 2009.

[34] H. A. Ghoneim and S. Stidham. Control of arrivals to two queues in series. *European Journal of Operational Research*, 21:399–409, 1985.

[35] L. V. Green, S. Savin, and B. Wang. Managing patient service in a diagnostic medical facility. *Operations Research*, 54(1):11–25, 2006.

[36] E. D. Güneş, S. E. Chick, and O. Z. Akşin. Breast cancer screening services: Trade-offs in quality, capacity, outreach, and centralization. *Health Care Management Science*, 7(4):291–303, 2004.

[37] B. Hajek. Optimal control of two interacting service stations. *IEEE Transactions on Automatic Control*, 29(6):491–499, 1984.

[38] P. R. Harper and S. K. Jones. Mathematical models for the early detection and treatment of colorectal cancer. *Health Care Management Science*, 8:101–109, 2005.

[39] J. M. Harrison. Dynamic scheduling of a multiclass queue: Discount optimality. *Operations Research*, 23(2):270–282, 1975.

[40] W. E. Helm and K. H. Waldmann. Optimal control of arrivals to multiservers queues in a random environment. *Journal of Applied Probability*, 21:602–615, 1984.

[41] H. W. Hethcote and J. A. Yorke. *Gonorrhea Transmission Dynamics and Control*. Lecture Notes in Biomathematics, 1984.

[42] A. Hordijk and G. Koole. On the assignment of customers to parallel queues. *Probability in the Engineering and Informational Sciences*, 6:495–511, 1992.

[43] N. C. I. Fact sheet: Colorectal cancer screening. *International Journal of Colorectal Disease*, 21(4), 2008.

[44] J. Ivy. Can we do better? optimization models for breast cancer screening. In *Handbook of Optimization in Medicine*, pages 1–28. Springer, 2009.

[45] S. G. Johansen and S. Stidham. Control of arrivals to a stochastic input-output system. *Advances in Applied Probability*, 12:972–999, 1980.

[46] E. H. Kaplan. What are the risks of risky sex? modeling the aids epidemic. *Operations Research*, 37(2):198–209, 1989.

[47] R. K. Khandker, J. D. Dulski, J. B. Kilpatrick, R. P. Ellis, J. B. Mitchell, and W. B. Baine. A decision model and cost-effectiveness analysis of colorectal cancer screening and surveillance guidelines for average-risk adults. *International Journal of Technology Assessment in Health Care*, 16(3):799–810, 2000.

[48] H. Koenig, M. McCullough, and D. Larson. *Handbook of Religion and Health*. Oxford, England: Oxford University Press, 2001.

[49] G. Kolata. 50 and ready for colonoscopy? doctors say wait is often long. *the New york Times Monday*, 2003.

[50] R. Kolisch and S. Sickinger. Providing radiology health care services to stochastic demand of different customer classes. *OR Spectrum*, 30:375–395, 2008.

[51] G. Koole. Structural results for the control of queueing systems using event-based dynamic programming. *Queueing Systems*, 30:323–339, 1998.

[52] U. Ladabaum, C. Chopra, G. Huang, J. Scheiman, and a. A. F. M.E. Chernew. Aspirin as an adjunct to screening for prevention of sporadic colorectal cancer. a cost-effectiveness analysis. *Ann Intern Med.*, 135:769–781, 2001.

[53] R. LAG and et al. (eds). Seer cancer statistics review, 1975-2005. *National Cancer Institute*, http://seer.cancer.gov/csr/1975_2005/, 2008.

[54] H. L. Lee and W. P. Pierskalla. Mass screening models for contagious diseases with no latent period. *Operations Research*, 36(6):917–928, 1988.

[55] H. L. Lee and W. P. Pierskalla. Theory and general models of mass screening for contagious and non-contagious diseases. *Operations Research*, 87:428–441, 1988.

[56] J. P. Lehoczky and D. P. Gaver. Diffusion approximations for the cooperative service of voice and data messages. *Journal of Applied Probability*, 18(3):660–671, 1981.

[57] M. Leshno, Z. Halpern, and N. Arber. Cost-effectiveness of colorectal cancer screening in the average risk population. *Health Care Management Science*, 6(3):165–174, 2003.

[58] M. E. Lewis. Average optimal policies in a controlled queueing system with dual admission control. *Journal of Applied Probability*, 38(2):369–385, 2001.

[59] W. Lin and P. R. Kumar. Optimal control of queueing system with two heterogeneous servers. *IEEE Transactions on Automatic Control*, 29(8):696–703, 1984.

[60] Z. Liu and P. Nain. Optimal scheduling in some multi-queue single-server systems. *IEEE Transactions on Automatic Control*, 37(2):247–252, 1992.

[61] C. Maglaras and J. A. V. Mieghem. Queueing systems with leadtime constraints: A fluid-model approach for admission and sequencing control. *European Journal of Operational Research*, 167:179–207, 2005.

[62] S. I. McClean and P. H. Millard. A three compartment model of the patient flows in a geriatric department: a decision support approach. *Health Care Management Science*, 1:159–163, 1998.

[63] A. Menges, J. Fischinger, B. Gärtner, T. Georg, D. Woerdehoff, M. Maier, M. Harloff, C. Stegmaier, J. Raedle, and M. Zeitz. Screening colonoscopy in 40- to 50-year-old first-degree relatives of patients with colorectal cancer is efficient: a controlled multicentre study. *International Journal of Colorectal Disease*, 21(4):301–307, 2006.

[64] S. V. Mondschein and G. Weintraub. Appointment policies in service operations: A critical analysis of the economic framework. *Production and Operations Management*, 12(2):266–286, 2003.

[65] P. Naor. The regulation of queue size by levying tolls. *Econometrica*, 37(1):15–24, 1969.

[66] M. F. Neuts. A queue subject to extraneous phase changes. *Advances in Applied Probability*, 3(1):78–119, 1971.

[67] E. L. Örmeci and J. van der Wal. Admission policies for a two class loss system with general interarrival times. *Stochastic Models*, 22(11), 2006.

[68] S. Özekici. Complex systems in random environments. In S. Özekici, editor, *Reliability and Maintenance of Complex Systems*. Springer, 1996.

[69] S. Özekici and M. Parlar. Inventory models with unreliable suppliers in a random environment. *Annals of Operations Research*, 91(6):123–136, 1999.

[70] S. Özekici and R. Soyer. Network reliability assessment in a random environment. *Naval Research Logistics*, 50:574–591, 2003.

[71] J. Patrick, M. Puterman, and M. Queyranne. Dynamic multi-priority patient scheduling for a diagnostic resource. *Operations Research*, 56(6):1507–1525, 2008.

[72] E. Plambeck, S. Kumar, and J. M. Harrison. A multiclass queue in heavy traffic with throughput time constraints: Asymptotically optimal dynamic controls. *Queueing Systems*, 39:23–54, 2001.

[73] N. U. Prabhu and Y. Zhu. Markov-modulated queueing systems. *Queueing Systems*, 5:215–246, 1989.

[74] P. Purdue. The m/m/1 queue in a markovian environment. *Operations Research*, 22(3):562–569, 1974.

[75] M. L. Puterman. *Discrete Stochastic Dynamic Programming*. John Wiley and Sons, USA, 1994.

[76] C. Revelle, F. Feldmann, and W. Lynn. An optimization model of tuberculosis epidemiology. *Management Science*, 16(4):190–211, 1969.

[77] D. Rex, D. Johnson, J. Anderson, P. Schoenfeld, C. Burke, and J. Inadomi. American college of gastroenterology guidelines for colorectal cancer screening 2008. *The American Journal of Gastroenterology*, 104:739–750, 2009.

[78] D. K. Rex and F. . S. Liangpunsakul. Colorectal cancer screening. *American College of Gastroenterology*, 2009.

[79] R. Righter. A resource allocation problem in a random environment. *Operations Research*, 37(2):329–338, 1989.

[80] R. Righter. Expulsion and scheduling control for multiclass queues with heterogeneous servers. *Queueing Systems*, 34:289–300, 2000.

[81] J. S. Song and P. Zipkin. Inventory control in a fluctuating demand environment. *Operations Research*, 41(2):351–370, 1993.

[82] A. Sonnenberg, F. Delcó, and J. M. Inadomi. Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Annals of Internal Medicine*, 133(8):573–584, 2000.

[83] A. Sonnenberg and M. W. Gavin. Crohns & colitis foundation of america, inc timing of surgery for enterovesical fistula in crohns disease: Decision analysis using a time-dependent compartment model. *Inflammatory Bowel Diseases*, 6(4):280–285, 2000.

[84] S. Stidham. Socially and individually optimal control of arrivals to a $gi/m/i$ queue. *Management Sciences*, 24(15):1598–1610, 1978.

[85] S. Stidham. Optimal control of admission to a queueing system. *IEEE Transactions on Automatic Control*, 30(8):705–713, 1985.

[86] S. Stidham and R. Weber. *Control of Service Rates in Cycles and Series of Queues*. Stat. Lab., Univ. Cambridge, Cambridge, 1983.

[87] S. Stidham and R. Weber. Control of service rates in networks of queues. *Advanced Applied Probability*, 19:202–218, 1987.

[88] USPSTF. Screening for colorectal cancer: Preventive services task force recommendation statement. *Annals of Internal Medicine*, 149:627–637, 2008.

[89] USPSTF. Screening for colorectal cancer, topic page, 2009. http://www.ahrq.gov/clinic/uspstf/uspscolo.htm.

[90] J. L. Wagner, R. C. Herdman, and S. Wadhwa. Cost effectiveness of colorectal cancer screening in the elderly. *Annals of Internal Medicine*, 115(10):807–817, 1991.

[91] R. R. Weber and S. Stidham. Optimal control of service rates in network queues. *Advanced Applied Probability*, 19:202–218, 1987.

[92] R. R. Weber and S. Stidham. A survey of markov decision models for control of networks of queues. *Queueing Systems*, 13:291–314, 1993.

[93] W.H.O. *Cancer Control Knowledge into Action WHO Guide for Effective Programmes*. WHO Press, Geneva, 2007.

[94] S. Winawer, R. Fletcher, D. Rex, J. Bond, R. Burt, J. Ferrucci, T. Ganiats, T. Levin, S. Woolf, D. Johnson, L. Kirk, S. Litin, and C. Simmang. Colorectal cancer screening and surveillance: Clinical guidelines and rationale update based on new evidence. *Gastroenterology*, 124:544–560, 2003.

[95] S. J. Winawer, A. G. Zauber, M. N. Ho, M. J. OBrien, L. Gottlieb, S. S. Sternberg, and et al. Prevention of colorectal cancer by colonoscopic polypectomy. *The New England Journal of Medicine*, 329(27):1977–1981, 1993.

[96] J. M. Wong, M. F. Yen, M. S. Lai, S. W. Duffy, R. A. Smith, and T. H. H. Chen. Progression rates of colorectal cancer by dukes' stage in a high-risk group: Analysis of selective colorectal cancer screening. *The Cancer Journal*, 10(3):160–169, 2004.

[97] S. H. Xu. A duality approach to admission and scheduling controls of queues. *Queueing Systems*, 18:273–300, 1994.

[98] S. H. Xu, R. Righter, and J. G. Shanthikumar. Optimal dynamic assignment of customers to heterogeneous servers in parallel. *Operations Research*, 40(6):1126–1138, 1992.

[99] S. H. Xu and J. G. Shanthikumar. Optimal expulsion control-a dual approach to admission control of an ordered-entry system. *Operations Research*, 41(6):1137–1152, 1993.

[100] U. Yechiali and P. Naor. Queuing problems with heterogeneous arrivals and service. *Operations Research*, 19(3):722–734, 1971.

[101] G. S. Zaric, M. L. Brandeau, and P. G. Barnett. Methadone maintenance and hiv prevention: A cost-effectiveness analysis. *Management Science*, 46(8):1013–1031, 2000.

Appendix A

PROOFS FOR CHAPTER 3

We will give a representative example of proofs for Chapter 3.

A.1 Monotonicity in x_H

We will prove that all operators defined in Chapter 3, preserve the monotonicity property of any monotone function f in x_H, such that the following inequality,

$$Tf(e, x_H, x_L) \leq Tf(e, x_H + 1, x_L)$$

holds for any non-decreasing function f in x_H and for all operators T.

Monotonicity in x_H *Preserved by* T_{ARR_H}

$$\begin{aligned}
& \alpha[f(e, x_H + 1, x_L) + c] + (1 - \alpha) f(e, x_H, x_L) \\
& -\alpha[f(e, x_H + 2, x_L) + c] - (1 - \alpha) f(e, x_H + 1, x_L) \\
= \; & \alpha[f(e, x_H + 1, x_L) - f(e, x_H + 2, x_L) + c - c] + (1 - \alpha)[f(e, x_H, x_L) - f(e, x_H + 1, x_L)] \\
\leq \; & 0,
\end{aligned}$$

is true by the monotonicity of $f(e, x_H, x_L)$ in x_H.

Monotonicity in x_H *Preserved by* T_{ADM_i}

Let $A(e, x_H, x_L)$ be the optimal action for admission control in state (e, x_H, x_L), define it as follows;

$$A(e, x_H, x_L) = \begin{cases} 0: & \text{asymptomatic patient is rejected in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is accepted in } (e, x_H, x_L) \end{cases}$$

We also define

$$AA_{x_H}(e, x_H, x_L) = \{(x, y) : x \in A(e, x_H, x_L), y \in A(e, x_H + 1, x_L)\}.$$

We consider when there is an arrival of asymptomatic patient and there are four possible cases due to the actions.

Case 1: $AA_{x_H} = \{(0, 0)\}$,

$$f(e, x_H, x_L) + r_{L_i} - f(e, x_H + 1, x_L) - r_{L_i} \leq 0,$$

holds due to the monotonicity of $f(e, x_H, x_L)$ in x_H.

Case 2: $AA_{x_H} = \{(0, 1)\}$,

Since an asymptomatic patient is rejected in state (e, x_H, x_L),

$$f(e, x_H, x_L) + r_{L_i} \leq f(e, x_H, x_L + 1).$$

By monotonicity of $f(e, x_H, x_L)$ in x_H,

$$f(e, x_H, x_L + 1) \leq f(e, x_H + 1, x_L + 1).$$

So

$$f(e, x_H, x_L) + r_{L_i} \leq f(e, x_H, x_L + 1) \leq f(e, x_H + 1, x_L + 1).$$

Therefore,

$$f(e, x_H, x_L) + r_{L_i} - f(e, x_H + 1, x_L + 1) \leq 0.$$

Case 3: $AA_{x_H} = \{(1, 0)\}$,

Admission of an asymptomatic patient in state (e, x_H, x_L) implies that

$$f(e, x_H, x_L + 1) \leq f(e, x_H, x_L) + r_{L_i}.$$

By monotonicity of $f(e, x_H, x_L)$ in x_H,

$$f(e, x_H, x_L) \leq f(e, x_H + 1, x_L).$$

So

$$f(e, x_H, x_L + 1) \leq f(e, x_H, x_L) + r_{L_i} \leq f(e, x_H + 1, x_L) + r_{L_i}.$$

Therefore,

$$f(e, x_H, x_L + 1) \leq f(e, x_H + 1, x_L) + r_{L_i}.$$

Case 4: $AA_{x_H} = \{(1, 1)\}$,
By monotonicity of $f(e, x_H, x_L)$ in x_H,

$$f(e, x_H, x_L + 1) - f(e, x_H + 1, x_L + 1) \leq 0,$$

Hence, the operator, T_{ADM_i}, preserves the monotonicity of a function f in x_H.

Monotonicity in x_H Preserved by T_{SCH}

We define $P(e, x_H, x_L)$ be the optimal action for scheduling control in state (e, x_H, x_L), such that;

$$P(e, x_H, x_L) = \begin{cases} 0: & \text{symptomatic patient is scheduled in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is scheduled in } (e, x_H, x_L) \end{cases}$$

We also define

$$PP_{x_H}(e, x_H, x_L) = \{(x, y) : x \in P(e, x_H, x_L), y \in P(e, x_H + 1, x_L)\}.$$

We consider scheduling of a patient and there are four possible cases due to the actions.

Case 1: $PP_{x_H} = \{(0,0)\}$,

By monotonicity of $f(e, x_H, x_L)$ in x_H,

$$f(e, x_H - 1, x_L) - f(e, x_H, x_L) \leq 0.$$

Case 2: $PP_{x_H} = \{(0,1)\}$,

Since symptomatic patient is served in state (e, x_H, x_L),

$$f(e, x_H - 1, x_L) \leq p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1).$$

By monotonicity of $f(e, x_H, x_L)$ in x_H,

$$f(e, x_H, x_L - 1) \leq f(e, x_H + 1, x_L - 1).$$

Therefore,

$$\begin{array}{c} p_{e,e-1} f(e-1, x_H, x_L - 1) \\ +(1 - p_{e,e-1}) f(e, x_H, x_L - 1) \end{array} \leq \begin{array}{c} p_{e,e-1} f(e-1, x_H + 1, x_L - 1) \\ +(1 - p_{e,e-1}) f(e, x_H + 1, x_L - 1). \end{array}$$

So,

$$f(e, x_H - 1, x_L) - p_{e,e-1} f(e-1, x_H + 1, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H + 1, x_L - 1) \leq 0.$$

Case 3: $PP_{x_H} = \{(1,0)\}$,

Since asymptomatic patient is served in state (e, x_H, x_L),

$$p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) \leq f(e, x_H - 1, x_L).$$

By monotonicity of $f(e, x_H, x_L)$ in x_H,

$$f(e, x_H - 1, x_L) \leq f(e, x_H, x_L).$$

Therefore,

$$p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) \leq f(e, x_H - 1, x_L) \leq f(e, x_H, x_L).$$

So,

$$p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) - f(e, x_H, x_L) \leq 0.$$

Case 4: $PP_{x_H} = \{(1,1)\}$,

$$\begin{aligned}
& p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) \\
& -p_{e,e-1} f(e-1, x_H + 1, x_L - 1) - (1 - p_{e,e-1}) f(e, x_H + 1, x_L - 1) \\
= \; & p_{e,e-1}[f(e-1, x_H, x_L - 1) - f(e-1, x_H + 1, x_L - 1)] \\
& +(1 - p_{e,e-1})[f(e, x_H, x_L - 1) - f(e, x_H + 1, x_L - 1)] \\
\leq \; & 0,
\end{aligned}$$

is valid by monotonicity of $f(e, x_H, x_L)$ in x_H. We show that T_{SCH} preserves monotonocity of a function f in x_H.

Monotonicity in x_H Preserved by T_{DET}

$$\begin{aligned}
& \tau f(e+1, x_H, x_L) + (1 - \tau) f(e, x_H, x_L) - \tau f(e+1, x_H + 1, x_L) - (1 - \tau) f(e, x_H + 1, x_L) \\
= \; & \tau[f(e+1, x_H, x_L) - f(e+1, x_H + 1, x_L)] + (1 - \tau)[f(e, x_H, x_L) - f(e, x_H + 1, x_L)] \\
\leq \; & 0,
\end{aligned}$$

is true due to the monotonicity of $f(e, x_H, x_L)$ in x_H.

A.1.0.1 Monotonicity in x_H Preserved by T_{UNIF}

Since, T_{UNIF} is the sum of T_{ARR_H}, T_{ADM_i}, T_{SCH} and T_{DET} which preserve monotonicity of $f(e, x_H, x_L)$ in x_H, T_{UNIF} preserves that monotonicity property too.

Monotonicity in x_H Preserved by T_{COST}

$$c_H x_H + c_L x_L - c_H(x_H + 1) - c_L x_L \leq 0,$$

since c_H is non-negative.
So h is non-decreasing in x_H which implies that T_{COST} preserves monotonicity property too.

A.2 Monotonicity in x_L

We will prove that all operators defined in Chapter 3, preserve the monotonicity property of any monotone function f in x_L, such that the following inequality,

$$Tf(e, x_H, x_L) \leq Tf(e, x_H, x_L + 1)$$

holds for any non-decreasing function f in x_L and for all operators T.
Proof is similar to the previous one.

A.3 Monotonicity in e

We will prove that all operators defined in Chapter 3, preserve the monotonicity property of any monotone function f in e, such that the following inequality,

$$Tf(e, x_H, x_L) \leq Tf(e + 1, x_H, x_L)$$

holds for any non-decreasing function f in e and for all operators T.

Monotonicity in e Preserved by $\bar{\lambda}_H T_{ARR_H}$

$$
\begin{aligned}
& \lambda_H(e)(f(e, x_H+1, x_L)+c) + (\bar{\lambda}_H - \lambda_H(e)) f(e, x_H, x_L) \\
& -\lambda_H(e+1)(f(e+1, x_H+1, x_L)+c) - (\bar{\lambda}_H - \lambda_H(e+1)) f(e+1, x_H, x_L) \\
= \;& \bar{\lambda}_H[f(e, x_H, x_L) - f(e+1, x_H, x_L)] \\
& +\lambda_H(e)[f(e, x_H+1, x_L)+c - f(e, x_H, x_L)] \\
& -\lambda_H(e+1)[f(e+1, x_H+1, x_L)+c - f(e+1, x_H, x_L)] \\
\leq \;& \bar{\lambda}_H[f(e, x_H, x_L) - f(e+1, x_H, x_L)] \\
& +\lambda_H(e)[f(e, x_H+1, x_L)+c - f(e, x_H, x_L)] \\
& -\lambda_H(e)[f(e+1, x_H+1, x_L)+c - f(e+1, x_H, x_L)] \\
= \;& (\bar{\lambda}_H - \lambda_H(e))[f(e, x_H, x_L) - f(e+1, x_H, x_L)] \\
& +\lambda_H(e)[f(e, x_H+1, x_L)+c - f(e+1, x_H+1, x_L) - c] \\
\leq \;& 0.
\end{aligned}
$$

The first inequality holds since $\lambda_H(e) < \lambda_H(e+1)$ and the second inequality holds by the monotonicity of f in e.

Monotonicity in e Preserved by T_{ADM_i}

Let $A(e, x_H, x_L)$ be the optimal action for admission control in state (e, x_H, x_L), define it as follows;

$$
A(e, x_H, x_L) = \begin{cases} 0: & \text{asymptomatic patient is rejected in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is accepted in } (e, x_H, x_L) \end{cases}
$$

We also define

$$
AA_e(e, x_H, x_L) = \{(x, y) : x \in A(e, x_H, x_L), y \in A(e+1, x_H, x_L)\}.
$$

We consider when there is an arrival of asymptomatic patient and there are four possible cases due to the actions.

Case 1: $AA_e = \{(0,0)\}$,

$$f(e, x_H, x_L) + r_{L_i} - f(e+1, x_H, x_L) - r_{L_i} \leq 0,$$

holds by the monotonicity of f in e.

Case 2: $AA_e = \{(0,1)\}$,

$$f(e, x_H, x_L) + r_{L_i} - f(e+1, x_H, x_L+1) \leq f(e, x_H, x_L+1) - f(e+1, x_H, x_L+1) \leq 0.$$

The first inequality holds since in state (e, x_H, x_L) the patient is rejected. That is:

$$f(e, x_H, x_L) + r_{L_i} \leq f(e, x_H, x_L+1),$$

and second inequality holds by monotonicity of f in e.

Case 3: $AA_e = \{(1,0)\}$,

$$f(e, x_H, x_L+1) - f(e+1, x_H, x_L) - r_{L_i} leq f(e, x_H, x_L) + r_{L_i} - f(e+1, x_H, x_L) - r_{L_i} \leq 0.$$

The first inequality holds since in state (e, x_H, x_L) the patient is admitted. That is;

$$f(e, x_H, x_L+1) \leq f(e, x_H, x_L) + r_{L_i},$$

and the second inequality holds by monotonicity of f in e.

Case 4: $AA_e = \{(1,1)\}$,

$$f(e, x_H, x_L+1) - f(e+1, x_H, x_L+1) \leq 0,$$

is true due to monotonicity of f in e.

Thus, T_{ADM_i} is non-decreasing in e.

Monotonicity in e Preserved by T_{SCH}

We define $P(e, x_H, x_L)$ be the optimal action for scheduling control in state (e, x_H, x_L), such that;

$$P(e, x_H, x_L) = \begin{cases} 0: & \text{symptomatic patient is scheduled in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is scheduled in } (e, x_H, x_L) \end{cases}$$

We also define

$$PP_e(e, x_H, x_L) = \{(x, y) : x \in P(e, x_H, x_L), y \in P(e+1, x_H, x_L)\}.$$

We consider scheduling of a patient and there are four possible cases due to the actions.

Case 1: $PP_e = \{(0, 0)\}$,

$$f(e, x_H - 1, x_L) - f(e+1, x_H - 1, x_L) \leq 0,$$

holds by monotonicity of f in e.

Case 2: $PP_e = \{(0, 1)\}$,

$$\begin{aligned} & f(e, x_H - 1, x_L) - p_{e+1,e} f(e, x_H, x_L - 1) - (1 - p_{e+1,e}) f(e+1, x_H, x_L - 1) \\ \leq\ & p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) \\ & - p_{e+1,e} f(e, x_H, x_L - 1) - (1 - p_{e+1,e}) f(e+1, x_H, x_L - 1) \end{aligned}$$

$$\begin{aligned} =\ & p_{e,e-1}[f(e-1, x_H, x_L - 1) - f(e, x_H, x_L - 1)] \\ & + (1 - p_{e+1,e})[f(e, x_H, x_L - 1) - f(e+1, x_H, x_L - 1)] \\ \leq\ & 0. \end{aligned}$$

First inequality holds since in state (e, x_H, x_L) symptomatic patient is served ($f(e, x_H - 1, x_L) \leq p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1)$).

Case 3: $PP_e = \{(1,0)\}$,

$$\begin{aligned} & p_{e,e-1}f(e-1,x_H,x_L-1) + (1-p_{e,e-1})f(e,x_H,x_L-1) - f(e+1,x_H-1,x_L) \\ \leq\ & f(e,x_H-1,x_L) - f(e+1,x_H-1,x_L)] \\ \leq\ & 0, \end{aligned}$$

First inequality holds since in state (e,x_H,x_L) asymptomatic patient is served ($p_{e,e-1}f(e-1,x_H,x_L-1) + (1-p_{e,e-1})f(e,x_H,x_L-1) \leq f(e,x_H-1,x_L)$).

Case 4: $PP_e = \{(1,1)\}$,

$$\begin{aligned} & p_{e,e-1}f(e-1,x_H,x_L-1) + (1-p_{e,e-1})f(e,x_H,x_L-1) \\ & -p_{e+1,e}f(e,x_H,x_L-1) - (1-p_{e+1,e})f(e+1,x_H,x_L-1) \\ =\ & p_{e,e-1}[f(e-1,x_H,x_L-1) - f(e,x_H,x_L-1)] \\ & +(1-p_{e+1,e})[f(e,x_H,x_L-1) - f(e+1,x_H,x_L-1)] \\ \leq\ & 0, \end{aligned}$$

holds by monotonicity of f in e where $0 \leq p \leq 1$.

Therefore, the operator, T_{SCH}, will be non-decreasing in e.

Monotonicity in x_H Preserved by $\bar{\gamma}T_{DET}$

$$\begin{aligned} & \gamma_{e,e+1}f(e+1,x_H,x_L) + (\bar{\gamma}-\gamma_{e,e+1})f(e,x_H,x_L) \\ & -\gamma_{e+1,e+2}f(e+2,x_H,x_L) - (\bar{\gamma}-\gamma_{e+1,e+2})f(e+1,x_H,x_L) \\ =\ & (\bar{\gamma}-\gamma_{e,e+1})[f(e,x_H,x_L) - f(e+1,x_H,x_L)] \\ & -\gamma_{e+1,e+2}[f(e+2,x_H,x_L) - f(e+1,x_H,x_L)] \\ \leq\ & 0, \end{aligned}$$

holds by monotonicity of f in e.

Therefore, $T_{DET}f$ is non-decreasing in e if f is a non-decreasing function of e.

Monotonicity in x_H Preserved by T_{UNIF}

Since, T_{UNIF} is the sum of T_{ARR_H}, T_{ADM_i}, T_{SCH} and T_{DET} which preserve monotonicity of $f(e, x_H, x_L)$ in e, T_{UNIF} preserves that monotonicity property too.

Monotonicity in x_H Preserved by T_{COST}

$$c_H x_H + c_L x_L - c_H x_H - c_L x_L \leq 0,$$

by simple algebra.

Since h is not a function of e, T_{COST} preserves monotonicity property too.

A.4 *Dec*(p)

We assume that $\underline{p} \leq \overline{p}$. Let $\underline{f}(e, x_H, x_L)$ be the corresponding function for $\underline{p}$, and $\overline{f}(e, x_H, x_L)$ be the corresponding function for $\overline{p}$. We assume that $\underline{f}$ and $\overline{f}$ satisfies the following inequality,

$$\underline{f}(e, x_H, x_L) - \overline{f}(e, x_H, x_L) \geq 0 \tag{A.1}$$

for any state (e, x_H, x_L) which implies *Dec*(p) property.

We will show that

$$T\underline{f}(e, x_H, x_L) - T\overline{f}(e, x_H, x_L) \geq 0$$

holds for any functions $\underline{f}$, $\overline{f}$ satisfying *Dec*(p) property and for all operators T.

Dec(p) Property Preserved by T_{ARR_H}

$$\begin{aligned} & \alpha \underline{f}(e, x_H + 1, x_L) + (1-\alpha)\underline{f}(e, x_H, x_L) \\ & -\alpha \overline{f}(e, x_H + 1, x_L) - (1-\alpha)\overline{f}(e, x_H, x_L) \\ = \; & \alpha[\underline{f}(e, x_H + 1, x_L) - \overline{f}(e, x_H + 1, x_L)] \\ & +(1-\alpha)[\underline{f}(e, x_H, x_L) - \overline{f}(e, x_H, x_L)] \\ \geq \; & 0, \end{aligned}$$

is true due to the $Dec(p)$ property for the state (e, x_H, x_L).

Dec(p) Property Preserved by T_{ADM_i}

Let $\underline{A}(e, x_H, x_L)(\overline{A}(e, x_H, x_L))$ be the optimal action for admission control in state (e, x_H, x_L) for probability $\underline{p}(\overline{p})$, define it as follows;

$$A(e, x_H, x_L)(\overline{A}(e, x_H, x_L)) = \begin{cases} 0: & \text{asymptomatic patient is rejected in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is accepted in } (e, x_H, x_L) \end{cases}$$

We also define

$$\underline{A}\overline{A}_{x_H}(e, x_H, x_L) = \{(x, y) : x \in \underline{A}(e, x_H, x_L), y \in \overline{A}(e, x_H, x_L)\}.$$

We consider when there is an arrival of asymptomatic patient and there are four possible cases due to the actions.

Case 1: $\underline{A}\overline{A}_{x_H} = \{(0, 0)\}$,

$$\underline{f}(e, x_H, x_L) + r_{L_i} - \overline{f}(e, x_H + 1, x_L) - r_{L_i} \geq 0,$$

holds due to the $Dec(p)$ property for the state (e, x_H, x_L).

Case 2: $\underline{A}\overline{A}_{x_H} = \{(0, 1)\}$,

Since an asymptomatic patient is admitted in $\overline{A}(e, x_H, x_L)$,

$$\overline{f}(e, x_H, x_L) + r_{L_i} \geq \overline{f}(e, x_H, x_L + 1).$$

By the $Dec(p)$ property for the state (e, x_H, x_L),

$$\underline{f}(e, x_H, x_L) \geq \overline{f}(e, x_H, x_L).$$

$$\underline{f}(e, x_H, x_L) + r_{L_i} \geq \overline{f}(e, x_H, x_L) + r_{L_i} \geq \overline{f}(e, x_H, x_L + 1).$$

Therefore,

$$\underline{f}(e, x_H, x_L) + r_{L_i} - \overline{f}(e, x_H, x_L + 1) \geq 0.$$

Case 3: $\underline{A}\overline{A}_{x_H} = \{(1, 0)\}$,

Rejection of an asymptomatic patient in $\overline{A}(e, x_H, x_L)$ implies that

$$\overline{f}(e, x_H, x_L + 1) \geq \overline{f}(e, x_H, x_L) + r_{L_i}.$$

By the $Dec(p)$ property for the state (e, x_H, x_L),

$$\underline{f}(e, x_H, x_L + 1) \geq \overline{f}(e, x_H, x_L + 1).$$

So

$$\underline{f}(e, x_H, x_L + 1) \geq \overline{f}(e, x_H, x_L + 1) \geq \overline{f}(e, x_H, x_L) + r_{L_i}.$$

Therefore,

$$\underline{f}(e, x_H, x_L + 1) - \overline{f}(e, x_H, x_L) - r_{L_i} \geq 0.$$

Case 4: $\underline{A}\overline{A}_{x_H} = \{(1, 1)\}$,

By the $Dec(p)$ property for the state $(e, x_H, x_L + 1)$,

$$\underline{f}(e, x_H, x_L + 1) - \overline{f}(e, x_H, x_L + 1) \geq 0,$$

Hence, the operator, T_{ADM_i}, preserves the property in (A.1).

Dec(p) Property Preserved by T_{SCH}

We define $\underline{P}(e, x_H, x_L)(\overline{P}(e, x_H, x_L))$ be the optimal action for scheduling control in state (e, x_H, x_L) for $\underline{p}(\overline{p})$, such that;

$$\underline{P}(e, x_H, x_L)(\overline{P}(e, x_H, x_L)) = \begin{cases} 0: & \text{symptomatic patient is scheduled in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is scheduled in } (e, x_H, x_L) \end{cases}$$

We also define

$$\underline{P}\overline{P}_{x_H}(e, x_H, x_L) = \{(x, y) : x \in \underline{P}(e, x_H, x_L), y \in \overline{P}(e, x_H, x_L)\}.$$

We consider scheduling of a patient and there are four possible cases due to the actions.

Case 1: $\underline{P}\overline{P}_{x_H} = \{(0, 0)\}$,

By the $Dec(p)$ property for the state $(e, x_H - 1, x_L)$,

$$\underline{f}(e, x_H - 1, x_L) - \overline{f}(e, x_H - 1, x_L) \geq 0.$$

Case 2: $\underline{P}\overline{P}_{x_H} = \{(0, 1)\}$,

Since asymptomatic patient is served $\overline{P}(e, x_H, x_L)$,

$$\overline{f}(e, x_H - 1, x_L) \geq \overline{p}\overline{f}(e - 1, x_H, x_L - 1) + (1 - \overline{p})\overline{f}(e, x_H, x_L - 1).$$

By the $Dec(p)$ property for the state $(e, x_H - 1, x_L)$,

$$\underline{f}(e, x_H - 1, x_L) \geq \overline{f}(e, x_H - 1, x_L).$$

Thus,

$$\underline{f}(e, x_H - 1, x_L) \geq \overline{f}(e, x_H - 1, x_L) \geq \overline{p}\overline{f}(e - 1, x_H, x_L - 1) + (1 - \overline{p})\overline{f}(e, x_H, x_L - 1),$$

which leads to

$$\underline{f}(e, x_H - 1, x_L) - \overline{p}\overline{f}(e-1, x_H, x_L - 1) - (1-\overline{p})\overline{f}(e, x_H, x_L - 1) \geq 0.$$

Case 3: $\underline{P}\overline{P}_{x_H} = \{(1,0)\}$,
Since symptomatic patient is served in state $\overline{P}(e, x_H, x_L)$,

$$\overline{p}\overline{f}(e-1, x_H, x_L - 1) + (1-\overline{p})\overline{f}(e, x_H, x_L - 1) \geq \overline{f}(e, x_H - 1, x_L).$$

Hence,

$$\begin{aligned} & \underline{p}\underline{f}(e-1, x_H, x_L - 1) + (1-\underline{p})\underline{f}(e, x_H, x_L - 1) - \overline{f}(e, x_H - 1, x_L) \\ \geq\ & \underline{p}\underline{f}(e-1, x_H, x_L - 1) + (1-\underline{p})\underline{f}(e, x_H, x_L - 1) \\ & -\overline{p}\overline{f}(e-1, x_H, x_L - 1) - (1-\overline{p})\overline{f}(e, x_H, x_L - 1) \end{aligned}$$

$$\begin{aligned} =\ & \underline{p}(\underline{f}(e-1, x_H, x_L - 1) - \overline{f}(e-1, x_H, x_L - 1)) \\ & +(\overline{p} - \underline{p})(\overline{f}(e, x_H, x_L - 1) - \overline{f}(e-1, x_H, x_L - 1)) \\ & +(1-\underline{p})(\underline{f}(e, x_H, x_L - 1) - \overline{f}(e, x_H, x_L - 1)) \\ \geq\ & 0, \end{aligned}$$

holds by $Inc(e)$ property and by the $Dec(p)$ property for the state $(e-1, x_H, x_L - 1)$ and $(e, x_H, x_L - 1)$.

Case 4: $\underline{P}\overline{P}_{x_H} = \{(1,1)\}$,

$$\begin{aligned} & \underline{p}\underline{f}(e-1, x_H, x_L - 1) + (1-\underline{p})\underline{f}(e, x_H, x_L - 1) \\ & -\overline{p}\overline{f}(e-1, x_H, x_L - 1) - (1-\overline{p})\overline{f}(e, x_H, x_L - 1 \end{aligned}$$

$$
\begin{aligned}
&= \underline{p}(\underline{f}(e-1, x_H, x_L-1) - \overline{f}(e-1, x_H, x_L-1)) \\
&\quad +(\overline{p}-\underline{p})(\overline{f}(e, x_H, x_L-1) - \overline{f}(e-1, x_H, x_L-1)) \\
&\quad +(1-\underline{p})(\underline{f}(e, x_H, x_L-1) - \overline{f}(e, x_H, x_L-1)) \\
&\geq 0,
\end{aligned}
$$

is valid by $Inc(e)$ property and by the $Dec(p)$ property for the state $(e-1, x_H, x_L-1)$ and (e, x_H, x_L-1).

We show that T_{SCH} preserves monotonocity of a function f in p.

$Dec(p)$ Property Preserved by T_{DET}

$$
\begin{aligned}
&\tau\underline{f}(e+1, x_H, x_L) + (1-\tau)\underline{f}(e, x_H, x_L) - \tau\overline{f}(e+1, x_H, x_L) - (1-\tau)\overline{f}(e, x_H, x_L) \\
= \;& \tau[\underline{f}(e+1, x_H, x_L) - \overline{f}(e+1, x_H, x_L)] + (1-\tau)[\underline{f}(e, x_H, x_L) - \overline{f}(e, x_H, x_L)] \\
\geq \;& 0,
\end{aligned}
$$

is true due to the $Dec(p)$ property for the state (e, x_H, x_L) and $(e+1, x_H, x_L)$.

$Dec(p)$ Property Preserved by T_{UNIF}

Since, T_{UNIF} is the sum of T_{ARR}, T_{FIC}, T_{ADM}, T_{SCH} and T_{DET} which preserve the property of $Dec(p)$, T_{UNIF} preserves that property too.

$Dec(p)$ Property Preserved by T_{COST}

$$c_H x_H + c_L x_L - c_H x_H - c_L x_L = 0,$$

which implies that T_{COST} preserves property of $Dec(p)$ too.

A.5 Monotonicity on the Diagonal

In this proof, we show that all operators preserve monotonicity on the diagonal property of function f. More explicitly, we will show the inequality for (x_H, x_L):

$$Tf(e, x_H - 1, x_L) \leq Tf(e, x_H, x_L - 1)$$

for any function f with the monotonicity on the diagonal property and for all operators T.

Monotonicity on the Diagonal Preserved by T_{ARR_H}

$$\begin{aligned}
& \alpha[f(e, x_H, x_L) + c] + (1-\alpha) f(e, x_H - 1, x_L) \\
& -\alpha[f(e, x_H + 1, x_L - 1) + c] + (1-\alpha) f(e, x_H, x_L - 1) \\
= \; & \alpha[f(e, x_H, x_L) + c - f(e, x_H + 1, x_L - 1) - c] \\
& +(1-\alpha)[f(e, x_H - 1, x_L) - f(e, x_H, x_L - 1)] \\
\leq \; & 0
\end{aligned}$$

holds by monotonicity on the diagonal of f in $(x_H + 1, x_L)$ and (x_H, x_L).

Monotonicity on the diagonal Preserved by T_{ADM_i}

Let $A(e, x_H, x_L)$ be the optimal action for admission control in state (e, x_H, x_L), define it as follows;

$$A(e, x_H, x_L) = \begin{cases} 0: & \text{asymptomatic patient is rejected in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is accepted in } (e, x_H, x_L) \end{cases}$$

We also define

$$AA_{Diag}(e, x_H, x_L) = \{(x, y) : x \in A(e, x_H - 1, x_L), y \in A(e, x_H, x_L - 1)\}.$$

We have several cases due to different optimal actions in states (e, x_H-1, x_L) and (e, x_H, x_L-1).
Case 1: $AA_{Diag} = \{(0,0)\}$,

$$\begin{aligned} & f(e, x_H - 1, x_L) + r_{L_i} - f(e, x_H, x_L - 1) - r_{L_i} \\ = \;& f(e, x_H - 1, x_L) - f(e, x_H, x_L - 1) \\ \leq \;& 0, \end{aligned}$$

holds by monotonicity on the diagonal property of f in (x_H, x_L).
Case 2: $AA_{Diag} = \{(0,1)\}$,

$$\begin{aligned} & f(e, x_H - 1, x_L) + r_{L_i} - f(e, x_H, x_L) \\ \leq \;& f(e, x_H - 1, x_L + 1) - f(e, x_H, x_L) \\ \leq \;& 0, \end{aligned}$$

First inequality holds since in state (e, x_H-1, x_L) the patient is rejected ($f(e, x_H-1, x_L)+r_{L_i} \leq f(e, x_H - 1, x_L + 1)$), and last inequality holds by monotonicity on the diagonal property of f in $(x_H, x_L + 1)$.
Case 3: $AA_{Diag} = \{(1,0)\}$,

$$\begin{aligned} & f(e, x_H - 1, x_L + 1) - f(e, x_H, x_L - 1) - r_{L_i} \\ \leq \;& f(e, x_H - 1, x_L) + r - f(e, x_H, x_L - 1) - r_{L_i} \\ = \;& f(e, x_H - 1, x_L) - f(e, x_H, x_L - 1) \\ \leq \;& 0, \end{aligned}$$

First inequality holds since in state (e, x_H-1, x_L) the patient is admitted ($f(e, x_H-1, x_L+1) \leq f(e, x_H - 1, x_L) + r_{L_i}$), and last inequality holds by monotonicity on the diagonal property of f in (x_H, x_L).

Case 4: $AA_{Diag} = \{(1,1)\}$,

$$f(e, x_H - 1, x_L + 1) - f(e, x_H, x_L) \leq 0,$$

holds by monotonicity on the diagonal property of f in $(x_H, x_L + 1)$.

Monotonicity on the Diagonal Preserved by T_{SCH}

We define $P(e, x_H, x_L)$ be the optimal action for scheduling control in state (e, x_H, x_L), such that;

$$P(e, x_H, x_L) = \begin{cases} 0: & \text{symptomatic patient is scheduled in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is scheduled in } (e, x_H, x_L) \end{cases}$$

We also define

$$PP_{Diag}(e, x_H, x_L) = \{(x, y) : x \in P(e, x_H - 1, x_L), y \in P(e, x_H, x_L - 1)\}.$$

Four possible cases arise with respect to the different optimal actions .

Case 1: $PP_{Diag} = \{(0,0)\}$,

$$f(e, x_H - 2, x_L) - f(1, x_H - 1, x_L - 1) \leq 0,$$

holds by monotonicity on the diagonal property of f in $(x_H - 1, x_L)$.

Case 2: $PP_{Diag} = \{(0,1)\}$,

$$\begin{aligned} & f(e, x_H - 2, x_L) - p_{e,e-1} f(e-1, x_H, x_L - 2) - (1 - p_{e,e-1}) f(e, x_H, x_L - 2) \\ \leq\ & p_{e,e-1} f(e-1, x_H - 1, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H - 1, x_L - 1) \\ -\ & p_{e,e-1} f(e-1, x_H, x_L - 2) - (1 - p_{e,e-1}) f(e, x_H, x_L - 2) \end{aligned}$$

$$\begin{aligned} &= p_{e,e-1}[f(e-1,x_H-1,x_L-1)-f(e-1,x_H,x_L-2)] \\ &+ (1-p_{e,e-1})[f(e,x_H-1,x_L-1)-f(e,x_H,x_L-2)] \\ &\leq 0, \end{aligned}$$

The first inequailty holds since in state (e, x_H-1, x_L) symptomatic patient is screened ($f(e, x_H-2, x_L) \leq p_{e,e-1}f(e-1, x_H-1, x_L-1) + (1-p_{e,e-1})f(e, x_H-1, x_L-1)$), and last inequality holds by monotonicity on the diagonal of f in (x_H, x_L-1).

Case 3: $PP_{Diag} = \{(1,0)\}$,

$$\begin{aligned} & p_{e,e-1}f(e-1,x_H-1,x_L-1)+(1-p_{e,e-1})f(e,x_H-1,x_L-1)-f(e,x_H-1,x_L-1) \\ = \; & p_{e,e-1}[f(e-1,x_H-1,x_L-1)-f(e,x_H-1,x_L-1)] \\ & +f(e,x_H-1,x_L-1)-f(e,x_H-1,x_L-1) \\ \leq \; & 0, \end{aligned}$$

is valid due to the monotonicity of f in e.

Case 4: $PP_{Diag} = \{(1,1)\}$,

$$\begin{aligned} & p_{e,e-1}f(e-1,x_H-1,x_L-1)+(1-p_{e,e-1})f(e,x_H-1,x_L-1) \\ - \; & p_{e,e-1}f(e-1,x_H,x_L-2)-(1-p_{e,e-1})f(e,x_H,x_L-2) \\ = \; & p_{e,e-1}[f(e-1,x_H-1,x_L-1)-f(e-1,x_H,x_L-2)] \\ + \; & (1-p_{e,e-1})[f(e,x_H-1,x_L-1)-f(e,x_H,x_L-2)] \\ \leq \; & 0, \end{aligned}$$

holds by monotonicity on the diagonal property of f in (x_H, x_L-1).

Monotonicity on the diagonal Preserved by T_{DET}

$$\begin{aligned} & \tau f(e+1, x_H-1, x_L) + (1-\tau) f(e, x_H-1, x_L) - \tau f(e+1, x_H, x_L-1) - (1-\tau) f(e, x_H, x_L-1) \\ = \; & \tau[f(e+1, x_H-1, x_L) - f(e+1, x_H, x_L-1)] + (1-\tau)[f(e, x_H-1, x_L) - f(e, x_H, x_L-1)] \\ \leq \; & 0, \end{aligned}$$

holds due to monotonicity on the diagonal of f in (x_H, x_L).

Monotonicity on the diagonal Preserved by T_{UNIF}

Since, T_{UNIF} is the sum of T_{ARR_H}, T_{ADM_i}, T_{SCH} and T_{DET} which preserve the monotonicity on the diagonal property of f, then T_{UNIF} preserves this property, too.

Monotonicity on the diagonal Preserved by T_{COST}

For holding costs,

$$\begin{aligned} h(e, x_H-1, x_L) - h(e, x_H, x_L-1) &= c_H x_H - c_H + c_L x_L - c_H x_H - c_L x_L + c_L \\ &= -c_H + c_L \leq 0 \end{aligned}$$

holds since $c_H \geq c_L$.

T_{COST}, being a sum of T_{UNIF} and holding costs, preserves the monotonicity of f on the diagonal.

A.6 $IDiag_e$ Property

In this proof, we show that all operators preserve the following property of function f for (x_H, x_L);

$$f(e-1, x_H, x_L-1) \leq f(e, x_H-1, x_L) \tag{A.2}$$

Explicitly, we will show the inequality:

$$Tf(e-1, x_H, x_L - 1) \leq Tf(e, x_H - 1, x_L)$$

for any function f satisfying Equation A.2 for all operators T.

$IDiag_e$ Property Preserved by T_{ARR_H}

$$\begin{aligned}
& \frac{\lambda_H(e-1)}{\bar{\lambda}_H}[f(e-1, x_H+1, x_L-1) + c] + (1 - \frac{\lambda_H(e-1)}{\bar{\lambda}_H}) f(e-1, x_H, x_L - 1) \\
& -\frac{\lambda_H(e)}{\bar{\lambda}_H}[f(e, x_H, x_L) + c] - (1 - \frac{\lambda_H(e)}{\bar{\lambda}_H}) f(e, x_H - 1, x_L) \\
= \; & [f(e-1, x_H, x_L - 1) - f(e, x_H - 1, x_L)] \\
& + \frac{\lambda_H(e-1)}{\bar{\lambda}_H}[f(e-1, x_H+1, x_L-1) - f(e-1, x_H, x_L-1)] \\
& + \frac{\lambda_H(e)}{\bar{\lambda}_H}[f(e, x_H - 1, x_L) - f(e, x_H, x_L)] + \frac{\lambda_H(e-1)}{\bar{\lambda}_H} c - \frac{\lambda_H(e)}{\bar{\lambda}_H} c \\
= \; & (1 - \frac{\lambda_H(e-1)}{\bar{\lambda}_H})[f(e-1, x_H, x_L - 1) - f(e, x_H - 1, x_L)] \\
& + \frac{\lambda_H(e-1)}{\bar{\lambda}_H}[f(e-1, x_H+1, x_L-1) - f(e, x_H, x_L)] \\
& + \frac{(\lambda_H(e) - \lambda_H(e-1))}{\bar{\lambda}_H}[f(e, x_H - 1, x_L) - f(e, x_H, x_L)] + \frac{(\lambda_H(e-1) - \lambda_H(e))}{\bar{\lambda}_H} c \\
\leq \; & 0,
\end{aligned}$$

holds due to $IDiag_e$ property of f on (x_H, x_L), and $(x_H + 1, x_L)$, and monotonicity of f in x_H respectively.

$IDiag_e$ Property Preserved by T_{ADM_i}

Let $A(e, x_H, x_L)$ be the optimal action for admission control in state (e, x_H, x_L), define it as follows;

$$A(e, x_H, x_L) = \begin{cases} 0: & \text{asymptomatic patient is rejected in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is accepted in } (e, x_H, x_L) \end{cases}$$

We also define

$$AA_{IDiag_e}(e, x_H, x_L) = \{(x, y) : x \in A(e-1, x_H, x_L-1), y \in A(e, x_H-1, x_L)\}.$$

There are four possible cases due to the actions on the states $(e-1, x_H, x_L-1)$, and (e, x_H-1, x_L).
Case 1: $AA_{IDiag_e} = \{(0,0)\}$,

$$\begin{aligned} & f(e-1, x_H, x_L-1) + r_{L_i} - f(e, x_H-1, x_L) - r_{L_i} \\ = \; & f(e-1, x_H, x_L-1) - f(e, x_H-1, x_L) \\ \leq \; & 0, \end{aligned}$$

holds due to the property of f in (x_H, x_L).
Case 2: $AA_{IDiag_e} = \{(0,1)\}$,

$$\begin{aligned} & f(e-1, x_H, x_L-1) + r_{L_i} - f(e, x_H-1, x_L+1) \\ \leq \; & f(e-1, x_H, x_L) - f(e, x_H-1, x_L+1) \\ = \; & f(e-1, x_H, x_L) - f(e, x_H-1, x_L+1) \\ \leq \; & 0, \end{aligned}$$

First inequality holds since in state $(e-1, x_H, x_L-1)$, the patient is rejected ($f(e-1, x_H, x_L-1) + r_{L_i} \leq f(e-1, x_H, x_L)$), and last inequality is due to the property of f in (x_H, x_L+1).
Case 3: $AA_{IDiag_e} = \{(1,0)\}$,

$$\begin{aligned} & f(e-1, x_H, x_L) - f(e, x_H-1, x_L) - r_{L_i} \\ \leq \; & f(e-1, x_H, x_L-1) + r_{L_i} - f(e, x_H-1, x_L) - r_{L_i} \\ = \; & f(e-1, x_H, x_L-1) - f(e, x_H-1, x_L) \\ \leq \; & 0, \end{aligned}$$

First inequality holds since in state $(e-1, x_H, x_L-1)$, the patient is admitted ($f(e-1, x_H, x_L) \leq f(e-1, x_H, x_L-1) + r_{L_i}$), and last inequality is due to the property of f in (x_H, x_L).

Case 4: $AA_{IDiag_e} = \{(1,1)\}$,

$$f(e-1, x_H, x_L) - f(e, x_H - 1, x_L + 1) \leq 0,$$

holds due to the property of f in $(x_H, x_L + 1)$.

$IDiag_e$ Property Preserved by T_{SCH}

We define $P(e, x_H, x_L)$ be the optimal action for scheduling control in state (e, x_H, x_L), such that;

$$P(e, x_H, x_L) = \begin{cases} 0: & \text{symptomatic patient is scheduled in } (e, x_H, x_L) \\ 1: & \text{asymptomatic patient is scheduled in } (e, x_H, x_L) \end{cases}$$

We also define

$$PP_{IDiag_e}(e, x_H, x_L) = \{(x, y) : x \in P(e-1, x_H, x_L - 1), y \in P(e, x_H - 1, x_L)\}.$$

We consider scheduling of a patient and there are four possible cases due to the actions.

Case 1: $PP_{IDiag_e} = \{(0,0)\}$,

$$f(e-1, x_H - 1, x_L - 1) - f(e, x_H - 2, x_L) \leq 0,$$

holds by the property of f in $(x_H - 1, x_L)$.

Case 2: $PP_{IDiag_e} = \{(0,1)\}$,

$$\begin{aligned} & f(e-1, x_H - 1, x_L - 1) - p_{e,e-1} f(e-1, x_H - 1, x_L - 1) - (1 - p_{e,e+1}) f(e, x_H - 1, x_L - 1) \\ \leq\ & (1 - p_{e,e-1})[f(e-1, x_H - 1, x_L - 1) - f(e, x_H - 1, x_L - 1)] \\ \leq\ & 0, \end{aligned}$$

holds due to monotonicity of f in e.

Case 3: $PP_{IDiag_e} = \{(1,0)\}$,

$$\begin{aligned} & p_{e-1,e-2} f(e-2, x_H, x_L - 2) + (1 - p_{e-1,e-2}) f(e-1, x_H, x_L - 2) - f(e, x_H - 2, x_L) \\ \leq \; & f(e-1, x_H - 1, x_L - 1) - f(e, x_H - 2, x_L) \\ \leq \; & 0. \end{aligned}$$

First inequality holds since in state $(e-1, x_H, x_L - 1)$, the asymptomatic patient is served $(p_{e-1,e-2} f(e-2, x_H, x_L - 2) + (1 - p_{e-1,e-2}) f(e-1, x_H, x_L - 2) \leq f(e-1, x_H - 1, x_L - 1))$, and last inequality is due to the property of f in $(x_H - 1, x_L)$.

Case 4: $PP_{IDiag_e} = \{(1,1)\}$,

$$\begin{aligned} & p_{e-1,e-2} f(e-2, x_H, x_L - 2) + (1 - p_{e-1,e-2}) f(e-1, x_H, x_L - 2) \\ & - p_{e,e-1} f(e-1, x_H - 1, x_L - 1) - (1 - p_{e,e-1}) f(e, x_H - 1, x_L - 1) \end{aligned}$$

$$\begin{aligned} \leq \; & f(e-1, x_H - 1, x_L - 1) - p_{e,e-1} f(e-1, x_H - 1, x_L - 1) - (1 - p_{e,e-1}) f(e, x_H - 1, x_L - 1) \\ \leq \; & (1 - p_{e,e-1})[f(e-1, x_H - 1, x_L - 1) - f(e, x_H - 1, x_L - 1)] \\ \leq \; & 0. \end{aligned}$$

First inequality holds since in state $(e-1, x_H, x_L - 1)$, the low patient is screened $(p_{e-1,e-2} f(e-2, x_H, x_L - 2) + (1 - p_{e-1,e-2}) f(e-1, x_H, x_L - 2) \leq f(e-1, x_H - 1, x_L - 1))$, and last inequality is due to monotonicity of f in e.

$IDiag_e$ Property Preserved by $\bar{\lambda}_H T_{ARR_H} + \bar{\gamma} T_{DET}$

T_{DET} alone does not preserve the mentioned property. However, if it is used with the operator T_{ARR_H}, then we could show that $\bar{\lambda}_H T_{ARR_H} + \bar{\gamma} T_{DET}$ preserves the property stated in Equation

A.2.

$$
\begin{aligned}
& \lambda_H(e-1)(f(e-1,x_H+1,x_L-1)) - \lambda_H(e)(f(e,x_H,x_L)) \\
& +(\bar{\lambda}_H - \lambda_H(e-1))f(e-1,x_H,x_L-1) \\
& -(\bar{\lambda}_H - \lambda_H(e))f(e,x_H-1,x_L) \\
& +\gamma_{e-1,e}f(e,x_H,x_L-1) - \gamma_{e,e+1}f(e+1,x_H-1,x_L) \\
& +(\bar{\gamma} - \gamma_{e-1,e})f(e-1,x_H,x_L-1) - (\bar{\gamma} - \gamma_{e,e+1})f(e,x_H-1,x_L) \\
= \; & \bar{\lambda}_H[f(e-1,x_H,x_L-1) - f(e,x_H-1,x_L)] \\
& +\lambda_H(e)[f(e,x_H-1,x_L) - f(e,x_H,x_L)] \\
& +\lambda_H(e-1)[f(e-1,x_H+1,x_L-1) - f(e-1,x_H,x_L-1)] \\
& +\bar{\gamma}[f(e-1,x_H,x_L-1) - f(e,x_H-1,x_L)] \\
& +\gamma_{e-1,e}[f(e,x_H,x_L-1) - f(e-1,x_H,x_L-1)] \\
& +\gamma_{e,e+1}[f(e,x_H-1,x_L) - f(e+1,x_H-1,x_L)] \\
= \; & (\bar{\lambda}_H - \lambda_H(e-1) - \gamma_{e-1,e})[f(e-1,x_H,x_L-1) - f(e,x_H-1,x_L)] \\
& +(\lambda_H(e) - \lambda_H(e-1) - \gamma_{e-1,e})[f(e,x_H-1,x_L) - f(e,x_H,x_L)] \\
& +\lambda_H(e-1)[f(e-1,x_H+1,x_L-1) - f(e,x_H,x_L)] \\
& +\bar{\gamma}[f(e-1,x_H,x_L-1) - f(e,x_H-1,x_L)] \\
& +\gamma_{e-1,e}[f(e,x_H,x_L-1) - f(e,x_H,x_L)] \\
& +\gamma_{e,e+1}[f(e,x_H-1,x_L) - f(e+1,x_H-1,x_L)] \\
\leq \; & 0. \qquad \text{(A.3)}
\end{aligned}
$$

$\lambda_H(e) - \lambda_H(e-1) \geq \gamma_{e-1,e}$ is satisfied by the assumption, and this also implies that $\bar{\lambda}_H - \lambda_H(e-1) \geq \gamma_{e-1,e}$.

Equation A.3 holds by $IDiag_e$ property of f on (x_H, x_L), monotonicity of f in x_H, $IDiag_e$ property of f on (x_H+1, x_L), $IDiag_e$ property of f on (x_H, x_L), monotonicity of f in x_L, and monotonicity of f in e respectively.

$IDiag_e$ Property Preserved by T_{UNIF}

Since, T_{UNIF} is the convex combination of T_{ARR_H}, T_{ADM_i}, T_{SCH} and T_{DET} which preserve the property of f, then T_{UNIF} preserves this property, too.

$IDiag_e$ Property Preserved by T_{COST}

In this part, we evaluate fixed cost c, and holding costs together.

$$\begin{aligned} & \lambda_H(e-1)c - \lambda_H(e)c - c_H x_H + c_L x_L - c_L - c_H x_H + c_H - c_L x_L \\ = \; & (\lambda_H(e-1) - \lambda_H(e))c - c_L + c_H \\ \leq \; & \frac{c_L - c_H}{c} c - c_L + c_H = 0, \end{aligned}$$

holds by the assumption.
T_{COST}, being a sum of T_{UNIF} and holding costs, preserves the property of f.

A.7 Convexity in x_L

In this proof, we show that all operators preserve convexity of function f. In other words, we will show that:

$$Tf(e, x_H, x_L) - Tf(e, x_H, x_L + 1) \geq Tf(e, x_H, x_L + 1) - Tf(e, x_H, x_L + 2)$$

holds for any convex function f for all operators T.

Convexity Preserved by T_{ARR_H}

$$\begin{aligned} & \alpha f(e, x_H + 1, x_L) + (1-\alpha) f(e, x_H, x_L) \\ & -2\alpha f(e, x_H + 1, x_L + 1) - 2(1-\alpha) f(e, x_H, x_L + 1) \\ & +\alpha f(e, x_H + 1, x_L + 2) + (1-\alpha) f(e, x_H, x_L + 2) \end{aligned}$$

$$\begin{aligned} &= \alpha[f(e,x_H+1,x_L)-2f(e,x_H+1,x_L+1)+f(e,x_H+1,x_L+2)] \\ &\quad +(1-\alpha)[f(e,x_H,x_L)-2f(e,x_H,x_L+1)-f(e,x_H,x_L+2)] \\ &\geq 0, \end{aligned}$$

is true due to the convexity of f in x_L. Hence, T_{ARR_H} preserves convexity of a function f.

Convexity Preserved by T_{ADM_i}

We would like to show that

$$\begin{matrix} min\{f(e,x_H,x_L)+r_{L_i},f(e,x_H,x_L+1)\} \\ -min\{f(e,x_H,x_L+1)+r_{L_i},f(e,x_H,x_L+2)\} \end{matrix} \geq \begin{matrix} min\{f(e,x_H,x_L+1)+r_{L_i},f(e,x_H,x_L+2)\} \\ -min\{f(e,x_H,x_L+2)+r_{L_i},f(e,x_H,x_L+3)\} \end{matrix}$$

We use coupling method. We let

$$\begin{aligned} \delta &= \min\{f(e,x_H,x_L)+r_{L_i},f(e,x_H,x_L+1)\} \\ &\quad -\min\{f(e,x_H,x_L+1)+r_{L_i},f(e,x_H,x_L+2)\} \\ &\quad -\min\{f(e,x_H,x_L+1)+r_{L_i},f(e,x_H,x_L+2)\} \\ &\quad +\min\{f(e,x_H,x_L+2)+r_{L_i},f(e,x_H,x_L+3)\}. \end{aligned}$$

We let systems A, B, C and D correspond to systems in states (e,x_H,x_L), (e,x_H,x_L+1),(e,x_H,x_L+1) and (e,x_H,x_L+2) in period n, respectively. We let system A and system D follow the optimal policy, and system B and system C imitate all the decisions of system A and system D.
First we assume that A and D accept to screen the patients. Since B and C imitate them, B and C accept the patients, too. So we have,

$$\begin{aligned} \delta &\geq f(e,x_H,x_L+1)-f(e,x_H,x_L+2) \\ &\quad -f(e,x_H,x_L+2)+f(e,x_H,x_L+3) \\ &= f(e,x_H,x_L+1)-2f(e,x_H,x_L+2)+f(e,x_H,x_L+3) \\ &\geq 0, \end{aligned}$$

which is true due to convexity of f.

Now, we assume that A and D reject the patients, and let B and C imitate them. Then we have,

$$\begin{aligned}\delta &\geq f(e,x_H,x_L)+r_{L_i}-f(e,x_H,x_L+1)-r_{L_i}\\ &\quad -f(e,x_H,x_L+1)-r_{L_i}+f(e,x_H,x_L+2)+r_{L_i}\\ &= f(e,x_H,x_L)-2f(e,x_H,x_L+1)+f(e,x_H,x_L+2)\\ &\geq 0\end{aligned}$$

which holds by convexity of f.

Now, we consider the cases where A and D make different decisions. We let system A accept to screen the patient and system D reject the patient. We assume system B imitates the decisions of system D, and system C imitates the decisions of system A.

$$\begin{aligned}\delta &\geq f(e,x_H,x_L+1)-f(e,x_H,x_L+1)-r_{L_i}\\ &\quad -f(e,x_H,x_L+2)+f(e,x_H,x_L+2)+r_{L_i}=0.\end{aligned}$$

Now, we let system A reject the patient and system D accept to screen the patient with the above assumption.

$$\begin{aligned}\delta &\geq f(e,x_H,x_L)+r_{L_i}-f(e,x_H,x_L+2)\\ &\quad -f(e,x_H,x_L+1)-r_{L_i}+f(e,x_H,x_L+3)\end{aligned}$$

$$\begin{aligned}&= f(e,x_H,x_L)-f(e,x_H,x_L+1)\\ &\quad -f(e,x_H,x_L+2)+f(e,x_H,x_L+3)\\ &\geq f(e,x_H,x_L+1)-f(e,x_H,x_L+2)\\ &\quad -f(e,x_H,x_L+2)+f(e,x_H,x_L+3)\\ &\geq 0,\end{aligned}$$

where the inequalities hold by convexity of f.

Convexity Preserved by T_{DEP}

There are two cases where $x_H = 0$ and $x_H \neq 0$.

Case 1: $x_H = 0$

To begin with, let

$$\Delta = \Psi(e, x_H, x_L - 1) - 2\Psi(e, x_H, x_L) + \Psi(e, x_H, x_L + 1).$$

For $e = 1$,

$$\Delta = f(1, x_H, x_L - 1) - 2f(1, x_H, x_L) + f(1, x_H, x_L + 1) \geq 0$$

holds by convexity of f.

For $e = 2, ..., E$,

$$\begin{aligned} \Delta &= p_{e,e-1}[f(e-1, x_H, x_L - 1) - 2f(e-1, x_H, x_L) + f(e-1, x_H, x_L + 1)] \\ &\quad +(1 - p_{e,e-1})[f(e, x_H, x_L - 1) - 2f(e, x_H, x_L) + f(e, x_H, x_L + 1)] \\ &\geq 0 \end{aligned}$$

is true due to convexity of f.

Case 2: $x_H \neq 0$

Since there are symptomatic patients in the system, first we serve them. So,

$$f(e, x_H - 1, x_L) - 2f(e, x_H - 1, x_L + 1) + f(e, x_H - 1, x_L + 2) \geq 0,$$

holds by convexity of f.

Convexity Preserved by T_{DET}

$$\begin{aligned}
& \tau f(e+1,x_H,x_L)+(1-\tau)f(e,x_H,x_L) \\
& -2\tau f(e+1,x_H,x_L+1)-2(1-\tau)f(e,x_H,x_L+1) \\
& +\tau f(e+1,x_H,x_L+2)+(1-\tau)f(e,x_H,x_L+2) \\
= \;& \tau[f(e+1,x_H,x_L)-2f(e+1,x_H,x_L+1+)f(e+1,x_H,x_L)] \\
& +(1-\tau)[f(e,x_H,x_L)-2f(e,x_H,x_L+1)+f(e,x_H,x_L)] \\
\geq \;& 0,
\end{aligned}$$

which is true due to the convexity of f in x_L. Hence, T_{DET} preserves convexity of a function f.

Convexity Preserved by T_{UNIF}

Here, we note that sum of convex functions are convex. Since, T_{UNIF} is the sum of T_{ARR_H}, T_{ADM_i}, T_{DEP} and T_{DET} which are convex operators, then T_{UNIF} is convex.

Convexity Preserved by T_{COST}

$$c_H x_H + c_L x_L - c_H x_H - c_L(x_L+1) - c_H x_H - c_L(x_L+1) + c_H x_H + c_L(x_L+2) \geq 0, \quad \text{(A.4)}$$

is true by simple algebra. Hence h is convex in x_L. Therefore, as a sum of two convex functions T_{UNIF} and h, T_{COST} is convex in x_L.

A.8 Supermodularity Property in (e, x_L)

In this proof, we show that all operators preserve supermodularity property of f in (e, x_L). Mathematically, we will show the inequalities for the pair (x_H, x_L):

$$Tf(e,x_H,x_L) - Tf(e,x_H,x_L+1) - Tf(e+1,x_H,x_L) - Tf(e+1,x_H,x_L+1) \geq 0. \quad \text{(A.5)}$$

for any supermodular function f in (e, x_L) and for all operators T.

Supermodularity Property in (e, x_L) Preserved by $\bar{\lambda}_H T_{ARR_H}$

$$
\begin{aligned}
& \lambda_H(e)[f(e, x_H+1, x_L) - f(e, x_H+1, x_L+1)] \\
- \; & \lambda_H(e+1)[f(e+1, x_H+1, x_L) - f(e+1, x_H+1, x_L+1)] \\
+ \; & [\bar{\lambda}_H - \lambda_H(e)][f(e, x_H, x_L) - f(e, x_H, x_L+1)] \\
- \; & [\bar{\lambda}_H - \lambda_H(e+1)][f(e+1, x_H, x_L) - f(e+1, x_H, x_L+1)] \\
= \; & \lambda_H(e)[f(e, x_H+1, x_L) - f(e, x_H+1, x_L+1) \\
- \; & f(e, x_H, x_L) + f(e, x_H, x_L+1)] \\
+ \; & \lambda_H(e+1)[f(e+1, x_H+1, x_L+1) - f(e+1, x_H+1, x_L) \\
- \; & f(e+1, x_H, x_L+1) + f(e+1, x_H, x_L)] \\
+ \; & \bar{\lambda}_H[f(e, x_H, x_L) - f(e, x_H, x_L+1) \\
- \; & f(e+1, x_H, x_L) + f(e+1, x_H, x_L+1)] \\
= \; & \lambda_H(e)[f(e, x_H+1, x_L) - f(e, x_H+1, x_L+1) \\
- \; & f(e, x_H, x_L) + f(e, x_H, x_L+1) + f(e+1, x_H+1, x_L+1) - f(e+1, x_H+1, x_L) \\
- \; & f(e+1, x_H, x_L+1) + f(e+1, x_H, x_L) + f(e, x_H, x_L) - f(e, x_H, x_L+1) \\
- \; & f(e+1, x_H, x_L) + f(e+1, x_H, x_L+1)] \\
+ \; & (\lambda_H(e+1) - \lambda_H(e))[f(e+1, x_H+1, x_L+1) - f(e+1, x_H+1, x_L) \\
- \; & f(e+1, x_H, x_L+1) + f(e+1, x_H, x_L)] \\
+ \; & (\bar{\lambda}_H - \lambda_H(e))[f(e, x_H, x_L) - f(e, x_H, x_L+1) \\
- \; & f(e+1, x_H, x_L) + f(e+1, x_H, x_L+1)] \\
= \; & \lambda_H(e)[f(e, x_H+1, x_L) - f(e, x_H+1, x_L+1) \\
- \; & f(e+1, x_H+1, x_L) + f(e+1, x_H+1, x_L+1)] \\
+ \; & (\lambda_H(e+1) - \lambda_H(e))[f(e+1, x_H+1, x_L+1) - f(e+1, x_H+1, x_L)
\end{aligned}
$$

$$
\begin{aligned}
& - \quad f(e+1, x_H, x_L+1) + f(e+1, x_H, x_L)] \\
& + \quad (\bar{\lambda}_H - \lambda_H(e))[f(e, x_H, x_L) - f(e, x_H, x_L+1) \\
& - \quad f(e+1, x_H, x_L) + f(e+1, x_H, x_L+1)] \\
& \geq \quad 0.
\end{aligned}
$$

Inequality holds due to supermodularity of f in (e, x_L) for fixed $x_H + 1$, supermodularity of f in (x_H, x_L) for fixed $e + 1$, and supermodularity of f in (e, x_L) for fixed x_H.

Supermodularity Property in (e, x_L) Preserved by T_{ADM_i}

We let

$$
\begin{aligned}
\delta \quad = \quad & \min\{f(e, x_H, x_L) + r_{L_i}, f(e, x_H, x_L+1) + s\} \\
& - \min\{f(e, x_H, x_L+1) + r_{L_i}, f(e, x_H, x_L+2) + s\} \\
& - \min\{f(e+1, x_H, x_L) + r_{L_i}, f(e+1, x_H, x_L+1) + s\} \\
& + \min\{f(e+1, x_H, x_L+1) + r_{L_i}, f(e+1, x_H, x_L+2) + s\}.
\end{aligned}
$$

Let systems A, B, C and D correspond to systems in states (e, x_H, x_L), $(e, x_H, x_L + 1)$,$(e + 1, x_H, x_L)$ and $(e+1, x_H, x_L+1)$ in period n, respectively. We let system A and system D follow the optimal policy, and system B and system C imitate all the decisions of system A and system D.

First we assume that A and D accept to screen the patients, so do B and C. Hence we have,

$$
\begin{aligned}
\delta \quad & \geq \quad f(e, x_H, x_L+1) - f(e, x_H, x_L+2) - f(e+1, x_H, x_L+1) + f(e+1, x_H, x_L+2) \\
& \geq \quad 0,
\end{aligned}
$$

which holds due to supermodularity of f in (e, x_L).

Now, we assume that A and D reject the patients, and since B and C imitate them we have,

$$\begin{aligned}
\delta &\geq f(e, x_H, x_L) + r_{L_i} - f(e, x_H, x_L + 1) - r_{L_i} \\
&\quad -f(e+1, x_H, x_L) - r_{L_i} + f(e+1, x_H, x_L + 1) + r_{L_i} \\
&= f(e, x_H, x_L) - f(e, x_H, x_L + 1) \\
&\quad -f(e+1, x_H, x_L) + f(e+1, x_H, x_L + 1) \\
&\geq 0,
\end{aligned}$$

holds by supermodularity of f in (e, x_L).

Now, we consider the cases where A and D make different decisions. We let system A accept to screen the patient and system D reject the patient. We let system B imitate the decisions of system D so that it rejects the same patient that system D rejects, and system C imitate the decisions of system A so that it accepts the same patient that system A accepts.

$$\begin{aligned}
\delta &\geq f(e, x_H, x_L + 1) - f(e, x_H, x_L + 1) - r_{L_i} \\
&\quad -f(e+1, x_H, x_L + 1) - f(e+1, x_H, x_L + 1) + r_{L_i} \\
&= f(e, x_H, x_L + 1) - f(e, x_H, x_L + 1) \\
&\quad -f(e+1, x_H, x_L + 1) - f(e+1, x_H, x_L + 1) \\
&= 0.
\end{aligned}$$

Now, we let system B and system C imitate the decisions of system A and system D respectively. We assume system A rejects the patient, whereas system D accepts to screen the patient.

$$\begin{aligned}
\delta &\geq f(e, x_H, x_L) + r_{L_i} - f(e, x_H, x_L + 1) - r_{L_i} \\
&\quad -f(e+1, x_H, x_L + 1) + f(e+1, x_H, x_L + 2)
\end{aligned}$$

$$\begin{aligned} &= f(e, x_H, x_L) - f(e, x_H, x_L + 1) \\ &\quad -f(e+1, x_H, x_L + 1) + f(e+1, x_H, x_L + 2) \\ &\geq f(e+1, x_H, x_L) - 2f(e+1, x_H, x_L + 1) + f(e+1, x_H, x_L + 2) \\ &\geq 0, \end{aligned}$$

where the first inequality holds by supermodularity of f in (e, x_L), and second inequality holds by convexity of f in x_L.

Supermodularity Property in (e, x_L) Preserved by T_{DEP}

We let

$$\Delta = \Psi(e, x_H, x_L) - \Psi(e, x_H, x_L + 1) - \Psi(e+1, x_H, x_L) + \Psi(e+1, x_H, x_L + 1).$$

There are two possible cases due to the number of symptomatic patients in the system.

Case 1: $x_H = 0$

In this case, there is no symptomatic patient in the system, so we serve asymptomatic patients. Since $x_H = 0$,

$$\begin{aligned} \Delta &= p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) \\ &\quad -p_{e,e-1} f(e-1, x_H, x_L) - (1 - p_{e,e-1}) f(e, x_H, x_L) \\ &\quad -p_{e+1,e} f(e, x_H, x_L - 1) - (1 - p_{e+1,e}) f(e+1, x_H, x_L - 1) \\ &\quad +p_{e+1,e} f(e, x_H, x_L) + (1 - p_{e+1,e}) f(e+1, x_H, x_L) \\ &= f(e, x_H, x_L - 1) - f(e, x_H, x_L) \\ &\quad -f(e+1, x_H, x_L - 1) + f(e+1, x_H, x_L) \\ &\quad +p_{e,e-1} f(e-1, x_H, x_L - 1) - p_{e,e-1} f(e, x_H, x_L - 1) \\ &\quad -p_{e,e-1} f(e-1, x_H, x_L) + p_{e,e-1} f(e, x_H, x_L) \end{aligned}$$

$$
\begin{aligned}
& -p_{e+1,e} f(e, x_H, x_L - 1) + p_{e+1,e} f(e+1, x_H, x_L - 1) \\
& +p_{e+1,e} f(e, x_H, x_L) - p_{e+1,e} f(e+1, x_H, x_L) \\
= \; & f(e, x_H, x_L - 1) - f(e, x_H, x_L) \\
& -f(e+1, x_H, x_L - 1) + f(e+1, x_H, x_L) \\
& -p_{e+1,e}[f(e, x_H, x_L - 1) - f(e+1, x_H, x_L - 1) \\
& -f(e, x_H, x_L) + f(e+1, x_H, x_L)] \\
& +p_{e,e-1}[f(e-1, x_H, x_L - 1) - f(e, x_H, x_L - 1) \\
& -f(e-1, x_H, x_L) + f(e, x_H, x_L)] \\
= \; & (1 - p_{e+1,e})[f(e, x_H, x_L - 1) - f(e, x_H, x_L) \\
& -f(e+1, x_H, x_L - 1) + f(e+1, x_H, x_L)] \\
& +p_{e,e-1}[f(e-1, x_H, x_L - 1) - f(e, x_H, x_L - 1) \\
& -f(e-1, x_H, x_L) + f(e, x_H, x_L)] \\
\geq \; & 0,
\end{aligned}
$$

holds due to supermodularity of f in $(e, x_L - 1)$ and $(e-1, x_L - 1)$.

Case 2: $x_H \neq 0$

Since there are symptomatic patients in the system, first we serve them. Then for any environment e,

$$
\begin{aligned}
\Delta \; &= \; f(e, x_H - 1, x_L) - f(e, x_H - 1, x_L + 1) - f(e+1, x_H - 1, x_L) + f(e+1, x_H - 1, x_L + 1) \\
&\geq \; 0,
\end{aligned}
$$

holds due to supermodularity of f in (e, x_L) for fixed x_H.

Supermodularity Property in (e, x_L) Preserved by $\bar{\gamma} T_{DET}$

$$
\begin{aligned}
& \gamma_{e,e+1}[f(e+1,x_H,x_L) - f(e+1,x_H,x_L+1)] \\
& -\gamma_{e+1,e+2}[f(e+2,x_H,x_L) - f(e+2,x_H,x_L+1)] \\
& +(\bar{\gamma} - \gamma_{e,e+1})[f(e,x_H,x_L) - f(e,x_H,x_L+1)] \\
& -(\bar{\gamma} - \gamma_{e+1,e+2})[f(e+1,x_H,x_L) - f(e+1,x_H,x_L+1)] \\
= \; & \gamma_{e,e+1}[f(e+1,x_H,x_L) - f(e+1,x_H,x_L+1) \\
& -f(e,x_H,x_L) + f(e,x_H,x_L+1)] \\
& +\bar{\gamma}[f(e,x_H,x_L) - f(e,x_H,x_L+1) \\
& -f(e+1,x_H,x_L) + f(e+1,x_H,x_L+1)] \\
& +\gamma_{e+1,e+2}[f(e+1,x_H,x_L) - f(e+1,x_H,x_L+1) \\
& -f(e+2,x_H,x_L) + f(e+2,x_H,x_L+1)] \\
= \; & (\bar{\gamma} - \gamma_{e,e+1})[f(e,x_H,x_L) - f(e,x_H,x_L+1) \\
& -f(e+1,x_H,x_L) + f(e+1,x_H,x_L+1)] \\
& +\gamma_{e+1,e+2}[f(e+1,x_H,x_L) - f(e+1,x_H,x_L+1) \\
& -f(e+2,x_H,x_L) + f(e+2,x_H,x_L+1)] \\
\geq \; & 0,
\end{aligned}
$$

holds due to supermodularity of f in (e, x_L) and $(e+1, x_L)$.

Supermodularity Property in (e, x_L) Preserved by T_{UNIF}

Since, T_{UNIF} is the sum of T_{ARR_H}, T_{ADM}, T_{DEP} and T_{DET} which preserve supermodularity of f in (e, x_L), then T_{UNIF} preserves this property, too.

Supermodularity Property in (e, x_L) Preserved by T_{COST}

$$c_H x_H + c_L x_L - c_H x_H - c_L(x_L + 1) - c_H x_H - c_L x_L + c_H x_H + c_L(x_L + 1) = 0.$$

is true by simple algebra. Hence h is supermodular in (e, x_L). Therefore, as a sum of two functions T_{UNIF} and h, T_{COST} preserves this property.

A.9 Supermodularity Property in (x_H, x_L)

In this proof, we show that all operators preserve supermodularity property of f in (x_H, x_L). More explicitly, we will show for the pair (x_H, x_L):

$$Tf(e, x_H, x_L) - Tf(e, x_H + 1, x_L) - Tf(e, x_H, x_L + 1) + Tf(e, x_H + 1, x_L + 1) \geq 0, \quad \text{(A.6)}$$

holds for any supermodular function f and for all operators T.

Supermodularity Property in (x_H, x_L) Preserved by T_{ARR_H}

$$\begin{aligned}
& \alpha f(e, x_H + 1, x_L) + (1 - \alpha) f(e, x_H, x_L) \\
- \; & \alpha f(e, x_H + 2, x_L+) + (1 - \alpha) f(e, x_H + 1, x_L) \\
- \; & \alpha f(e, x_H + 1, x_L + 1) + (1 - \alpha) f(e, x_H, x_L + 1) \\
+ \; & \alpha f(e, x_H + 2, x_L + 1) + (1 - \alpha) f(e, x_H + 1, x_L + 1) \\
= \; & \alpha [f(e, x_H + 1, x_L) - f(e, x_H + 2, x_L) - f(e, x_H + 1, x_L + 1) + f(e, x_H + 2, x_L + 1)] \\
+ \; & (1 - \alpha)[f(e, x_H, x_L) - f(e, x_H + 1, x_L) - f(e, x_H, x_L + 1) + f(e, x_H + 1, x_L + 1)] \\
\geq \; & 0,
\end{aligned}$$

holds by supermodularity of f in $(x_H + 1, x_L)$ and (x_H, x_L).

Supermodularity Property in (x_H, x_L) Preserved by T_{ADM_i}

We let

$$\begin{aligned}\delta &= \min\{f(e,x_H,x_L)+r_{L_i}, f(e,x_H,x_L+1)\} \\ &\quad -\min\{f(e,x_H+1,x_L)+r_{L_i}, f(e,x_H+1,x_L+1)\}\end{aligned}$$

$$\begin{aligned}&-\min\{f(e,x_H,x_L+1)+r_{L_i}, f(e,x_H,x_L+2)\} \\ &+\min\{f(e,x_H+1,x_L+1)+r_{L_i}, f(e,x_H+1,x_L+2)\}.\end{aligned}$$

Let systems A, B, C and D correspond to systems in states (e,x_H,x_L), (e,x_H+1,x_L),(e,x_H,x_L+1) and (e,x_H+1,x_L+1) in period n, respectively. We let system A and system D follow the optimal policy, and system B and system C imitate all the decisions of system A and system D. First we assume that A and D accept to screen the patients. Therefore, all systems accept to screen the patients. So we have,

$$\begin{aligned}\delta &\geq f(e,x_H,x_L+1)-f(e,x_H+1,x_L+1)-f(e,x_H,x_L+2)+f(e,x_H+1,x_L+2) \\ &\geq 0,\end{aligned}$$

which holds by supermodularity of f in (x_H, x_L+1).
Now, we assume that A and D reject the patients. Since systems B and C imitate them, we have,

$$\begin{aligned}\delta &\geq f(e,x_H,x_L)+r_{L_i}-f(e,x_H+1,x_L)-r_{L_i}-f(e,x_H,x_L+1)-r_{L_i}+f(e,x_H+1,x_L+1)+r_{L_i} \\ &= f(e,x_H,x_L)-f(e,x_H+1,x_L)-f(e,x_H,x_L+1)+f(e,x_H+1,x_L+1) \\ &\geq 0,\end{aligned}$$

which holds by supermodularity of f in (x_H, x_L).
Now, we consider the cases where A and D make different decisions. We let system A reject

the patient and system D accept to screen the patient. We assume that system B imitates the decision of system A and system C imitates the decision of system D.

$$\begin{aligned}\delta &\geq f(e,x_H,x_L)+r_{L_i}-f(e,x_H+1,x_L)-r_{L_i}-f(e,x_H,x_L+2)+f(e,x_H+1,x_L+2)\\ &= f(e,x_H,x_L)-f(e,x_H+1,x_L)-f(e,x_H,x_L+2)+f(e,x_H+1,x_L+2)\\ &\geq f(e,x_H,x_L+1)-f(e,x_H+1,x_L+1)-f(e,x_H,x_L+2)+f(e,x_H+1,x_L+2)\\ &\geq 0,\end{aligned}$$

where the first inequality is due to the supermodularity of f in (x_H, x_L), and second inequality is due to the supermodularity of f in (x_H, x_L+1).
Now, we let system B imitate the decision of system A so that it accepts the patient that system A accepts, whereas system C imitate the decision of system D so that it rejects the patient that system D rejects.

$$\begin{aligned}\delta &\geq f(e,x_H,x_L+1)-f(e,x_H+1,x_L+1)-f(e,x_H,x_L+1)-r_{L_i}+f(e,x_H+1,x_L+1)+r_{L_i}\\ &= f(e,x_H,x_L+1)-f(e,x_H+1,x_L+1)-f(e,x_H,x_L+1)+f(e,x_H+1,x_L+1)\\ &= 0.\end{aligned}$$

Supermodularity Property in (x_H, x_L) Preserved by T_{DEP}

To begin with, we let

$$\Delta = \Psi(e,x_H,x_L)-\Psi(e,x_H+1,x_L)-\Psi(e,x_H,x_L+1)+\Psi(e,x_H+1,x_L+1).$$

Case 1: $x_H = 0$

In this case, the system has no symptomatic patient, hence we serve asymptomatic patients.

Then for $e = 1$,

$$\begin{aligned}\Delta &= f(1, x_H, x_L - 1) - f(1, x_H, x_L) - f(1, x_H, x_L) + f(1, x_H, x_L + 1) \\ &= f(1, x_H, x_L - 1) - 2f(1, x_H, x_L) + f(1, x_H, x_L + 1) \\ &\geq 0,\end{aligned}$$

holds due to convexity of f in $x_L - 1$ for fixed x_H.

For $e \in \{2, \cdots, E\}$,

$$\begin{aligned}\Delta &= p_{e,e-1} f(e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) f(e, x_H, x_L - 1) - f(e, x_H, x_L) \\ &- p_{e,e-1} f(e-1, x_H, x_L) - (1 - p_{e,e-1}) f(e, x_H, x_L) + f(e, x_H, x_L + 1) \\ &= p_{e,e-1}[f(e-1, x_H, x_L - 1) - f(e, x_H, x_L - 1) - f(e-1, x_H, x_L) + f(e, x_H, x_L)] \\ &+ \quad [f(e, x_H, x_L - 1) - f(e, x_H, x_L) - f(e, x_H, x_L) + f(e, x_H, x_L + 1)] \\ &\geq 0.\end{aligned}$$

The terms in the first bracket is positive due to $Sup(e, x_L)$ property of f, and remaining terms are positive due to convexity of f in $x_L - 1$ for fixed x_H.

Case 2: $x_H \neq 0$

Since there are symptomatic patients in the system, first we serve them. So,

$$\begin{aligned}\Delta &= f(e, x_H - 1, x_L) - f(e, x_H, x_L) - f(e, x_H - 1, x_L + 1) + f(e, x_H, x_L + 1) \\ &\geq 0,\end{aligned}$$

and holds by supermodularity of f in $(x_H - 1, x_L)$.

Supermodularity Property in (x_H, x_L) Preserved by T_{DET}

$$\begin{aligned}
&\tau f(e+1, x_H, x_L) + (1-\tau) f(e, x_H, x_L) \\
&-\tau f(e+1, x_H+1, x_L) - (1-\tau) f(e, x_H+1, x_L) \\
&-\tau f(e+1, x_H, x_L+1) - (1-\tau) f(e, x_H, x_L+1) \\
&-\tau f(e+1, x_H+1, x_L+1) - (1-\tau) f(e, x_H+1, x_L+1)
\end{aligned}$$

$$\begin{aligned}
= \quad & \tau[f(e+1, x_H, x_L) - f(e+1, x_H+1, x_L) - f(e+1, x_H, x_L+1) + f(e+1, x_H+1, x_L+1)] \\
& +(1-\tau)[f(e, x_H, x_L) - f(e, x_H+1, x_L) - f(e, x_H, x_L+1) + f(e, x_H+1, x_L+1)] \\
\geq \quad & 0,
\end{aligned}$$

holds by supermodularity of f in (x_H, x_L).

Supermodularity Property in (x_H, x_L) Preserved by T_{UNIF}

Since, T_{UNIF} is the sum of T_{ARR_H}, T_{ADM_i}, T_{DEP} and T_{DET} which preserve supermodularity property of f in (x_H, x_L), then T_{UNIF} preserves this property, too.

Supermodularity Property in (x_H, x_L) Preserved by T_{COST}

$$c_H x_H + c_L x_L - c_H x_H - c_L(x_L+1) - c_H(x_H+1) - c_L x_L + c_H(x_H+1) + c_L(x_L+1) = 0.$$

is true by simple algebra. Hence h is supermodular. Therefore, as a sum of two functions T_{UNIF} and h, T_{COST} preserves this property.

Appendix B

PROOFS FOR CHAPTER 7

B.1 The Properties Preserved by T_{WORSE}

We introduce a new operator, T_{WORSE}, and let it represent the process of developing cancer while waiting for the service in the queue where M is the fixed cost of developing cancer. We define it as follows:

$$T_{WORSE}f(e, x_H, x_L) = f(e, x_H + 1, x_L - 1) + M$$

We investigate which properties it preserves. We will show that the operator, T_{WORSE}, preserves the following properties;

(a)

$$Inc(x_H) : T_{WORSE}f(e, x_H, x_L) \leq T_{WORSE}f(e, x_H + 1, x_L),$$

for any non-decreasing function f in x_H.

$$f(e, x_H + 1, x_L - 1) + M - f(e, x_H + 2, x_L - 1) - M \leq 0,$$

holds by monotonicity of f in x_H.

(b)

$$Inc(x_L) : T_{WORSE}f(e, x_H, x_L) \leq T_{WORSE}f(e, x_H, x_L + 1),$$

for any non-decreasing function f in x_L.

$$f(e, x_H + 1, x_L - 1) + M - f(e, x_H + 1, x_L) - M \leq 0,$$

holds by monotonicity of f in x_L.

(c)

$$Inc(e) : T_{WORSE} f(e, x_H, x_L) \leq T_{WORSE} f(e + 1, x_H, x_L),$$

for any non-decreasing function f in e.

$$f(e, x_H + 1, x_L - 1) + M - f(e + 1, x_H + 1, x_L - 1) - M \leq 0,$$

holds by monotonicity of f in e.

(d)

$$Dec(p) : T_{WORSE} \underline{f}(e, x_H, x_L) - T_{WORSE} \overline{f}(e, x_H, x_L) \geq 0,$$

for any two function $\underline{f}$ and $\overline{f}$ (with $\underline{p}$ and $\overline{p}$ respectively where $\underline{p} \leq \overline{p}$) satisfying the following inequality,

$$\underline{f}(e, x_H, x_L) - \overline{f}(e, x_H, x_L) \geq 0.$$

$$\underline{f}(e, x_H + 1, x_L - 1) + M - \overline{f}(e, x_H + 1, x_L - 1) - M \geq 0,$$

is true due to the $Dec(p)$ property for the state $(e, x_H + 1, x_L - 1)$.

(e)

$$Diag : T_{WORSE} f(e, x_H - 1, x_L) \leq T_{WORSE} f(e, x_H, x_L - 1).$$

for any function f with the monotonicity on the diagonal property.

$$f(e, x_H, x_L - 1) + M - f(e, x_H + 1, x_L - 2) - M \leq 0,$$

holds by monotonicity on the diagonal property of f.

(f)

$$IDiag_e : T_{WORSE} f(e - 1, x_H, x_L - 1) \leq T_{WORSE} f(e, x_H - 1, x_L),$$

for any function f satisfying $IDiag_e$ property.

$$f(e - 1, x_H + 1, x_L - 2) + M - f(e, x_H, x_L - 1) - M \leq 0,$$

holds by $IDiag_e$ property of f.

(g)

$$\begin{aligned} Conv(x_L) \quad &: \quad T_{WORSE} f(e, x_H, x_L) - T_{WORSE} f(e, x_H, x_L + 1) \\ &\geq \quad T_{WORSE} f(e, x_H, x_L + 1) - T_{WORSE} f(e, x_H, x_L + 2), \end{aligned}$$

holds for any convex function f.

$$f(e, x_H+1, x_L-1)+M-f(e, x_H+1, x_L)-M-f(e, x_H+1, x_L)-M+f(e, x_H+1, x_L+1)+M \geq 0,$$

holds by convexity of f.

(h)

$$\begin{aligned} Sup(e, x_L) \quad &: \quad T_{WORSE} f(e, x_H, x_L) - T_{WORSE} f(e, x_H, x_L + 1) \\ &\geq \quad T_{WORSE} f(e + 1, x_H, x_L) - T_{WORSE} f(e + 1, x_H, x_L + 1), \end{aligned}$$

for any supermodular function f in (e, x_L).

$$f(e, x_H{+}1, x_L{-}1){+}M{-}f(e, x_H{+}1, x_L){-}M{-}f(e{+}1, x_H{+}1, x_L{-}1){-}M{+}f(e{+}1, x_H{+}1, x_L){-}M \geq 0,$$

holds by supermodularity of f in (e, x_L).

(i)

$$\begin{aligned} Sup(x_H, x_L) \quad &: \quad T_{WORSE}f(e, x_H, x_L) - T_{WORSE}f(e, x_H + 1, x_L) \\ &\geq \quad T_{WORSE}f(e, x_H, x_L + 1) - T_{WORSE}f(e, x_H + 1, x_L + 1), \end{aligned}$$

holds for any supermodular function f in (x_H, x_L).

$$f(e, x_H{+}1, x_L{-}1){+}M{-}f(e, x_H{+}2, x_L{-}1){-}M{-}f(e, x_H{+}1, x_L){-}M{+}f(e, x_H{+}2, x_L){+}M \geq 0,$$

is true by supermodularity of f in (x_H, x_L).

B.2 The Properties Preserved by the State Space S_{new}

For S_{new}, the optimality equation is given as follows:

$$\begin{aligned} v_{n+1}(A, e, x_H, x_L) \quad = \quad & T_{COST}(T_{UNIF}(\{T_{ARR_H}v_n(A, e, x_H, x_L), \{T_{ADM_i}v_n(A, e, x_H, x_L)\}_i, \\ & T_{SCH}v_n(A, e, x_H, x_L), T_{DET}v_n(A, e, x_H, x_L)\}; \{\bar{\lambda}_H, \{\lambda_{L_i}\}_i, \mu, \bar{\gamma}\})), \end{aligned}$$

and

$$T_{SCH}v(A, e, x_H, x_L) = \min\{v(H, e, x_H - 1, x_L), g(L, e, x_H, x_L - 1)\} + s, \tag{B.1}$$

where

$$g(L, e, x, y) = \begin{cases} v(L, 1, x, y) & \text{if} \quad e = 1 \\ p_{e,e-1}v(L, e-1, x_H, x_L - 1) + (1 - p_{e,e-1})v(L, e, x_H, x_L - 1) & \text{otherwise.} \end{cases}$$

For this section, it is enough to analyze the multiplier of μ since the results for other components are similar to the proofs with state space S. We make the following assumption.

Assumption 6 *The value functions preserve the corresponding property.*

$Inc(x_H)$ Property

In this section, we will show that the following inequality is true:

$$T_{SCH}v(A, e, x_H, x_L) \leq T_{SCH}v(A, e, x_H + 1, x_L).$$

Assuming that in both states ((A, e, x_H, x_L) and $(A, e, x_H + 1, x_L)$), symptomatic patients are scheduled:

$$v(H, e, x_H - 1, x_L) - v(H, e, x_H, x_L) \leq 0, \tag{B.2}$$

is true by the assumption.
Assuming that in both states ((A, e, x_H, x_L) and $(A, e, x_H + 1, x_L)$), asymptomatic patients are scheduled:

$$\begin{aligned}
& p_{e,e-1}v(L, e-1, x_H, x_L - 1) + (1 - p_{e,e-1})v(L, e, x_H, x_L - 1) \\
& -p_{e,e-1}v(L, e-1, x_H + 1, x_L - 1) - (1 - p_{e,e-1})v(L, e, x_H + 1, x_L - 1) \\
\leq\ & p_{e,e-1}(v(L, e-1, x_H, x_L - 1) - v(L, e-1, x_H + 1, x_L - 1)) \\
& +(1 - p_{e,e-1})(v(L, e, x_H, x_L - 1) - v(L, e, x_H + 1, x_L - 1)) \\
\leq\ & 0,
\end{aligned} \tag{B.3}$$

is true by the assumption.
Let H and L be the optimal scheduling actions. We define costs of optimal actions

$$(X, Y) =$$
$$\{T_{SCH}v(A, e, x_H, x_L) - T_{SCH}v(A, e, x_H + 1, x_L) \wedge X \in (A, e, x_H, x_L), Y \in (A, e, x_H + 1, x_L)\}$$

Let us assume that it is optimal to schedule symptomatic patient in state (A, e, x_H, x_L), and it is optimal to schedule asymptomatic patient in state $((A, e, x_H + 1, x_L)$. Therefore the following expression

$$(H, L) \tag{B.4}$$

is less than

$$(L, L), \tag{B.5}$$

since the cost of optimal actions will give the minimum value. Expression (B.5) is less than 0 by equation (B.3), which ensures that Expression (B.4) is less than 0.
Assuming that it is optimal to schedule asymptomatic patient in state (A, e, x_H, x_L), and it is optimal to schedule symptomatic patient in state $((A, e, x_H + 1, x_L)$, we obtain

$$(L, H) \leq (H, H) \leq 0,$$

by equation (B.2).

$Inc(x_L)$ Property

Similar to previous proof.

$Inc(e)$ Property

We will show that the following inequality holds

$$T_{SCH}v(A, e, x_H, x_L) \leq T_{SCH}v(A, e + 1, x_H, x_L).$$

Assuming that in both states $((A, e, x_H, x_L)$ and $(A, e+1, x_H, x_L))$, symptomatic patients are scheduled:

$$v(H, e, x_H - 1, x_L) - v(H, e+1, x_H - 1, x_L) \leq 0, \tag{B.6}$$

is true by the assumption.
Assuming that in both states $((A, e, x_H, x_L)$ and $(A, e+1, x_H, x_L))$, asymptomatic patients are scheduled:

$$\begin{aligned}
& p_{e,e-1} v(L, e-1, x_H, x_L - 1) + (1 - p_{e,e-1}) v(L, e, x_H, x_L - 1) \\
& -p_{e+1,e} v(L, e, x_H, x_L - 1) - (1 - p_{e+1,e}) v(L, e+1, x_H, x_L - 1) \\
= \; & p_{e+1,e}(v(L, e-1, x_H, x_L - 1) - v(L, e, x_H, x_L - 1)) \\
& +(1 - p_{e,e-1})(v(L, e, x_H, x_L - 1) - v(L, e+1, x_H, x_L - 1)) \\
\leq \; & 0,
\end{aligned} \tag{B.7}$$

is true by the assumption.
Let H and L be the optimal scheduling actions. We define costs of optimal actions

$$\begin{aligned}
& (X, Y) = \\
& \{T_{SCH} v(A, e, x_H, x_L) - T_{SCH} v(A, e+1, x_H, x_L) \wedge X \in (A, e, x_H, x_L), Y \in (A, e, x_H, x_L + 1)\}
\end{aligned}$$

Let us assume that it is optimal to schedule symptomatic patient in state (A, e, x_H, x_L), and it is optimal to schedule asymptomatic patient in state $((A, e+1, x_H, x_L)$. Therefore the following expression

$$(H, L) \tag{B.8}$$

is less than

$$(L, L), \tag{B.9}$$

since the cost of optimal actions will give the minimum value. Expression (B.9) is less than 0 by equation (B.7), which ensures that Expression (B.8) is less than 0.
Assuming that it is optimal to schedule asymptomatic patient in state (A, e, x_H, x_L), and it is optimal to schedule symptomatic patient in state $((A, e, x_H + 1, x_L)$, we obtain

$$(L, H) \leq (H, H) \leq 0,$$

by equation (B.6).

Monotonicity on the Diagonal Property

We will prove that the following inequality holds.

$$T_{SCH}v(A, e, x_H - 1, x_L) \leq T_{SCH}v(A, e, x_H, x_L - 1).$$

Let H and L be the optimal scheduling actions. We define costs of optimal actions

$$(X, Y)_{Diag} =$$
$$\{T_{SCH}(A, e, x_H - 1, x_L) - T_{SCH}v(A, e, x_H, x_L - 1) \wedge X \in (A, e, x_H - 1, x_L), Y \in (A, e, x_H, x_L - 1)\}.$$

There are four cases to be considered.

$$(H, H)_{Diag} = v(H, e, x_H - 2, x_L) - v(H, e, x_H - 1, x_L - 1) \leq 0, \tag{B.10}$$

holds by monotonicity on the diagonal property of value functions.

$$\begin{aligned}(L,L)_{Diag} &= p_{e,e-1}v(L,e-1,x_H-1,x_L-1)+(1-p_{e,e-1})v(L,e,x_H-1,x_L-1)\\ &\quad -p_{e,e-1}v(L,e-1,x_H,x_L-2)-(1-p_{e,e-1})v(L,e,x_H,x_L-2)\\ &= p_{e,e-1}(v(L,e-1,x_H-1,x_L-1)-v(L,e-1,x_H,x_L-2))\\ &\quad +(1-p_{e,e-1})(v(L,e,x_H-1,x_L-1)-(1-p)v(L,e,x_H,x_L-2))\\ &\leq 0, \end{aligned} \tag{B.11}$$

holds by monotonicity on the diagonal property of value functions.

$$(L,H)_{Diag} \leq (H,H)_{Diag} \leq 0,$$

holds by equation (B.10).

$$(H,L)_{Diag} \leq (L,L)_{Diag} \leq 0,$$

holds by equation (B.11).

$IDiag_e$ Property

We will show that we can extend the result of the operator T_{SCH} in the case of new state space. Mathematically, for the value function preserving $IDiag_e$ property, the following inequality holds.

$$T_{SCH}v(A,e-1,x_H,x_L-1) \leq T_{SCH}v(A,e,x_H-1,x_L).$$

Let H and L be the optimal scheduling actions. We define costs of optimal actions

$$\begin{aligned}(X,Y)_{IDiag_e} &= \{T_{SCH}(A,e-1,x_H,x_L-1)-T_{SCH}v(A,e,x_H-1,x_L)\wedge\\ &\quad X\in(A,e-1,x_H,x_L-1), Y\in(A,e,x_H-1,x_L)\}\end{aligned}$$

There are four cases due to available actions.

$$(H,H)_{IDiag_e} = v(H, e-1, x_H-1, x_L-1) - v(H, e, x_H-2, x_L) \leq 0, \tag{B.12}$$

holds by $IDiag_e$ property of value functions.

$$\begin{aligned}
(L,L)_{IDiag_e} &= p_{e-1,e-2}v(L, e-2, x_H, x_L-2) + (1-p_{e-1,e-2})v(L, e-1, x_H, x_L-2) \\
&\quad -p_{e,e-1}v(L, e-1, x_H-1, x_L-1) - (1-p_{e,e-1})v(L, e, x_H-1, x_L-1) \\
&= p_{e-1,e-2}(v(L, e-2, x_H, x_L-2) - v(L, e-1, x_H-1, x_L-2)) \\
&\quad +(p_{e,e-1}-1)(v(L, e-1, x_H, x_L-2) - v(L, e-1, x_H-1, x_L-1)) \\
&\quad +v(L, e-1, x_H, x_L-2) - v(L, e, x_H-1, x_L-1) \\
&\leq 0,
\end{aligned} \tag{B.13}$$

holds by $IDiag_e$ and $Diag$ property of value functions.

$$(L,H)_{IDiag_e} \leq (H,H)_{IDiag_e} \leq 0,$$

holds by equation (B.12).

$$(H,L)_{IDiag_e} \leq (L,L)_{IDiag_e} \leq 0,$$

holds by equation (B.13).

Convexity in x_L

We will need to show that the operator T_{DEP} preserves the convexity property which is equal to

$$T_{DEP}v(A, e, x_H, x_L) - T_{DEP}v(A, e, x_H, x_L+1) \geq T_{DEP}v(A, e, x_H, x_L+1) - T_{DEP}v(A, e, x_H, x_L+2)$$

for convex value functions.

We define

$$\begin{aligned}(X,Y,Z,W) \quad = \quad & \{T_{DEP}v(A,e,x_H,x_L) - T_{DEP}v(A,e,x_H,x_L+1) - T_{DEP}v(A,e,x_H,x_L+1) \\ & +T_{DEP}v(A,e,x_H,x_L+2)|X \in (A,e,x_H,x_L), Y \in (A,e,x_H,x_L+1), \\ & Z \in (A,e,x_H,x_L+1), W \in (A,e,x_H,x_L+2)\}\end{aligned}$$

There are two cases where $x_H = 0$ and $x_H \neq 0$.

Case 1: $x_H \neq 0$

$$\begin{aligned}(H,H,H,H) &= v(H,e,x_H-1,x_L) - v(H,e,x_H-1,x_L+1) \\ &\quad -v(H,e,x_H-1,x_L+1) + v(H,e,x_H-1,x_L+2) \\ &\geq 0,\end{aligned}$$

holds by the assumption. **Case 2:** $x_H = 0$

$$\begin{aligned}(L,L,L,L) &= p_{e,e-1}v(L,e-1,x_H,x_L-1) + (1-p_{e,e-1})v(L,e,x_H,x_L-1) \\ &\quad -2p_{e,e-1}v(L,e-1,x_H,x_L) - 2(1-p_{e,e-1})v(L,e,x_H,x_L) \\ &\quad +p_{e,e-1}v(L,e-1,x_H,x_L+1) + (1-p_{e,e-1})v(L,e,x_H,x_L+1) \\ &= p_{e,e-1}(v(L,e-1,x_H,x_L-1) - 2v(L,e-1,x_H,x_L) + v(L,e-1,x_H,x_L+1)) \\ &\quad +(1-p_{e,e-1})(v(L,e,x_H,x_L-1) - 2v(L,e,x_H,x_L) + v(L,e,x_H,x_L+1)) \\ &\geq 0,\end{aligned}$$

holds by the assumption.

B.3 Non-linear Holding Costs

We assume increasing and convex costs for x_H, and x_L.

Assumption 7

i) $h(x_H, x_L)$ *is non-decreasing in* x_H*, and* x_L*.*

ii) $h(x_H, x_L)$ *is convex in* x_H*, and* x_L*.*

$Inc(x_H)$

$$h(x_H, x_L) \leq h(x_H + 1, x_L)$$

holds by i).

$Inc(x_L)$

$$h(x_H, x_L) \leq h(x_H, x_L + 1)$$

holds by i).

$Inc(e)$

$$h(x_H, x_L) \leq h(x_H, x_L)$$

holds by simple algebra.

Convexity in x_L

$$h(x_H, x_L) - h(x_H, x_L + 1) \geq h(x_H, x_L + 1) - h(x_H, x_L + 2)$$

holds by convexity of holding costs in x_L.

VITA

DERYA KUNDUZCU was born in Ayvalık, Balıkesir, Turkey on January 28, 1984. She graduated from Ayvalık Anadolu Lisesi in 2002. She received her B.Sc. degree in Mathematics and B.S. degree in Economics from Koç University, Istanbul, in 2007. In September 2007, she joined the Department of Industrial Engineering at Koç University, Istanbul. From 2004 to 2009, she worked as a teaching and research assistant in Koç University, Turkey.