A PARTIALLY OBSERVABLE MARKOV DECISION PROCESS APPROACH FOR CLINICAL DECISION SUPPORT IN CANCER TREATMENT: IMPLEMENTATION FOR COLON CANCER

KANSER TEDAVİSİNDE KLİNİK KARAR DESTEĞİ İÇİN KISMİ GÖZLEMLENEBİLİR MARKOV KARAR SÜRECİ YAKLAŞIMI: KOLON KANSERİ UYGULAMASI

AYŞE SEVDE EDİZER

ASSIST. PROF. BANU YÜKSEL ÖZKAYA

Supervisor

Submitted to

Graduate School of Science and Engineering of Hacettepe University

as a Partial Fulfillment to the Requirements

for the Award of the Degree of Master of Science

in Industrial Engineering.

2022

Dedication

This thesis is dedicated to my beloved father,

whom I lost in 2016 due to colon cancer.

ÖZET

KANSER TEDAVİSİNDE KLİNİK KARAR DESTEĞİ İÇİN KISMİ GÖZLEMLENEBİLİR MARKOV KARAR SÜRECİ YAKLAŞIMI: KOLON KANSERİ UYGULAMASI

Ayşe Sevde EDİZER

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Banu YÜKSEL ÖZKAYA

Mayıs 2022, 90 sayfa

Kanser tedavi süreci birçok yönden belirsizlik içermektedir. Hastalığın dinamik değişen doğası ve hastalığı evrelemek için kullanılan testlerin belirli bir oranda sapma içermesi hastalığın gerçek durumunu bilmeyi zorlaştırmaktadır. Hekimler, diğer birçok hastalıkta olduğu gibi kanser tedavi kararlarını da stokastik bir ortamda vermektedir. Bu çalışma, kısmi gözlemlenebilir Markov karar süreçlerini kullanarak kolorektal kanseri tedavi sürecini matematiksel olarak modellemeyi amaçlamaktadır. Kısmi gözlemlenebilir çevrenin tedavi seçenekleri üzerindeki etkisini anlamak için hastanın gerçek sağlık durumunun gözlem durumu olarak tanımlanan kandaki karsinoembriyonik antijen seviyesi değişimi ve bilgisayarlı tomografi sonuçları üzerinden tahmin edildiği kısmi

gözlemlenebilir Markov karar süreci modelinin sonuçlarını hastanın gerçek sağlık durumunun tam olarak bilindiğini varsayan temel bir Markov karar süreci modeliyle karşılaştırılmalı değerlendirmesi yapılmıştır. Önerilen modelin çıktıları Surveillance, Epidemiology ve End Results veritabanındaki 5 yıllık sağkalım sonuçlarıyla karşılaştırılmıştır. Modelin etkinliğinin anlaşılması için bir dizi varsayımsal senaryo sunulmuş olup modelleme sürecinde karşılaşılan bazı kısıtlamalardan bahsedilerek gelecekteki çalışmalar için önerilerde bulunulmuştur.

Anahtar Kelimeler: Kısmi Gözlemlenebilir Markov Karar Süreçleri, Dinamik Karar Modelleri, Kolon Kanseri, Karar Destek Sistemleri, Markov Karar Süreçleri

ABSTRACT

A PARTIALLY OBSERVABLE MARKOV DECISION PROCESS APPROACH FOR CLINICAL DECISION SUPPORT IN CANCER TREATMENT: IMPLEMENTATION FOR COLON CANCER

Ayşe Sevde EDİZER

Master of Science, Department of Industrial Engineering

Supervisor: Assist. Prof. Banu YÜKSEL ÖZKAYA

May 2022, 90 pages

The cancer treatment process involves uncertainty by its nature. Since the disease evolves continuously and diagnostic tests used to detect the level of the disease are not totally accurate, the actual state of the disease remains unknown. Therefore, physicians should make treatment decisions in a stochastic environment. This study aims to develop a mathematical model of the history of the colorectal cancer treatment process by using partially observable Markov decision process. To understand the impact of the partially observable environment on modeling the history of the disease, a comparative analysis of the outputs of the partially observable Markov decision process model, in which the patient's actual health status is estimated from the blood carcinoembryonic antigen level change and computed tomography results as observational states, with a basic Markov decision process model that assumes the patient's actual health status is fully known. has

been made. The output of the proposed model has been compared to 5-year survival outcomes that come from Surveillance, Epidemiology, and End Results database. A series of hypothetical scenarios have been presented to understand the effectiveness of the model and some limitations encountered in the modeling process have been mentioned along with suggestions for future studies will be made.

Keywords: Partially Observable Markov Decision Process, Dynamic Decision Models, Colorectal Cancer, Clinical Decision Support, Markov Decision Process

ACKNOWLEDGMENT

I would like to express my special thanks to my esteemed Academic Supervisor “Assist. Prof. Banu YUKSEL OZKAYA” who believed and supported me in this difficult task that I will remember with respect throughout my life.

I would also like to express my gratitude to my beloved husband “Ahmed EDIZER, M.D.”, who has always been by my side on this long journey.

I am also grateful to the physicians who guided me to build and understand this study amid the extraordinarily increased workload due to the global pandemic that coincided with the thesis process.

CONTENTS

ÖZET i

ABSTRACT iii

ACKNOWLEDGMENT v

FIGURES ix

TABLES x

SYMBOLS AND ABBREVIATIONS xi

1.INTRODUCTION 1

1.1. Aim of Study 2

1.2. Outline of Thesis 3

2.LITERATURE REVIEW 4

2.1. Cancer-Related Applications 4

2.1.1. Diagnostic Models 5

2.1.2. Treatment Models 6

2.1.3. Screening Models 8

2.2. Non-Cancer-Related Applications 11

2.2.1. Diagnostic Models 12

2.2.2. Treatment Models 14

2.2.3. Screening Models 15

2.3. Summary of Literature 17

3.BACKGROUND ON COLORECTAL CANCER 22

3.1. Overview for Colorectal Cancer 22

3.1.1. Molecular Pathogenesis 23

3.1.2. Risk Factors of *CRC* 24

3.2. Diagnosis and Treatment Methods 25

3.2.1. Cancer Staging 25

3.2.2. Treatment Procedure of *CRC* 27
3.3. Screening Policies 28
4.THEORATICAL BACKGROUND 29
4.1. Markov Decision Process 29
4.1.1.MDP Model Formulation 29
4.1.2. Value Function 31
4.1.3. Optimal Policy 32
4.1.4. Applications of MDP 32
4.2. Partially Observable Markov Decision Process 33
4.2.1. MDP Model Formulation 33
4.2.2. Value Function and Optimal Policy 34
4.2.3. Representation Methods of POMDP 36
4.3. Solution Algorithms 40
4.3.1. Exact Solution Algorithms 41
4.3.2. Approximate Solution Algorithms 45
4.4. The Quality Adjusted Life Years 46
4.5. Applications of POMDP 48
5.METHODOLOGY 49
5.1. Model Formulation 50
5.1.1. System State (Actual / Core State) 50
5.1.2. Observation State 51
5.1.3. Available Actions 52
5.1.4. Transition Probabilities 53
5.1.5. Observation Probabilities 55
5.1.6. Reward Function 57
5.1.7. Initial Belief State 59
5.1.8. Value Function 59
5.2. Colorectal Cancer Treatment Model 61
5.3. Solution Method 61
6.SCENARIO ANALYSIS 63
6.1. General Assumptions 64

6.2. The Effect of Unobservable Environment 65
6.2.1. POMDP Treatment Model 66
6.2.2. MDP Treatment Model 66
6.2.3. Computational Results 67
6.3. Optimal Policies 70
6.4. Comparison with Authorities 73
6.4.1. Comparison with SEER Survival 73
6.4.2. Additional Comparison 75
6.5. Robustness Analysis 76
6.5.1. The Effect of Diagnosing Stage 76
6.5.2. Changing Utilities 77
6.5.3. The Effect of the Clinical Experience 79
7.CONCLUSION 81
7.1. Recommendations for Future Study 82
REFERENCES 84
ÖZGEÇMİŞ 90

FIGURES

Figure 3. 1. Parts of the colon (National Cancer Institute, 2022) 23
Figure 3. 2. Adenocarcinoma sequence from normal colon to carcinoma 24
Figure 3. 3. Colorectal cancer staging (Mayer,2018) 26
Figure 4. 1. A finite part of an MDP 31
Figure 4. 2. Decision network of a finite part of a POMDP 35
Figure 4. 3. Belief space representation for a two state POMDP 37
Figure 4. 4. Discretized belief space demonstration 38
Figure 4. 5. One dimensional belief state - space demonstration 38
Figure 4. 6. Representation of a t-step policy tree for an action. 39
Figure 4. 7. One dimensional belief state - space demonstration 41
Figure 4. 8. 2- dimensional belief state - space demonstration 43
Figure 4. 9. The optimal policy representation with region 43
Figure 4. 10. PBVI demonstration 46
Figure 5. 1. Transition probabilities of POMDP model as a function of actions 55
Figure 5. 2. Point based value iteration algorithm 62

TABLES

Table 2. 1. Reviewed studies in the literature with the contribution of our work.....18
Table 5. 1. Stage and treatment dependent quality of life utilities from literature.....56
Table 5. 2. Stage and treatment dependent quality of life utilities from literature.....57
Table 5. 3. Combined utilities.....58
Table 5. 4. Used QALYs in the model.....58
Table 6. 1. The outputs of POMDP and MDP model.....68
Table 6. 2. %Difference between two model.....68
Table 6. 3. Policy map sample.....71
Table 6. 4. 5-year Relative Survival (%) (SEER based Results).....73
Table 6. 5. Survival results and %Differences.....75
Table 6. 6. QALY results with different utility functions.....78
Table 6. 7. QALYs based on different expert groups.....79

SYMBOLS AND ABBREVIATIONS

Symbols

A	Set of actions
a	Action index
A_S	Set of available actions in state s
a_t	Selected action at time t
B	Set of beliefs
b	Belief state index
b_0	The initial belief state
$B(s)$	Belief state space of state s
$b\ (s')$	The next belief state
h	Process history index
h_i	Health utility index for health state i
h_t	History of the process at time t
N	The last decision time index
o	Observation state index
o_t	Observation state at time t
P	Set of policy trees
p_i	Policy tree index
R	Reward index
$r\ (b, a)$	Rewards of being belief b and selecting action a
$R(\ s, a)$	Rewards of selecting action a in state s

$R(s, s', a)$	Rewards of selecting reaching state s' from state s with choosing action a
$R_N(s)$	Rewards at the last decision time N
S	Set of system states
s	System state index
s_t	System state at time t
s'	System state at time $t + 1$
T	Set of transition probabilities
t	Decision time index
u_t^*	Optimal utility at time t
$v^*(s)$	Optimal value in state s
$v_p(b)$	Value function of policy tree p in belief b
$v_p(s)$	Value function of policy tree p in state s
$v_t^*(b)$	Optimal value at t-step policy tree in belief b
$v_t^*(s)$	Optimal value in state s at time t
$v^\pi(s)$	Value of the policy π
$v^{\pi'}(s)$	Value of the next policy π'
Z	Set of observations
α_p	Alpha vector space for policy p
π	Policy index
π'	The next policy
π_0	The initial policy
π_t	The t-step policy
$\pi^*(s)$	Optimal policy in state s

Abbreviations

POMDP	Partially Observable Markov Decision Process
MDP	Markov Decision Process
CRC	Colorectal Cancer
QALYs	Quality Adjusted Life Years
DM	Decision Maker
SEER	Surveillance, Epidemiology, and End Results Program
MCRC	Metachronous Colorectal Cancer
LYs	Life Years
UNOS	United Network for Organ Sharing
NLST	National Lung Screening Trial
GLOBOCAN	Global Cancer Observatory
NCI	National Cancer Institute
PET/CT	Positron Emission Tomography / Computed Tomography
AJCC	American Joint Committee on Cancer
CEA	Carcinoembryonic Antigen
PSPACE	Polynomial Space
PBVI	Point Based Value Iteration
HR-QOL	Health Related Quality of Life
NCNN	National Comprehensive Cancer Network
ACS	American Cancer Society

1.INTRODUCTION

In healthcare, physicians often make stochastic decisions when faced with decision-making problems regarding the diagnosis, treatment, and screening of a disease. Because such decisions are fraught with uncertainty, predicting the long-term consequences of the prescribed medication as well as the future consequences of the decisions made is challenging. In addition to the stochastic decisions of the physicians, the patient's responses to the treatment protocol are also among the main causes of uncertainty. Thanks to the ability to overcome these challenges, the use of quantitative models becomes more and more important. Furthermore, according to 1999 research by the Institute of Medicine, medical errors were the major cause that led to death in the United States, with roughly 100,000 deaths each year. The annual cost of medical errors was estimated to be over $ 37.6 billion, with $ 17 billion of that being attributed to preventable errors (Donaldson et al., 2000). Therefore, making efficient, and accurate medical decisions is one of the major concerns in the healthcare system.

In theoretical framework and in practice, many mathematical models, such as stochastic, Bayesian networks, and decision trees, are being studied to support clinical decision-making. Among others, *Partially Observable Markov Decision Process (POMDP)* are one of the most convenient modeling methods, with the ability to represent hidden states as medical decisions are often made in dynamic and highly stochastic environments. For example, medical cases such as “drug infusion”, “ischemic heart disease”, and “breast cancer screening & treatment” have been modeled with *POMDP* (Schaefer et al., 2005). In order to perform robust modeling of diagnosis, treatment, and screening policies, it is essential to represent the patient's actual health conditions in the model. Actual health condition refers to the patient's overall health state and includes all known and undiscovered and/or unknown diseases and potential diseases with a high risk of development. In other words, it is an indicator of the patient's entire medical history. It is hardly possible to determine the actual health condition precisely. Because it depends on several variables such as the margin of error in the tests applied, the experience and knowledge level of the examining doctor, and the patient's statement of symptoms and

any yet unknown factors and/or measurements of an undiscovered characteristics Therefore, physical examination, detailed anamnesis, and various biomedical tests specific to the suspected disease are used to predict the patient's actual health condition. Thanks to *POMDP* models having the ability to represent hidden states it allows us to model the actual health state in a partially observable environment. The use of *POMDP* models is becoming more common for the more realistic modeling of the imperfect (partially observable) environment as *Markov Decision Process (MDP)* models assume that the actual state is fully known, which is unlikely in the real world.

1.1. Aim of Study

This thesis uses a *POMDP* framework to investigate the history of *Colorectal Cancer (CRC)* in order to address the following research questions:

1. What is the impact of unobservable health conditions on treatment policies and survival?
2. Does the proposed *POMDP* model provide an accurate representation for *CRC* treatment process?
3. Are the results obtained from the proposed *POMDP* model compatible with the literature and guidelines?
4. Is the model robust to model the different scenarios?

In order to investigate these research questions, Firstly, we propose a novel *POMDP* model to find the optimal treatment policies. We then solve the *POMDP* model to present our performance metrics (Quality adjusted life years, life years and 5-year survival rates). Next, we compare our model with current guidelines, studies and authorities in the literature. Finally, we present some hypothetical scenarios to analyze the robustness of the model.

1.2. Outline of Thesis

A summary of each section of the thesis is given as below:

In Chapter 2, we present a methodological literature review of the *POMDP/MDP* studies in healthcare. To deeply understand *CRC* modeling, we also review some colorectal cancer-based studies that used methods other than *POMDP/MDP* for modeling. We review the literature in two main categories as cancer-related and non-cancer-related studies within the framework of the thesis. We use three sub-categories as diagnostic, treatment, and screening to show the range and the context of the studies clearly.

In Chapter 3, we present a brief summary of *CRC*. We provide a background information on *CRC* and how the disease progresses over time. We explain the prognostic factors of the disease and give some epidemiological statistics about the disease. We also mention diagnosis procedures, treatment options, and screening policies.

Chapter 4 provides the general *POMDP/MDP* modeling details and mentions the most commonly used solution algorithms to understand the theoretical background of the study.

In Chapter 5, we present our *POMDP* model of the colorectal cancer treatment process. We explain the model parameters in detail and provide the assumptions to clearly frame the thesis. Chapter 6 presents our numerical results. In particular, we provide scenario analyses to investigate the robustness of the model and perform detailed sensitivity analysis for further analysis and investigation.

And finally, in Chapter 7, we discuss the results of the model, compare them with the findings in the literature, and current guidelines and present several recommendations for further studies.

2.LITERATURE REVIEW

In medicine, most decisions are made under conditions of uncertainty. Some known sources of uncertainty include a patient's actual health status, treatment response, measurement device sensitivity and precision, and physicians' clinical experience. Thanks to the ability to deal with uncertainty, the use of decision models in healthcare has increased significantly over the last two decades. *Markov Models*, *Bayesian Networks*, and *Decision Trees* are some of the commonly used decision-support models. Among the others, *POMDP* is one of the most appropriate clinical decision support models due to their capacity to represent the unobservable actual health status. Because the course of the disease is still unknown and there are too many treatment options, cancer is one of the diseases that are suitable to mode with *POMDP*. In fact, all stages of disease management, which are screening, diagnosis, treatment, and surveillance, can be modeled with Markov models. Since the detection time and age, treatment options, screening, and surveillance procedures can affect the patients' mortality and morbidity, mathematical modeling and finding the implementable good solutions can help improve the patients' quality of life and increase the survival rates. In this chapter, studies using *POMDP* for clinical decision support were carefully reviewed and categorized. Although the literature covers a wide variety of such models, we focused on two main categories which can support the framework of the thesis and make it clearer. These categories are cancer-related applications and non – cancer-related applications of *POMDP*. Although the literature presents these applications in a variety of contexts, we primarily focused on diagnostic, treatment, and screening models. We present all reviewed studies in Table 2.1 to frame the concept of review clearly and the contribution of the thesis.

2.1. Cancer-Related Applications

The science of oncology consists of uncertainty that includes the sensitivity and precision of diagnostic tests, actual disease stage, possible comorbidities, side-effects, and treatment decisions affected by patients, family members, and oncologists. Predictive decisions can be made in all stages of disease management, from screening to hospitalization. Although periodic screening can provide early diagnosis, the question of whom and when to screen is an important and major part of predictive decisions due to factors such as high cost and possible complications. Identifying the stage of the disease accurately for diagnosed cancer patients is also a prediction problem that is affected by the

physician's experience and the accuracy of monitoring agents. Treatment policies, such as risk-based choices regarding surgery, chemotherapy, and radiotherapy, are also affected by the predictions. In bladder cancer, “radical cystectomy” versus “transurethral resection”, in lymphoma “single-agent” versus “combination chemotherapy”, for gastric cancer “post-resection monitoring” versus “adjuvant chemotherapy”, and for breast cancer “modified radical mastectomy” versus “breast-conserving surgery” are all examples of risk-related decisions. (Vickers, 2011). Decision to initiate palliative therapy is survival prediction for patients who are in the terminal stage. Besides the probabilistic decisions, the cost of cancer is an issue that needs to be taken into consideration (Bullement et al., 2019). Decision models can represent the stochastic framework and offer cost-effective optimal or near-optimal solutions. Therefore, the integration of mathematical models in cancer management processes increases day by day. In this section, we focus on studies related with *POMDP* applications as clinical decision support. The applications in the literature are examined under three categories as diagnostic, treatment, and screening models.

2.1.1. Diagnostic Models

Decision models can assist physicians in diagnosing the disease and determining the right time to start the treatment by providing a myopic view. Diagnostic models are often intertwined with screening models. That is because screening models monitor disease progression from the origin. However, we found it more convenient to examine models under separate categories that aim to find the right time to start treatment and confirm the diagnosis of the disease. In this section, some diagnostic studies in the literature are reviewed.

Based on mammographic characteristics and demographic factors, Chhatwal et al. (2010) suggested a *POMDP* model to determine the optimal time to send a woman for biopsy. They formulated the problem as a discrete-time *POMDP* with a finite period. The decision epochs begin at the age of forty. The *Mammography Bayesian Network* was used to predict the likelihood of cancer related on mammographic characteristics and demographic factors. The risk scores (probability of cancer) used in the study were determined according to the probabilities obtained from *Mammography Bayesian Network*. The actions that the radiologist can take are specified as a “biopsy” or “an annual mammogram”. They defined the reward function as the total expected *Quality Adjusted Life Years (QALYs)*. They have found that the decision for biopsy should be made by considering the age of the

patient. Additionally, they compared the model with radiologist decisions and claimed that the model created outperformed radiologists in the biopsy decision-making problem.

Ayvaci et al. (2018) presented a stochastic modeling approach to optimize "risk-sensitive diagnostic decisions" following a mammography examination. The study aimed to find the maximum utility, defined as quality-adjusted survival time. They developed a methodology for assessing patients' risk preferences based on post-mammography decisions. In the model, "biopsy", "short-term follow-up", and "routine mammography" are options that a radiologist can suggest. They used a discrete-time *MDP* model. To ensure the *MDP* model is risk sensitive, the reward function of the model is created using "invertible utility functions". The model aims to find the maximum expected survival time. The state space represents "the risk status of cancer". They stated that *Decision-Maker (DM)* should consider the fact that loss of welfare is unavoidable for survival in situations where risk preferences are involved. Instead of using a risk-neutral approach, they suggested using an intensive "follow up" and "biopsy" plan, in which the myopic nature of medical judgments is more compatible.

Zhang (2012) proposed a *POMDP* model to investigate bladder cancer surveillance policies. The proposed model aims to find the optimal surveillance policy. They concluded that "age" and "comorbidity" have a substantial impact on the optimal surveillance policy. They used *QALYs* metrics as reward criteria. They also stated that as the age of the patient decreases, the follow-up period should be frequent, and patients with comorbidities should be followed less frequently, and hence less disutility of cystoscopy is, the more intensive surveillance is required. They also modified the *POMDP* model to include a novel "urine-based biomarker test" in the surveillance and compared the optimal policy to heuristic policies. They concluded that adding a biomarker does not improve optimal policy, but that biomarkers can significantly improve the heuristics policies.

2.1.2. Treatment Models

In oncology, it is vital to give the right treatment to the right patient at the right time. An oncologist should decide the efficient treatment options for each patient while considering the comorbidities if any exist, and side effects of treatment. All the available treatment options have serious potential for side effects. Physicians should decide whether chemotherapy, radiotherapy, surgery, palliative care, or combination therapy should be given to each patient in cancer treatment. Meanwhile, if a patient

develops drug sensitivity, physicians can dynamically change the treatment option. Thanks to the dynamic nature and myopic view, *POMDP* models are suitable for cancer treatment modeling. In this section, some of the treatment models in the literature are presented.

Cubbage (2004) modeled "the natural history of colorectal cancer" with a discrete vent simulation. He compared simulation models with Markov models and concluded that the discrete event simulation is more appropriate for modeling course and treatment of *CRC*. He set the model as a population study in which he demonstrated a patient lifetime in the simulation. He also considered adenoma incidents in the model as well as cancer incidents. He compared the model's results to the cumulative cancer risk in *Surveillance, Epidemiology, and End Results Program (SEER)* database and preexisting models in the literature. According to this study, about 2.5 percent of the population loses almost 10 years of their expected lifespan, meaning that colon cancer shortens the average lifespan by 0.24 years. He emphasized that the model was robust when compared to a similar model in the literature.

Erenay et al. (2011) studied "the natural history of *Metachronous Colorectal Cancer (MCRC)*" to predict some of the unknown parameters of the disease. In this study, they used discrete event simulation to model the course of the disease. They used the real patient data of 284 *CRC* patients from Mayo Clinic. In the model, they simulate the progression of cancer for 5-years as post-treatment of primary *CRC*. In conclusion, even the *CRC*- related and all-cause mortalities are significantly associated with age (they found that 5- years *MCRC* incidence is gender related (P = 0.005). They found estimated annual probabilities of transition from "adenomatous polyp" to *MCRC*" and from "*MCRC* to metastatic *MCRC*", "annual mortality probabilities of *MCRC*", and metastatic *MCRC* treatments within the study framework. They emphasized that the gender factor (more common in women) is also as important as an age factor for *MCRC*.

Joranger et al. (2020) proposed a *CRC* model using a semi- Markov model to estimate the survival rate, lifetime treatment costs, and changing treatment strategies costs. They used real patient data of 2049 *CRC* patients. The study was a cohort study, and they used patients with age>70. They described health states in the model as "alive without relapse", "alive with relapse", and "dead". They used *QALY* as an objective criterion. They considered recurrence and palliative treatment in their model. They constructed different treatment scenarios and compared them in terms of cost, survival, and

QALY gained. They computed the estimated treatment cost for a 70-years old *CRC* patient and found out that the use of palliative therapy increases the cost by up to %29. They found that cancer treatment increased *Life Years (LYs)* by 6.05 years. In conclusion, according to their model, the overall *CRC* treatment costs are low to health gain, and palliative therapy is a major part of the treatment cost.

Goulionis and Koutsiumaris (2010) proposed a *POMDP* model to provide "Early Prostate Cancer". treatment support. This study developed an extended policy iteration model for prostate cancer treatment. In the model actions were defined as "watchful waiting", "radiotherapy", and "surgery". The aim of the model is minimizing the total discounted costs. The estimated overall cost is defined from the range of the approximate costs of prostate cancer screening and treatment. They reconstructed the *POMDP* model as a *belief MDP,* as control policies are information vectors containing all the information necessary to choose an action at a given time. The authors did not recommend that the model results be used to directly influence treatment policy. They did, however, present proposed models that could be useful for testing potential treatment alternatives at a low cost, in a short time, and without putting patients at risk.

Nouri (2019) studied the *POMDP* method to optimize radiotherapy planning. Cancer patients are given radiation therapy to kill tumor cells and prevent them from spreading. Usually, the prescribed radiation therapy is given to the patient over several treatment sessions (fractional treatment plan) to prevent fatal damage to surrounding healthy organs, called organs at risk. According to the current guidelines, treatment is prescribed as an equal dose of radiation to the patient over multiple treatment sessions. This procedure ignores, some uncertainties associated with tumor dynamics, biological response to radiation, and organ movement that occur during radiation therapy. The proposed *POMDP* model is implemented in two cases of prostate cancer and pediatric ependymoma. The model aims to maximize the expected bioequivalent dose of the tumor under the organ at risk survival constraint. According to the model findings, the author suggested using lower doses for earlier sessions and higher doses for subsequent sessions.

2.1.3. Screening Models

In cancer treatment, the time of diagnosis and treatment epoch is vital for many aspects. Early diagnosis can increase the chances of survival while reducing the treatment costs. For example, the

10-year survival rate for Stage 3 colon cancer is 52% whereas it is 92% for Stage 1 colon cancer (McLeish et al., 2002). Because early detection of cancer is crucial to the outcome of treatment processes, researchers' efforts to develop effective screening technologies to diagnose cancer at an early stage have led to the development of many screening methods, such as “clinical breast examination”, “mammography”, and “magnetic resonance imaging”, “colonoscopy”, “fecal occult blood test”, “sigmoidoscopy” etc. The development of computer-aided modeling technologies has increased the use of clinical decision support systems. For this reason, decision support models have been widely used in cancer screening. In all decision support models such as Bayesian Network, Decision Trees, etc. *POMDP* models provide an efficient framework to optimize screening decisions since their nature allows that the representation of the unobservable health status of a patient. Alagoz (2011) presented a brief tutorial for mammography screening to demonstrate the development and application of a POMDP model for cancer screening. In this section, some of the screening models in the literature are reviewed.

Leshno et al. (2003) proposed a *POMDP* model to evaluate the cost-effectiveness analysis of *CRC* screening for the average-risk population, which includes the age range over 50 years old. They compared the effectiveness of the following strategies: “No screening”, “one-time colonoscopy screening”, “colonoscopy (10-year interval following colonoscopy)”, “annual fecal occult blood testing”, “annual fecal occult blood testing and sigmoidoscopy in the 5-year interval”, “annual detection of altered human DNA in the stool”. They evaluated the incremental average cost-effectiveness ratio for each strategy (additional expected cost divided by additionally expected effectiveness). The authors suggested that screening the average-risk asymptomatic individuals is highly cost-effective. “One-time colonoscopy” or “annual fecal occult blood testing and sigmoidoscopy in the 5-year interval” would be the preferred test for screening, starting screening at age 50.

Erenay et al. (2014) presented a *POMDP* model to optimize the colonoscopy screening policies. The goal of the model was to maximize total *QALYs*. In the model, they represented the patient’s health states such as without lesion, having adenomatous polyp, and having *CRC.* The model led to the conclusion that the proposed optimal policies recommend more frequent colonoscopy screening than the related guidelines. In addition, the proposed model recommended that women with a history of *CRC* should be screened more frequently than men whereas women without a history of *CRC* should

be screened less frequently than men. This result emphasized the role of gender in optimal *CRC* screening decisions. According to the model younger patients should be screened more frequently than older patients. This result contradicts with guidelines that recommend screening starting at the age of 50.

Zhu (2010) modeled the natural history of *CRC* with *POMDP*. The model's states are probabilistic representations of a patient's health status. At the beginning of each year, the patient can choose whether to get a colonoscopy in the consecutive year. The *POMDP* model was used to include the uncertainty since the patient's initial condition is not known by the *DM*. In the model, the planning horizon is between age 40 and 100. The state was defined as the size and the number of adenomas. The actions that can be taken by a physician is either perform colonoscopy or wait until the next decision epoch. If the action is a colonoscopy observation state is defined according to the type of adenomas. Otherwise, the observation state is defined as either normal or symptomatic. They used the incremental cost-effectiveness ratio to evaluate the *CRC* screening policies whereas the effectiveness of a screening policy is measured by the total *QALYs*. They concluded that the optimal start age for patients, who receive a one-shot colonoscopy, is 65. According to the referenced guidelines, the optimal start age has been given as 61.

Ayer et al. (2012) has developed a *POMDP* model for a personalized mammography screening policy based on previous screening results and individual risk characteristics. Mammography is known as the most efficient method for breast cancer screening. The model contains unobservable disease evolution, mammography test features, and two detection strategies as "self" and "screen". As an objective function, they used the maximization of the total expected *QALYs*. The study considers ages over 40. The state set of the model is defined to include the presence of cancer if any, the prognosis of the disease, and the state of death. They compared the model's result with the current guidelines and showed that the proposed personalized screening model gave better results compared to the guidelines in terms of total expected *QALYs*.

In lung cancer detection, the low dose computed tomography is used to detect individuals at high risk. However, the outputs of this tomography contain a significant level of false positives. Petousis et al. (2019b) proposed a new approach to reduce the false positive rate of lung cancer screening while improving true positive rate. They combined machine learning and sequential decision-making

methods to construct a predictive model of personalized cancer screening. They presented a *POMDP* framework for learning that gradually improves the screening action selection based on previous observations. They trained a dynamic Bayesian network and applied inverse reinforcement learning to generate a reward function. In the model, they used three state which are "no cancer" which implies healthy status, "uncertain" which implies suspicious state with some abnormalities, and "lung cancer" which implies diagnosed lung cancer. Defined actions are to "continue screening with a follow-up tomography" or to "recommend an intervention". Observations in the model are "annual screens and interpretation" and "diagnostic intervention findings". A Dynamic Bayesian Network was used to determine the observations as a probability. They used inverse reinforcement learning to determine the reward function. (Petousis et al., 2019a). In conclusion, the proposed model reduced false positive rates while keeping true positive rate high. Furthermore, the model diagnosed a significant number of cancers earlier than the doctors.

"Cytology" and "Human Papillomaviruses -DNA" are primary screening tests for cervical cancer. (Obulaney et al., 2016). Because pre-cancerous lesions evolve slowly, cervical cancer is preventable with an efficient screening policy. Human Papillomaviruses -DNA test results are more accurate than cytology, yet it is a high-cost test and inaccessible in many communities. Namen Leon et al. (2015) proposed a finite horizon *POMDP* model to define the starting age and test frequency for a patient both tests. In the model, they defined the state space with virus infection and cancer stage to represent the actual health state. The decision-maker can take one of the three actions as "wait", "cytology", and "Human Papillomaviruses -DNA" tests for all decision epochs. They solved the model by using Monahan's algorithm with Eagle's reduction. In conclusion, the results outperformed the guidelines in terms of *QALYs.*

2.2. Non-Cancer-Related Applications

In medicine, all decision-making steps involve some type of uncertainty. First of all, each patient is unique in terms of treatment response. Therefore, physicians should make treatment decisions considering the common side effects, rare but highly mortal side effects, comorbidities -if any exist, age, gender, and weight of the patient. The diagnosis and treatment approaches of physicians vary depending on their theoretical knowledge and clinical experience. In addition, the etiology of most diseases is still unknown. Therefore, treatment policies are usually symptomatic rather than oriented towards eliminating the causes. The time of diagnosis, the age of the patient, and the treatment policy

may affect mortality and morbidity. Physicians are often intuitive when deciding on a patient-specific diagnostic and treatment. Especially for chronic diseases like diabetes mellitus, hypertension, and heart failure early diagnosis has a crucial effect on survival and quality of life. For such diseases, diagnosis and screening models can help to increase the effectiveness of treatment and surveillance and decreasing the cost of the procedure. However, there is always a treatment refusal probability for the patient. There might also be a shortage of resources for the decision-maker to consider i.e., blood, cadaveric organs, bone marrow, etc. Considering all, determining the most appropriate treatment policy for each patient in this uncertain environment turns into a very complex decision-making problem. Mathematical models can be used as supportive agents to close the gap in the clinical experience of physicians. Quantitative models can help track patients' medical history, improve personalized medicine, and eventually reduce mortality. Thanks to the ability to deal with uncertainty, MDP/POMDP are also suitable modeling methods for many diseases other than cancer. In this section, some studies in the literature are reviewed under three categories.

2.2.1. Diagnostic Models

Treating a disease begins with diagnosis. Since the diseases progress over time, the time of diagnosis affects the success rate. Early diagnosis can help improve the survival. However, detecting disease at the beginning of its existence is not an easy task. Especially in Turkey, due to the high number of patients per physician and the high patient density, early diagnosis may be missed in some diseases where the examination time per patient is short. Due to the increase in the workload of the physicians with the Covid-19 pandemic, it has become difficult for the patient to access the specialist. All these have further reduced the possibility of early diagnosis in diseases such as cancer, where early diagnosis is of vital importance. Physicians use up-to-date guidelines to make the correct diagnosis. Disease definitions, diagnostic criteria, and diagnostic tests can change dynamically depending on scientific developments. Also, a diagnostic test can give false-negative or false-positive results. Therefore, no single diagnostic test is sufficient to diagnose a particular disease. Mathematical models are suitable tools to support the decisions for diagnosis and hence they can help reduce the burden of physicians by assisting them mathematically and statistically. In this section, some of the studies that used *POMDP* for diagnostic modeling are presented.

Sehr and Bitmead (2017) investigated a version of Stochastic Model Predictive Control with the dual optimal policy structure. They propose a *POMDP* modeling approach to system dynamics with a

finite horizon stochastic optimal control problem and demonstrated that the problem can be solved with a dual optimal control policy. They explained that the benefit of the dual approach could maintain optimality in contrast with the common version of Stochastic Model Predictive Control. They presented a case scenario of a hypothetical patient who was being treated for an illness that could be controlled but not cured. They described the disease in three-stages with respect to the disease level. They demonstrated the use of a costly diagnostic test to update current information without impacting patient status in the proposed scenario. They stated that if there was no duality in control input, diagnostic tests would result badly for treatment recommendations.

Latha and Vetrivelan (2019) studied on heart disease prediction model using *POMDP*. Heart related disease is one of the most common cardiovascular diseases. The risk factors for heart related disease are high blood pressure, increased blood viscosity, diabetes, obesity, etc. The patient's health condition is monitored using observable values and belief states and evaluated with the integrated *POMDP* directed graph. The model aims to accomplish early diagnosis by alerting the patient and sending an ambulance by means of fog computing. Physicians can access and track patient situation, heart disease risk factors, medical history via "iFogSim". In the model, they track the patient's blood viscosity as a state and the goal is maximizing the expected long-term rewards.

The Covid-19 pandemic has brought some problems with diagnostic tests. Testing is essential to control the spread of the disease. With common testing of infected individuals and their close contacts, the environment can be quarantined on time to reduce the spread. However, the source of test kits is not limitless and needs to be used smartly. Therewith, not all infected individuals are symptomatic and asymptomatic individuals are highly risky to spread of disease. Singh et al., (2020) studied the development of efficient testing strategies for Covid-19 using *POMDP*. In the model, they aimed to find the minimum number of infected individuals to be tested as soon as possible, under the test capacity constraint. They modeled the social contacts as a time-varying, weighted, and undirected graph. An individual's disease status is an unknown parameter until confirmed by observations. The available actions are defined as test and quarantine. They assumed that the virus could infect only one person during two consecutives periods and, they ignore some limitations of the real world like test accuracy, actual contact graphing limitations. Even so, they mentioned some sub-optimal solutions.

2.2.2. Treatment Models

After diagnosing the disease, physicians should decide on the treatment procedure that is to be implemented for the patient. Since treatment policies may change with age, comorbidities, and the disease situation it is also essential to give the right treatment to the right patient. To accomplish that, physicians should evaluate the demographic characteristics and medical history of the patients as well as possible side effects of the treatment. The treatment process is stochastic and dynamic so could change in every step according to the patient's response or new treatment options. For physicians, it is crucial to track current guidelines to keep up to date. Treatment models can help to find optimal personal treatment policy, and artificial intelligence-supported models can help to keep up to date. In this chapter, some of the studies in the literature using *POMDP* are mentioned briefly.

Ibrahim et al. (2016) developed a mixed Markov model to personalize "anticoagulation therapy" for stroke prevention. There are two stages to the research. The doctor analyzes the patient's sensitivity to warfarin during in the initiation stage, which is modeled with *POMDP*. That is the model's unobservable feature. Warfarin is an anticoagulant medication that is widely prescribed in worldwide. Adjusting a stable dose of warfarin is an important issue for cardiology.

The physicians use their established belief about patient sensitivity to estimate the optimum warfarin dose in the maintenance stage, which is modeled using an *MDP.* The purpose of model is minimizing the total discounted expected risk. The length of the initiation stage is substantially influenced by the initial belief regarding the patient's health state, according to the findings. It was also reported that the myopically optimal policy performs similarly to the optimal *POMDP* strategy in terms of average risk and time in therapeutic range.

Bennett and Hauser (2013a) developed a "non-disease-specific computational/artificial intelligence" framework to investigate optimal treatment methodology. This framework consists of two potential functions: "various health policies, payment methodologies", and a clinical artificial intelligence framework - "an artificial intelligence that can think like a doctor". They used *POMDP* and dynamic decision networks to build a model that learns from clinical data and can establish complex strategies with the simulation of alternative scenarios. The model maintains beliefs about the patient's state of health and acting as an online agent obtaining new observations. The aim of the model is to maximizing improvements in patient health while minimizing the cost of treatment. They claimed

that, despite its in early stages, the study outperformed established case rate/cost per service models. This highlights the applicability of these approaches to human decision-making performance.

Patients suffering from “End-stage liver disease” apply to a waiting list to reach a cadaver liver transplant. Patients can get partial information about the “Liver transplant waiting list” through the *United Network for Organ Sharing (UNOS)* liver allocation system's website. This system aims to help patients make better predictions about their queue order on the waiting list. (Sandikçi et al., 2008) proposed a *POMDP* model to decide which potential liver would be accepted and which would be rejected. The decision to admit a transplant is based on the actual health status of a patient and the components of the waiting list. They explained the concept of the privacy price of the patient, which is the expected number of life lost days related to a lack of accurate waiting list information. They proposed two versions of the *POMDP* model. The first version assumes excellent backlog information and gives upper limits on the price of real privacy in comparison. The second version relaxes the assumption of “perfect information” and gives an admissible representation of the waiting list. The study highlighted that when a patient knows her location on the list is near the top, she is considerably more selective, and as her position drops, she gets less selective. However, the study has some limitations regarding the traceability of liver transplantation procedures.

2.2.3. Screening Models

The main purpose of screening a disease is to provide early diagnosis. Because it can prolong the survival time and meanwhile reduce the cost of treatment. Some diseases such as “cancer”, “ischemic heart disease”, and “diabetes” can be prevented, at least their prognosis can be controlled with early diagnosis. One of the advantages of periodic screening of the population is that it can reduce treatment costs and saves physician time. It's hard to find solutions to questions about who should be screened and when. The screening procedure is very costly. In the meantime, it is not always possible to reach people who have not yet had the disease. Some diseases are inherited genetically. In such cases, the family should be screened after the diagnosis is made. Some disease risk increases with age or gender even if it is not genetic i.e., osteoporosis, heart failure, hypertension, etc. For such diseases, a screening policy can be followed from a certain age, or screening policies based on gender can be implemented. Screening models aim to determine the optimal screening interval, and most models are compared to current guidelines from cost and age range points of view. In this section, some of the screening models in the literature are reviewed.

Vozikis and Goulionis (2009) proposed a *POMDP* model Parkinson's disease. Parkinson's disease is a common neurological disorder that can cause a decrease in quality of life for both the patients and their families. The study aimed to determine the critical threshold value required for a surgical operation to minimize the total lifetime cost. In the model, actions that physicians could choose was defined as medical treatment with incomplete monitoring and surgical operation. The proposed model creates a simulation of the disease monitoring. In conclusion, they were able to find the optimum average cost policy. In the study, it was also emphasized that the model is sufficient in terms of estimation the clinicians' decisions by implementing it in clinical practice.

Tuberculosis is a contagious disease and can cause death if not treated properly. Early diagnosis is vital as it can help prevent the spread of the disease. There are two types of diagnostic tests for Tuberculosis as skin test and the blood test. The skin test is more cost-effective but less sensitive than a blood test. Kiani et al. (2020) proposed an *MDP* model to decide on the type of Tuberculosis test to minimize the overall costs. For this purpose, they grouped healthcare workers into multiple risk groups according to their job and department, and the birth country. They used Approximate Dynamic Programming to find a "near-optimal" solution and they suggested a simple policy that could be used by healthcare facilities. The proposed policy outperformed the guidelines in terms of costs.

Cipriano et al. (2018) searched for optimal information collection policies within the *MDP* framework on a "birth-cohort hepatitis C virus" screening model. The study was designed as a cohort study that focuses on 50 year old patients at a causal hospital visit. In each period, the policymaker can choose between actions defined as screening the current 50 year old group and purchasing sample information on the parameters that influence the decision. The goal is maximizing the expected net monetary benefit and they suggested that it may be optimal to delay information gathering until information affect decision-making more rapidly. They emphasized that for hepatitis C virus screening, the proposed cost-optimal policy outperformed current guidelines, given the initial beliefs.

2.3. Summary of Literature

With the widespread transfer of medical data to the electronic environment, the use of mathematical models in medical science is increasing. Mathematical models not only contribute to the understanding of modeled diseases but also enable obtaining useful statistics about the course of the disease and evaluation of treatment policies over many criteria. In the previous section, we gave a brief summary of the studies on Markov modeling in the medical field. We have shown the studies mentioned in this section by categorizing them into several categories. These categories are the type of disease, category of the study (diagnostic, screening, and treatment), type of study based on target data (population or cohort), objective function, and methodology used. In addition, we give the year of the study, and the data type of the real or non-real data as R, N, respectively. In *Table 2.1,* we present our proposed model with the specified categories to show the place of our study in the literature.

Table 2. 1. Reviewed studies in the literature with the contribution of our work

Study	**Disease**	**Category of Study**	**Type of Study (Data)**	**Objective Function (Cost/Reward)**	**Method**	**Data (R/N)**
Leshno et al. (2003)	Colorectal Cancer	Screening	Cohort (age>50)	Discounted LYs earned, costs (cost effectiveness analysis)	POMDP	SEER
Cubbage (2004)	Colorectal Cancer	Treatment	Population	Years of life loss	Discrete Event Simulation	SEER
Sandikçi et al. (2008)	End-stage liver disease	Treatment	Cohort	Maximizing the patient's expected total remaining life	POMDP	UNOS
Vozikis and Goulionis (2009)	Parkinson's Disease	Screening	Cohort	Minimizing the total lifetime cost	POMDP	R
Goulionis and Koutsiumaris (2010)	Prostate Cancer	Treatment	-	Minimizing the expected total discounted cost	Discrete-time POMDP	N
Zhu (2010)	Colorectal Cancer	Screening	Population	QALYs/ Cost Effectiveness Ratio	POMDP	From many sources

Chhatwal et al. (2010)	Breast Cancer	Diagnostic	Cohort Study (age >40)	QALYs	Finite-period discrete-time POMDP	R
Erenay et al. (2011)	Colorectal Cancer	Treatment	Cohort (50<age<79)	Estimating unknown parameters (survival, incidence, etc.)	Discrete Event Simulation	R
Zhang, (2012)	Bladder Cancer	Diagnostic	Cohort	QALYs	POMDP	R
Ayer et al. (2012)	Breast Cancer	Screening	Population (age>40)	QALYs	Finite Horizon POMDP	From many sources
Bennett and Hauser (2013a)	Non-disease specific	Treatment	Population	Maximizing the improvement in patient health, Minimizing cost of treatment	POMDP/DDN/AI	R
Erenay et al. (2014)	Colorectal Cancer	Screening	Cohort (age>50)	QALYs	POMDP	From many sources
Namen Leon et al. (2015)	Cervical Cancer	Screening	Population	QALYs	Finite Horizon POMDP	N

Ibrahim et al. (2016)	Atrial Fibrillation	Treatment	Cohort	Minimizing the cumulative discounted expected risk	POMDP/MDP	R
Sehr and Bitmead (2017)	Hypothetical Disease	Diagnostic	-	Minimizing total expected costs	POMDP	N
Cipriano et al. (2018)	Hepatitis C Virus	Screening	Cohort (age >50)	Maximize the expected net monetary benefit	MDP	N
Ayvaci et al. (2018)	Breast Cancer	Diagnostic	Cohort (age > 40)	Maximize the total expected quality adjusted longevity	Finite-horizon discrete-time MDP	R
Nouri (2019)	Prostate cancer/ Pediatric ependymoma	Treatment	Population	Maximizing the expected biological equivalent dose of tumor	POMDP	R
Petousis et al. (2019b)	Lung Cancer	Screening	Cohort (nodules>6mm)	Maximizing early disease detection while minimizing false positives	POMDP/Dynamic Bayesian Network	NLST
Latha and Vetrivelan (2019)	Heart Disease	Diagnostic	Population	Maximizing the expected long-term rewards	POMDP	N

Kiani et al. (2020)	Tuberculosis	Screening	Cohort	Minimizing the overall costs	MDP	R
Singh et al. (2020)	COVID-19	Diagnostic	Population	Minimizing total number of infected individuals	POMDP	N
Joranger et al. (2020)	Colorectal Cancer	Treatment	Cohort (Average age =70)	Estimate expected lifetime cost, survival & QALYs	Semi Markov Model	R
Our work	Colon Cancer	Treatment	Cohort (age >50)	Maximizing the expected LYs, QALYs	POMDP	N

3.BACKGROUND ON COLORECTAL CANCER

Colorectal cancer is the world's third most frequent cancer as the second highest cause of cancer-related fatalities. According to *Global Cancer Observatory (GLOBOCAN)* 2020 data, new colorectal cancer cases were 1,880,725; and the number of deaths due to *CRC* were 915,880, the incidence and mortality ranking of *CRC* in Turkey is the same as the worldwide ranking (Sung et al., 2021a). Despite the high incidence and mortality rates, in early detection, colorectal cancer is not dangerous and is highly curable. Thanks to developing screening methods and increasing treatment options, mortality rates have decreased by approximately 25% in the past 25 years (Mayer, 2018).The course of the disease and survival rates depend on many factors such as tumor grade, gender, age, and family history. The five-year relative survival is 90.6% in patients diagnosed at a localized stage, and in a regional stage the rate is 72.2% and in patients with distant metastases is only 14.7% (Surveillance Epidemiology and End Results Program, 2018). The probability of developing *CRC* in the whole population is 4.2%. The risk for men (4.3%) is slightly higher than for women (4%) (Sung et al., 2021b) The *CRC* incidence among people under 50 is extremely low and increases with age. Individuals who have a first-degree relative with *CRC* are twice as likely to develop it (Altekruse et al., 2010). In this section, we briefly explain the fundamental characteristics of colorectal cancer with the help of textbooks, current guidelines, and experts' help and we also present several measures and methods to analyze the lifetime of patients.

3.1. Overview for Colorectal Cancer

In this thesis, our objective is to develop and evaluate a mathematical model in order to demonstrate the colorectal cancer treatment history. To deeply understand the model, it is essential to understand the nature of colorectal cancer and current treatment protocols. Therefore, we give a brief background for *CRC.*

The colon and rectum are the last parts of the human digestive system where the absorption and excretion of digested food take place. The sections of the colon are shown in *Figure 3.1.*

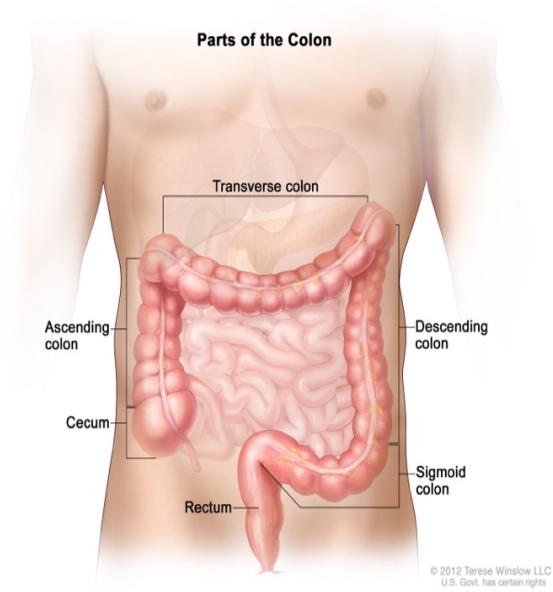

Figure 3. 1. Parts of the colon (National Cancer Institute, 2022)

While the exact cause is still unknown, the transformation of healthy colonic mucosa into invasive carcinoma is thought to result from a series of molecular size disruptions that include microsatellite instability and chromosomal instability. In most cases, colorectal adenomas are at the origin of cancer (Mayer, 2018). The following section describes the formation and progression of cancer in detail.

3.1.1. Molecular Pathogenesis

A polyp is a protrusion that can be seen from the surface of the mucosal surface. Polyps are pathologically divided into three groups as “juvenile polyps”, “hyperplastic polyps”, and “adenomatous polyps”. Only “adenomatous polyps” are categorized as precancerous among these three types of groups, and only about 1% of “adenomatous polyps” progress to malignancy. “Adenomatous polyps” are found in the colons of 30% of middle-aged people and 50% of the elderly. Once adenomas are detected, colonoscopy should be performed periodically in such patients as the risk of developing colorectal carcinoma is higher than average. Since adenomatous polyps become clinically significant in about 5 years, the recommended frequency of colonoscopy is every 3 years (Mayer, 2018). The following *Figure 3.2* describes the progression of the normal colon to carcinoma.

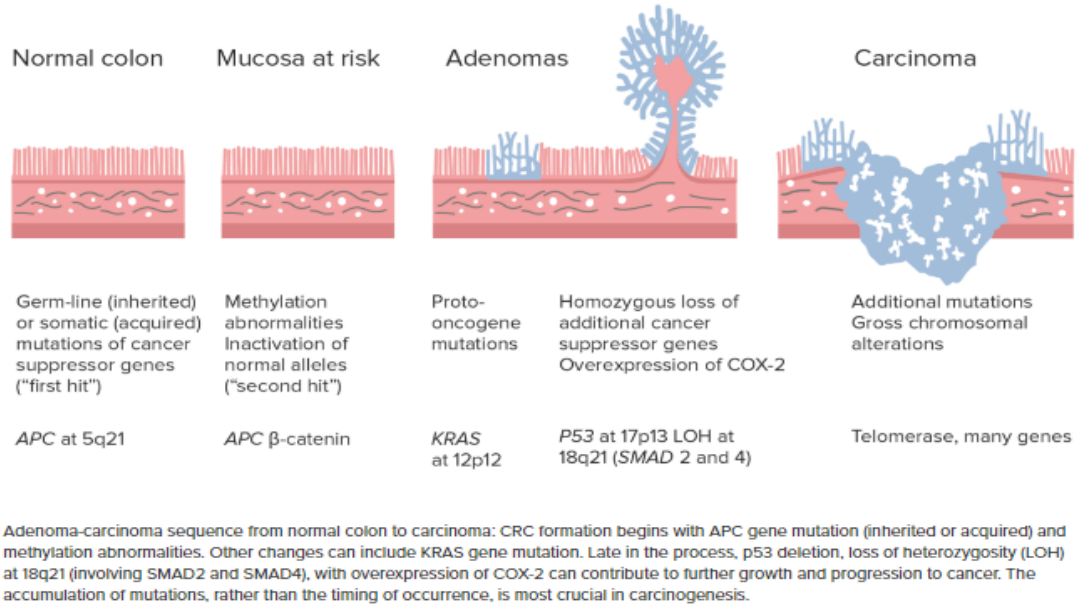

Figure 3. 2. Adenocarcinoma sequence from normal colon to carcinoma (Kumar et al., 2013)

Figure 3.2. shows the transformation of a normal colon to adenocarcinoma in four steps. (Kumar et al., 2013) In the first step, the normal colon with cancer suppressor genes mutation is shown. This step is the beginning of the history of *CRC*, and it is almost impossible to detect cancer at this step. Mutations continue cumulatively, which puts the mucosa at risk. Later, the process continues with additional loss of cancer suppressor genes that lead to adenoma formation. Eventually, cumulative gene mutations develop into adenocarcinoma.

3.1.2. Risk Factors of *CRC*

Most occurrences of colorectal cancer are caused by environmental factors rather than a genetic background. Nonetheless, up to 25% of patients have a family history (Mayer, 2018). Regarding the incidence, geographical differences are mostly related to dietary habits and unrelated to genetic differences. Diet, particularly diets rich in animal fat and calories, is an important causative role. Therefore, the incidence rate is higher in upper socioeconomic populations living in urban areas than in lower socioeconomic populations. Polyps can become cancerous due to factors including such animal fat consumption, insulin resistance (obesity), and fiber deficiencies. Smoking is one of the

reasons, which is thought to be associated with the incidence of colorectal adenoma, especially in smokers over the age of 35 (Mayer, 2018). Individuals with chronic "inflammatory bowel disease" are also at an increased risk of cancer. Both diseases have the same symptoms, such as "bloody diarrhea", "abdominal cramps", and "congestion". Therefore, it is difficult to detect cancer for patients suffering from inflammatory bowel disease".

3.2. Diagnosis and Treatment Methods

At the onset of the cancer, the majority of patients usually do not show any symptoms until the disease progresses and metastasis occurs. This is one of the reasons why screening is so vital for the early detection of precancerous and/or invasive cancer. Depending on where the tumor is located, the symptoms may vary. The most common symptom is iron deficiency anemia with an unknown cause (for postmenopausal women and all men). The other common symptoms are changes in bowel habits, abdominal cramping, occasional obstruction, and fatigue. These symptoms may lead physicians to another disease. For example, the rectal bleeding and altered bowel habits can lead the physicians to both *CRC* and hemorrhoids. Therefore, differential diagnosis is crucial to make the accurate and correct diagnosis. For the diagnosis of colorectal cancer, physicians make a physical examination, take clinical anamnesis, and examine blood measurement (complete blood count, biochemical tests) and *Positron Emission Tomography and Computed Tomography (PET/CT)* results, and then they perform colonoscopy and take a sample for biopsy. If any cancerous lesion is detected, then the treatment protocol begins.

3.2.1. Cancer Staging

There are several ways to classify cancer as pathological. The most used ones are *American Joint Committee on Cancer (AJCC)* classification, Modified Astler-Coller Duke's, and *SEER* classification. Duke's classification classifies cancer as A, B, C, D. *SEER* uses a more general three-fold classification as regional, distant, and metastasis. Among these classifications, the most detailed and widely used one is the *AJCC* classification. There are several editions as 5^{th}, 7^{th}, and 8^{th}. The 7^{th} and 8^{th} editions are

extended versions of the 5th edition. In Turkey, oncologists are frequently using the 5th edition. Therefore, we used the *AJCC* 5th edition in the thesis for classification.

TNM Classification Method: In the TNM abbreviation, T, N and M refer to "depth of tumor penetration", "lymph node involvement", and "distant metastases", respectively. The stage of the disease is assumed to be determined by surgical resection and pathological analysis (biopsy) of samples. Colorectal cancer often metastasizes to the liver, and other common sites of metastasis are the lungs and bones (Kumar et al., 2013).

Stage 0 (Tis; N0; M0): The newly formed malignancy is isolated in the "inner lining of the colon".

Stage I (T1–2; N0; M0): Superficial lesions, regional lymph nodes are not involved, and tumors cannot extend beyond the "submucosa (T1)" or "muscularis (T2)".

Stage II (T3-4; N0; M0): Tumors pass through "the muscularis propria" however, spread to lymph nodes does not occur

Stage III (TX; N1-2; M0): Regional lymph nodes involvement are observed.

Stage IV (TX; NX; M1): Metastatic spread to areas such as liver, lungs, or bones

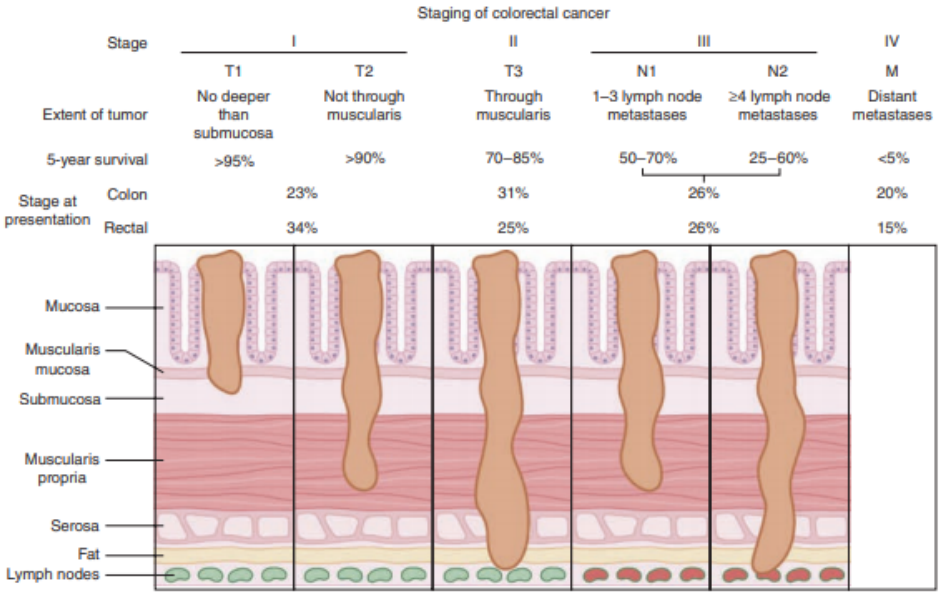

Figure 3. 3. Colorectal cancer staging (Mayer,2018)

Figure 3.3 shows pathologic staging of *CRC* (Mayer, 2018). In Stage I, II, and III the cancer is local. In Stage IV, it requires a multidimensional treatment protocol as the tumor spreads throughout the body.

3.2.2. Treatment Procedure of *CRC*

The treatment procedure may differ according to the tumor stage, patients' age, comorbidities, and medical history. The primary treatment of colorectal cancer is a surgical operation, if possible. If a patient is not suitable for immediate surgery, chemotherapy, radiation therapy or combination therapy can be applied. Recently, targeted therapy has been being used in cancer treatment, but this option is still in the trial phase.

Surgical Operation: The primary treatment for colorectal cancer is the complete removal of the tumor, if possible. The surgery is performed to remove malignant tissue and clean the colon. Before surgery, physical examination, *Carcinoembryonic Antigen (CEA)* level measurement, biochemical assessment of liver function, comprehensive *CT* scan, and colonoscopy (if possible) are done to detect concurrent polyps. A careful surveillance policy should be followed for 5 years after resection which includes *CEA* levels measurement, and *CT* scan. The examinations should be performed at 3–6-month intervals. The probability of recurrence of colon cancer in recovered patients is 3-5%, and the probability of developing adenomatous polyps is > 15%. Since recurrences after surgery mostly occur within the first 4 years, it is reliable to use 5-year survival as an indicator (Mayer, 2018).

Chemoradiation/radiation therapy: The purpose of chemoradiation therapy is to prevent the cancer spread to other organs and/or lymph nodes. The use of radiation therapy alone as a primary treatment is not suggested. Radiation therapy combined with chemotherapy, on the other hand, can substantially reduce relapse and enhance overall survival (Mayer, 2018).

Chemotherapy: The purpose of chemotherapy is to prevent cancer cells from feeding and to enable the cancer cells to shrink and disappear. Chemotherapy cures can provide a median survival of 2 years in metastatic disease. In addition, monoclonal antibodies (targeted therapy) are effective in advanced cases by targeting the epidermal growth factor receptor. Targeted therapy responds particularly well in recurrent cases (Mayer, 2018).

Adjuvant Therapy: The purpose of adjuvant therapy is to increase the effect of surgery with combining it with chemo and/or radiation therapy. Adjuvant therapy can be applied pre-and post-surgery. It is recommended in patients with a high probability of recurrence. Adjuvant therapy can reduce the risk of recurrence and improve response rates in patients with stage II and III tumors (Mayer, 2018)

In the thesis, we considered the treatment protocol as a treatment type, not on a drug basis. Also, we ignored radiation therapy for computational simplicity. We determined treatment methods as surgery, chemotherapy, adjuvant chemotherapy, and chemoradiotherapy.

3.3. Screening Policies

Effective screening programs can prevent the formation of colorectal cancer by eliminating adenomatous polyps, and early detection of cancerous polyps increases the success rate of treatment. The cost of screening is much lower than the cost of treatment and healthy individuals are the most important issue. Therefore, it is an important step in cancer treatment. Individuals with a family history of cancer and the elderly population over the age of 50 are the targets of these programs. Screening methods include rectal examination, stool test, imaging, and endoscopy. According to guidelines, healthy persons with no family history should be screened without colonoscopy every five years. After the age of 50, the entire population should be screened and colonoscopied every ten years.(National Comprehensive Cancer Network, 2022)

4.THEORATICAL BACKGROUND

This chapter first introduced the basics of the *MDP* model, which has the same theoretical background as the *POMDP* model. As explained later in this section, *POMDP* is a relaxed version of *MDP* where system states are not fully observable. After the *MDP* model explanation, we present the theory of the *POMDP* models in this framework. After explaining the theory, frequently used exact and approximate solution methods are given. Finally, we briefly explained *QALY* metrics and give some methods for utility calculation.

4.1. Markov Decision Process

MDPs, also known as "Stochastic Dynamic Programming" or "Stochastic Control Processes", are sequential decision-making models used to determine the optimal solutions when the environment changes over time and/or the outcomes of the decisions are uncertain. *MDPs* are an extended version of Markov chains with an integrated decision set and state-based reward function (Braziunas, 2003). At each decision epoch, an agent or a *DM* chooses an action based on information gathered from the current system state that is conditionally independent of the history. Therefore, the decision-making procedure conforms to the Markov Property (Memoryless; depends on the history of the process only through the current state). *MDP* is a powerful mathematical modeling tool that can be used to solve a wide variety of real-world issues due to its capability to handle stochastic and dynamic nature. Thanks to the state-based immediate one-stage reward function that allows the agent to learn the environment online, *MDP* is also applicable for reinforcement learning.

4.1.1.MDP Model Formulation

An *MDP* can be characterized by a 4- Tuple process (*S, A, T, R*) with states, actions, transition probabilities, and reward function. The process progresses with all its components at a specified point over time known as the decision epoch. The decision period can be either discrete or continuous in a finite or infinite horizon, depending on the scope of the model (Puterman, 2014). For the sake of clarity, the finite-horizon discrete-time model is explained below:

The elements of *MDP* model for time period t, *where* $t = \{1, 2, ..., N\}$ are:

- a set of states, S: Representatives of the environment/ the system in the model ($s_t \in S$)
- a set of actions, A: All possible actions that a *DM* can choose at state s ($a \in A_s$)
- the transition probability function, $T: S \times A \rightarrow \mathrm{P}(s, a, s') =$

  $\mathrm{P}(s_{t+1} = s' | s_t = s, a_t = a)$: The probability of reaching the next state, s' at time $t + 1$ upon choosing action a in state s at time t. All probabilities are stationary, so it is independent of t.
- the reward function, R: $S \times A \rightarrow R(s, a)$: The expected reward or cost of choosing action a whenever in state s.

The scope of the model is explained as follows: At a certain time t, the *DM*/the agent chooses an action from the action set of the state s, A_s. The declaration of the chosen action generates a reward based on the state-dependent reward function and moves the agent to the next state based on transition probabilities. This decision-making process repeats for each decision point over time. (See *Figure 4.1.*) The model aims to determine the optimal policy structure. ($\pi: S \rightarrow A$) that guides the *DM* to choose the most efficient action in each time period t to minimize/maximize a particular objective function such as "the average reward per unit time", "the total reward" or "the expected total discounted rewards". The optimal policy (π) is an action-selection strategy that leads *DM* to the optimal objective function, called the value function. Optimal policies may be stochastic or deterministic, stationary, or non-stationary according to the model framework.

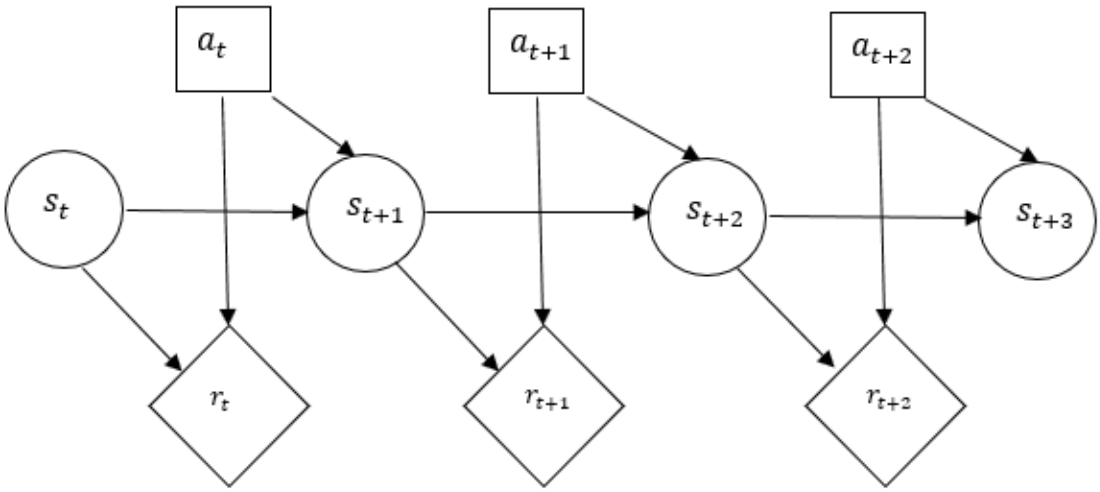

Figure 4. 1. A finite part of an MDP

4.1.2. Value Function

The aim of the model is to find a policy π, which gives us the optimal objective function namely the optimal value. The most frequently used optimization criteria are "the expected total rewards", "the average reward per unit time", or "the expected total discounted rewards". For a maximization problem, the following Bellman equations can be recursively solved to generate value functions. (Bellman and Kalaba, 1965)

$$u_t^*(s) = \max_{a \epsilon A_S} \{R(s,a) + \sum_{s' \epsilon S} \mathrm{P}(s,a,s').u_{t+1}^*(s')\} \text{ for } t \epsilon \{1, 2, ..., N\text{-}1\} \text{ and } s \epsilon S \quad \textbf{(4.1)}$$

$$u_N^*(s) = R_N(s) \text{ for all } s \epsilon S \quad \textbf{(4.2)}$$

The Bellman equations give the optimal utility (value) function, u^* for a particular system state, s at time t. Applying equations for each state along the planning horizon gives an optimal policy map. The optimal action a^* is chosen to maximize the right-hand side of the Equation (4.1) while Equation (4.2) is the boundary condition where $R_N(s)$ represents the expected total reward in state s and in period N. The period N represents the end of the planning horizon, in which the overall future results obtained. Therefore, usually no decisions are made in period N.

The optimal value function that maximizes the expected total discounted reward for a finite horizon *MDP* problem can be obtained by iteratively solving Equation (4.3) (Shani et al., 2013). For calculating the expected total discounted rewards, a predefined discount factor, λ where $0 \leq \lambda < 1$ is applied to value function. When $\lambda = 1$, the optimization criterion turns into the expected total reward.

$$v_t^*(s) = \max_{a \epsilon A_S} \{R(s, a) + \lambda \sum_{s' \epsilon S} P(s, a, s').\, v_{t+1}^*(s')\} \quad \text{for } t \epsilon \{1, 2, \ldots, N\text{-}1\} \text{ and } s \epsilon S \quad \textbf{(4.3)}$$

where $v_t^*(s)$ is the optimal value function of state *s*.

4.1.3. Optimal Policy

The policy $\pi^*(s)$ is the optimal stationary action selection strategy when the system is in state *s*. In order to find the optimal policy π*, Equation (4.5) must be solved recursively by using Equation (4.4).

$$v^*(s) = \max_{a \epsilon A_S} \{R(s, a) + \lambda \sum_{s' \epsilon S} P(s, a, s').\, v^*(s')\} \qquad \text{for } s \epsilon S \qquad \textbf{(4.4)}$$

$$\pi^*(s) = \text{argmax}_{\pi}\, v^*(s) \qquad \text{for } s \epsilon S \qquad \textbf{(4.5)}$$

where $\pi^*(s)$ is the optimal policy and $v^*(s)$ is the optimal value function for state *s*.

4.1.4. Applications of MDP

Because of the Markovian and myopic features, *MDP* models are efficient tools for a wide range of real-life applications. Thanks to the advances in artificial intelligence, which has led to real-time learning and improvement more accurately, the *MDPs* have become suitable for different areas such as medical decision-making (Schaefer et al., 2005), the robot industry (Spaan and Vlassis, 2004). For further reading see Chapter 2.

4.2. Partially Observable Markov Decision Process

The state that reflects the environment we are examining with 100% accuracy is called the actual state. *MDPs* assume that the system state is the actual state. However, in most real-world applications, the environment is affected by a variety of hidden and observable parameters. *POMDPs* assume that the environment is not fully observable as it is in real life. Therefore, it adds a new expression to the model, the observation state, which is an estimation of the actual state. The observation state is based on the collected observations about the system state. In other words, *POMDP* is the general form of *MDPs* with the integrated observation state when the system state is not fully known. The observations state, o_t is a probabilistic function of actual (core) system state, s_t. ($o_t = y. s_t$ where y is the probability factor between 0 and 1and if $y = 1$, *POMDP* is reduced to the classical *MDP* model). Unlike *MDP*, the *POMDP* model also keeps a record on an internal information (belief) state as a representation of the history of the process which is called the belief state. *POMDP* models can be reduced to continuous belief- state *MDPs* by substituting *POMDP's* belief states for *MDP's* system states. The need to keep track of the complete history through the belief state concept makes the process non-Markovian. On the other hand, maintaining a belief state is still Markovian (memoryless) since the next belief state is determined only by the current belief state and current action and observation.

4.2.1. MDP Model Formulation

A *POMDP* is an n- Tuple process (*S, A, Z, T, O, R)* with states, actions, observations, transition probabilities, observation probabilities, and reward function. *S, A, T* and *R* are *MDP* elements as defined in Section 4.1.1, known as the underlying *MDP* of the *POMDP*. The decisions can be taken at discrete points or continuously and in a finite or infinite horizon, depending on the scope of the model (Monahan, 1982). For the sake of clarity, the finite-horizon discrete-time *POMDP* model with a total of *N* periods is explained below:

The elements of *POMDP* model for time period t, *where* $t = \{1, 2, ..., N\}$ are:

- a set of states, S: Represents the actual (core) states of the environment in the model, $(s_t \in S)$
- a set of actions, A: All possible actions that a DM can choose in state s. $(a \in A_s)$
- a set of observations, Z: Set of all possible observations that can be realized during the process. $(o_t \in Z)$
- a transition probability function, $T: S \times A \rightarrow \mathrm{P}(s, a, s') =$

  $\mathrm{P}(s_{t+1} = s' | s_t = s, a_t = a)$: The probability of reaching the next state s' at time $t+1$ by choosing action a when in state s at time t.
- an observation function, $Z: S \times A \rightarrow Z(s', a, o) = \mathrm{P}\ (o_{t+1} = o \mid a_t = a, s_{t+1} = s')$: The probability of observing o, given the chosen action a, and reaching state s'.
- a reward function, R: $S \times A \rightarrow R(s, a)$: The expected reward/cost of choosing action a in state s.

4.2.2. Value Function and Optimal Policy

After adding the observation state and observation probabilities to the *MDP* value function, the discounted *POMDP* value function takes the form in Equation (4.6).

$$v_t^*(s) = \max_{a \epsilon A_S} \{R(s, a) + \lambda \sum_{s' \epsilon S} \mathrm{P}(s, a, s') . \sum_{s' \epsilon S} \mathrm{Z}(s', a, o) . v_{t+1}^*(s')\} \quad \textbf{(4.6)}$$

for $t \in \{1, 2, ..., N\text{-}1\}$ and $s \in S$

The policy $\pi^*(s)$ is the optimal stationary action selection strategy when the system is in state s. In order to find the optimal policy π^*, Equation (4.7) and Equation (4.8) must be solved recursively.

$$v^*(s) = \max_{a \epsilon A_S} \{R(s, a) + \lambda \sum_{s' \epsilon S} \mathrm{P}(s, a, s') . \sum_{s' \epsilon S} \mathrm{Z}(s', a, o) . v^*(s')\} \text{ for } s \in S \quad \textbf{(4.7)}$$

$$\pi^*(s) = \mathrm{argmax}_{\pi}\, v^*(s) \qquad \text{for } s \in S \quad \textbf{(4.8)}$$

where $\pi^*(s)$ is the optimal policy and $v^*(s)$ is the optimal value function for state s.

The scope of the model is explained as follows: At specified time period *t*, the *DM*/the agent chooses an action from the action set of the actual(core) state *s*. With the declaration of the chosen action the system generates a reward, and the agent moves to the next state according to the transition probabilities and the *DM* acquires an observation associated with the actual state as an output of previous action. Since the actual state is not fully observable, the *DM* receives an estimation of the actual state based on the observation received and the previous action, which is the belief state, and the belief state is updated. This decision-making process repeats for each decision point over time (See *Figure 4.2*). The purpose of the model is to find an optimal policy structure ($\pi: S \rightarrow A$) that maps the states to the actions. Optimal policies may be stochastic or deterministic, stationary or non-stationary according to the model framework. The decision network of *POMDP* is shown in *Figure 4.2.*

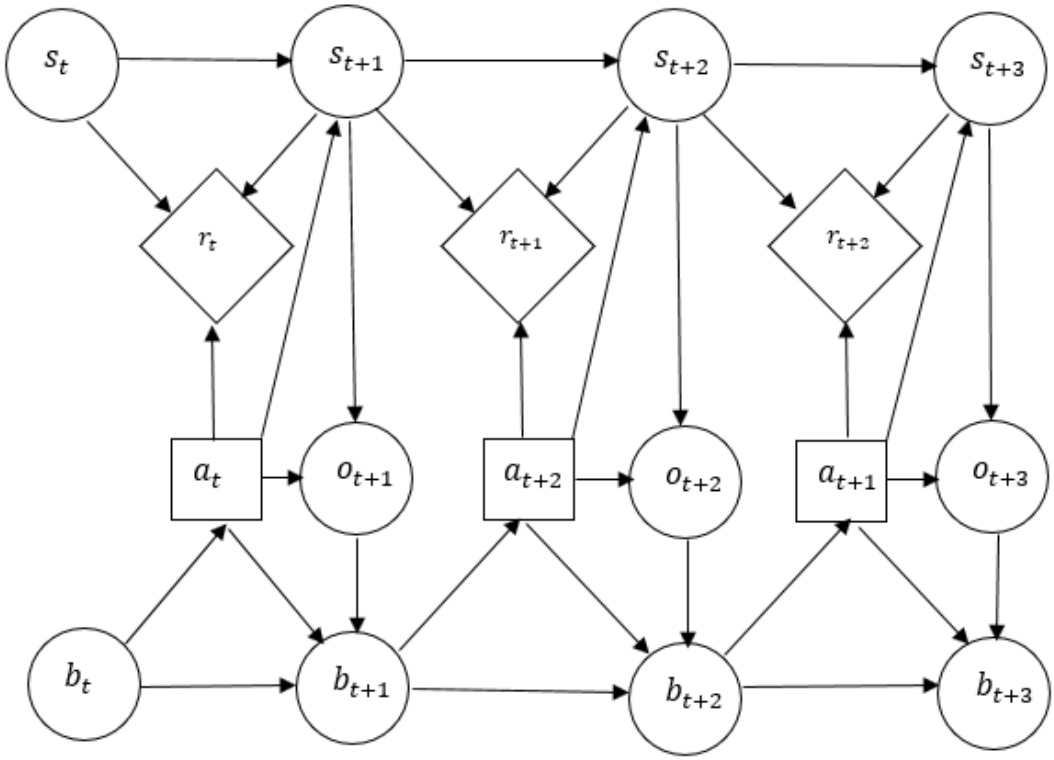

Figure 4. 2. Decision network of a finite part of a POMDP

For a more detailed background on the concept of *POMDP* theory, one can refer to Lovejoy (1991) and Monahan (1982)

4.2.3. Representation Methods of POMDP

Model representation is an important step in solving *POMDP* efficiently. Two methods are commonly used in the literature. The first is the continuous *belief state MDP* approach, which is based on modelling *POMDPs* as *MDPs.* Although this approach transforms a finite-state *POMDP* model into a continuous state *MDP* model, it still reduces the computation time. The second is the policy tree approach, where the *POMDP* model is represented by policy trees. It also eases the calculation, and the policy tree representation is generally used in policy iteration and heuristic search algorithms.

4.2.3.1. Belief-MDP

To understand the *belief state MDP*, we must first understand the concept of belief status. Unlike *MDP,* a *POMDP* model also keeps a record on an internal belief, b (information) state as a representation of the history of the process.

- a belief state, *b:* b = P ($s \mid h$) After witnessing history *h,* the probability of being in state s

where h is: $h_t = (a_0, o_1, a_1, o_2, a_2, o_3, \ldots\ldots, a_{t-1}, o_t)$

The belief state $b \in B$ is a probability function of the actual state. For $s \in S$, $b(s) \in [0, 1]$, and $\sum_{s \in S} b(s) = 1$. The belief state can be updated online according to the Bayes rule (state estimator). At any time period t, the agent evaluates the belief state, b_t using a state estimator based on the prior belief state b_{t-1}, the last action a_{t-1}, and the current observation o_t as presented in Equation (4.9).

$$b(s') = \mathrm{P}(s' \mid o, a, b) = \frac{\mathrm{Z}(s', a, o) . \sum_{s \prime \epsilon S} \mathrm{P}(s, a, s') . b(s)}{\sum_{s \prime \epsilon S} \mathrm{Z}(s', a, o) \sum_{s \prime \epsilon S} \mathrm{P}(s, a, s') . b(s)} \quad \text{for } s \in S \qquad \textbf{(4.9)}$$

The initial belief state b_0 is given by the *DM* or chosen at random. Since the belief state is the probability distribution over the system states, the belief space *B(s)* is also a probability space. If the probability of being state $s_1 = p$, then probability of being state s_2 is 1 - p to satisfy $\sum_{s \in S} b(s) = 1$.

Therefore, the belief space of two state *POMDP* is a line segment as presented in *Figure 4.3.* As the number of states increases, the belief space becomes hyperplanes. To keep it simple, the rest of this study uses one-dimensional belief space for representations.

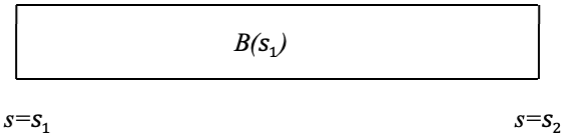

Figure 4. 3. Belief space representation for a two state POMDP

POMDP models can be written as *MDP* models whose system states are belief states. The optimal solution of the *POMDP* is the optimal solution of the belief *MDP*. This *belief MDP* is a 4- Tuple processes (*B, A, T, R*) where:

- infinite set of belief states, *B*: ($b \in B$)
- a set of actions, *A*: ($a \in A_s$)
- transition function, *T*: $\mathrm{P}(b, a, b') = \sum_{o \in Z} P(b'|b, a, o).P(o|a, b)$ *where* $P(b'|b, a, o)$ = 1 if state estimator (b, a, o) = b', $P(b'|b, a, o)$ = 0 otherwise.
- the reward function, *R*: $r(b, a) = \sum_{s \in S} b(s)R(s, a)$

The belief state is a probability distribution. Therefore, the *belief MDP* is in continuous space even if it is the discrete state space of the respective *POMDP.* For this version of *POMDP,* policy π maps belief states to actions (π: $B \rightarrow A$). For *belief-MDP*, the value function is a set of $|S|$ dimensional vectors, called *α-vectors* (See Equation 4.10). Each *α-vector* is a value that created by selecting a particular action a, ($\alpha \in V$). The corresponding value computed by Equation 4.11. Since the reward function $r(b, a)$ is linear. "For the finite-horizon *POMDPs*, the optimal value function is piecewise-linear and convex" (Sondik, 1971). The purpose is finding the optimal policy π^*that gives the maximum expected total future value. The optimal policy π^* is found by recursively solving Equations (11) and (12).

$$v(b) = \alpha \cdot b = \sum_{s \in S} \alpha(s) b(s) \quad \textbf{(4.10)}$$

$$v^*(b) = \max_{a \in A_S} \{r(b,a) + \lambda \sum_{o \in Z} \mathrm{P}(b,a,b').v^*(b')\} \quad \text{for } o \in Z \quad \textbf{(4.11)}$$

$$\pi^* = \mathrm{argmax}_\pi v^*(b_0) \quad \textit{where } b_0 \text{ is the initial belief} \quad \textbf{(4.12)}$$

To better illustrate the concept of the *belief MDP*, we give a belief space demonstration for a case involving two actions and two observations (actions as a_1, a_2 and observations as o_1, o_2). In *Figure 4.4.* the red point represents the current belief state, and the black points represent the next possible belief states based on actions, and observations. For a given action, the sum of the probability of subsequent belief states is always 1.

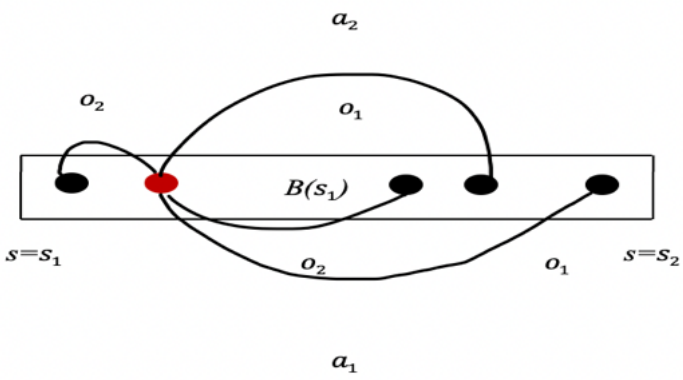

Figure 4. 4. Discretized belief space demonstration

In belief space, transition from the current belief state to the next belief state satisfies the Markovian property indicating it is independent from the history (only depends on current state). Since the space is continues the value function is an arbitrary function over belief space. In *Figure 4.5.* the value function over belief space *B(s)* is shown.

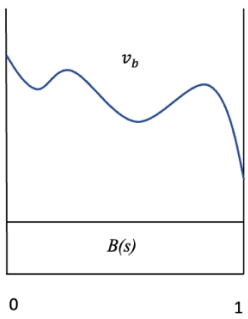

Figure 4. 5. One dimensional belief state - space demonstration

4.2.3.2. Policy Trees

A finite horizon *POMDP* can be expressed by a finite number of policy trees, as shown in *Figure 4.6.* In a t-step(horizon) *POMDP*, nodes represent the current actions and branches represent the observations. For the t-step policy tree with one step remaining, *DM* chooses an action, with two steps remaining *DM* chooses an action, collects an observation, and chooses another action. Using this logic, the total number of observations is given by Equation (4.13), *where* T equals the last step and O is number of observations.

$$\sum_{t=0}^{t=T-1} |O|^t = \frac{|O|^T - 1}{|O| - 1} \quad \textbf{(4.13)}$$

A policy tree is constructed for each action in the action set, where |A| is the number of actions the number of all possible trees is given by Equation 4.14.

$$|A|^{\frac{|O|^T - 1}{|O| - 1}} \quad \textbf{(4.14)}$$

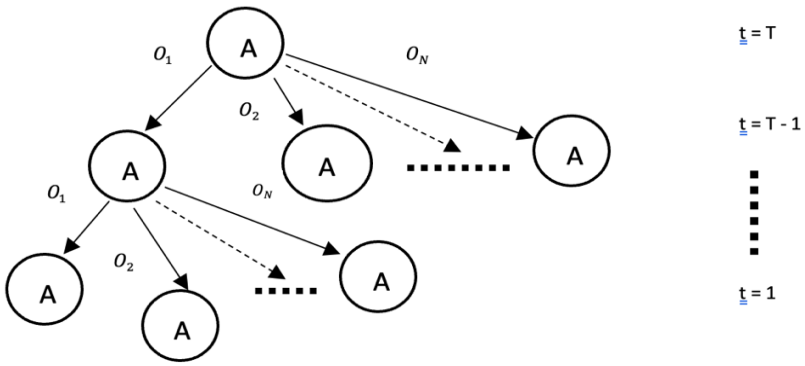

Figure 4. 6. Representation of a t-step policy tree for an action.

As p_i represents a policy tree and P is the set of all policy trees $P = \{p_1,\ p_2,\ p_3, .., p_n\}$, $v_p(s)$ is the expected future value in state *s, where* $v_p(s)$ is calculated by Equation (4.7) for each $p \epsilon\ P$. As mentioned in previous section, *POMDP* can be represented as a *belief*

MDP, where $v_p(b)$ is the expected future value starting from the belief *b.* Therefore, we can write the value functions of policy tree p for each state with *α-vectors*, *where* $v_p(s)$ is value function of policy tree p for state *s*.

$$\alpha_p = < v_p(s_1), v_p(s_2), v_p(s_3), \dots, v_p(s_N) > \qquad \text{for } p \in P, s \in S \qquad \textbf{(4.15)}$$

$v_p(s)$ is multiplied by the belief state to convert it to the belief version. As given in Equation (4.16), the modified value function $v_p(b)$ is:

$$v_p(b) = \sum_{s \in S} b(s) v_p(s) \qquad \text{for } p \in P \qquad \textbf{(4.16)}$$

Where $v_p(b)$ is value function of policy tree p for belief *b*. Finally, the optimal value function of a t-step policy tree from the belief *b* is given in Equation (4.17).

$$v_t^*(b) = \max_{p \in P} b.\ \alpha_p \qquad \textbf{(4.17)}$$

As seen in Equation (4.16), $v_p(b)$ is linear for each policy $p \in P$. Therefore, $v_p^*(b)$ is piecewise linear and convex by the Equation (4.16).

4.3. Solution Algorithms

In an *MDP*, the model includes uncertainties arising from selected actions, e.g., immediate rewards and/or subsequent states can be random (myopic feature). In a *POMDP*, the model contains some other uncertainties in addition to those that already exist in the *MDP* model. These additional uncertainties are information about the initial state and the actual(core) state in the partially observable environment. Therefore, the solution of *POMDP* models is much more difficult and more complex than that of *MDPs*. For the continuous system/belief state model, the value function grows exponentially which makes the exact solution of *POMDP* is *PSPACE - Complete*. It means in input size the problem is in *polynomial space (PSPACE)* and every other problem in *PSPACE* can be transformed to it in polynomial time (Papadimitriou and Tsitsiklis, 1987). For ease of

computation, discretization of the belief space is usually an efficient preparation method that can result in near-optimal solutions with a good performance. As we mentioned in Section 4.2.3.1, *POMDP* can be represented as a continuous space *belief MDP*, and the finite horizon value function is "piecewise linear and convex" (Sondik, 1971). Therefore, we can discretize belief space with finite number of α - *vectors*. In *Figure 4.7.* discretized one-dimensional belief space is presented. In figure, the corresponding value functions of each action a_1, a_2, a_3 are shown by the vectors v_1, v_2, v_3 referred as α - *vectors*.

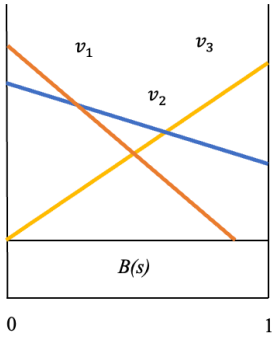

Figure 4. 7. One dimensional belief state - space demonstration

4.3.1. Exact Solution Algorithms

The computation time is quite high and grows exponentially as new states are added for finding the exact solution of *POMDP* (*PSPACE-Complete*). However, *belief MDP* notation provides *MDP* solution algorithms that can also be applicable to *POMDP*. These are policy iteration and value iteration. The logic of these two algorithms is explained in this section and some of the other most used exact solution algorithms which used to reduce complexity, mentioned at the end of this section. For more detailed information, you can refer to Shani et al. (2013), Pineau et al. (2006).

4.3.1.1. Policy Iteration

Policy iteration algorithms iteratively increments the policies until they find the optimal or near-optimal policy. For a maximization problem, the policy π' is superior of policy π

if and only if the expected overall value of π' is greater than the expected overall value of π for all possible states:

$$\pi' > \pi \Leftrightarrow \forall s, v^{\pi'}(s) > v^{\pi}(s) \qquad \text{for} \qquad s \in S \tag{4.18}$$

where, $v^{\pi}(s)$ is the value function of state s corresponding to the policy π.

In the policy iteration algorithm, a finite state controller t, is defined to keep track of the iterations and an initial policy is determined on a random or rule basis, called "*the policy evaluation step*". Starting from this initial policy, the method used to determine an optimal or near-optimal policy is applied iteratively, called "*the policy improvement step*". The policy iteration algorithm can be summarized as follows:

Step 1. Set $t = 0$, choose an initial policy π_0

Step 2. "*Policy Evaluation*": Compute the value function of π_t, *where* π_t is the policy of the step t

Step 3. "*Policy Improvement*": Search a better policy which will provide higher value function

Step 4. If a better policy does not exist stop, *otherwise*, increment t by 1, and return to step 2.

Restriction by discretization or by selecting restricted region of the belief space can lead to near-optimal solutions. For the infinite horizon model, the decision epoch, t is infinite. As $t \rightarrow \infty$ policies $\pi_0, \pi_1, \pi_2, \ldots, \pi_t$ converges to the optimal policy π^*. In the literature, (Sondik, 1971) proposed a policy iteration algorithm for finite horizon *POMDP*. Since the reward function of *belief MDP*, r (*b, a*) is linear, due to (Sondik, 1971), the optimal t-step value function of belief *b,* $v_t^*(b)$ is a "piecewise linear and convex" function. Therefore, the value function is representable with lines, planes, and hyper-planes (α *vectors*) according to the number of states (Sondik, 1971). The optimal value function is found by selecting the upper surface of the *α-vectors*. In *Figure 4.8.,* every line represents a value function for a specific action, the red line presents the optimal value surface.

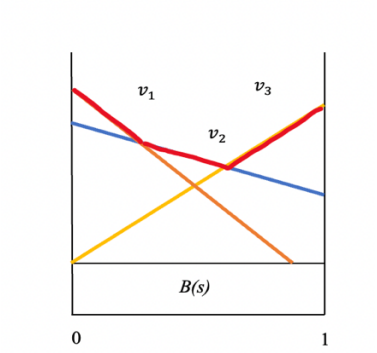

Figure 4. 8. 2- dimensional belief state - space demonstration

In the *Figure 4.9.,* the optimal policy is presented. The rectangular filled areas over the horizontal axis are called regions and they represent the optimal actions that gives higher value for each belief.

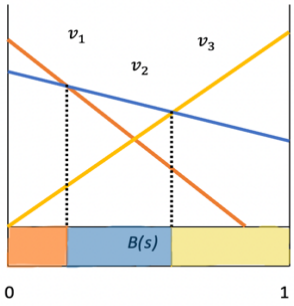

Figure 4. 9. The optimal policy representation with region

4.3.1.2. Value Iteration

In the value iteration method, finite horizon optimal value functions $v_0^*, v_1^*, v_2^*, \dots, v_t^*$ are computed to obtain the infinite horizon optimal policy π^*. As in the policy iteration method, a finite state controller t is defined to keep track of the iterations. As $t \rightarrow \infty$ the gap between the optimal value function and the optimal t-horizon value function converges to 0 (Howard, 1960) as presented in Equation (4.19).

$$\lim_{t \rightarrow \infty} \max_{s \in S} |v^*(s) - v_t^*(s)| = 0 \qquad \textbf{(4.19)}$$

To get the optimal value function, the Bellman error ε, the maximum difference between two consecutive finite horizon value functions, is used. For t = 0, $v_0(s) = 0$ for $s \in S$ by definition. The optimal value functions calculated iteratively by Equation (4.20) while $\max_{s \in S} |v_{t+1}(s) - v_t(s)| > \varepsilon$. As explained Section 4.2.3.1, we can rewrite *POMDP* as *belief MDP*. The value iteration algorithm for a *belief MDP* can be summarized as follows:

Step 1. Set $t = 0$ and $v_0(b) = 0$ for all $b \in B$.

Step 2. If $\max_{b \in B} |v_{t+1}(b) - v_t(b)| > \varepsilon$, calculate $v_{t+1}(b)$ for all $b \in B$

$$v_t^*(s) = \max_{\alpha} [R(s,a) + \sum_{s'} P(s,a,s') v_{t+1}^*(s')] \quad \textbf{(4.20)}$$

$$v_t^*(b) = \max_{\alpha} [\sum_s b(s) R(s,a) + \sum_{b'} P(b'|b,a) v_{t+1}^*(b')] \quad \textbf{(4.21)}$$

Where $v_t^*(s)$ and $v_t^*(b)$ are calculated by Equation (4.20) and Equation (4.21) respectively. Since the optimal finite-horizon value function is "piecewise linear, and convex", the value function is representable with *α-vectors* (Sondik, 1971). For a more detailed background about value iteration algorithms, you can refer to Monahan (1982), Bellman (1966), and Howard (1960)

4.3.1.3. Other Exact Solution Algorithms

In this section, we summarize some other commonly used exact solution algorithms in the literature. These algorithms can give good results, especially for small-size problems.

The first method is Monahan Enumeration (Monahan, 1982), which computes all vectors in the belief space, including redundant ones. The algorithm uses the following domination condition: For a vector not to be redundant, there must be at least one belief point over the belief space with a greater value than the others. The algorithm works as follows: Create a list that contains every vector in the belief space and mark all the vectors, then select a marked vector from the list. If the selected vector satisfies the domination condition, unmark this vector and keep the vector in the list, otherwise remove the vector from the list. Do this until all vectors in the list are unmarked.

The second is Sondik's One Pass Algorithm (Sondik, 1971), which starts with the random selection of a set of belief points over the belief space. The algorithm works as follows: First, an empty target list and a search list containing all selected beliefs are defined, and a point is selected from the search list. Then, the optimal value vector and corresponding action for this point are found and the selected vector is added to the target list. Next, a region around the selected point is defined where the selected vector is dominant. Finally, neighboring points at the corners of the selected region are selected and they are added to the search list. This procedure is continued until the call list is empty.

There are some other commonly used exact solution algorithms in the literature. One is the Witness Algorithm which is an action-based forward policy tree search algorithm (Littman, 1994).The other one is the Incremental Pruning algorithm which is also an action-based algorithm, and it considers one observation at a time. The algorithm allows pruning redundant vectors (Zhang and Liu, 1996)

4.3.2. Approximate Solution Algorithms

In this section, we give some approximate solution algorithms used to reduce the computation time. These algorithms can provide sub-optimal or near-optimal solutions. Firstly, we introduce a basic approximation method in the literature, namely "*Point-Based Value Iteration (PBVI)".* We also mention some other algorithms at the end of this section.

4.3.2.1. Point Based Value Iteration

The *PBVI* is a value iteration algorithm that only considers a small set of belief points and applies value iteration to these points (Pineau et al., 2006). The *PBVI* algorithm is an anytime algorithm that combines value iteration steps and belief set expansion steps. It starts with an initial set of belief points to which it applies the first set of backups. In *PBVI*, only one (best) vector for each belief point is maintained. It then enlarges the set of belief points by choosing reachable belief points and finds a new solution for the expanded set. To do this, it uses each action to produce new beliefs. Then, calculate the L1 distance from all belief points. The new point is then furthest from every point in the

set. It generates one belief from each belief. *PBVI* provides a set of solutions that interpolate value backup iterations with extensions of the belief set and gradually reduces computation time and solution quality. *Figure 4.10.* shows a *PBVI* demonstration.

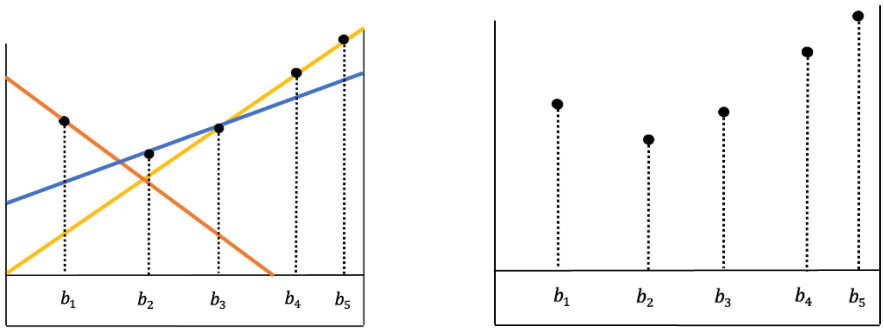

Figure 4. 10. PBVI demonstration

4.3.2.2. Other Approximate Solution Algorithms

Reinforcement learning is one of the most recent methods proposed to solve *POMDP* approximately. The goal of reinforcement learning is for an agent to operate in the world in order to maximize rewards. The environment is represented as a "stochastic finite state machine" with inputs as "agent actions" and outputs as "observations and rewards the agent gets". The agent's goal is finding a policy and a state-update function that will calculate the total of the maximum expected future rewards. (Doshi et al., 2008)

There are also some greedy approaches to solve *POMDP* such as the Monte Carlo Method, useful for continuous state space, help to find an approximate belief space. (Thrun, 1999)

4.4. The Quality Adjusted Life Years

The Quality Adjusted Life Years (QALY) is a type of cost utility analysis metric (for more information, please refer to Scuffham et al. (2008) which allows us to measure the morbidity as well as the mortality. To put it more clearly, it allows for the calculation of survival and quality of life at the same time. Mathematically:

$$QALY = Time \times Utility$$

Here, the "*Time*" represents the length of time that a patient is expected to spend in a particular health condition, and the "*Utility*" represents the well-being of the patient also known as "*Health Related Quality of Life (HR-QOL)"*. For *HR-QOL*,1 represents a perfect health condition and 0 means death. Values between 0 and 1 represent a person's quality of life. For instance, if the expected duration in a specified health state is 5 years and utility is 0.8, then the expected quality adjusted life year for this specified health state is 5 x 0.8 = 4. The perception of quality of life may differ geographically, genetically, and sociologically. Moreover, even if all these are the same, they may differ depending on age and gender. In the literature, there are direct and indirect elicitation methods for the computation of *QALY* utilities. "*Time Trade-off"*, "*Standard Gamble*", and "*Visual Analogue Scale*" are the most widely used direct elicitation methods which are explained below:

Time Trade-off: In this method (Gudex, 1994), the patient is asked to choose between two scenarios which are: (i) impaired health state for t years and (ii) perfect health state for x years, *where* x < t. The patient is asked the same question for varying x values until reaching the indifference point. At this point *HR-QOL* utility for health state i (h_i) is:

$$h_i = \frac{x}{t} \qquad \textbf{(4.20)}$$

Standard Gamble: In this method (Gafni, 1994) the patient is asked to gamble between certain remaining life years and either perfect health (with probability p) or death (with probability q), *where* p + q = 1. The patient is asked to gamble for varying q probabilities until reaching the indifference point. At this point, the *HR -QOL* utility is determined as the p value.

Visual Analogue Scale: In this method (Crichton and Clin Nurs, 2001) the patient is asked directly to give a utility weight for specified health state in the given scale. For example: For 0 is the worst and 10 is the best, patient assign a number for health state.

There are also indirect methods known as generic preference- based measures to elicit utility weights. Which includes standardized generic utility questionnaires. For

preference- based measures time trade off or standard gamble methods used to calculate utilities. The most frequently used generic utility questionnaire is "*EQ-5D*" (Shaw et al., 2005). The "*EQ-5D questionnaire*" is consisted of five dimension which are: "*mobility*", "*self-care*", "*usual activities*", "*pain/discomfort*" and "*anxiety/depression*". Each dimension includes three statement that patients asked to choose one of them. Some of the other generic utility questionnaires are: "*Health Utility Index*" (Feeny et al., 2002) and "*QLQ- C30"* (Fayers and Bottomly, 2002) for cancer specific utilities. The target group for *HR-QOL* questionnaires can be general population, healthcare professionals, patients, or patient's relatives.

4.5. Applications of POMDP

The real-world environment, which requires myopic planning in uncertain and partially observable situations, can be accurately modeled by *POMDPs.* Therefore, in the literature, *POMDP* studies progress in a wide spectrum due to its suitability in many different areas. Robot planning (Spaan and Vlassis, 2004), spoken dialogue systems (Williams et al., 2006), reinforcement learning (Doshi et al., 2008), artificial intelligence (Bennett and Hauser, 2013) healthcare (Schaefer et al., 2005) are some of the common and frequent application areas in the literature. For more detailed information on the application areas see Cassandra (1998)

5.METHODOLOGY

In the literature, although there are some studies for *POMDP* modeling of colorectal cancer screening and surveillance (Zhu, 2010), (Erenay et al., 2014), (Leshno et al., 2003) as far as we know there is no study that specifically addresses the colorectal cancer treatment model with a *POMDP* modeling approach. In addition, there is a recent study modeling the *CRC* treatment process with semi-Markov models by (Joranger et al., 2020). On the other hand, there are a few studies related to the treatment of some other types of cancer, for example, the prostate cancer treatment model, (Goulionis and Koutsiumaris, 2010)

In this study, the colorectal cancer treatment procedure is examined within the *POMDP* modeling framework. In medicine, physicians make medical decisions about patients' health based on their theoretical knowledge and mostly on their professional experience. The theoretical knowledge usually consists of current guidelines and textbooks which are universal. However, environmental factors such as geography, genetic background, social and economic structure affect the course and progress of the disease and hence, the available and useful treatment options. Physicians learn about the effects of these environmental factors in the local background from the first day they start their profession. The level of professional experience that includes the perception of hidden environmental factors affects the physician's heuristic decisions. Therefore, based on the professional experience and knowledge, the decisions may vary among the doctors. In this chapter, we present the methodology of our *POMDP* model, and used solution algorithm. Since we have no actual data to evaluate the model properly, we tried to present an efficient scenario analysis to examine the effect of the doctor's perceptions about partially observable states, which we will explain in the next chapter.

5.1. Model Formulation

The characteristics of the *POMDP* model have been constructed using the current guidelines, studies from the literature, and expert opinions of the oncologists having experience in *CRC.* We used 6-month follow-up periods as decision epochs and run the model by taking $t = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ to calculate 5-year *QALYs* and *LYs.* We did not have access to real patient data to establish the model. For this reason, we obtained the transition and observation probabilities with the help of experts. We used the transition probabilities from experts to see the perception of clinical expertise. Finally, we got help from the evaluators check to see if the output of our model was proper.

5.1.1. System State (Actual / Core State)

In medicine, patients' actual health status contains uncertainties from many points of view. Firstly, as mentioned earlier, the experts' perception and knowledge are one of the important dimensions of uncertainty. Moreover, measurement tests used for decisions about patients' health status have a certain level of sensitivity and accuracy. Still with recent medical knowledge, there are unknown or unrevealed factors and/or indicators about a human's health. Therefore, understanding of a patient's actual health status may differ from time to time as well as from doctor to doctor. Quantitative models are efficient tools to model such kind of uncertainty. Thanks to the uncertainty to be included, *POMDP* models showed significant improvement in healthcare modeling recently. In this thesis, we define a patient's health state s_t with cancer stages (explained below), recovery, and death situations. $s_t \in S$ *where* $S \in \{$*Cancer Stage*, *Recovery*, *Death*$\}$

The first element of the system state is the cancer stage as represented by the 5th edition of *AJCC.* As explained in Chapter 3, in *AJCC 5th* edition, cancer is examined in 5 stages which are defined with TNM classification. Stage 0 means carcinoma in situ (there is no current cancer but may occur in the future). Stage I means the cancer is localized. Stage II means cancer has progressed locally but in an early stage. Stage III means cancer progressed locally but in late stages. Stage IV means that distant metastases occurred. In this thesis, we used a 4-staged demonstration (See Assumption 2). Therefore, *Cancer Stage* $\in \{$*Stage I*, *Stage II*, *Stage III*, *Stage IV*$\}$

The second element is the state of recovery. *Recovery* state is an absorbing state. If the patient in the system has responded well to the treatment and completely recovered from cancer the treatment process is terminated.

The last element is the death situation. *Death* state is also an absorbing state. If the patient dies during the treatment, the treatment process is obviously terminated. When all elements are included, the final version of the system state is $s \in S$: {*Stage I, Stage II, Stage III, Stage IV, Recovery, Death*}

5.1.2. Observation State

In situations where the environment is not fully observable, the *DM* makes decisions on actions according to the previous observations and the past actions. For most healthcare applications, the observation states of the *POMDP* models are defined as the symptoms as specified by the patients and several laboratory results. Since the symptoms are indicators of the patient's health state. After doctors decide on the treatment option for a patient, they monitor some of the specified symptoms and laboratory findings and the results. These symptoms could be "headache", "lassitude", "nausea", "vomiting", "sleep disorder", etc., and the laboratory findings identified as disease specific. In this thesis, the observations of the model are defined as the change in the blood measurement result of *Carcinoembryonic Antigen (CEA)* and *CT* scan evaluated during the treatment and surveillance procedure. Which are commonly used to understand the patient's health status.

The first element of the observation state is the *CEA,* a general cancer marker for all cancers and is the most used blood measure for the surveillance of *CRC.* Higher *CEA* means a higher tumor burden. However, the *CEA* value itself does not make sense when evaluated alone. Physicians check the change in the level of *CEA* to understand the condition of the disease. If the *CEA* level rises by more than 10 units, it means that the patient's condition has worsened and the disease has progressed, otherwise this means that the patient may be in stable or in a better condition. In this thesis, we represent *CEA*

with two values. We use $CEA+$ to represent increments greater than 10 units and $CEA-$ to represent insignificant changes or a possible decrease. $CEA \in \{CEA+, CEA-\}$

The second element of the observation state is the *CT* scan, which is a frequently used technique to study the body composition of patients. The physician evaluates the *CT* result in two ways. Therefore, we used two elements to represent the *CT.* We use $CT+$ if the patient's condition worsens (positive findings) and use $CT-$ if the patient's condition does not change significantly (negative findings). $CT \in \{CT+, CT-\}$

Lastly, the final version of the space of the observation states is found by the Cartesian product of the two subsets. Hence, $o_t \in Z$: $\{(CEA+, CT+), (CEA-, CT+), (CEA+, CT-), (CEA-, CT-)\}$

After a selected treatment option is applied at a certain decision point and the patient reaches the next decision epoch, physicians receive an observation as output and estimate the patient's health status based on this observation. If the observation status is o_t is $(CEA+, CT+)$ the cancer stage of the patient is estimated to progress, and treatment policies are adjusted accordingly. For all other observations, the patient is considered to be in the regular state and the follow-up process continues.

5.1.3. Available Actions

In colorectal cancer, the main purpose of the treatment procedure is to perform a surgical operation. However, depending on the clinical condition of the patient and the stage of the disease, it is not always possible to implement this option. If a patient has some comorbidities such as "diabetes", "heart disease", "chronic obstructive pulmonary disease" or another active type of cancer, the surgical operation can be taken out from the available options and oncologists continue with other medical options. In case the patient has late-stage distant metastases, palliative care is applied to relieve the patient's pain. For the rest, the goal of the treatment is to make the patient operable. In the early stage of the disease, oncologists prescribe combination therapy like adjuvant chemotherapy or chemoradiotherapy to control tumor spread. In this thesis, we defined the available

actions that oncologists can choose for the treatment of colorectal cancer as $a \in A_s$: {*Surgery, Chemotherapy, Adjuvant Chemotherapy, Chemoradiotherapy*} for $s \in S$. Available actions are summarized below (See Chapter 3 for more detailed information)

***Surgery*:** Surgery is the main treatment option for *CRC* e.g., it can be performed as colectomy (removal of all colon) and block resection of lymph nodes.

***Chemotherapy*:** Chemotherapy is used to reduce the spread of cancer by disrupting tumor metabolism. In Stage IV, chemotherapy may be used as neoadjuvant therapy to restore the respectability of initially unresectable metastatic disease or as palliative care.

***Adjuvant Chemotherapy*:** Adjuvant therapy is a combination therapy in which pre-and / or postoperative chemotherapy can be applied to prepare the patient for surgery or to support the postoperative recovery process. Particularly for Stage II and Stage III, adjuvant chemotherapy is a treatment option that can help increase the likelihood of survival.

***Chemoradiotherapy*:** Chemoradiotherapy is another combination therapy that combines chemotherapy and radiotherapy. Chemoradiotherapy is used to slow the spread of the disease in the treatment of *CRC.*

Surgery is the only recommended treatment for *Stage I* in the treatment of *CRC* (National Comprehensive Cancer Network, 2022). Therefore, we defined surgery as the only available action for *Stage 1* when constructing the model. For Stage II, III and IV all actions are available.

5.1.4. Transition Probabilities

As explained in the previous section, the transition probability P (s, a, s') is P $(s_{t+1} = s' | s_t = s, a_t = a)$. *Recovery* and *Death* states are absorbing states (there are no exit transitions from these two states). Therefore, a total of 60 possible transitions are defined

and transition probabilities for all action and states have been calculated based on expert knowledge and presented in *Figure 5.1.*

In determining the probabilities, we get help from a total of five physicians. We used the simple mean method to calculate the final probabilities. We also grouped the experts as oncologists and non-oncologists to understand the impact of clinical experience in the scenario analysis (see Chapter 6.5.3) and calculated the probabilities separately. The oncologist group consists of two oncologists having10-15 years of experience, working in two different geographical areas who face *CRC* patients every day. The non-oncologist group consists of two physicians having 3-5 years of experience who encounter *CRC* patients through consultations (One in internal medicine and one in emergency medicine). Their knowledge is mostly from textbooks, medical education, and clinical from active *CRC* patients who need additional treatment for additional issues. For example, *CRC* patients with diabetes, high blood pressure, cardiac arrest, etc. Finally, we get help from an additional oncologist to check the model outputs to see if the model adequately represented the *CRC* environment in Turkey. The transition probabilities were obtained from each of the four physicians through an online survey and hence the expert opinions were taken independently and separately.

The transition probabilities presented in *Figure 5.1.* were computed by taking the simple average of 4 transitions probabilities obtained. In Chapter 6, we will also consider the transition probabilities computed only from the oncologist and from the non-oncologist group.

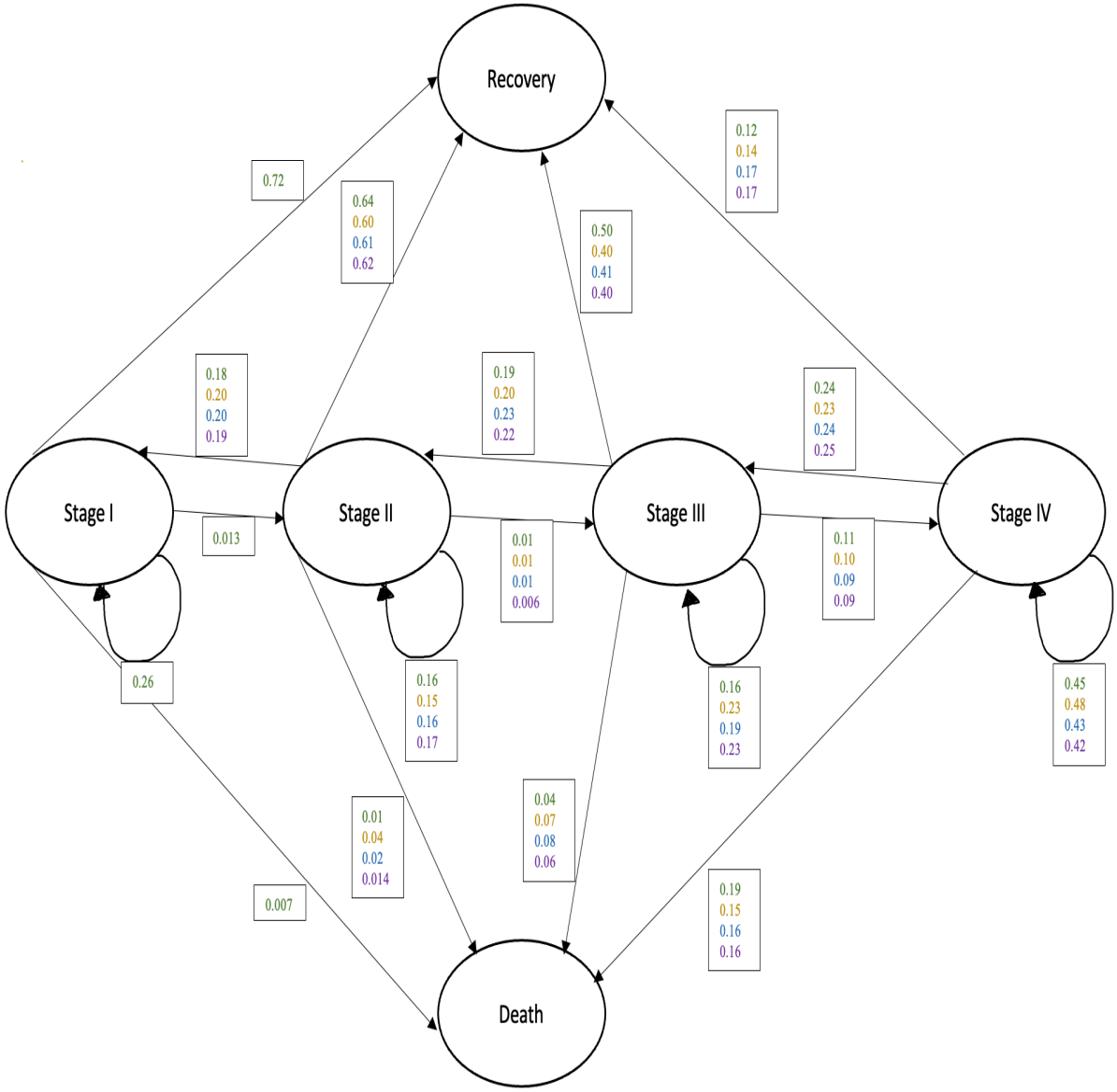

Figure 5. 1. Transition probabilities of POMDP model as a function of actions

(Actions: Surgery, Chemotherapy, Chemoradiotherapy, Adjuvant Chemotherapy)

5.1.5. Observation Probabilities

As explained in previous section, the observation probability. $Z(s', a, o) = \mathrm{P}\,(o_{t+1} = o \mid a_t = a, s_{t+1} = s')$ and depends on action a and the next state s'. The physicians will observe the change in the *CEA* levels and *CT* findings as observation about patient actual health status. Similar to the transition probabilities, the observation probabilities in the model are also defined with the help of expert knowledge and shown in the *Table 5.1.* For *Stage I,* where surgery is the only available action, the observation probabilities for other

treatment options are determined uniformly. In addition, for *Recovery* and *Death* states the observation probabilities are determined uniformly

Table 5. 1. Stage and treatment dependent quality of life utilities from literature

	(*CEA-, CT-*)	**(*CEA-, CT+*)**	**(*CEA+, CT-*)**	***(CEA+, CT+*)**
Surgery				
Stage I	0.94	0.06	0	0
Stage II	0.88	0.06	0.06	0
Stage III	0.75	0.06	0.13	0.06
Stage IV	0.32	0.31	0.06	0.31
Chemotherapy				
Stage II	0.88	0.06	0.06	0
Stage III	0.811	0.063	0.063	0.063
Stage IV	0.13	0.44	0.19	0.24
Adjuvant Chemotherapy				
Stage II	0.88	0.06	0.06	0
Stage III	0.811	0.063	0.063	0.063
Stage IV	0.37	0.33	0.14	0.16
Chemoradiotherapy				
Stage II	0.88	0.06	0.06	0
Stage III	0.811	0.063	0.063	0.063
Stage IV	0.38	0.26	0.20	0.16

5.1.6. Reward Function

In this thesis, we use two types of reward functions. One of them is "*The Quality-Adjusted Life Years (QALYs)",* and the other one is "*The Life-Years (LYs)".* We assume that there is no cost to perform the actions from the patients' point of view for computational simplicity.

For every decision epoch *LYs* equals to 0.5. To calculate *QALYs* we obtain the utilities of choosing action a from the state *s* from the studies by Huang et al. (2021) and Huang et al. (2018) as presented in *Table 5.2.*

Table 5. 2. Stage and treatment dependent quality of life utilities from literature

	Utility		**References**
Stage dependent	Stage I	0.77	Huang et al. (2018)
	Stage II	0.66	
	Stage III	0.56	
	Stage IV	0.50	
Treatment dependent	Surgery	0.65	Huang et al. (2021)
	Chemotherapy	0.83	
	Adjuvant Chemotherapy	0.80	
	Chemoradiotherapy	0.72	

In our model, we defined utilities as combination of stage and treatment dependent utilities. To obtain combined utilities we multiplicate stage and treatment dependent utilities. In *Table 5.3.* we present the recovery and death included combined utilities corresponding all the stages and actions.

Table 5. 3. Combined utilities

	Stage I	**Stage II**	**Stage III**	**Stage IV**	**Recovery**	**Death**
Surgery	0.5005	0.429	0.364	0.3185	1	0
Chemotherapy	0.6391	0.5478	0.4648	0.4067	1	0
Adjuvant Chemotherapy	0.616	0.528	0.448	0.392	1	0
Chemoradiotherapy	0.5544	0.4752	0.4032	0.3528	1	0

As we mentioned in Chapter 4.4 *QALY* is *Utility* x *Time*. Since we use 6 months follow up periods as decision epochs we multiply each utilities by 0.5. *QALYs* corresponding combined utilities are shown in Table *5.4.*

Table 5. 4. Used QALYs in the model

	Stage I	**Stage II**	**Stage III**	**Stage IV**	**Recovery**	**Death**
Surgery	0.25025	0.2145	0.182	0.15925	0.5	0
Chemotherapy	0.31955	0.2739	0.2324	0.20335	0.5	0
Adjuvant Chemotherapy	0.308	0.264	0.224	0.196	0.5	0
Chemoradiotherapy	0.2772	0.2376	0.2016	0.1764	0.5	0

In the model framework the reward function $R(s, s', a)$ works as follows. When the action a chosen, a patient moves from state s to state s' and the *DM* receives an observation o as an indicator of patient actual health state. The patient gets the reward according to observation o and $R(s, s', a)$.

5.1.7. Initial Belief State

In a partially observable environment, defining the initial belief is one of the most important issues in modeling. As explained in the previous chapter, the initial belief is a probability function over system states. If the initial belief is deterministic (the initial belief is known with certainty), then the belief probability for this particular state is equal to 1, and the others are 0. If the initial belief is stochastic, then the initial belief is the probability distribution over the states and its sum is equal to 1.

In this thesis, we run our model for two initial belief scenarios. Since the *Recovery* and *Death* are absorbing states, the initial probabilities of *Recovery* and *Death* are taken as 0 in both scenarios. Which are:

1. Definite value for all cancer stages separately for $s \in S$– {*Recovery, Death*}

   $b_0 \in$ {(1,0,0,0,0,0), (0,1,0,0,0,0), (0,0,1,0,0,0), (0,0,0,1,0,0)}
2. Probabilistic, in which we use the probability of incidence, prevalence and overall stage distribution probabilities of *CRC* from *SEER* database as b_0.
   a. Initial belief of incidence $s \in S$– {*Recovery, Death*} = (0.07, 0.26, 0.35, 0.32)
   b. Initial belief of prevalence $s \in S$– {*Recovery, Death*} = (0.32, 0.25, 0.23, 0.20)
   c. Initial belief of overall $s \in S$– {*Recovery, Death*} = (0.17, 0.22, 0.38, 0.23)

5.1.8. Value Function

In this thesis, we use the maximization of total expected future rewards as the objective function where the reward function uses *QALYs* and *LYs* as metrics. Therefore, our value function is undiscounted maximum total expected *QALYs* and the undiscounted maximum total expected *LYs.* Recall that in a finite horizon *POMDP*, the value function can in general be written as:

$$v_t^*(s) = \max_{a \in A_S} \{R(s, a) + \lambda \sum_{s' \in S} P(s, a, s') . \sum_{s' \in S} Z(s', a, o) . v_{t+1}^*(s')\}$$
for $t \in \{1, 2, \ldots, N\text{-}1\}$ and $s \in S$ **(4.6)**

Since our model is undiscounted, λ in Equation (4.6) is equal to 1. As mentioned earlier, each epoch is a 6-month follow-up period. Since we are looking for 5-year outputs, we use 10 decision epochs as discrete horizons in the model. Therefore, the value function:

$$v_t^*(s) = \max_{a \epsilon A_S} \{R(s, a) + \lambda \sum_{s' \epsilon S} P(s, a, s') . \sum_{s' \epsilon S} Z(s', a, o) . v_{t+1}^*(s')\}$$

for t= {1, 2, …, *10*} and $s \in S$: {*State I, State II, State III, State IV, Recovery, Death*}

where,

$a \in A_s$:{*Surgery, Chemotherapy, Adjuvant Chemotherapy, Chemoradiotherapy*}

$o_t \in Z$: {(*CEA+, CT+), (CEA-, CT+), (CEA+, CT-), (CEA-, CT-)*}

And the optimal policy $\pi^*$is:

$$\pi^*(s) = \text{argmax}_\pi \, v^*(s) \qquad \text{for } s \in S$$

We solved the model with belief *MDP* representation. In which:

- States as belief states: $b \in B$
- $a \in A_s$:{*Surgery, Chemotherapy, Adjuvant Chemotherapy, Chemoradiotherapy*}
- transition function, T: $P(b, a, b') = \sum_{o \in Z} P(b'|b, a, o) . P(o|a, b)$

  where $P(b'|b, a, o)$ = 1 if state estimator $(b, a, o) = b'$, $P(b'|b, a, o)$ = 0 otherwise.
- the reward function, R: $r(b, a) = \sum_{s \in S} b(s) R(s, a)$

And the value function and optimal policy $\pi^*$of *belief MDP*:

$$v^*(b) = \max_{a \epsilon A_S} \{r(b, a) + \sum_{o \epsilon Z} P(b, a, b') . v^*(b')\} \qquad \text{for } o \in Z$$

$$\pi^* = \text{argmax}_\pi \, v^*(b_0)$$

where,

$b_0 \in$ {(1,0,0,0,0,0), (0,1,0,0,0,0), (0,0,1,0,0,0), (0,0,0,1,0,0), (0.07,0.26,0.35,0.32,0,0), (0.32,0.25,0.23,0.20,0,0), (0.17,0.22,0.38,0.23,0 0)}

5.2. Colorectal Cancer Treatment Model

After explaining the components of the *POMDP* model in the previous section, we are ready to present how the model works. When the patient is diagnosed with cancer as a result of the pathological examination, the treatment process begins immediately. Depending on the stage of cancer, treatment options may differ as described in Chapter 3. During each decision period, *CEA* levels and *CT* results are examined as a result of the previous action and used to determine the current belief states. If the patient dies or recovers from the disease, the treatment period ends. Otherwise, the process continues between 10 determined horizons. After the process is finished, the optimal total *QALYs* or optimal total LYs are returned as a value function.

5.3. Solution Method

We use the *"Pomdp"* package in R programming language to model and solve our proposed model. After the model is built, we use the *"solve_pomdp"* function by setting "*Point Based Value Iteration Algorithm*" as method to solve the *POMDP* model and "*solve_mdp"* function by value iteration method to solve reduced *MDP*. Pseudo code of general *PBVI* is presented in *Figure 5.2.*

PBVI Algorithm

Function PBVI

$B \leftarrow \{b_0\}$

while v is not converged to v^* *do*

Improve (v, B)

$B \leftarrow Expand\ (B)$

Function Improve

repeat

for each $b \in B$ *do*

$\alpha \leftarrow$ backup (b, v)

$v \leftarrow v \cup \{\alpha\}$

until v is converged

Function Expand (B)

Initialize $B_{new} \leftarrow B$

for each $b \in B$ *do*

Successors $(b) \leftarrow \{ b' > 0\}$

$B_{new} \leftarrow B_{new} \cup \underset{b' \in \text{Successors}(b)}{argmax} ||b' - B||_{L1}$

return B_{new}

Figure 5. 2. Point based value iteration algorithm

6.SCENARIO ANALYSIS

In this chapter, we present the results of our numerical study where some hypothetical scenarios are generated to answer several research questions stated in Chapter 1. In particular, we want to investigate the impact of the unobservable health conditions on treatment policies and *QALYs.* We also want to see whether the proposed *POMDP* model is an accurate representation for the colon cancer treatment process and the results obtained from the proposed *POMDP* model are compatible with the literature and the guidelines. Lastly, we want to examine the robustness of the model for different scenarios.

With these research questions in mind, we use the *LY*, *QALY* and 5- year cancer related survival metrics to evaluate the performances of the *POMDP* model. In order to make a fair comparison throughout the analysis, we use the 5-year cancer-related survival information obtained from *SEER* data. As a base case, we run the treatment model generated with the help of current guidelines, expert opinions, and literature. We compare the base case with the established scenarios, selected studies in the literature, and the guidelines. We examine the model outputs for different diagnosis stages, different belief states, and different treatment periods. We also perform a robustness analysis to examine the model validity for different hypothetical scenarios. Modeling the *CRC* treatment history has a high level of complexity (O (|A| x |S| x |B| x$|S|^2$x |Z|)) where, |A| denotes the total number of actions, |S| denotes the total number of system states, |B| denotes the total number of belief states, and |Z| denotes the total number of observations. Therefore, we make some assumptions for computational simplicity which we explain in the next subsection.

In the Section 6.2, we first propose the *POMDP* model for the base case and then present the corresponding *MDP* model with no unobservable state as a part of the first research question, in which we investigate the impact of unobservable health conditions on *LY*, *QALY*, and 5-year cancer related survival. In the Section 6.3, to answer the second research question, we examine whether our model accurately represents the *CRC* treatment process. Next, in the framework of third research question, we investigate whether the output of our model is compatible with the latest guidelines in Section 6.4.

Finally, as a part of the final research question, we examine whether our model is robust for different scenarios in Section 6.5

6.1. General Assumptions

Assumption 1: In real life, comorbidities such as heart disease, hypertension, diabetes, etc. may affect the choice of the treatment policy, especially when it is to be decided whether a patient is suitable for surgery. Likewise, synchronous cancers may affect the treatment policy to follow. For example, the extent of the liver metastasis provides clues about patient survival and is vital when deciding whether to choose curative or palliative therapy. Since adding comorbidities and synchronous cancers to the state space increases computational complexity, we assume that there is neither comorbidity nor synchronous cancer with *CRC*.

Assumption 2: *AJCC* 5^{th} edition includes Stage 0 means the newly formed cancer. For Stage 0, available actions are watchful waiting or removing nodules during colonoscopy. After consulting with the experts, we excluded Stage 0 in our model because this stage is more suitable for screening modeling and not suitable for treatment modeling, and hence no treatment policy needs to be followed strictly.

Assumption 3: In real life practice, the quality of life naturally decreases as age increases. In our model, we assume that utilities *(HR-QOL)* are not age-dependent. Utilities obtained depend only on treatment and cancer stage. We combined stage and treatment-dependent utilities.

Assumption 4: In real life practice, treatment policies are age-dependent, and model outcomes such as survival, mortality, and quality of life may differ according to age and gender. The utility values used in this thesis do not depend on age and gender, so we assume that age and gender do not affect model outputs. Treatment policies depend only on stage and observations.

Assumption 5: Physicians decide follow-up times based on patients and guidelines – *National Comprehensive Cancer Network (NCNN)* guideline recommends 3-6 months of follow-up. For this reason, in our model, with the help of the expert opinion. we assume that there is a 6-month follow-up period. Since we build a discrete time model for computational simplicity, one period in our model represents 6-month.

Assumption 6: There is a treatment option as palliative therapy for *CRC* patients of Stage IV. In this case, the goal of the treatment is not to cure, as the patient is in the terminal (untreatable phase) stage. This cure is given to reduce the patient's pain. Due to the lack of data, we assume that palliative therapy is not an available option.

Assumption 7: Cancer is a recurrent disease. When a patient recovers from cancer, a 5-year follow-up is applied due to the risk of recurrence. Relapse modeling needs follow-up data and also increases the size of the model extensively. We exclude the recurrence case in our model for computational simplicity. We assume that when a patient recovers, it is a complete recovery, and the patient leaves the system that the model is working at.

6.2. The Effect of Unobservable Environment

Recall that our first research question tries to investigate the impact of the unobservable environment on treatment policies and survival. To answer this question, we generate the corresponding *MDP* model to compare with the proposed *POMDP* model, e.g., the *MDP* model assumes that there is no uncertainty regarding the states and that the environment is fully known. The *POMDP* model is based on the most recent guidelines, expert opinions, and literature, as detailed in the previous chapter. We first present a recall to the proposed *POMDP* model and then present the *MDP* model. Finally, we compare the *LYs*, *QALYs*, and 5-year cancer-related survival results of the two models with each other and the *SEER*-based survival results to examine how the unobservable environment affects the *CRC* treatment policies.

6.2.1. POMDP Treatment Model

We got help from experts and current guidelines to model the *CRC* treatment process as a *POMDP*. In Turkey, the medical board decides on the treatment policy according to the patient's health status (cancer stage, *CT* results, blood measurement results (i.e., *CEA*), clinical characteristics (age, gender, comorbidities, medical history, drug response, etc.), guidelines, and clinical experience. According to the guidelines, comorbidities, age, and other social and physical parameters affect treatment policy. Recall that, for ease of calculation, we assume the proposed model is not age/gender-dependent and that other physical conditions (i.e., smoking status, diet, etc.), comorbidities, and social status have no influence on the treatment decisions. In the model framework, we assume that physicians make decisions based on only the current cancer stage, *CEA* level, and *CT* scan. In most of the studies in literature, stage transitions are considered only one way. Some studies consider progression-free conditions. In fact, the cancer stage is a dynamic parameter, not a static one. If treatment is successful, the actual stage can be readjusted. Therefore, we have considered multidirectional transitions in our model.

6.2.2. MDP Treatment Model

As mentioned in Chapter 4.2, the observation state, o_t is a probabilistic function of the actual (core) system state, s_t ($o_t = y.\ s_t$ → where $0 \leq y \leq 1$). If y =1 then the state is fully known, which means the *POMDP* turns into the *MDP*. If we interpret it within the framework of the *CRC* treatment model, $o_t = s_t$ means that the environment is perfectly known, so the cancer stages are accurate, and all possible type I errors are excluded. In other words, the radiological *(PET, CT)* and pathological diagnostic methods work perfectly with no margin of error. In this fully known ideal and hypothetical environment, treatment policies are independent of physicians' experience and theoretical knowledge. Physicians can follow the current guidelines without a need for intuitive and case-specific decisions. In the proposed *POMDP* model, we define the observation states as the change in the level of *Carcinoembryonic Antigen (CEA),* which is highly correlated with cancer stages, as well as *CT* results, which is a method used to describe the health status of patients. To construct the *MDP* model, we assume that treatment selection and reward function can only be made based on the current cancer stages only.

6.2.3. Computational Results

Recall that we try to investigate the impact of the unobservable environment on treatment policies and survival. We also evaluate the effect of the physician's experience and theoretical knowledge on treatment policies. With these objectives, we run a base-case *POMDP* model and the corresponding *MDP* model. We compare two models based on *LYs, QALYs,* and 5-year cancer related survival results. Along with the absolute values, we also present the percentage difference between the results provided by *POMDP* model and *MDP* model. For a particular measure (x), we define the percentage difference for x (Δ_x%) as follows:

$$\Delta_x\ \% = \frac{(x(POMDP) - x(MDP))}{x(POMDP)}\ \text{x}\ 100$$

where x(i) represents the value of x performance measure under model i.

The results are presented in *Table 6.1* and *6.2*

Table 6. 1. The outputs of POMDP and MDP model

	Initial belief	**LYs (5 year)**	**Initial belief**	**QALYs (5 year)**	**Initial belief**	**5-year Survival**
POMDP Model	Stage I	4.97	Stage I	4.87	Stage I	%98.8
	Stage II	4.95	Stage II	4.81	Stage II	%98
	Stage III	4.61	Stage III	4.33	Stage III	%84
	Stage IV	3.58	Stage IV	3.1	Stage IV	%43
	Incidence based	4.57	Incidence based	4.33	Incidence based	%83
	Prevalence based	4.33	Prevalence based	4.02	Prevalence based	%73
MDP Model	Stage I	4.98	Stage I	4.63	Stage I	%99.2
	Stage II	4.95	Stage II	4.58	Stage II	%98
	Stage III	4.67	Stage III	4.1	Stage III	%87
	Stage IV	3.77	Stage IV	3.01	Stage IV	%50
	Incidence based	4.63	Incidence based	4.12	Incidence based	%85
	Prevalence based	4.43	Prevalence based	3.84	Prevalence based	%77

Table 6. 2. %Difference between two models

	Initial belief	**LYs (5 year)**	**Initial belief**	**QALYs (5 year)**	**Initial belief**	**5-year Survival**
Δ_x%	Stage I	-0.2	Stage I	4.9	Stage I	-0.4
	Stage II	0	Stage II	4.8	Stage II	0
	Stage III	-1.3	Stage III	5.3	Stage III	-3.6
	Stage IV	-5.3	Stage IV	2.99	Stage IV	-16
	Incidence based	-1.3	Incidence based	4.8	Incidence based	-2.4
	Prevalence based	-2.3	Prevalence based	4.5	Prevalence based	-5.5

In *Table 6.1* and *6.2* it can be observed that both models have similar outputs in terms of LYs. For the early stages (Stage I and Stage II), the difference is hardly noticeable, and both models give almost the same results. However, in the late stages (Stage III and Stage IV), the difference increases as the stage increases. We think that the higher result of *MDP* in terms of *LYs* in the late stages indicates the effect of unobserved health factors in the life year estimation. In the *POMDP* model, progression of the disease due to unobservable factors can be represented and we think it is more effective in representing stage changes. Therefore, a decrease in life-year results can be expected in the *POMDP* model. This may be an indication of the increasing importance of observational effect as the stage progresses.

We think that the higher outcomes in terms of *QALYs* in *POMDP* may be due to the benefit derived from observational treatment policies. For example, when a patient's condition is observed in a lower stage while it seems to be in an upper stage, the treatment to be applied may be given more mildly than in that upper stage. This can result in an increase in the patient's quality of life. As stage increase the difference between two model is reduced.

The biggest difference in survival is also observed at Stage IV. As the cancer progresses, new dimensions are added to the uncertainty of the patient's overall health status. For example, newly occurred additional diseases, deterioration of psychological health, gradual deterioration of the patient's body functions, side effects of severe cancer treatment all affect the health status of a *CRC* patient. Looking at the results, it can be said that the difference between the two models widens as the uncertainty increases. The dramatic reduction in *POMDP* for Stage IV may be due to its ability to better reflect the uncertainty in the environment than *MDP*. Finally, the difference between *LYs* and *QALYs* outcomes indicates a loss in quality of life. In *POMDP* model this difference is smaller than in *MDP* model. We can say that small gains in the quality of life of cancer patients, especially in advanced stages, make huge differences. This may also indicate that clinical expertise is an important factor in improving the patient's quality of life. Therefore, the results of the *POMDP* model have positive returns.

In this section, where we investigated the effect of clinical experience on patient life expectancy, it was observed that the outputs of the two models showed similar trends. We recommend further research involving more experts for more significant results.

6.3. Optimal Policies

To explore our second research question and to find out whether our model provides an accurate representation for *CRC* treatment process with the guidelines, we compare the optimal actions chosen by the proposed POMDP model according to the *QALYs* and *LYs* metrics with the treatment policies recommended by the "*ACS (American Cancer Society)*" ACS and *NCNN* guidelines. As we mentioned in Chapter 3, the goal of treating *CRC* is to (have the patient ready to) perform surgery and hence, the only action available for Stage I is surgery.

According to the guidelines, adjuvant chemotherapy is being tried for Stage II. Our model chooses adjuvant chemotherapy for Stage II under the *QALY* metric. This result shows that adjuvant chemotherapy gives satisfying results in terms of quality of life.

For Stage III, guidelines recommend adjuvant chemotherapy over chemotherapy, if possible since it can help prevent the recurrence of the disease. However, state that chemotherapy is also an appropriate treatment. The proposed *POMDP* model chooses chemotherapy in Stage III. As stated in Assumption 7, in our model, we ignore the possible recurrence. This might probably explain the output that our *POMDP* model chooses chemotherapy over adjuvant chemotherapy.

Finally, guidelines recommend adjuvant chemotherapy for Stage IV if the patient is suitable for surgical operation, and chemotherapy for other conditions. For Stage IV, our model chooses the adjuvant chemotherapy in accordance with model guidelines.

The proposed model does not select chemoradiotherapy as an action within the model. This result is consistent with the fact that guidelines do not recommend

chemoradiotherapy as the first option. This is probably due to the *HR-QOL* coefficients and the fact that expert opinions are not biased towards chemoradiotherapy in accordance with guidelines. Moreover, when we changed the utilities in favor of chemoradiotherapy, our model also incorporates it into the optimal policy map. Hence, it can be said that our model outputs, in general, conform to the guidelines.

For the *LYs* metric, our model selects surgery for all stages except Stage IV. This trend is also appropriate for the primary treatment of *CRC* to be surgery which cannot always be applicable due to some complications and its effects on quality of life. In cases where metastasis has not yet developed, surgery helps prolong the life expectancy. We present a small part of the model output to facilitate understanding. *Table 6.3* shows a part of the policy map belonging to the 10^{th} epoch *α-vectors* values for the 3 separate vectors. For this small piece of vector space, *Surgery* appears to yield the best results for Stages I and Stage III. For Stage II, it appears *Chemotherapy* and *Adjuvant Chemotherapy* have the same gains and outperform the *Surgery.* For Stage IV, the best output belongs to *Adjuvant Chemotherapy.* The effect of changing treatment policies (in terms of % Differences) between *Chemotherapy* and *Surgery* -2.5%, -2.9%, and -16.6% between *Adjuvant Chemotherapy* and *Surgery* -2.5%, 5.8%, and -22%, and between *Adjuvant Chemotherapy* and *Chemotherapy* 0%, -3%, 4.5% respectively. In *Stage IV* the effect of changing the treatment policy on life expectancy is higher than other stages. It may mean that treatment policy decisions for Stage IV are more specific. It may also be an indication that the quality-of-life coefficients for Stage IV have a greater effect on treatment selection.

Table 6. 3. Policy map sample

α-vector	**Stage I**	**Stage II**	**Stage III**	**Stage IV**	**Recovery**	**Death**	**Action**
1	0.00	0.41	0.33	0.21	0.5	0	Chemotherapy
2	0.00	0.41	0.32	0.22	0.5	0	Adjuvant Chemotherapy
3	0.43	0.40	0.34	0.18	0.5	0	Surgery

We also present a 2-dimensional belief space of policies shown in *Table 6.3* between Stage II and Stage III. As we explained in Chapter 4 each lines indicate an *α-vector* and corresponding values of boundary stages which are Stage II and Stage III. As can be seen in the *Figure 6.2 Surgery* and *Adjuvant Chemotherapy* is on the upper surface. But *Chemotherapy* is a dominated vector for this small part of policy map.

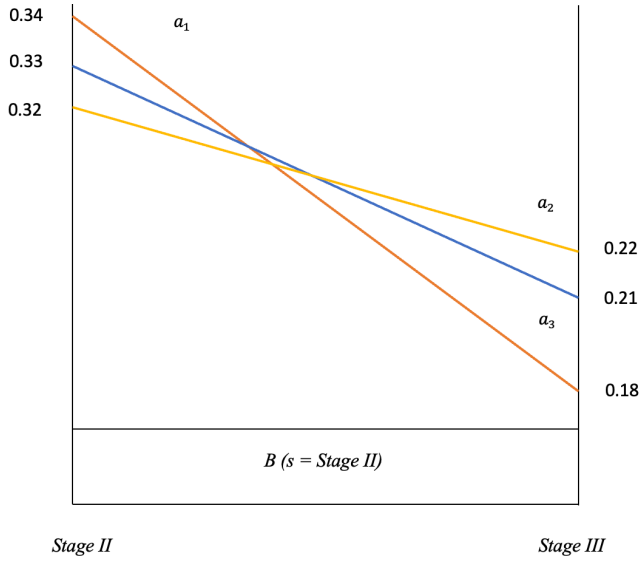

Figure 6. 1. 2-Dimensional belief space sample

(a_1= Surgery, a_2= Adjuvant Chemotherapy, a_3= Chemotherapy)

Considering the results of all these analyses, we conclude that the proposed model complies well with the guidelines. However, we need real data to train and test with the objective of building a more robust model.

6.4. Comparison with Authorities

In this section, we wanted to analyze the effectiveness of the proposed model to answer our 3rd research question in which if the results of the proposed *POMDP* model compatible with authorities. Within this framework, we present a comprehensive comparison with outputs obtained from the *SEER* in Section 6.4.1 and we present an additional comparison including *SEER* data, the *Turkey Cancer Statistics (TCS)* (Turkyilmaz et al., 2021), and a local study (Gulsen Unal et al., 2019) in Section 6.4.2

6.4.1. Comparison with SEER Survival

We compare 5-year survival results with *SEER* survival data to see if the proposed *POMDP* is suitable for modeling *CRC* treatment history. With this objective, we used *SEER*Explorer* to obtain the *SEER* results (National Cancer Institute, 2022b). *SEER* is an enormous US based cancer database. In particular, the life expectancy data is based on the US population. However, *SEER* data is used all over the world to get statistics on all types of cancer. *SEER* used a different cancer staging system from the *AJCC 5th* edition, which includes three stages as localized, regional, and distant. Localized stage is a combination of Stage I and Stage II of *AJCC 5th* edition. Regional corresponds to Stage III and distant is the Stage IV. To make a fair comparison of the results of the proposed *POMDP* model and the *SEER* Data, we selected data that include the non-Hispanic white people having age > 50 from both genders in *Table 6.4.*

Table 6. 4. 5-year Relative Survival (%) (SEER based Results)

	5-year Relative Survival (%)	**95% C.I. Lower Limit**	**95% C. I. Upper Limit**
Localized	94.7	94.1	95.3
Regional	78.3	77.4	79.1
Distant	18	17.0	19.0
Overall	70.2	69.6	70.7

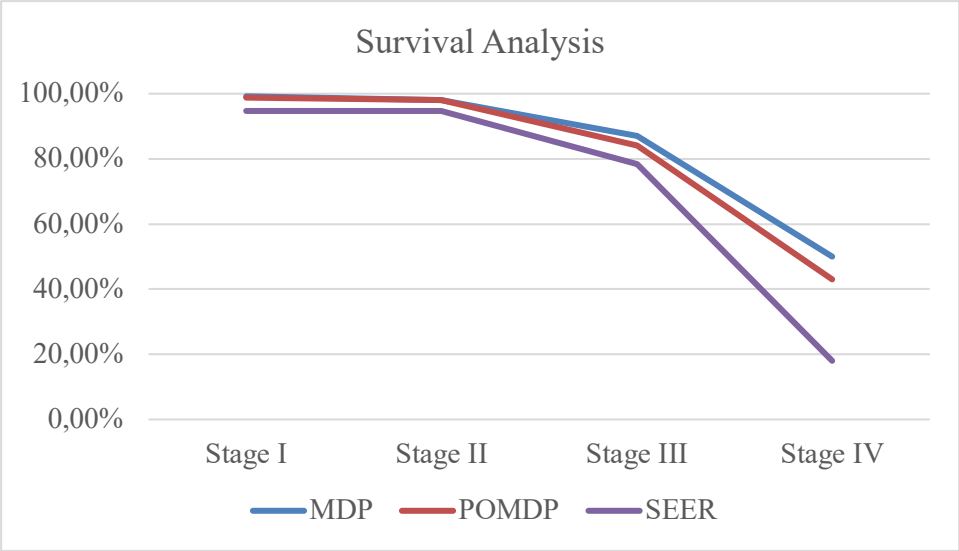

Figure 6. 2. Survival Analysis (5 year)

In *Figure 6.2* we present survival outputs of proposed *POMDP*, corresponding *MDP* and *SEER* data. As can be observed from the *Figure 6.2* both models have similar trends with the *SEER* based results. However, *POMDP* is closer to *SEER's* relative survival than the *MDP*. This may indicate that clinical expertise is an effective factor when modeling on treatment procedures due to the unobservable nature of the disease. We can also say that the 5-year survival rates found from the *MDP* model may provide an upper bound to the 5-year survival rates found from the proposed *POMDP* model.

The difference between the SEER results and the proposed *POMDP* can be attributed to geographical factors. The socioeconomic level of the population also affects the benefit from treatment. As stated before, *SEER* is a US-based database, but our model represents the history of the *CRC* in Turkey. In addition, we calculated cancer-related survival, but *SEER* data that have been used consider the relative survival of all *CRC* patients (including *non-CRC*-related deaths of *CRC* patients). Therefore, the difference may also indicate loss of life due to other mortality factors.

Even if both US and Turkey populations are similar in terms of accessing the treatment, the patient's psychological condition (stress factor), access to quality food and physical activity affect the outcome of the treatment and therefore life expectancy and quality.

Moreover, the fact that health services are more accessible in Turkey compared to the US, for example, that cancer treatment is covered by social security insurance, may also be an explanation for the high survival rates for distant stage. However, our model has some limitations and yet not sufficient to talk about this with certainty. More research needs to be done to be sure in terms of data and opinion collection.

Finally, as can be seen from the *Figure 6.2* our model has similar trends with *SEER* result. Therefore, it can be sad that our model has promising outputs and provide a good representation of survival.

6.4.2. Additional Comparison

In order to evaluate the performance of the model, we now examine the obtained survival outputs by using the stage distribution of the targeted data as the initial belief status of proposed *POMDP.* We think the *POMDP* model is promising compared to the *SEER* and (Turkey Cancer Statistic, 2017) results. As % differences with target values are acceptable with 3.55 and 4, respectively. We also present an additional comparison with a study conducted at Ege University Faculty of Medicine, in which they used patient data admitted the university hospital and the %difference is 8. As we mentioned before, environmental conditions are very effective in the course of *CRC*. We think that these differences between the results of Ege University and the proposed model may be due to the fact that the used data is more local (Includes one hospital data.)

Table 6. 5. Survival results and %Differences

Study	**5-year Survival**	**POMDP**	**Δ_x%**
SEER	65.1%	67.5%	3.55
TCS	55.5%	57.9%	4
Ege University	54.5%	59.4%	8

In general, it can be thought that the transition and observation probabilities we defined with a limited number of experts perform well. The performance of the model may improve if the probability collection process involves more experts. Based on these results, we think that the proposed *POMDP* is a suitable tool for modeling the environment of colon cancer treatment. For further research, we recommend deriving probabilities from actual data or conducting extensive studies involving more experts.

6.5. Robustness Analysis

In this section we run the proposed *POMDP* model in different scenarios to understand if our model is robust enough with respect to the parameters used in the model. To answer our last research question, we mainly focused on *QALYs*.

6.5.1. The Effect of Diagnosing Stage

In the base-case *POMDP* model in Section 6.4.2, we define the initial belief with the stage distributions obtained from the corresponding datasets to calculate the overall results. In this section, we present a new scenario and run a model initiated separately for each stage to see how the diagnostic stage affects the *QALYs* obtained.

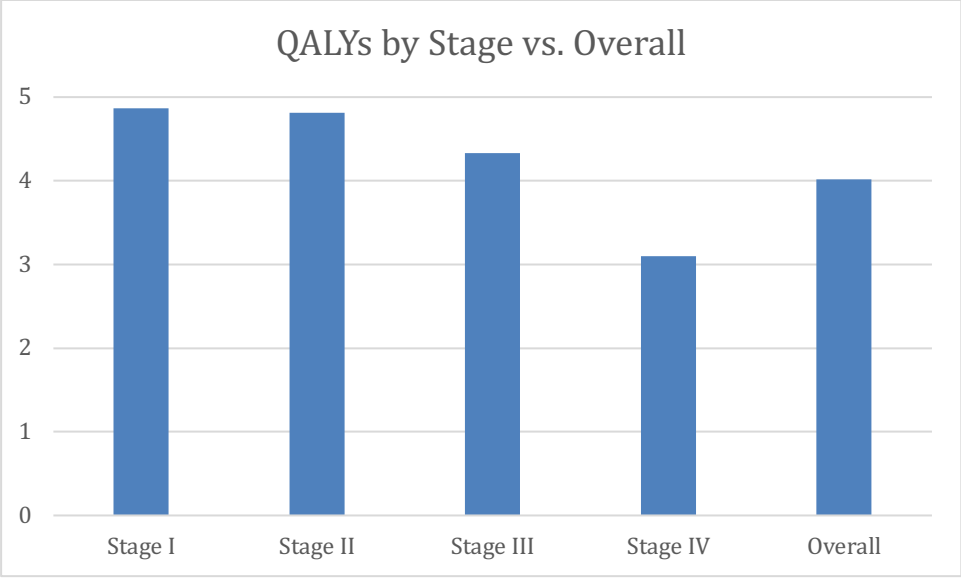

Figure 6. 3. 5-year QALY values by stage and overall survival

Colon cancer progress very quietly because the colon is in an isolated part of the body. The transition time from Stage I to Stage II takes an average of 10 years. This situation is very restrictive in terms of diagnosis of the disease. Many patients do not experience or do not notice symptoms until they reach Stage III or even Stage IV. This means that there is no significant decrease in the patient's quality of life in the initial stages (Stage I and Stage II). When it comes to Stage IV a dramatic decrease in the patient's quality of life occurs with the cancer spread through the body.

As can be observed from *Figure 6.3* the *QALYs* decrease as the cancer stage increases. The lowest value belongs to Stage IV as expected. When we look at the general results, we see that the 5-year life expectancy of a CRC patient is equal to 4.02 *QALYs.* Therefore, it can be said that *CRC* causes 0.98 quality adjusted lifetime loss on the average.

For a more comprehensive analysis and to understand quality perception of Turkey population we recommend using quality coefficients from patients in Turkey and transitions gathered from more experts. Due to the lack of data and the COVID-19 pandemic (reaching experts and getting help was much more difficult and riskier than during non-pandemic periods) we are unable to do any further analysis.

6.5.2. Changing Utilities

In the base case scenario, we combined the stage-dependent and action-dependent utility functions obtained from the literature. Actually, there are many versions of utility functions in the literature that depend on the stage, action, age, gender, or some combination of several of these. As we have explained in Assumption 3 in Section 6.1., we do not consider age- and gender-dependent utilities in our model. To understand the impact of changing utilities on the model output, we run two new scenarios using only the stage-dependent utilities and treatment-dependent utilities for each stage and overall. We compared both models with the base case *POMDP.* The *QALYs* for both scenarios and proposed *POMDP* are shown in *Table 6.6* below.

Table 6. 6. QALY results with different utility functions

	Proposed Model	**Stage-dependent**	**Treatment-dependent**
Stage I	4.87	4.93	4.91
Stage II	4.81	4.87	4.88
Stage III	4.33	4.42	4.5
Stage IV	3.1	3.2	3.41
Overall	4.02	4.11	4.21

As shown in *Table 6.6,* the treatment-dependent utility model and the stage-dependent utility model produce more optimistic *QALY* results than the proposed model in which we use combined utilities. We think this may be due to ignoring one dimension of quality of life in each scenario.

For Stage I, the stage-dependent model yields the best *QALYs* among all which makes sense since the disease is highly curable in this stage and has minimal impact on the patient's quality of life. The stage dependent model outputs -1.23% differs from the proposed *POMDP* and -0.4% with treatment dependent model.

For Stage II, Stage III and Stage IV treatment dependent utilities has better *QALYs* among all and QALYs deviate from the proposed MDP model with a difference of -1.5%, -3.9%, and -10%, respectively. We think this may be because more than one available action is on the table for these stages (recall the only action available for Stage I is surgery), and each of them has better utility coefficients than surgery. We can think that *QALY* results, which decrease as the stage progresses, may be an indication that the effect of the disease and the side-effects of the treatments applied on the patient's quality of life increases as the stage progresses.

For the overall expected *QALYs* outcomes, we can say that the stage-dependent utility model yields closer *QALY* results to the proposed *POMDP* than the treatment-dependent one, with a difference of -2.2% and -4.7%, respectively. For a more realistic model, we also recommend using a *HR-QOL* that depends on both stage and treatment. In addition, we think that the inclusion of age and gender-related utilities from the targeted patient population (for our case *CRC* patients in Turkey) in the model may increase the ability of representation of the history of the *CRC* treatment process.

6.5.3. The Effect of the Clinical Experience

In the scope of this study, we now present a new scenario to understand the effect of clinical expertise on the *QALYs.* With this objective, we collect opinions on the transition probabilities from two separate group of physicians. The first group is an oncologist group who are expert in specifically cancer. The second group is also a group of physicians not working in the oncology field. However, their knowledge about *CRC* is also theoretical and clinical. The theoretical knowledge comes from medicine school, and clinical knowledge is obtained through communications with oncologist within the hospital (for patient consultations) and their internship in oncology department. In *Table 6.7* we present the *QALY* results obtained from the *POMDP* model using the opinions of the transition probabilities of both groups.

Table 6. 7. QALYs based on different expert groups

	Proposed Model	**Oncologist Group**	**Non-Oncologist Group**
Stage I	4.87	4.88	4.91
Stage II	4.81	4.85	4.8
Stage III	4.33	4.65	3.88
Stage IV	3.1	4.17	1.17
Overall	4.02	4.45	3.61

For Stage I, the proposed model and oncologist group model is not different with -0.2% difference. The best outcomes come from non- oncologist group with 1.2% difference with proposed model. For this stage we can say that both groups have similar knowledge.

For Stage II, Stage III, and Stage IV non-oncologist group has a more pessimistic perception about the course of the disease than the oncologist group with %1, 16.5% and 71% differences, respectively. The difference of both model is increase as stage progresses. The knowledge of non-oncologist group is mostly theoretical and comes from textbooks which are mostly based on US data. On the other hand, the oncologist group has obviously more control and more knowledge over Turkey *CRC* patient data. We think their experience had some effect, particularly on the odds of death and recovery. The non-oncologist group estimated higher death probabilities than the oncologist group.

Therefore, we think the differences of both models might be an indicator of geographical effect. In addition, medicine is a multidisciplinary field in which many specialists work together for patients with many diseases. When a *CRC* patient needs to be evaluated by another field specialists, the patient's condition may worsen, and treatment may be required not only for cancer but also for comorbidities or side effects. We think that the non-oncologist group is facing with *CRC* patients in bad situations. For example, an *ER* specialist evaluates a *CRC* patient when the patient is in a terminal stage and is close to death. This may be due to the more pessimistic perception of the course of the disease by non-oncologists. But our study is not sufficient to make certain statements. Future works that involve more experts from different geographic areas and different levels of expertise may be more useful.

7.CONCLUSION

In this thesis, we developed a *POMDP* model for clinical decision support of *CRC* treatment. The model was constructed to represent the environment that is partially observable due to the dynamic nature of the disease and imperfect diagnostic tests. We made some assumptions for computational simplicity. The model represents the natural history of *CRC*, which includes the four cancer stages, recovery, and death conditions. The transition probabilities for this model were determined with the help of experts. Due to the COVID-19 pandemic, the number of experts providing opinions is not high. Although we reached experts working in different regions of the country, it was not enough to reach a general opinion and the expert opinions showed some variability. Despite this, our model can be considered an accurate representation of *CRC* history. Because we considered observations such as *CEA* level and *CT* results in our model that had not been addressed by previous studies although they are used very frequently in practice, this has increased the representation level of the model.

In this study, we also have shown the difference between the model that considers observations (proposed *POMDP*) and the model that does not consider the observations (*MDP* model). The results showed that adding the aforementioned observations to the model had some effect on the QALYs gained, but further analysis was required.

We have shown that our model provides treatment recommendations similar to the guidelines, and we consider the proposed model can be considered as a promising representation of the history of the *CRC* treatment process and our model have complied well with guidelines.

We compared the outputs of proposed model with *SEER* and *TCI* results. We have shown that our model overall results are not much different with 3.55% and 4% respectively. This indicate that transition and observation probabilities we defined with expert' help is acceptable and provide an accurate model.

We proposed different scenarios to understand model robustness. We investigated the effect of the utilities, reward function, and transitions on the modeling *CRC.*

With this study, we present 5-year *QALY* and *LY* estimations for patients with *CRC.* We think that it can support the clinical decisions of physicians in determining the direction of treatment according to the patient's life expectancy.

As the staging system and observation states are common to various cancer types, we think our model is adaptable for other cancers with the integration of specific probabilities.

As a result, in this study, *POMDP* has been shown to be a convenient and a suitable method for modeling the history of *CRC* and can be used to provide clinical decision support. Obviously, there are several limitations of this study, which we try to explain in the next section.

7.1. Recommendations for Future Study

The integration of country-based utilities to the model is the key area for future research. Currently, the model is a good representation of the natural history of *CRC* but cannot be said to be an accurate representation for *CRC* patients in Turkey.

For a more realistic model, we used HR-*QOL* that depends on both stage and treatment. In addition, we recommend the inclusion of age and gender-related utilities from the targeted patient population may increase the model’s ability of representation of the history of the *CRC* treatment process.

The current model takes a homogeneous group of people >50 years old. Future studies could consider diverse cohorts of models from many different age ranges and genders to examine whether they might affect *QALYs.*

We think further research involving more experts may yield more meaningful results. In addition, the use of real data integrated with expert opinion may be more useful in obtaining transition probabilities. The cost effectiveness ratio metrics can be used by adding treatment costs to the model for more meaningful outcomes. This could help determine thresholds for treatment options. Expanding the frame by adding the screening process to the model allows for a more periscopic view of the *CRC* history. This approach can be used to support future medical research, particularly epidemiology studies.

For the proposed *POMDP* model, we assumed that there are no comorbidities and synchronous cancers. Comorbidities and synchronous cancer susceptibility analysis may be included in future research. We used traditional treatment methods such as chemotherapy surgery adjuvant chemotherapy and chemoradiotherapy in our model. We suggest more alternative options can be added in future works.

We also constrained the model by ignoring the recurrence of the disease. Therefore, we assume that a patient who reaches the recovery state is completely done with the cancer and never becomes a CRC patient again. Since colon cancer is a slowly progressing disease and we examined the 5-year time period, we assumed that the effect of this assumption on the outputs of the model was negligible. But, in real life there is always a risk of recurrence. Therefore, we think that adding the probabilities that a patient can transition from recovery to cancer stages will increase representational accuracy of the model. Especially, when the planning horizon is longer, adding the recurrence possibility will be quite important.

We think that relaxing all these constraints will increase the ability of representation of the model. A more powerful and robust *POMDP* model may be possible if the modeling methodology can be integrated with reinforcement learning.

REFERENCES

Alagoz, O., 2011. Optimizing cancer screening using partially observable markov decision processes. Transforming Research into Action 75–89. https://doi.org/10.1287/educ.1110.0087

Altekruse, S.F., Kosary, C.L., Krapcho, M., Neyman, N., Aminou, R., Waldron, W., Ruhi, J., Howlader, N., Tatalovich, Z., Cho, H., Mariotto, A., Eisner, M.P., Lewis, D.R., Cronin, K., Chen, H.S., Feuer, E.J., Stinchcomb, D.G., Edwards, B.K., 2010. SEER Cancer Statistics Review, 1975-2007.

Ayer, T., Alagoz, O., Stout, N.K., 2012. OR forum-A POMDP approach to personalize mammography screening decisions. Operations Research 60, 1019–1034. https://doi.org/10.1287/OPRE.1110.1019

Ayvaci, M.U.S., Alagoz, O., Eren Ahsen, M., Burnside, E.S., 2018. Preference-Sensitive Management of Post-Mammography Decisions in Breast Cancer Diagnosis. Wiley Online Library 27, 2313–2338. https://doi.org/10.1111/poms.12897

Bellman, R., 1966. Dynamic Programming. Science (1979) 153, 34–37. https://doi.org/10.1126/SCIENCE.153.3731.34

Bellman, R., Kalaba, R., 1965. Dynamic programming and modern control theory.

Bennett, C.C., Hauser, K., 2013a. Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. Artificial Intelligence in Medicine 57, 9–19. https://doi.org/10.1016/J.ARTMED.2012.12.003

Bennett, C.C., Hauser, K., 2013b. Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. Artificial Intelligence Med.

Braziunas, D., 2003. POMDP solution methods.

Bullement, A., Cranmer, H.L., Shields, G.E., 2019. A Review of Recent Decision-Analytic Models Used to Evaluate the Economic Value of Cancer Treatments. Applied Health Economics and Health Policy 17, 771–780. https://doi.org/10.1007/S40258-019-00513-3

Cassandra, A.R., 1998. A survey of POMDP applications, in: Planning with Partially Observable Markov Decision Processes.

Chhatwal, J., Alagoz, O., Burnside, E.S., 2010. Optimal breast biopsy decision-making based on mammographic features and demographic factors. Operations Research 58, 1577–1591. https://doi.org/10.1287/OPRE.1100.0877

Cipriano, L.E., Goldhaber-Fiebert, J.D., Liu, S., Weber, T.A., 2018. Optimal Information Collection Policies in a Markov Decision Process Framework. Medical Decision Making 38, 797–809. https://doi.org/10.1177/0272989X18793401

Crichton, N., Clin Nurs, J., 2001. Visual analogue scale (VAS).

Cubbage, D., 2004. Simulation of colorectal cancer: The natural history of disease.

Donaldson, M.S., Corrigan, J.M., Kohn, L.T., 2000. To Err Is Human: Building a Safer Health System, Vol. 6. ed. National Academies Press.

Doshi, F., Pineau, J., Roy, N., 2008. Reinforcement learning with limited reinforcement 256–263. https://doi.org/10.1145/1390156.1390189

Erenay, F., Alagoz, O., Service, A.S.-M.&, 2014, undefined, 2014. Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. pubsonline.informs.org 16, 381–400. https://doi.org/10.1287/msom.2014.0484

Erenay, F.S., Alagoz, O., Banerjee, R., Cima, R.R., 2011. Estimating the unknown parameters of the natural history of metachronous colorectal cancer using discrete-event simulation. Medical Decision Making 31, 611–624. https://doi.org/10.1177/0272989X10391809

Fayers, P., Bottomly, A., 2002. Quality of life research within the EORTC—the EORTC QLQ-C30. European Journal of Cancer 38, 125–133.

Feeny, D., Furlong, W., Torrance, GW., Goldsmith CH., 2002. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care.

Gafni, A., 1994. The standard gamble method: what is being measured and how it is interpreted. Health Services Research.

Goulionis, J.E., Koutsiumaris, B.K., 2010. Partially observable Markov decision model for the treatment of early Prostate Cancer. OPSEARCH 47, 105–117. https://doi.org/10.1007/S12597-010-0015-0

Gudex, C., 1994. Time trade-off user manual: props and self-completion methods.

Gülşen Ünal, N., Coşgun, G., Korkut, M., Ömer Özütemiz, A., Doğanavşargil, B., Çakar, B., Özkök, S., Yalman, D., Sert, F., Sezak, M., Yoldaş, T., Çalışkan, C., Akgün, E., Osmanoğlu, H., Haydaroğlu, A., 2019. Epidemiologic and survival characteristics of patients with colorectal cancer in Ege University medical faculty database. Ege Journal of Medicine 58, 68–77.

Howard, R.A., 1960. Dynamic programming and Markov processes [WWW Document]. URL https://psycnet.apa.org/record/1961-01474-000 (accessed 5.2.22).

Huang, H.Y., Wang, H., Shi, J.F., Bai, Y.N., Wang, L., Liu, C.C., Yan, X.X., Liu, G.X., Liao, X.Z., Shi, D., Mao, A.Y., Ren, J.S., Sun, X.J., Guan, H.J., Huang, W.D., Lei, H.K., Gong, J.Y., Liu, Y.Y., Zhu, L., Song, B.B., Du, L.B., Guo, L.W., Liu, Y.Q., Ren, Y., Lan, L., Zhou, J.Y., Qi, X., Sun, X.H., Lou, P.A., Wu, S.L., Tang, Y., Li, Ni, Zhang, K., Li, Ning, Chen, W.Q., Dai, M., He, J., 2021. Health-related quality of life of patients with colorectal neoplasms in China: A multicenter cross-sectional survey. Journal of Gastroenterology and Hepatology (Australia) 36, 1197–1207. https://doi.org/10.1111/jgh.15238

Huang, W., Yang, J., Liu, Y., Liu, C., Zhang, X., Fu, W., Shi, L., Liu, G., 2018. Assessing health-related quality of life of patients with colorectal cancer using EQ-5D-5L: A cross-sectional study in Heilongjiang of China. BMJ Open 8. https://doi.org/10.1136/bmjopen-2018-022711

Ibrahim, R., Kucukyazici, B., Verter, V., Gendreau, M., Blostein, M., 2016. Designing Personalized Treatment: An Application to Anticoagulation Therapy. Production and Operations Management 25, 902–918. https://doi.org/10.1111/POMS.12514

Joranger, P., Nesbakken, A., Sorbye, H., Hoff, G., Oshaug, A., Aas, E., emeritus, professor, 2020. Survival and costs of colorectal cancer treatment and effects of changing treatment strategies: a model approach. Springer 21, 321–334. https://doi.org/10.1007/s10198-019-01130-6

Kiani, M., Isik, T., Eksioglu, B., 2020. Dynamic Tuberculosis Screening for Healthcare Employees. https://doi.org/10.48550/arxiv.2007.05568

Kumar, V., Abbas, A.K., Aster, J.C., Perkins, J.A., 2013. Robbins basic pathology, 9th ed. Elsevier, Philadelphia.

Latha, R., Vetrivelan, P., 2019. Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing. 2019 11th International Conference on Communication Systems and Networks, COMSNETS 2019 833–837. https://doi.org/10.1109/COMSNETS.2019.8711358

Leshno, M., Halpern, Z., Arber, N., 2003. Cost-effectiveness of colorectal cancer screening in the average risk population. Health Care Management Science 6, 165–174. https://doi.org/10.1023/A:1024488007043

Littman, M.L., 1994. The witness algorithm: Solving partially observable Markov decision processes. 128.148.32.111.

Lovejoy, W.S., 1991. A survey of algorithmic methods for partially observed Markov decision processes. Annals of Operations Research 28, 47–65. https://doi.org/10.1007/BF02055574

Mayer, R.J., 2018. Lower Gastrointestinal Cancers, in: Jameson, J.L., Kasper, D.L., Hauser, S.L., Longo, D.L., Loscalzo, J.L. (Eds.), Harrison's Principles of Internal Medicine. pp. 572–578.

McLeish, J.A., Thursfield, V.J., Giles, G.G., 2002. Survival from colorectal cancer in Victoria: 10-year follow up of the 1987 management survey. ANZ Journal of Surgery 72, 352–356. https://doi.org/10.1046/J.1445-2197.2002.02407.X

Monahan, G.E., 1982. State of the art—a survey of partially observable Markov decision processes: theory, models, and algorithms, in: Pubsonline.Informs.Org. Management Science, pp. 1–111. https://doi.org/10.1287/mnsc.28.1.1

Namen Leon, M.I., Akhavan Tabatabaei, R., Gel, E., 2015. A POMDP approach to age-dependent primary screening policies for cervical cancer in Colombia.

National Cancer Institute, 2022a. Parts of the colon [WWW Document]. National Cancer Institute. URL https://nci-media.cancer.gov/pdq/media/images/741570.jpg (accessed 5.2.22).

National Cancer Institute, 2022b. SEER*Explorer Application [WWW Document]. URL https://seer.cancer.gov/statistics-network/explorer/application.html?site=20&data_type=4&graph_type=5&compareBy=stage&chk_stage_101=101&chk_stage_104=104&chk_stage_105=105&chk_stage_106=106&series=9&sex=1&race=8&age_range=141&advopt_precision=1&advopt_show_ci=on&advopt_display=2 (accessed 5.16.22).

National Comprehensive Cancer Network, 2022. Colon Cancer [WWW Document]. URL https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf (accessed 5.2.22).

Nouri, N., 2019. Fractionated Treatment Planning of Radiation Therapy Considering Biological Response. University of Houston.

Obulaney, P.A., Gilliland, I., Cassells, H., 2016. Increasing Cervical Cancer and Human Papillomavirus Prevention Knowledge and HPV Vaccine Uptake through Mother/Daughter Education. Journal of Community Health Nursing 33, 54–67. https://doi.org/10.1080/07370016.2016.1120595

Papadimitriou, C., Tsitsiklis, J., 1987. The complexity of Markov decision processes. Mathematics of Operations Research.

Petousis, P., Han, D.X., Hsu, W., Bui, A.A., 2019a. Generating reward functions using IRL towards individualized cancer screening. International Workshop on Artificial Intelligence in Health 11326 LNAI, 213–227. https://doi.org/10.1007/978-3-030-12738-1_16

Petousis, P., Winter, A., Speier, W., Aberle, D.R., Hsu, W., Bui, A.A., 2019b. Using sequential decision making to improve lung cancer screening performance. IEEE Access 7.

Pineau, J., Gordon, G., Thrun, S., 2006. Anytime Point-Based Approximations for Large POMDPs. Journal of Artificial Intelligence Research 27, 335–380. https://doi.org/10.1613/jair.2078

Puterman, M.L., 2014. Markov Decision Processes: Discrete Stochastic Dynamic Programming [WWW Document]. John Wiley & Sons. URL (accessed 5.2.22).

Sandikçi, B., Maillart, L.M., Schaefer, A.J., Alagoz, O., Roberts, M.S., 2008. Estimating the Patient's Price of Privacy in Liver Transplantation. https://doi.org/10.1287/opre.1080.0648 56, 1393–1410. https://doi.org/10.1287/OPRE.1080.0648

Schaefer, A.J., Bailey, M.D., Shechter, S.M., Roberts, M.S., 2005. MODELING MEDICAL TREATMENT USING MARKOV DECISION PROCESSES. Springer, Boston.

Scuffham, P.A., Whitty, J.A., Mitchell, A., Viney, R., 2008. The use of QALY weights for QALY calculations: a review of industry submissions requesting listing on the Australian Pharmaceutical Benefits Scheme 2002-4. Pharmacoeconomics 26, 297–310. https://doi.org/10.2165/00019053-200826040-00003

Sehr, M., Bitmead, R., 2017. Tractable dual optimal stochastic model predictive control: An example in healthcare. ieeexplore.ieee.org.

Shani, G., Pineau, J., Multi-Agent, R.K.-A.A. and, 2013, undefined, 2013. A survey of point-based POMDP solvers. Springer 27, 1–51. https://doi.org/10.1007/s10458-012-9200-2

Shaw, J., Johnson, J., Coons, S., 2005. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. JSTOR.

Singh, R., Liu, F., Shroff, N.B., 2020. A Partially Observable MDP Approach for Sequential Testing for Infectious Diseases such as COVID-19. https://doi.org/10.48550/arxiv.2007.13023

Sondik, E., 1971. The optimal control of partially observable Markov processes.

Spaan, M., Vlassis, N., 2004. A point-based POMDP algorithm for robot planning. IEEE International Conference, Amsterdam.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021a. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians 71, 209–249. https://doi.org/10.3322/CAAC.21660

Surveillance Epidemiology and End Results Program, 2018. Colorectal Cancer Statistic [WWW Document]. URL https://seer.cancer.gov/statfacts/html/colorect.html (accessed 5.2.22).

Thrun, S., 1999. Monte Carlo POMDPs.

Turkyilmaz, M., Ozturk, M., Dundar, S., Kavak Ergun, A., Sevinc, A., Tutuncu, S., Seymen, E., 2021. Türkiye Kanser İstatistikleri (2017). Ankara.

Vickers, A.J., 2011. Prediction models in cancer care. CA: A Cancer Journal for Clinicians n/a-n/a. https://doi.org/10.3322/CAAC.20118

Vozikis, A., Goulionis, J.E., 2009. Medical decision making for patients with Parkinson disease under Average Cost Criterion. Australia and New Zealand Health Policy 6. https://doi.org/10.1071/HP090615

Williams, J., Language, S.Y.-C.S.&, 2007, undefined, 2006. Partially observable Markov decision processes for spoken dialog systems. Elsevier. https://doi.org/10.1016/j.csl.2006.06.008

Zhang, N., Liu, W., 1996. Planning in stochastic domains: Problem characteristics and approximation.

Zhang, Y., 2012. A Partially Observable Markov Decision for Optimal Design of Surveillance Policies for Bladder Cancer. North Carolina State University.

Zhu, M., 2010. Assessing colorectal cancer screening and surveillance strategies with colonoscopy: A Partially Observable Markov Decision Processes (POMDP) approach. Prude University.