MODELING DISEASE MANAGEMENT DECISIONS FOR PATIENTS WITH PNEUMONIA-RELATED SEPSIS

by

Jennifer E. Kreke

BSIE, University of Pittsburgh, 2001

MSIE, University of Pittsburgh, 2002

Submitted to the Graduate Faculty of

the School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

SCHOOL OF ENGINEERING

This dissertation was presented

by

Jennifer E. Kreke

It was defended on

July 9, 2007

and approved by

Andrew J. Schaefer, Associate Professor, Department of Industrial Engineering

Matthew D. Bailey, Assistant Professor, Department of Industrial Engineering

Larry J. Shuman, Professor & Associate Dean for Academic Affairs, School of Engineering

Mainak Mazumdar, Professor Emeritus, Department of Industrial Engineering

Mark S. Roberts, Associate Professor, Department of Medicine

Dissertation Director: Andrew J. Schaefer, Associate Professor, Department of Industrial Engineering

MODELING DISEASE MANAGEMENT DECISIONS FOR PATIENTS WITH PNEUMONIA-RELATED SEPSIS

Jennifer E. Kreke, PhD

University of Pittsburgh, 2007

Sepsis, the tenth-leading cause of death in the United States, accounts for more than $16.7 billion in annual health care spending. A significant factor in these costs are unnecessarily long hospital lengths of stay, which stem from the lack of optimal hospital discharge policies and the inability to assess a patient's true underlying health state effectively. Researchers have explored ways of standardizing hospital discharge policies by comparing various strategies, but have not been able to determine optimal policies due to the large number of treatment options.

Furthering the state of research into decisions made in the management of patients with sepsis, this dissertation presents clinically based optimization models of pneumonia-related sepsis that use patient data to model disease progression over time. Formulated using Markov Decision Process (MDP) and Partially Observable Markov Decision Process (POMDP) techniques, these models consider the clinician's decisions of when to test for additional information about the patient's underlying health state and when to discharge the patient from the hospital.

This work utilizes data from the Genetic and Inflammatory Markers for Sepsis (GenIMS) study, a large multi-center clinical trial led by the University of Pittsburgh School of Medicine. A key aim of the GenIMS trial is to demonstrate that the levels of certain cytokines are predictors of patient survival. Utilizing these results, the models presented in this dissertation consider the question of when to test for cytokine levels using testing procedures that may be costly and inaccurate. A significant result of this dissertation demonstrates

that testing should be performed when a clinician is considering the decision to discharge the patient from the hospital.

This study characterizes optimal testing and hospital discharge policies for multiple problem instances. In particular, multi-region control-limit policies are demonstrated for specific patient cohorts defined by age and race. It is shown that these control-limit policies depend on the patient's length of stay in the hospital. The effects of testing cost and accuracy on the optimal testing and discharge policies are also explored. Finally, clinical interpretations of the optimal policies are provided to demonstrate how these models can be used to inform clinical practice.

Keywords: Markov decision processes, partially observable Markov decision processes, medical decision making, sepsis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am grateful for the encouragement and support of my colleagues, family, and friends.

I am thankful for the mentoring and significant financial support I received from Yoni Levy and the AT&T Labs Fellowship Program.

I appreciate the many contributions of my dissertation committee:

My advisor and dissertation committee chair, Andrew J. Schaefer, provided valuable guidance and direction.

I am indebted to Matthew D. Bailey for his mathematical expertise and considerable effort in reviewing this dissertation.

Larry J. Shuman, Mainak Mazumdar, and Mark S. Roberts shared many insights into applying engineering techniques to the field of medicine.

1.0 INTRODUCTION

1.1 CURRENT SEPSIS MANAGEMENT OPTIONS

Sepsis results from an overwhelming inflammatory response to infection. Under normal circumstances, the human body mounts a potent, complex immunologic response when invaded by a pathogen, ensuring adequate protection against infection. For some patients, however, a deficient immunologic defense may allow infection to become established. On the other hand, an excessive or poorly regulated response may actually harm the body [117]. By overproducing or producing the wrong proportions of inflammatory and anti-inflammatory molecules (also known as cytokines), the body may negatively impact one or more of its organ systems, leading to organ dysfunction and possibly death. This serious condition, sometimes referred to as severe sepsis, septic shock, or septicemia, is the tenth leading cause of death in the United States, as illustrated in Table 1.1.

Researchers continue to explore ways of reducing patient mortality through improving treatment efficacy at all stages of the disease, yet current therapy options are still mainly ad-hoc [53] and highly depend on the severity of the disease [117]. Initially, antibiotics may be used to treat the underlying infection; however, factors such as polymicrobial infections and antimicrobial-resistant organisms make prompt diagnosis and treatment of infection difficult [37]. If the administered antibiotics are ineffective, organ support therapies such as fluid replacement, mechanical ventilation, and blood transfusions, may be needed to prevent organ failure.

In addition to support treatments, researchers are investigating means by which to control the body's inflammatory response. Despite considerable advances in medicine, researchers still do not have a complete understanding of sepsis at the molecular level [117]. Though

Table 1.1: Leading Causes of Death, 2003 [54]

Rank	Cause of Death	Number	Percent of Total Deaths
···	All Causes	2,448,288	100.0
1	Diseases of the Heart	685,089	28.0
2	Malignant neoplasms	556,902	22.7
3	Cerebrovascular diseases	157,689	6.4
4	Chronic lower respiratory diseases	126,382	5.2
5	Accidents (unintentional injuries)	109,277	4.5
6	Diabetes mellitus	74,219	3.0
7	Influenza and pneumonia	65,163	2.7
8	Alzheimer's disease	63,457	2.6
9	Nephritis, nephrotic syndrome, and nephrosis	42,453	1.7
10	**Septicemia**	**34,069**	**1.4**
11	Intentional self-harm (suicide)	31,484	1.3
12	Chronic liver disease and cirrhosis	27,503	1.1
13	Essential (primary) hypertension and hypertensive renal disease	21,940	0.9
14	Parkinson's disease	17,997	0.7
15	Assault (homicide)	17,732	0.7
···	All other causes	416, 932	17.0

many studies using cells and animals have greatly improved knowledge of the pathophysiology of sepsis, it is still not clear what factors are advantageous or deleterious in the progression of sepsis within an individual. As a result, experimental medications aimed at controlling the body's inflammatory response have had limited success [15, 16, 77, 119].

1.2 GENETIC AND INFLAMMATORY MARKERS FOR SEPSIS (GENIMS)

In an attempt to learn more about sepsis, research efforts such as the GenIMS trial led by the Department of Critical Care Medicine at the University of Pittsburgh and funded by the National Institutes of Health [45], are trying to understand the body's inflammatory response at a molecular level. The work presented in this dissertation utilizes data obtained through the GenIMS study, which is the largest study to date of this kind.

The study of specific patient populations has been suggested as a way to improve the value of results from clinical trials of novel anti-sepsis strategies [6]. For this reason, the GenIMS study chose to focus on pneumonia, the leading cause of sepsis [91]. Further restricting the cohort, the study includes only those patients who are admitted to the hospital with pneumonia (i.e., community-acquired pneumonia). The main goal of the GenIMS study is to determine the extent to which specific genetic, inflammatory, and clinical factors influence the development of infection and progression to sepsis, organ dysfunction and death. This goal is achieved through three specific aims: (1) to determine whether specific DNA polymorphisms for key inflammatory molecules are associated with the risk of developing pneumonia and progressing to severe sepsis, septic shock, organ dysfunction, and death, (2) to investigate the relationships among specific DNA polymorphisms, inflammatory molecule expression, and clinical course and outcome in infection and sepsis, and, (3) to develop and evaluate clinical decision tools that include genetic and inflammatory response information.

The work presented in this dissertation extends Aims (2) and (3) by developing and solving optimization models that consider decisions made sequentially and dynamically in the care of sepsis patients. Due to the complexity of the disease and its treatment, this work

chooses to focus on decisions made in two areas of sepsis management: cytokine testing and hospital discharge policies. These decisions are key disease management problems that regularly impact sepsis treatment. In particular, this dissertation investigates the questions of when to test for patient cytokine levels and how the information can be used to optimally discharge patients from the hospital.

1.3 PATIENT-SPECIFIC HOSPITAL DISCHARGE POLICIES

All models presented in this dissertation are parameterized via patient-based information obtained through the GenIMS trial to model the clinician's decision-making process for the treatment of patients with severe sepsis. In particular, these models focus on developing standard decision-making policies that can provide clinical guidelines for patient treatment. For example, the last decision a clinician makes during sepsis treatment is when to discharge a patient from the hospital.

Evidence has shown that the discharge decision is rarely made using patient-based standards [50]. With the annual costs associated with severe sepsis exceeding $16.7 billion in the United States [7], it is highly desirable to avoid unnecessary days in the hospital. Yet, despite attempts to decrease costs by reducing hospital length of stay [41, 78], concerns about the morbidity and mortality associated with premature hospital discharge have resulted in substantial differences in length of stay between and within institutions [21, 42]. These differences suggest that decisions are made in an ad-hoc fashion, often due to physician intuition and clinical uncertainty [17, 78].

Recent studies have focused on standardizing discharge procedures to reduce cost without increasing the risk of patient morbidity and mortality. In an attempt to develop patient-based discharge policies, these studies have explored modeling techniques that consider the cost-benefit tradeoff underlying the discharge decision. For example, Clermont et al. [24] developed a dynamic microsimulation model to predict various outcomes for critically ill patients, including day of discharge from the intensive care unit (ICU). While this model can be used as a predictive tool, it does not provide patient-specific optimal discharge policies.

Similarly, Halm et al.[50] used statistical modeling to measure the time to clinical stability in patients with community-acquired pneumonia. The authors discuss how their results can be used to improve the efficiency of in-patient management by providing an evidence-based estimate for optimal length of stay. These estimates, however, cannot be easily translated into health-based discharge policies.

This dissertation extends these and other previous modeling efforts by presenting a model and analysis of the hospital discharge decision using a Markov decision process (MDP) approach. Historically, MDPs [13, 89] have been applied in areas such as inventory control [29, 56] and production planning [14], but have recently seen increased application in medicine [98], including organ transplantation [3] and HIV therapy planning [100]. Within the limits of its assumptions, MDPs provide optimal decision policies. In addition, MDPs can more efficiently evaluate a larger number of policy alternatives than other modeling techniques used in sepsis research to date, such as statistics [62], Markov modeling [91], and simulation [22, 24].

1.4 USING CYTOKINE INFORMATION TO IMPROVE PATIENT SURVIVAL

A key aim of the GenIMS investigation is to demonstrate that the levels of certain cytokines are strongly correlated with patient survival. It has been discovered that in some cases, a patient may appear to be well, but the patient's cytokine levels indicate that the patient has a higher probability of death should the patient be discharged from the hospital than if the patient were to remain in the hospital receiving standard treatment [63]. Assuming that cytokines are correlated with patient survival, it is clear that knowing the levels of these cytokines will change how the clinician makes decisions, for example, when to discharge the patient from the hospital.

The decision of when to test for cytokine information is not obvious for several reasons. First, the test may be costly both from an economic standpoint and in terms of the time spent by the clinician in performing the test and analyzing the results. Given this cost, the

clinician may not want to order a test in each time period leading to periods of time when no test results are received. Secondly, the testing procedure may be inaccurate. For example, the test results may have an associated measurement error. Even when the measured test results are accurate, there may be an error associated with how the clinician interprets the results with respect to the patient's true underlying health state. Each of these situations creates an environment of partial observability, in that the patient's underlying true health state is not known with certainty by the clinician.

This dissertation presents a Partially Observable Markov Decision Process (POMDP) model [81, 82, 103] to explore how testing decisions influence the hospital discharge decision. POMDPs extend the modeling framework of the MDP by allowing the current state of the patient to be partially observable. In this case, the true underlying health state can only be observed through a testing procedure that may have associated error and cost.

POMDPs have been previously applied to medical decision making questions in the areas of heart disease treatment [51, 85], efficient dosage policies for medical drug therapy [55], and breast cancer treatment [57], but this is the first study to utilize patient-based data from a large scale clinical trial to develop optimal policies that can inform clinical decision making guidelines.

1.5 GENERAL PROBLEM STATEMENT

The cytokine testing and hospital discharge decisions made in the management of a patient with sepsis can be described more formally as follows.

A patient is admitted to the hospital. At some point at or after hospital admission, the patient is suspected of having developed sepsis. Once suspected of having sepsis, the clinician then treats the patient following some general treatment process or strategy. The data used in this study consider only those patients that are suspected of having sepsis, based on a variety of criteria utilized by the GenIMS investigator team. This dissertation does not consider issues related to the diagnosis of sepsis.

Treatment decisions are made based upon the patient's current health state, which is comprised of completely and partially observable elements. Completely observable elements of the patient's health state can be directly observed by the clinician. Partially observable elements, however, are not directly observable, usually because they can only be obtained through testing procedures that have associated costs and degrees of accuracy.

It is assumed that at regular intervals during the treatment process, the clinician must make decisions about patient care. Due to the complexity of the disease and resulting treatment options, this dissertation considers only a subset of the decisions made during the treatment process, namely, when to order various tests for more information about the patient's underlying health state and when to discharge the patient from the hospital. It is assumed that all other decisions follow the "standard care" process, although as has already been discussed, opportunities exist throughout sepsis treatment for the standardization of treatment policies.

At each decision point the clinician can readily observe various completely observable elements of the patient's health state. If tests were ordered previously, the clinician may also have test results to observe. Based on this information, the clinician then chooses either to discharge the patient from the hospital or to keep the patient in the hospital for continued treatment. If the clinician chooses to keep the patient in the hospital, then the clinician can also choose, for a cost, to order one or more tests to obtain additional information about the partially observable elements of the patient's health state. It is assumed that the test results are received at the beginning of the next time period and are therefore available before the next decision is made.

The clinician's goal at each decision point is to maximize the patient's expected survival over a finite observation horizon. Since sepsis is an acute, short-term disease, patient death due to sepsis usually occurs within 90 days of hospital admission. As a result, the models presented in this dissertation measure the value of a decision policy in terms of a patient's 90-day expected survival as measured from hospital admission. Sepsis treatment typically occurs over an even shorter treatment horizon. As observed in the GenIMS study, treatment rarely last longer than 30 days; therefore, the models considered in this dissertation utilize a treatment horizon of 30 days.

1.6 DISSERTATION AIMS

This dissertation models this general problem as a POMDP. Due to the complexity of analyzing and solving large-scale POMDPs [18, 103], this dissertation then considers two simplified model variants. For each of these model variants, the model structure and optimal policies are analyzed and computational experiments are conducted using patient data from the GenIMS trial. The overall goal of this dissertation is to provide insight into decision making strategies for the management of patients with sepsis. In addition, the insight gained from the application of these novel modeling techniques to the field of medical decision making will further the roles of industrial engineering and operations research in formalizing decision making strategies in clinical practice. These goals are further described in the following sections.

1.6.1 Gaining insight into optimal policy structure

In addition to formulating the clinician's hospital discharge and testing decisions, the dissertation investigates the mathematical structure of these models.

For example, one goal of this dissertation is to characterize non-stationary optimal policies for the hospital discharge problem. A non-stationary control-limit policy takes the following form: treat the patient in the hospital until the patient's health improves and reaches a time-dependent control limit state, then discharge the patient from the hospital. These types of policies are appealing since they are easy to understand and can be implemented as part of a general discharge strategy. In addition to defining optimal policies, properties of the input parameters and other model components are also evaluated.

Another goal of this dissertation is to explore the impact of test cost and accuracy on testing and discharge decisions. In particular, the effects of increasing test accuracy and decreasing test costs on the decision of when to test for additional cytokine information will be used to motivate the need for more accurate and less expensive testing methods.

From a research perspective, these structural properties provide mathematically interesting insights into a new application of the MDP and POMDP modeling methodologies. From

an application standpoint, the results help to provide structure to a decision making strategy that is mainly ad-hoc in practice. As will be discussed in later chapters, these results are particularly interesting in that they suggest that standard policies are possible. Even though the models are fairly simplistic from a clinician's viewpoint, they provide a starting point for future analysis. These policy structures suggest baseline strategies to inform future policy decisions within the medical decision making community.

1.6.2 Calibrating complex models with actual patient data

This dissertation also uses patient data from the GenIMS trial to solve various problem instances based on a variety of patient characteristics. It will be shown that the optimal policies for many problem instances are similar to control-limit type policies. For the POMDP instances, the results pertaining to changes in test cost and accuracy are also validated. While the specific results are not yet at the level that can be directly implemented in practice, they do validate many of the results that are shown through the structural analysis. A comparison of results based on varying patient cohorts is also presented.

From a stochastic optimization research perspective, computational experiments demonstrate the effectiveness of using the MDP and POMDP modeling techniques to determine optimal policies for clinical decisions made in the treatment of patients with sepsis. From an application perspective, computational experiments tie the mathematical models back to the underlying clinical problem by presenting concrete examples of input parameters and resulting optimal policies. As will be seen for each instance tested, the optimal policy can be translated into an actual optimal decision that should be made for each possible state and stage. This information is particularly useful when explaining the model structure and results to the medical community in general and to the clinicians that actually make these treatment decisions in practice.

1.7 DISSERTATION ORGANIZATION

The remainder of this dissertation is organized as follows: Chapter 2 reviews the literature concerning the current state of medical decision making in general and as related to sepsis research in particular. Background information on the modeling methodologies used throughout this dissertation is also provided. Chapter 3 then presents a formulation of the general model as presented in the problem description. An overview of the model, its variants, and its relationship to the remainder of the dissertation are discussed. Chapter 4 then presents a simplified variant of the general model in which the hospital discharge decision is considered in an MDP model that uses the Sepsis-related Organ Failure Assessment (SOFA) score to describe the patient's completely observable health state. Chapter 5 extends this model with another simplified variant of the general model in which both the hospital discharge decision and the cytokine testing decision are considered in a POMDP model that uses the value of a single cytokine to describe the patient's observable health state. In both Chapters 4 and 5, structural results of the model formulations and computational experiments utilizing the GenIMS data are presented and discussed. Chapter 5 also discusses heuristic approaches for combining the results from the MDP and POMDP models to develop more complex SOFA- and cytokine-based testing and discharge policies. Finally, Chapter 6 discusses the contributions of this dissertation from both the medical application and the methodological research perspectives and provides directions for future research.

2.0 LITERATURE REVIEW

2.1 CURRENT METHODS IN SEPSIS MODELING

Until recently, clinical trials aimed at assessing the efficacy of treatment options for sepsis have met with limited success due both to trial design [6] and treatment options [20]. By the year 2000 there were over thirty large randomized controlled trials (RCTs) of novel anti-sepsis strategies that failed to demonstrate any clinical impact [6]. Since then, numerous other RCTs have met with the same fate [25, 53]. Due to the complexity of the disease, RCTs have been unable to adequately compare the virtually limitless management possibilities. These failures provide strong evidence of the need for advanced modeling techniques that allow for the evaluation of a large number of treatment decisions made over time without the need for large RCTs. With the advent of clinical studies like GenIMS, aimed at gathering data concerning the underlying progression of the disease, the use of mathematical modeling techniques to analyze and interpret therapeutic options has become increasingly important. Mathematical modeling serves to extend current research efforts by providing the capability to analyze both current and new treatment strategies.

Many clinical researchers have begun to develop models to compare available management options [33] based on clinical data from large-scale trials like GenIMS. These studies are aimed at assessing the effectiveness of various anti-sepsis treatments and strategies and demonstrating a relationship between one or more clinical, biological, inflammatory, and genetic factors and patient survival. The models presented in this dissertation extend these efforts by allowing for the implicit comparison of virtually all decision options to determine the optimal strategy.

This section both reviews the literature surrounding sepsis modeling as well as presents the current state of medical decision making methodology including cost-effectiveness analysis, decision trees, influence diagrams, Markov modeling, Bayesian analysis, simulation analysis, and Monte Carlo simulation. Relevant studies involving other medical decision making questions are discussed as needed to describe the application of a specific methodology. Most importantly, the motivation for the modeling techniques used in this dissertation, namely Markov decision processes and partially observable Markov decision processes, is presented. Since the models presented in this dissertation are the first applications of these techniques to questions in sepsis management, applications of these methods to other areas in medical decision making are also reviewed.

2.1.1 Cost-effectiveness analysis

Cost-effectiveness analysis (CEA) [48, 86] is a commonly used analysis technique that involves the calculation and comparison of the costs and effects of various disease management options. The relative cost-effectiveness of each option can then be assessed by calculating and comparing their respective incremental cost-effectiveness ratios. In the area of sepsis research, Wang et al. [115] conducted a cost-effectiveness analysis to compare the use of various treatments in the management of sepsis. The authors concluded that the epidemiology of the disease state is an important factor in cost-effectiveness analysis and recommended an infection-specific approach to modeling treatment options in sepsis. Burchardi and Schneider [16] compared intensive care unit (ICU) care versus non-ICU care for the treatment of patients with sepsis. They also investigated the cost-effectiveness of potential new therapies and concluded that new therapies should be directed at patients that are the most likely to benefit from the costly intervention. Both of these studies demonstrate the ability of cost-effectiveness analysis to compare known treatment strategies and possibly suggest direction for new research. However these studies cannot be used to suggest what these new strategies should be.

2.1.2 Decision trees and influence diagrams

To evaluate the effects of a particular disease management strategy, medical decision makers often employ graphical analysis techniques such as decision trees [86] and influence diagrams [33]. Decision trees use a logical, cause-and-effect framework to depict possible decision choices and all possible outcomes for each choice. The outcomes can be deterministic or based upon a known probability distribution. Marchetti et al. [76] used a decision tree framework to demonstrate that a combined prophylactic splenectomy and cholecystectomy provides a substantial gain in quality-adjusted life expectancy in certain patient cohorts as compared to the no surgery option in the treatment of patients with mild hereditary spherocytosis.

Unfortunately, as the number of possible decisions increases, the decision framework can become unmanageable and difficult to analyze. This problem occurs, in particular, when a decision or series of decisions must be made sequentially throughout time with the effects of earlier decisions influencing later decisions. The MDP framework discussed in the next section will improve this framework, even though it too suffers from what has become known as the "curse of dimensionality" [12], though at a much later stage than the decision tree.

While a decision tree describes the causes and effects of decisions made in a specific management strategy, an influence diagram is a network with directed arcs and no cycles that is used to show relationships between random variables and possible decisions. Unlike in a decision tree, where the probability distributions that characterize the effects of making a particular decision are necessary to evaluate a decision policy, the influence diagram can actually be used to gain insight into the value of the transition probabilities themselves. This technique, however, requires an in-depth understanding of the relationships between variables that comprise the system under analysis. In the field of sepsis treatment, researchers are still working to develop a fundamental understanding of the disease [53]. A specific aim of the GenIMS study is to "investigate the relationships among specific DNA polymorphisms, inflammatory molecule expression, and clinical course and outcome in infection and sepsis." This information will be useful in constructing an influence diagram of sepsis as part of future research efforts.

As the relationships between patient health states become more complicated, influence diagrams will often employ Bayesian methods to capture conditional probability distributions between random variables. Spiegerlhalter presented graphical Bayesian methods that are basically complex influence diagrams [108]. Computational methods were also discussed. Computational methods for using influence diagrams as input to more advanced modeling techniques also exist [73, 99]. In particular, Magni and Bellazzi [73] developed a software package called DT-Planner that used an "influence view" to depict the probabilistic relationships between variables that is then used to specify an MDP formulation of the model.

2.1.3 Markov modeling

Extending the capabilities of decision trees and influence diagrams, a common structure used to calculate the effects of a particular disease management option is the Markov model [83, 107]. Markov modeling is a decision-analytic technique in which all of the relevant conditions in a particular problem are represented as a set of states that are mutually exclusive and exhaustive. Patients (or portions of an entire cohort of patients) move through the model according to probabilities that govern how likely it is to transition from one state to another. The long-term behavior of the model provides insight into the expected behavior of the patient or group of patients under the current system as modeled. For more information on the general application of Markov models to medical decision making, the reader is referred to Sonnenberg and Beck [107].

Markov modeling has been used to understand the progression of sepsis in patients. Rangel-Frausto et al. [92] conducted a nine-month prospective cohort study, using their results to develop a Markov model of the natural history of sepsis. This model has the ability to predict the probability of movement to and from varying stages of sepsis (sepsis, severe sepsis, and septic shock) and can be used to predict the reduction in end-organ dysfunction and mortality resulting from the use of increasingly effective antisepsis agents. Bauerle et al. [11] developed a three-state (well, septic, and dead) Markov model to describe the course of the disease in critically ill patients. They used this model to develop risk profiles of various patient groups, allowing for the comparison between age- and gender-specific survival rates.

In addition to the study of sepsis, Markov modeling has been used in other areas of medical decision making, such as the study of patient preferences for various treatment options. For example, Ahn and Hornberger [1] developed a Markov model to asses organ quality from the patient's perspective in the allocation of organs for transplant to patients with End-Stage Renal Disease (ESRD). In their paper, Ahn and Hornberger presented a decision model that considers patient preferences for specified health states that then influence the patients' decisions about which organ are acceptable for transplantation. The authors demonstrated that patients with favorable health characteristics can afford to be more selective about the quality of the transplant organ whereas less healthy patients may be inclined to select a lower-quality organ. In another interesting application of Markov modeling, Kao [60] presented a semi-Markov process model that looks at patient paths through the hospital based on various arrival rates to determine care requirements based on patient characteristics. The semi-Markov nature refers to the incorporate of time into the state description, since patient length of stay in the hospital was seen to be a significant factor in care requirements.

2.1.4 Simulation

In recent years, simulation has become a popular and commonly used modeling method in medical decision making. Simulation analysis [68] is a method by which logical rules are used to replicate, or imitate, a system in order to gain understanding and insight into how that system behaves. For example, Clermont et al. [24] developed a dynamic microsimulation model to predict various outcomes for critically ill patients, including day of discharge from the intensive care unit (ICU). While this model can be used as a predictive tool, it does not provide patient-specific optimal discharge policies. Similarly, Saka et al. [96] developed a simulation model of sepsis in order to study the rates of hospital discharge and patient death based on the patient's changing health state over time as well as static variables such as age and race. The authors calibrated the model with clinical data from the GenIMS trial. As will be discussed in future chapters, the work presented in this dissertation extends this simulation model by modeling the clinician's decision making process and how it affects the patient's health state transitions. Other simulation models in the medical literature

can be found in the areas of critical care medicine [67], HIV modeling [102, 104], organ transplantation [66, 96, 101], and infectious disease modeling [70].

Monte Carlo simulation [68] is a specific simulation tool that considers a static framework, or a system in which the passing of time does not need to be explicitly modeled. Similar to Markov modeling, the model consists of possible states with transition probabilities that govern how patient transitions between these states. In a Monte Carlo simulation, however, patients are sent through the model one at a time and random numbers are used at each transition point to determine how the patient transitions through the model. Studying the results for large cohorts of patients lends insight into the functioning of the system. For example, Su et al. [111] developed a Monte Carlo simulation model to compare various organ allocation policies for kidney transplantation to study the effects of incorporating patient preferences into the allocation policy. Related works by these authors include [120, 121].

Cost-effectiveness analysis, decision trees, Markov modeling, and simulation are effective methods of evaluating and comparing specific management options, but they cannot be used to *compute* an optimal management strategy comprised of dynamic decisions made throughout time. Even using these methods to evaluate a large number of management options can become computationally prohibitive.

2.1.5 Motivation for advanced techniques

This dissertation extends current research efforts in the study of sepsis treatment by utilizing an modeling technique called the Markov decision process (MDP) to consider not only the stochastic progression of the disease, but the resulting effects of sequential decisions made throughout time as well. MDPs extend the framework of the Markov process model by introducing a control process that directly influences how a patient transitions between states. The solution to an MDP is an optimal decision making policy that can be used to inform clinical practice, not merely an evaluation of a few pre-defined policies. Section 2.2 describes the MDP modeling structure in more detail.

This dissertation presents the first MDP model to consider decisions made in the treatment of sepsis. MDPs, however, have been successfully applied to medical decision making

questions in the areas of organ transplantation [3, 4, 26], hereditary spherocytosis treatment [74], and the control of infectious disease epidemics [69]. A general review of MDPs in medicine is provided in [98].

Alagoz et al. [3] presented an MDP model that addresses the clinical question of when to conduct a living donor liver transplant in order to maximize the patient's life expectancy (or quality-adjusted life expectancy). The authors presented clinically intuitive conditions on the input parameters that ensure that the optimal policy will be of a particular type. Computational results using patient-based data were included to support these policies for specific patient cohorts. In [4], Alagoz et al. extended their earlier work by modeling the patient's choice between a living donor transplant and a cadaveric liver transplant in addition to the clinician's decision of when to conduct the transplant procedure. Structural results demonstrated specific policy types, which were confirmed through computational experiments using clinical data. Another liver transplant study by these authors is [2].

David and Yechiali [26] used an MDP to model the decision of whether to accept or reject a kidney transplant offer. The authors considered how the length of the patient's time under medical care affects the optimal policy. Various organ offer rates were tested and a numerical example using clinical data was interpreted in the context of the model.

Magni et al. [74] demonstrated the improvements that the MDP offers over traditional modeling approaches such as decision trees and influence diagrams, mainly in its ability to consider dynamic, sequential decisions. They presented an MDP model of when to perform prophylactic surgery in patients with mild hereditary spherocytosis. Using data from the literature and clinical expertise, the authors modeled the dynamic progression of the disease and used this understanding of the disease to solve both their MDP model as well as a traditional static model. Comparing model results, they were able to demonstrate significant gains in patient quality-adjusted life days by delaying surgery in some cases according to the MDP optimal policy versus the static policy.

Lefevre [69] modeled the spread of an infectious disease in a closed population. The continuous-time Markov decision process considered decisions such as quarantining section of the population and implementing medical care programs to control the spread of the disease. In addition to presenting the model formulation, Lefevre analyzed the dependence of

the optimal policy on the input parameters and provided conditions on the input parameters that ensure that the optimal quarantine and medical program levels do not increase as the size of the infected population increases.

Markov decision problems are based upon the assumption that the current state is completely observable, i.e., all information about the state can be observed or known with complete certainty. The partially observable Markov decision process (POMDP) [13, 82], a generalization of the MDP, relaxes this assumption and allows the model to consider patient health states that can only be observed through an inspection or testing procedure. This state, therefore, is said to be partially observable because the observation procedure is either inaccurate (due to testing and/or test interpretation error) or costly (and therefore it may not be performed at every decision point). The POMDP framework, described in detail in Section 2.3, describes the patient health state not by the values of the patient variables themselves; rather, the state is described as the clinician's *belief* in what the variables are, which is based upon the values of the observations received up until the current point. While this framework is much more general than the MDP, its data requirements make it difficult to obtain practical results.

This dissertation is the first work to utilize POMDPs in the study of sepsis management. It is also one of the first studies to use patient-based data to derive optimal decision policies in the hopes of informing clinical practice. POMDPs have been previously applied in other areas of medical decision making, such as ischemic heart disease treatment [51], congenital heart disease treatment [85], efficient dosage policies for medical drug therapy [55], and breast cancer treatment [57].

Hauskrecht and Fraser [51] presented a POMDP model for the management of patients with ischemic heart disease. The authors constructed a hierarchical Bayesian belief network based on data from the literature and clinical expertise to represent the disease dynamics. Their work demonstrated the ability of the modeling framework to provide clinical insight, but they discussed issues with increasing computational complexity as the model size increases.

As in [51], Peek [85] presented an influence diagram to describe the underlying relationships between state variables in patients with ventricular septum defect, a disorder with

characteristics for congenital heart disease. The author utilized this diagram to construct a POMDP model that considers the various treatment decisions such as when to perform a chest X-ray and when to perform surgery. While the model captured many aspects of the disease and its treatment, the author does not present structural policy results or solutions obtained from clinical data.

Hu et al. [55] presented a POMDP model to determine efficient dosage policies from medical drug therapy. Specifically, they considered the effects of various information gathering policies, such as myopic policies and active learning policies. Computational results were presented to compare policy types and a passive information gathering strategy was suggested for use in clinical practice.

The work presented in this dissertation investigates questions in sepsis management using both MDPs and POMDPs. In addition to presenting the model formulation, the models are solved using patient-based data obtain from the GenIMS trial. Model results are interpreted and general policy recommendations are presented to inform clinical practice. The main contribution of this work is its ability to provide insight based on an analysis of model structure and results calculated from patient-based data.

Unfortunately, while the models discussed in this review have made great strides in presenting and motivating the use of these frameworks and in demonstrating the complexity they can incorporate, the models have not led to implementable results, largely due to insufficient data availability. As additional data become available, many of the models and techniques in the literature will prove to be increasingly valuable to clinical practice. To begin to bridge the gap between theory and practice, the model presented in this dissertation attempts to incorporate sufficient complexity to capture the dynamic nature of disease progression, while still allowing for solutions to be obtained with available data. The next sections provide background information on the mathematical structure of MDPs and POMDPs before moving on to present the general model.

2.2 MARKOV DECISION PROCESSES

The Markov decision process [12, 13, 89] is a general modeling technique used to formulate a problem involving sequential decisions made over time. The objective is to maximize a reward function that quantifies the effects of all possible outcomes of these decisions. A basic MDP model has two main features: a discrete-time dynamic and stochastic system that underlies the entire problem and a reward function that is additive over time.

The underlying dynamic system describes how the system changes as decisions are made at discrete points in time called *stages*. At each stage, the decision maker observes the *state* of the system and chooses one *action* from the set of all actions available at that specific point in time. Based on the state and action chosen, the decision maker receives some *reward* and then the system's state changes based on specified *transition probabilities*. A *policy* is a decision rule that tells the decision maker which action to take when a patient is in a given state at a given time. The value of this decision rule is calculated through the *value function*. The optimal policy is the policy that maximizes the total expected reward received by the decision maker; or, in other words, the policy that maximizes the value function for each starting state.

A basic assumption of the Markov decision process is that the state is Markovian. In other words, the current state in the model is assumed to capture all information necessary to make a decision moving forward. While this assumption may seem unreasonable, the state description can be altered to incorporate any historical information that may be needed when making a decision in the current stage. Unfortunately, data requirements necessary for the solution of the model increase exponentially with the size of the state space, which is often referred to as the "curse of dimensionality" [12]. These issues will be discussed in more detail as they related to the models presented in later chapters.

The definition of stages allows for the separation of the class of MDPs into finite- and infinite-horizon problems. In finite-horizon problems, rewards are received over a finite number of stages while infinite-horizon problems allow for the accumulation of rewards over an uncertain or indefinite horizon. The models presented throughout this dissertation are finite-horizon problems, though finite-horizon problems can be reformulated as infinite-horizon

problems through a standard augmentation of the state with the stage [32]. A general description of a finite-horizon problem is provided.

Following a modified version of the notation from Puterman [89], let N be the time-horizon and let S be the defined state space of the MDP. For every stage t and state $s_t \in S$, let the set of feasible decisions or actions be $A(s_t)$, where for every action $a_t \in A(s_t)$, the decision maker receives reward $r_t(s_t, a_t)$. It is assumed that the rewards are bounded and that S and $A(s_t)$ are discrete and finite. A transition from state s_t to state s_{t+1} when action $a_t \in A(s_t)$ is chosen occurs with probability $p_t(s_{t+1}|s_t, a_t)$.

Let a policy $d = \{d_1, d_2, \ldots, d_N\}$ be a sequence of decision rules, where a decision rule d_t is a function mapping states into actions at stage t such that $d_t(s_t) \in A(s_t)$. The application of such a policy induces a Markov chain where X_t is the state of the system at stage t and Y_t is the action chosen in state X_t, so that $Y_t = d_t(X_t)$.

The objective of an MDP is to find an optimal policy d^* that maximizes one of three criteria: the total expected reward, the total discounted expected reward, or the average reward per stage. The total expected reward criterion is often used when the reward received in later stages of the model has the same value as those received in earlier stages, which is often the case for finite-horizon problems with a short time horizon. This criterion is used in the models presented in this dissertation as the time horizon is relatively short as compared to the lifetime of the patient.

In infinite-horizon problems, particularly those that consider decisions that may take place very far in the future, the total discounted expected reward criterion is used to give more importance to decisions made in the near future than those made at a later point in time. For more information on the total discounted expected reward criterion, the reader is referred to [13, 95]. Studies, such as [38], have also explored the effects of various weighting mechanisms for discrete time, infinite horizon MDPs.

For infinite-horizon applications where discounting is inappropriate and there is no naturally occurring cost-free state the system eventually enters, then the total expected reward criterion may not be applicable because the total cost is not guaranteed to be finite. In such cases, it is often advisable to use the average reward per stage criterion [13].

The total expected reward criterion is further explained as it will be used in the models presented throughout this dissertation. Let $V_t^d(s_t)$ represent the total expected reward when policy d is used and the system starts in state s_t where,

$$V_t^d(s_t) = \sum_{t=1}^{N} r_t(X_t, Y_t).$$

Under the assumptions of bounded rewards and finite S and $A(s_t)$, $V_t^d(s_t)$ exists for each $d \in D^{MD}$, the set of all deterministic, Markovian policies [89]. Let $V_t^*(s_t)$ denote the optimal total expected reward for each state $s_t \in S$ where,

$$V_t^*(s_t) = \max_d V_t^d(s_t).$$

Then, by the principle of optimality [89], $V_t^*(s_t)$ can be found by solving the standard set of optimality equations, also known as the Bellman equations [12]:

$$V_t^*(s_t) = \max_{a_t \in A(s_t)} \left[r_t(s_t, a_t) + \sum_{s_{t+1} \in S} p_t(s_{t+1}|s_t, a_t) V_t^*(s_{t+1}) \right], \text{ for } t = 1, \ldots, N-1, \text{ and}$$

$$V_t^*(s_N) = r_N(s_N), \text{ for all } s_N \in S.$$

2.2.1 MDP structural results from the literature

Each proposed model in this dissertation is formulated as an optimal stopping problem with ordered states. Optimal stopping problems are discussed in greater detail in Section 2.4. In this problem, the decision maker can either decide to continue or to stop the process based on the current state of the system. A general description of the optimal stopping problem with a completely observable state space can be found in [31, 34].

Assuming an ordering on the state space, the most significant result for an optimal stopping problem is to demonstrate that the optimal policy is of control limit type. Using the notation introduced at the beginning of Section 2.2, a *control limit policy* [89] is composed of decision rules, $d_t(s_t)$, of the form:

$$d_t(s_t) = \begin{cases} a_1 & s_t < s_t^* \\ a_2 & s_t \geq s_t^*. \end{cases}$$

This equation says that if the value of the state at time t is less than some value s_t^*, which may or may not be time-dependent, then the optimal action is a_1. Otherwise, the optimal action is a_2. Bertsekas [13] shows several optimal stopping problem applications (e.g. asset selling, purchasing with a deadline) for which the optimal policy is of control limit type.

Puterman [89] presents general conditions for the existence of a control limit policy, which include:

1. Inductively showing that the optimal value functions from t onward are nonincreasing or nondecreasing in the state, and
2. then showing that the value function itself is superadditive or subadditive.

These conditions are further discussed in Chapter 4 as they relate to the models presented in this dissertation. Examples of control limit policies can be found in [13, 89]. In addition to satisfying these general conditions, application-specific conditions have been demonstrated in the literature. For example, Alagoz et al. [3] provide clinically realistic conditions under which it is optimal to perform a living donor liver transplant.

2.2.2 MDP solution procedures

The separability of the MDP decisions allows for the decomposition of the above problem into smaller related subproblems. As a result, such decisions can be solved using a simple backward induction algorithm [89], which is presented in Appendix C. Backward induction is used to solve the MDP models presented in this dissertation. A variety of techniques exist for solving infinite-horizon problems, including value iteration, policy iteration, or modified policy iteration [13, 89].

An MDP can also be converted to an equivalent linear program and solved using standard linear programming techniques [28, 32, 75]. This solution method also has advantages from a modeling perspective in that it allows for the incorporation of constraints [5, 52]. Simulation can be used to determine suboptimal decision policies, particularly for the case of Semi-Markov models in which the underlying stochastic process cannot be characterized by the Markov chain alone [49].

The Markov decision process extends commonly used methodologies such as Markov modeling and simulation in that it can be used to compute optimal policies rather the merely evaluate a prespecified policy. One potential disadvantage of the MDP is that it assumes that the state of the system is completely observable at each decision point. In many applications in medicine, this is often not the case as the clinician observed available information and then tries to determine the patient's true underlying health state from this observed information. Partially observed Markov decision processes, as discussed in the next section, incorporate a partially observable state space and an observation process to more accurately model these types of situations.

2.3 PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES

The POMDP generalizes the MDP structure in that the patient's state is no longer required to be fully observable. As a result, the basic POMDP model includes an observation process in addition to the discrete-time dynamic and stochastic system and additive reward function that comprise a standard MDP. This observation process relates information that can be readily observed by the decision maker to the true system state through a known probability distribution.

The structure of a POMDP includes the five basic components of an MDP: stages, states (also referred to as *core states*), actions, transition probabilities, and rewards. In addition, a POMDP includes *observed states* (or *information states*), which describe the information about a patient that a clinician can directly observe (such as test results); a *belief vector*, which describes the probability that a patient is in a given core state given the patient's current observed state; and *observation probabilities*, which relate the observation states to the core states. The observation probabilities in the context of the models presented in this dissertation can also be described as test error (or test interpretation error).

Recalling the notation from the previous section, N represents the time-horizon and S represents the defined *core* state space of the POMDP. A key difference between the MDP and the POMDP is that the core state s_t cannot be directly observed by the decision maker.

Define O as the set of *observed* states. The probability of observing state $o_{t+1} \in O$ given that the underlying core state is s_t, $z_t(o_{t+1}|s_t, a_t)$, depends both on the probabilistic relationship between o_{t+1} and s_t, called the observation probability, as well as the action chosen. Since the observation process may have an associated cost or error, the decision maker may not choose to make an observation in every stage. As a result, let $\pi_t(s_t)$ denote the decision-maker's belief that the patient is in core state s_t at stage t. Let π_t denote the belief vector that defines the probability distribution over all $s_t \in S$, where Π defines the set of all possible belief vectors.

Using an initial estimate of the probability distribution over the true core states (called the *prior distribution*), the current observation of the patient, knowledge of the last action taken, and a distribution for the observation error (test error) if an observation (test result) was just received, Bayesian updating is used to form new estimates of the core states (called the *posterior distribution*). In other words, the following updating function is used to update $\pi_t(s_t)$ to $\pi_{t+1}(s_{t+1})$:

$$U(\pi_{t+1}(s_{t+1})|o_{t+1}, a_t, \pi_t(s_t)) \equiv \pi_{t+1}(s_{t+1})$$
$$= \frac{z(o_{t+1}|s_{t+1}, a_t) \sum_{s_t \in S} p_t(s_{t+1}|s_t, a_t)\pi_t(s_t)}{\sum_{s_{t+1} \in O} z(o_{t+1}|s_{t+1}, a_t) \sum_{s_t \in S} p_t(s_{t+1}|s_t, a_t)\pi_t(s_t)}.$$

It has been established that π_t summarizes all of the information necessary for making a decision as stage t [13, 106]. For every stage t and belief vector $\pi_t \in \Pi$, let the set of feasible decisions or actions be $A(\pi_t)$. Note that the MDP in which the core state was completely observable was defined on a finite state space. Since the core process must now be observed through an observation process, the POMDP can be defined as an equivalent MDP on an uncountable state space defined by the set of all possible belief vectors.

For notational convenience, define the probability of receiving observation o_{t+1} at time $t+1$ given that the belief vector was π_t at time t as $\gamma_t(o_{t+1}|\pi_t, a_t)$, where

$$\gamma_t(o_{t+1}|\pi_t(s_t), a_t) = \sum_{s_{t+1} \in O} z(o_{t+1}|s_{t+1}, a_t) \sum_{s_t \in S} p_t(s_{t+1}|s_t, a_t)\pi_t(s_t).$$

For every action $a_t \in A(\pi_t)$ define a reward function $r_t(s_t, a_t)$ that describes the immediate reward received when action a_t is taken at time t when the system is in core state

s_t. The resulting reward function $r_t(\pi_t(s_t), a_t)$ can then be calculated as $r_t(\pi_t(s_t), a_t) = \sum_{s_t \in S} r_t(s_t, a_t)\pi_t(s_t)$.

As in the MDP, let a policy $d = \{d_1, d_2, \ldots, d_N\}$ be a sequence of decision rules, where a decision rule d_t is a function mapping states into actions at stage t such that $d_t(\pi_t) \in A(\pi_t)$. The application of such a policy induces a Markov chain where X_t is the state of the system at stage t and Y_t is the action chosen in state X_t, so that $Y_t = d_t(X_t)$.

Let $V_t^d(\pi_t)$ represent the total expected reward when policy d is used and the system starts in belief state π_t where, $V_t^d(\pi_t) = \sum_{t=1}^{N} r_t(X_t, Y_t)$. Let $V_t^*(\pi_t)$ denote the optimal total expected reward for each state $\pi_t \in \Pi$ where,

$$V_t^*(\pi_t) = \max_d V_t^d(\pi_t) \qquad \text{for all } \pi_t \in \Pi.$$

As shown in [82, 103], $V_t^*(\pi_t)$ satisfies the following recursion:

$$V_N^*(\pi_N) = r_N(\pi_N),$$
$$V_t^*(\pi_t) = \max \left\{ r_t(\pi_t, a_t) + \sum_{o_{t+1} \in O} \gamma_t(o_{t+1}|\pi_t(s_t), a_t) V_{t+1}^*[U(\pi_{t+1}|o_{t+1}, a_t, \pi_t(s_t))]. \right\}$$

for all $\pi_t \in [0, 1]$.

POMDP models present a great challenge as they are in general more difficult to analyze than their MDP counterparts and their data requirements are significantly greater. The optimal policies of the more general models typically lack structure and the added uncertainty in the problem due to the incorporation of partial observability results in additional computational difficulties.

2.3.1 POMDP structural results from the literature

Monahan [81] considers partial observations and presents a general model of the optimal stopping problem where complete information can be purchased by testing. Unlike the completely observable optimal stopping problem referred to in Section 2.2.1, Monahan is able to demonstrate through an example that the optimal policy for this case may not have special structure. In particular, he shows that control-limit policies may not exist, the set of states for which it is optimal to purchase information (test) may not be convex, and

the optimal policies may not be monotone. In previous work, Monahan [79, 80] shows the optimal policy to be well structured for specific cases of the finite horizon partially observable stopping problem such as under conditions of perfect observability and complete uncertainty.

Monahan [82] explores the effects of partial observability on the optimal solution. Chapter 5 expands his results by exploring the effects of test cost and accuracy on the optimal decisions of when to test for cytokine levels and when to discharge the patient from the hospital.

In a more general paper, Smallwood and Sondik [103] show that for any finite-horizon POMDP, the optimal value function, $V_n(\pi)$, is piecewise linear and convex. Define N as the total number of decision periods and n as the number of decision periods remaining. Let π_i be the probability that the current internal state of the system is i, where the belief vector is $\pi = [\pi_1, \ldots, \pi_N]$.

Theorem 2.3.1 (Smallwood and Sondik (1973)) *$V_n(\pi)$ is piecewise-linear and convex, and can thus be written as*

$$V_n(\pi) = \max_k \left[\sum_{i=1}^{N} \alpha_i^k(n) \pi_i \right]$$

for some set of vectors $\alpha^{k(n)} = [\alpha_1^{k(n)}, \ldots, \alpha_N^{k(n)}]$, $k = 1, 2, \ldots$.

This result demonstrates that the state space can be partitioned into a finite number of convex regions within which the value function is linear. Smallwood and Sondik use this partitioning in an algorithm for solving finite-horizon POMDPs as will be discussed in the following section.

2.3.2 POMDP solution procedures

POMDPs are in general more difficult to analyze than their MDP counterparts. The standard approach to solving POMDPs involves transforming the POMDP into an equivalent, fully observable MDP, over all possible probability distributions on the original core state space [9, 105]. The continuous state space of the resulting MDP is computationally difficult to handle, resulting in limited and complicated solution algorithms [72].

Feasible numerical methods often involve reducing the infinite number of possible states of the system to a finite grid of points [71]. As stated previously, Smallwood and Sondik [103] show that for a finite-horizon problem, the optimal value function is piecewise linear and convex. Thus, the state space can be partitioned into a finite number of convex regions within each of which the value function is linear. Unfortunately, these regions must be reconstructed at each iteration of their proposed algorithm. Not only is the additional computational effort significant, but the number of regions necessary for an exact solution can grow exponentially with time.

Sondik's One-Pass algorithm [103] is the basis for the majority of the algorithms in the artificial intelligence literature to-date. The algorithm proceeds from an arbitrary belief point, constructs a set of vectors that describe the optimal value function based on that point, and then determines the set of constraints over the belief space where that vector is guaranteed to be dominant. Linear programming is then used to define points for the next iteration of the algorithm as it proceeds to calculate the value function for the next stage. The number of vectors generated to describe the value function from one stage to the next can become prohibitively large as the state space and time horizon increase. In addition, the data structures used to store the vector information can be cumbersome.

More recent algorithms have explored dominance criteria and pruning mechanisms to reduce the number of vectors needed to completely describe the value function. Examples include Cheng's Linear Support algorithm [18] and Littman et al.'s Witness algorithm [18]. For an in-depth review of these and other POMDP solution procedures, the reader is referred to Cassandra [18]. While the artificial intelligence community continues to research efficient algorithmic procedures, current algorithms require an in-depth understanding of data structures and computer programming for successful implementation.

The overwhelming computational burden associated with solving POMDPs demonstrated Sondik's One-Pass algorithm and more recent algorithms [106, 118] precludes their application to problems of practical size [84]. As a result, many heuristics and approximation methods for solving POMDPs have been proposed [18, 72]. For example, Lovejoy [71] proposes an approximation method that applies a bounding procedure to the Smallwood and Sondik algorithm, allowing for the solution of larger problems. Platzman [88] proposes an

approximation method where decisions are made based on a finite memory of the most recent decisions and observations. Yet, he notes that this memory can grow prohibitively large before the approximation is acceptable. Kakalik [59] and Eckles [35] discuss using an approximate value function based upon linear interpolation between fixed, discrete points in the continuous state space, but do not present results. Sondik [105] later provides bounds for Eckles' method.

A full-scale implementation and comparison of existing complex algorithmic methods is outside the scope of this dissertation. Rather, a heuristic approach is proposed and used to solve the POMDP model presented in Chapter 5. This heuristic takes advantage of the fact that the POMDP models presented in this dissertation are formulated as MDPs with a continuous state space. The heuristic first discretizes the belief vector and then incorporates the updating function into the standard backward induction algorithm as described in Appendix C. One area for future research would be to further explore existing exact and approximate solution procedures in conjunction with researchers from the artificial intelligence community, utilizing the available GenIMS data.

2.4 OPTIMAL STOPPING PROBLEMS

Both the MDPs and the POMDPs described in this dissertation can be formulated as *optimal stopping problems* [31]. In an optimal stopping problem, a decision maker views rewards sequentially at discrete points in time. In the MDP, the decision maker can either accept the reward (i.e., choose to discharge the patient and "receive" the patient's current expected survival) or reject the reward (i.e., keep the patient in the hospital for one more time period). In the POMDP, before accepting or rejecting the current reward, the decision maker may purchase information regarding its true value. Once the gathering of information regarding the current reward has ended, the decision maker must either decide to accept the current expected reward, thus ending the decision process, or to reject the current expected reward, pay a fixed continuation cost (which in the models presented in this dissertation could represent the additional day of survival in the hospital as well as the costs of care, the

risk of nosocomial infection, or even the patient's quality of life), and then imperfectly view the expected reward in the next time period. In either the MDP or the POMDP case, the decision makers face the same problem at each decision point. The objective of the optimal stopping problem is to determine a decision rule that indicates which action to take (continue or stop), based on the information available at each decision point to maximize the total expected reward.

A classic optimal stopping problem is the *secretary problem* [43, 44]. In its simplest form, the secretary problem has the following features [44]. There is a single secretarial position available and the total number of applicants for this position, n, is known. The applicants are interviewed sequentially in a random order. These applicants can be ranked in order from best to worst, without ties, but the decision to accept or reject the current candidate can only be made based on the relative ranks of the candidates viewed so far.

The decision maker wants to choose the very best applicant for the position, but once an applicant is rejected, she cannot later be accepted. It has been shown that the optimal policy for this case is of the following form. For large n, it is optimal to wait until approximately 37% of the applicants have been interviewed and then to select the next relatively best candidate [44, 46]. Therefore, if the best candidate was in the first 37%, then the optimal solution will be to choose the n^{th} candidate.

There are many applications of this problem. It was first proposed by Cayley [19] in the context of determining an optimal policy for playing the lottery. Other applications include hypothesis testing [8, 113, 114], asset selling [13, 61, 89], and purchasing call options [89]. Considering the decision problem of when to discharge a patient from the hospital can also be considered in the context of the secretary problem. In this case, there are a fixed number of days on which the patient can be released from the hospital. Associated with each of these days is a stochastic reward. The decision maker sequentially views these rewards and must choose to accept them (discharge the patient) or reject them (keep the patient in the hospital). Once the decision maker rejects a reward, the decision maker cannot later accept it (i.e., the decision maker cannot go back and release a patient on a previous day).

Extensions of the simple secretary problem include considering correlations between the ranks of sequential candidates [13], retaining the option to accept past candidates [13, 39],

and the incorporation of partial information [79, 80, 81, 87, 97, 109, 110]. This last case is of particular interest as the clinician's problem can also be formulated as a POMDP.

Of the literature reviewed, the most relevant case is that presented by Monahan [81] who considers partial observations and presents a general model of the optimal stopping problem where complete information can be purchased by testing. His model assumes that the time required to perform the test procedure is instantaneous and that multiple tests can be performed before making a decision at any decision point. The POMDP models presented in this dissertation consider not only which tests to order, but when they should be ordered.

Unlike in the completely observable optimal stopping problem, Monahan is able to demonstrate through an example that the optimal policy for this case may not have special structure: control-limit policies may not exist, the set of states at which it is optimal to purchase information may not be convex, and the optimal policies may not be monotone. However, Monahan has also shown the optimal policy to be well structured for specific cases of the partially observable stopping problem, such as under conditions of perfect observability and complete uncertainty [79, 80].

2.5 MACHINE MAINTENANCE AND REPLACEMENT PROBLEMS

The POMDP models formulated to address the management of severe sepsis have some similarities to problems found in the machine maintenance literature; however, the differences between the problem structures are also readily apparent.

The general form of the machine maintenance problem can be described as follows [36, 64, 94]. A production process produces items at the beginning of distinct time periods. It is supposed that at any time point the production process may be in any one of a countable number of states and that the quality of the item produced is a function of this underlying state. It is also supposed that the state of the process at any time point is not known and can only be determined by sampling the item produced. A cost is associated with sampling the item. This cost may be a function of the current state. The purpose of sampling is not to replace a poor item with a good one, but rather to check the manufacturing process.

At each time point, the decision maker selects an action from the set of possible actions including: replacement, repair, sampling (inspection), and do nothing. The decision maker would like to know the decision rule that minimizes total expected cost.

While the machine maintenance problem includes elements of partial observability where testing (inspection) must be performed to gain information about the system, which is similar to the sepsis management POMDP, other aspects of the problem greatly differentiate it from this model. Most importantly, machine maintenance typically deals with a system that is *deteriorating*. In contrast, the clinician is maintaining a system (patient health) that may be deteriorating, improving, or staying the same. The set of possible actions also differs in that the clinician does not consider repair (as the models to be described only consider testing and discharge decisions at this time, not treatment decisions).

Derman [30] does not consider repair and demonstrates optimal inspection schedules for equipment whose life is of a random length. However, the assumed deterioration of the system plays a key role. In addition, replacement of the equipment returns the system to its original new state. It is difficult to draw parallels between the replacement action and the patient discharge action because the problems have differing objectives. The objective of the models present in this dissertation is to maximize total expected life for an individual patient over a finite horizon while the machine maintenance problem considers minimizing the total cost of a production system over an infinite horizon.

3.0 MODELING TESTING AND HOSPITAL DISCHARGE DECISIONS (GENERAL MODEL)

As presented in the problem description in Chapter 1, the general model considers a patient that at some point at or after hospital admission is suspected of having developed sepsis. This model assumes that this patient is treated according to standard care while simultaneously considering a subset of decisions made by the clinician in the treatment of this patient.

More specifically, the general model considers the clinician's decision problem of when to test for additional information about a patient's health state in addition to the decision problem of when to discharge the patient from the hospital in order to maximize the patient's expected survival over a finite horizon as measured from hospital admission. The model is formulated as a finite-horizon POMDP, where the patient's health state is characterized by two vectors pertaining to completely and partially observable health state information. It is assumed that throughout the patient's stay in the hospital, the patient is being treated according to standard methods of care.

At each stage before a decision is made, the clinician first observes the patient's completely observable health state and the results of any tests that were ordered in the previous stage. Based on these observations, the clinician then either decides to discharge the patient from the hospital or to keep the patient in the hospital for one more stage. If the clinician decides to keep the patient in the hospital, then the clinician must also decide whether or not to order any tests to obtain additional information about the patient's partially observable health state variables. If a set of tests are ordered, their results are not known until the beginning of the next time period.

The following model formulation can be used to determine the optimal action and resulting expected survival over the remainder of the specified finite observation horizon for

a patient at any stage in the patient's treatment, based on the patient's observable health variables and the clinician's belief as to what the patient's partially observable health state variables are based on previous test results and the patient's current completely observable health state. First, the following assumptions are made for this general model and all model variants presented in this dissertation.

3.1 ASSUMPTIONS

Assumption 3.1.1 ***Markov property****: It is assumed that the patient's health variables are Markovian in that their values at a specific decision epoch only depend on the patient's state and the action taken at the previous decision epoch.*

Future research could consider expanding the state description to include additional history.

Assumption 3.1.2 ***Finite horizon****: A finite-horizon model is utilized to facilitate the incorporation of time-dependent factors into the model. The model incorporates both a decision horizon and an observation horizon. The decision horizon is used to reflect the short treatment horizon associated with acute diseases like sepsis. The observation horizon considers the time post-discharge during which death can be attributed to sepsis. Both of these horizons are demonstrably finite and short.*

Assumption 3.1.3 ***Finite, discrete state space****: All patient health variables can be represented as finite, discrete values.*

Although many variables may appear to be continuous, they are often discretized naturally in practice as a result of the measurement techniques used to assess their value.

Assumption 3.1.4 ***Testing delay****: It is assumed that all test results are received at the beginning of the stage immediately following the stage when the test(s) were ordered. Therefore, if one or more tests are ordered, the patient cannot be discharged until the next time period or later.*

Assumption 3.1.5 ***Test independence****: It is assumed that the results of a test for a particular health variable are only dependent on the true value of that health variable and are independent of the true and observed values of all other health variables.*

3.2 GENERAL MODEL FORMULATION

The following notation is used:

- $\mathcal{N} = \{1, 2, \ldots, N\}$: discrete stages at which a decision must be made by the clinician. If a patient has not died and has not been discharged by stage N, it is assumed that the patient is discharged at stage N. This dissertation defines a stage as one day; however, the model is flexible enough to consider smaller time intervals (hours, for example) as the data for solving such a model become available. Let t denote the current stage in the model.
- T: the observation horizon, as measured from admission to the hospital, in which a patient's death is attributable to sepsis.
- h_t: a vector describing the completely observable components of the patient's health at stage t. Let $\mathcal{H}$ be the set of all possible realizations of h_t. The ordered elements of $\mathcal{H}$ are represented as $1, 2, \ldots, H, H+1$, where $H+1$ represents the patient being dead and is an absorbing state.
- y_t: a vector describing the true values of the partially observable components of the patient's health at stage t. Let $\mathcal{Y}$ be the set of all possible realizations of y_t. The ordered elements of $\mathcal{Y}$ are represented as $1, 2, \ldots, Y$.
- δ_t: the set of tests ordered at time t. Let Δ be the set of all possible combinations of available tests, including $\emptyset$.
- $c(\delta_t)$: a scalar representing the cost of ordering test set δ_t at stage t (converted to patient life days using methods from cost-effectiveness analysis [47]).
- o_{t+1}: a vector describing the observed values of the partially observable components of the patient's health at stage $t+1$ for tests ordered in stage t. Let $\mathcal{O}$ be the set of all

possible realizations of o_{t+1}. The ordered elements of $\mathcal{O}$ are represented as $1, 2, \ldots, O$. If no tests were ordered in stage t, then $o_{t+1} = \emptyset$.

- a_t: the action chosen at stage t. Possible actions are to discharge the patient from the hospital (D) or to continue treating the patient in the hospital according to standard methods of care and order test set $\delta_t \in \Delta$ (C_{δ_t}). Note that the option to order no tests ($\emptyset$) is contained in Δ. Recall the assumption that when tests are ordered, the test results are received at the beginning of the next period before the next decision is made. Therefore, the patient cannot be discharged before the next period.
- $f_t(h_t, y_t, D)$: the expected $(T-t)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage t with completely observable health vector h_t and partial observable health vector y_t.
- $r_t(h_t, \pi_t, D)$: the expected $(T-t)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage t with completely observable health vector h_t and belief vector π_t, where $r_t(h_t, \pi_t, D) = \sum_{y_t \in \mathcal{Y}} f_t(h_t, y_t, D)\pi_t(y_t)$.
- $f_t(h_t, y_t, C_{\delta_t})$: the expected reward (in patient life days) received for keeping a patient with completely observable health vector h_t and partially observable health vector y_t at time t in the hospital for one more stage and ordering test set δ_t.
- $r_t(h_t, \pi_t, C_{\delta_t})$: the expected reward (in patient life days) received for keeping a patient with completely observable health vector h_t and belief vector π_t at time t in the hospital for one more stage and ordering test set δ_t, where $r_t(h_t, \pi_t, C_{\delta_t}) = \sum_{y_t \in \mathcal{Y}} f_t(h_t, y_t, C_{\delta_t})\pi_t(y_t)$.
- $f_N(h_N, y_N)$: the expected $(T-N)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage N with with completely observable health vector h_N and partially observable health vector y_N.
- $r_N(h_N, \pi_N)$: the expected $(T-N)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage N with with completely observable health vector h_N and belief vector π_N, where $r_t(h_N, \pi_N) = \sum_{y_N \in \mathcal{Y}} f_N(h_N, y_N)\pi_N(y_N)$.
- $p_t(h_{t+1}, y_{t+1} | h_t, y_t, a_t)$: the joint probability that the true values of the patient's completely observable and partial observable heath vectors are h_{t+1} and y_{t+1}, respectively, at stage $t+1$ given that their respective values were h_t and y_t and action a_t was chosen at stage t.

- $z(o_{t+1}|y_{t+1}, a_t)$: the probability of observing test results o_{t+1} at stage $t+1$ given that the patient's true partially observable health vector is y_{t+1} and action a_t was chosen at stage t.
- π_t: the belief vector. Let Π_t denote all possible realizations of π_t, where $\Pi_1 = \Pi_2 = \ldots = \Pi_{\mathcal{N}} = \Pi$. Let $\pi_t(y_t)$ denote the component of the belief vector corresponding to the probability that the patient's true partially observable health vector is $y_t \in \mathcal{Y}$ at stage t.
- $\beta_t(h_{t+1}|h_t, \pi_t, a_t)$: the probability of observing completely observable health state h_{t+1} at time $t+1$ given that at time t, the patient's completely observable health state was h_t, the belief vector was π_t, and action a_t was chosen, where

$$\beta_t(h_{t+1}|h_t, \pi_t, a_t) = \sum_{y_{t+1} \in \mathcal{Y}} \sum_{y_t \in \mathcal{Y}} p_t(h_{t+1}, y_{t+1}|h_t, y_t, a_t)\pi_t(y_t). \tag{3.1}$$

- $\gamma_t(o_{t+1}|h_t, \pi_t, a_t)$: the probability of receiving observation vector o_{t+1} at time $t+1$ given that the patient's true completely observable health state was h_t, the belief vector was π_t, and action a_t was chosen at time t, where

$$\gamma_t(o_{t+1}|h_t, \pi_t, a_t) = \sum_{y_{t+1} \in \mathcal{Y}} z(o_{t+1}|y_{t+1}, a_t) \sum_{h_{t+1} \in \mathcal{H}} \sum_{h_t \in \mathcal{H}} \sum_{y_t \in \mathcal{Y}} p_t(h_{t+1}, y_{t+1}|h_t, y_t, a_t)\pi_t(y_t). \tag{3.2}$$

- $U(\pi_{t+1}|o_{t+1}, h_{t+1}, \pi_t, a_t)$: the updating function used to update the belief vector π_{t+1} based on o_{t+1}, the observation vector at stage $t+1$, h_{t+1}, the patient's completely observable health vector at stage $t+1$, π_t, the belief vector at stage t, and a_t, the action chosen at time t. Let $U(\pi_{t+1}(y_{t+1})|o_{t+1}, h_{t+1}, \pi_t, a_t)$ denote the updating function used to update component y_{t+1} of the belief vector, where

$$U(\pi_{t+1}(y_{t+1})|o_{t+1}, h_{t+1}, \pi_t, a_t) \tag{3.3}$$
$$= \left\{ \frac{z(o_{t+1}|y_{t+1}, a_t) \sum_{h_{t+1} \in \mathcal{H}} \sum_{h_t \in \mathcal{H}} \sum_{y_t \in \mathcal{Y}} p_t(h_{t+1}, y_{t+1}|h_t, y_t, a_t)\pi_t(y_t)}{\gamma_t(o_{t+1}|h_t, \pi_t, a_t)}. \right.$$

- $V_t(h_t, \pi_t)$: the value function used to calculate the total expected reward (in patient life days) at stage t when the patient's truly observable health state is h_t and the belief vector is π_t, where $V_t^*(h_t, \pi_t)$ denotes the optimal value function value.

- $\mathcal{A}_t^*(h_t, \pi_t)$: the set of optimal actions at stage t when the patient's truly observable health state is h_t and the belief vector is π_t, where $a_t^*(h_t, \pi_t) \in \mathcal{A}_t^*(h_t, \pi_t)$ is an action that maximizes the value function $V_t(h_t, \pi_t)$.

After action a_t is taken at stage t, an immediate expected reward $r_t(h_t, \pi_t, a_t)$ is received. If $a_t = D$, the patient is discharged from the hospital and receives an expected reward, $r_t(h_t, \pi_t, D)$. If $a_t = C_{\delta_t}$, the patient remains in the hospital and receives an expected reward $r_t(h_t, \pi_t, C_{\delta_t})$. At the same time the clinician orders test set δ_t, the results of which will be received at the beginning of the next stage. After receiving the immediate expected reward, the patient's completely and partially observable health vectors transition to new values. At the beginning of the next stage, if the patient has not died then the patient's completely observable health state and the results of any test ordered at stage t are observed by the clinician.

Based on this information, each component of the belief vector, π_t, is updated from stage t to stage $t+1$ using the updating function, $U(\pi_{t+1}|o_{t+1}, h_{t+1}, \pi_t, a_t)$ as described in (3.3). Note that if one or more tests were ordered at stage t, then the resulting observation vector, o_{t+1}, which denotes the values of the test results received at the beginning of stage $t+1$, are used when updating the belief vector π_t to π_{t+1}, *before* a decision is made in stage $t+1$.

Let the optimal value function, $V_t^*(h_t, \pi_t)$, be the total expected reward for time t onward for a patient with completely observable health vector h_t and belief vector π_t at time t. $V_t^*(h_t, \pi_t)$ can then be defined recursively as follows.

$$V_N^*(h_N, \pi_N) = r_N(h_N, \pi_N), \text{ for all } \pi_N \in \Pi \text{ and } h_N = 1, \ldots, H, \tag{3.4}$$

$$V_t^*(h_t, \pi_t) = \max \begin{cases} r_t(h_t, \pi_t, D), \\ r_t(h_t, \pi_t, C_{\delta_t}) - c(\delta_t) + \sum_{h_{t+1} \in \mathcal{H}} \beta_t(h_{t+1}|h_t, \pi_t, a_t) \\ \quad \cdot \sum_{o_{t+1} \in \mathcal{O}} \gamma_t(o_{t+1}|h_t, \pi_t, a_t) V_{t+1}^*(h_{t+1}, U(\pi_{t+1}|h_{t+1}, o_{t+1}, \pi_t, a_t)), \end{cases}$$

for all $\pi_t \in \Pi$, $h_t = 1, \ldots, H$, and $t = 1, \ldots, N-1$, and

$$V_t^*(H+1, \pi_t) = 0, \text{ for all } \pi_t \in \Pi \text{ and } t \in \mathcal{N}. \tag{3.5}$$

This model is the most general problem considered in this dissertation. However, due to the data requirements needed to solve even a modestly sized MDP or POMDP, simplified

variants of this general model were developed to further explore both structural properties of the model and computation experiments using patient-based data from the GenIMS study. These variants are described in the following section.

3.3 SIMPLIFIED MODEL VARIANTS

Chapter 4 considers the most simplified variant of the general model in which the patient's health state is described by a single, completely observable variable, the patient's Sepsis-related Organ Failure Assessment (SOFA) score. This score considers many aspects of the patient's health state and is therefore an appropriate measure of completely observable patient health. This model is general enough to consider any single measure of patient health in future research.

Chapter 5 extends Chapter 4 by considering a partially observable patient health state. Similar to the previous chapter, however, the patient health state is still confined to a single health variable, the value of a single cytokine level. Since the interactions between cytokines are still under investigation and have not been completely analyzed for statistical dependencies, it is not possible at this time to develop a more sophisticated POMDP model utilizing multiple cytokine values. Also, due to the extensive data requirements needed to calculate joint probability distributions for a multi-state model, exact solutions to a SOFA- and cytokine-based model were not explored as part of this dissertation. Chapter 5 does, however, explore heuristic approaches to developing SOFA- and cytokine-base decision policies.

Note that the MDP model presented in Chapter 4 is general enough to also consider a cytokine-based MDP model. This is not explored in this dissertation, however, because it will not provide clinically useful results. Similarly, while the POMDP model presented in Chapter 5 is general enough to consider a SOFA-based POMDP, this model was not considered because of its lack of clinical relevance.

3.4 MODELING TEST ACCURACY

The general model and the simplified POMDP model presented in Chapter 5 utilize an observation process that updates the clinician's belief of the patient's true health state based on the values of one or more test results.

In both models, the clinician does not know the patient's true health state with complete certainty for at least one of two reasons. First, if the clinician decides not to order tests in a given period, then the clinician will not have any information on the patient's partially observable health state, causing uncertainty. Second, even if the clinician orders a test and observes its result, there may be error associated with the result.

Test error can be interpreted in two different ways. First, the result itself may be inaccurate from a measurement standpoint. Accuracy in this case refers to the sensitivity and specificity of the test. Alternatively, the interpretation of the test results may be inaccurate. In other words, even if the test result is numerically correct, the interpretation of the numerical result may not be completely accurate.

Since a test result may not be received in every stage and the test results that are received may either be inaccurate themselves or interpreted inaccurately, the models use a belief vector to describe a probability distribution over the possible true patient health states, which corresponds the the clinician's belief that the patient's true health is in each of the possible states. The model utilizes an observation probability matrix to relate the observed values to the underlying health state. Then, using an initial estimate of the belief variable, the current test results, and knowledge of the last action taken, Bayesian updating is used to form a new estimate of the belief vector. The clinician's decision is made based on the value of this belief vector at each decision point.

The model presented in Chapter 5 considers the accuracy of cytokine testing from the measurement perspective. Therefore, test error refers only to the inaccuracy between receiving a test result and it relationship to the true value of the patient's cytokine level. A more accurate model of testing inaccuracy would consider the clinician's interpretation error in terms of translating the actual cytokine level to the underlying patient health state. Unfortunately, the true relationship between these cytokine levels and the patient's true

underlying health state is not understood well enough to model this relationship, which is needed in order to incorporate interpretation error into the model. As these relationships are better understood, future research could consider interpretation error as an extension of the models presented in this dissertation.

3.5 QUANTIFYING TESTING COST

Testing for cytokines has not yet been quantified in the literature as this is a new procedure that has not yet been incorporated into standard practice. The cost of testing should include not only the cost of the materials needed to perform the test and process the test results, but also the time spent administering the test and reviewing and interpreting the results. One could even consider the quality of life implications associated with frequent testing.

The focus of this dissertation is to investigate general testing and hospital discharge strategies. Since testing cost is an important factor in these decisions, but has not yet been quantified, the computational experiments presented in Section 5.4 provide results for a range of testing costs. The reward function is calculated in units of patient life days; therefore, the testing cost utilized in the problem instances is also represented in patient life days. To convert actual dollars to life days, one could use a standard dollars-to-life days conversion rate, as is commonly done in cost-effective analysis [48].

4.0 MODELING HOSPITAL DISCHARGE POLICIES WITHOUT CONSIDERING TESTING DECISIONS (A MARKOV DECISION PROCESS APPROACH)

A simplified variant of the general model presented in Chapter 3, this model considers the clinician's decision problem of when to discharge an individual patient from the hospital in order to maximize that patient's expected survival over a finite observation horizon as measured from hospital admission, where common values of this observation horizon include 30, 60, and 90 days [27, 90, 116]. The problem is formulated as a finite-horizon Markov decision process to capture time dependencies among state transitions and rewards. In addition to only considering the discharge decision, this model also assumes that a single measure of patient health characterizes the health state and standard methods of care guide patient treatment throughout the patient's hospital stay. Therefore, at each decision point, the clinician can choose either to continue treating the patient in the hospital with standard care or to discharge the patient from the hospital. It is assumed that decisions are made at the end of each time period. The work presented in this Chapter has been submitted for publication [65].

4.1 MDP NOTATION

The following notation is used:

- $\mathcal{N} = \{1, 2, \ldots, N\}$: discrete stages at which a decision is made by the clinician, where N is the treatment horizon. If a patient has not died and has not been discharged by

stage N, it is assumed the patient is discharged at stage N. The finite-horizon model captures time dependencies among state transitions, where the value of N depends on the input data for computational experiments. This dissertation defines a stage as one day; however, as the data for solving such a model become available, the model is flexible enough to consider smaller time intervals (hours, for example).

- T: the observation horizon, as measured from admission to the hospital, in which a patient's death is attributable to sepsis.
- h: the patient's health state vector. Let $\mathcal{H}$ be the set of all possible realizations of h in order of decreasing health. The ordered elements of $\mathcal{H}$ are represented as $1, 2, \ldots, H, H+1$, where $H+1$ represents the patient being dead and is an absorbing state.
- a_t: the action taken at time t. The possible actions are to continue treating the patient in the hospital (C) and to discharge the patient from the hospital (D).
- $r_t(h, D)$: the expected $(T-t)$-day survival (in patient life days) of a patient that is discharged from the hospital on day t in health state h.
- $r_t(h, C)$: the expected reward (in patient life days) received for deciding at stage t to keep a patient in health state h in the hospital for one more stage. This model uses an expected reward for continuing of one day for all stages and states.
- $r_N(h)$: the expected $(T-N)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage N in health state h.
- $p_t(j|h, a)$: the probability that the patient's health state is j in stage $t+1$ given that the patient's health state is h in stage t and action a is chosen. Note that the process will terminate with reward $r_t(H+1, C) = 0$ if a patient transitions to the dead state or with reward $r_t(h, D)$ if action D is chosen.
- $V_t(h)$: the value function used to calculate the total expected reward (in patient life days) for stage t onward when the system is in state h, where $V_t^*(h)$ denotes the optimal value function value. Thus, $V_1^*(h)$ will be the optimal total expected $T-1$-day survival (in days) of a patient that has just been admitted to the hospital in state h.
- $\mathcal{A}_t^*(h)$: the set of optimal actions at stage t when the system is in state h, where $a_t^*(h) \in \mathcal{A}_t^*(h)$ is an action that maximizes the value function $V_t(h)$.

4.2 MDP MODEL FORMULATION

This problem can be formulated as the following optimality equations:

$$V_t(h) = \max\left\{ r_t(h, D),\ r_t(h, C) + \sum_{j=1}^{H+1} p_t(j|h, C) V_{t+1}(j) \right\} \tag{4.1}$$

$$\text{for } h = 1, \ldots, H \text{ and } t = 1, \ldots, N-1,$$

$$V_N(h) = r_N(h), \quad \text{for } h = 1, \ldots, H, \tag{4.2}$$

$$V_t(H+1) = 0, \quad \text{for } t = 1, \ldots, N. \tag{4.3}$$

The next section discusses the structure of the optimal value function in addition to presenting clinical conditions under which optimal non-stationary control-limit policies exist.

4.3 ANALYZING OPTIMAL HOSPITAL DISCHARGE POLICY STRUCTURE FOR THE MDP MODEL

The mathematical framework of the MDP model allows for the analysis of the structure of the model parameters and its optimal solution, providing insight into results that can be expected in practice. For example, this section demonstrates the monotonicity of the value function as formulated in (4.1) through (4.3). It will be demonstrated that as a patient becomes sicker, the patient's $(T-t)$-day expected survival does not increase. Conditions for the existence of an optimal non-stationary control-limit policy are also presented. Relevant proofs are included in the Appendix. First, the following definitions are provided.

4.3.1 Definitions

Definition [10] The $N \times N$ transition probability matrix $P(t)$, with entries $[P(t)]_{hj}$, is said to be *IFR* (Increasing Failure Rate) if the rows of $P(t)$ are in increasing stochastic order, that is, $z(h) = \sum_{j=k}^{N} [P(t)]_{hj}$ is monotonically increasing in h for $k = 1, \ldots, N$.

Definition [89] Let X and Y be partially ordered sets and $g(x, y)$ a real-valued function on $X \times Y$. The function g is said to be *superadditive* if for $x^+ \geq x^-$ in X and $y^+ \geq y^-$ in Y,

$$g(x^+, y^+) + g(x^-, y^-) \geq g(x^+, y^-) + g(x^-, y^+).$$

If the reverse inequality holds, the function is said to be *subadditive*.

The following assumptions are later verified in Section 4.4.4 for each of the problem instances presented in Section 4.4.3.

4.3.2 Additional assumptions for the MDP model

Assumption 4.3.1 *The patient health transition probability matrix $P(t)$, with entries $[P(t)]_{hj} = p_t(j|h, C)$, is IFR for all $t \in \mathcal{N}$.*

Assumption 4.3.1 implies that for two patients in health states h and $h+1$, respectively, the patient in health state $h+1$ is more likely to transition to a health state worse than h in the next stage. In other words, sicker patients are more likely to progress to being even sicker than are healthier patients.

Assumption 4.3.2 *The reward function $r_t(h, D)$ is nonnegative and monotone decreasing in h for all $t \in \mathcal{N}$. It follows that the reward function $r_N(h)$ is also nonnegative and monotone decreasing in h since a patient that has not died or been discharged by stage N must be discharged at stage N.*

Assumption 4.3.2 says that sicker patients have worse survival after discharge than healthier patients.

Assumption 4.3.3 *The reward function $r_t(h, C)$ is monotone decreasing in h for all $t \in \mathcal{N}$.*

Assumptions 4.3.2 and 4.3.3 imply that as a patient's health degrades, the value of remaining in the hospital for one additional day and the patient's expected $(T - t)$-day survival after discharge on day t do not increase.

4.3.3 Mathematical structure of optimal hospital discharge policies

Under these assumptions, it can be shown that the optimal value function, $V_t^*(h)$, is monotonically decreasing in h. First, two supporting lemmas are introduced.

Lemma 4.3.4 *(Adapted from Lemma 1 in [3]) Given Assumption 4.3.1 and a function, $V_{t+1}^*(h)$, that is monotone decreasing in h, the following inequalities hold for $h = 1, \ldots, H$ and $t = 1, \ldots, N-1$,*

$$\sum_{j=1}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \geq \sum_{j=1}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h), \tag{4.4}$$

$$\sum_{j=h+1}^{H+1}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \geq \sum_{j=h+1}^{H+1}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h+1). \tag{4.5}$$

Proof For inequality (4.4): Assumption (4.3.1) requires that

$$\sum_{j=0}^{h} p_t(j|h) \geq \sum_{j=0}^{h} p_t(j|h+1)$$

for $h = 0 \ldots, H$. Therefore,

$$\begin{aligned}
&\sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \\
&= [p_t(0|h,C) - p_t(0|h+1,C)]V_{t+1}^*(0) + \sum_{j=1}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \\
&\geq [p_t(0|h,C) - p_t(0|h+1,C)]V_{t+1}^*(1) + \sum_{j=1}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \qquad (4.6) \\
&= [p_t(0|h,C) - p_t(0|h+1,C)]V_{t+1}^*(1) + [p_t(1|h,C) - p_t(1|h+1,C)]V_{t+1}^*(1) \\
&\quad + \sum_{j=2}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \\
&= [p_t(0|h,C) + p_t(1|h,C) - p_t(0|h+1,C) - p_t(1|h+1,C)]V_{t+1}^*(1) \\
&\quad + \sum_{j=2}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \\
&\geq [p_t(0|h,C) + p_t(1|h,C) - p_t(0|h+1,C) - p_t(1|h+1,C]V_{t+1}^*(2) \\
&\quad + \sum_{j=2}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j), \qquad (4.7)
\end{aligned}$$

where (4.6) follows because $p_t(0|h,C) \geq p_t(0|h+1,C)$, by Assumption (4.3.1), and $V^*_{t+1}(0) \geq V^*_{t+1}(1)$, by the initial assumption on $V^*_{t+1}(h)$. Similarly, (4.7) holds because $p_t(0|h,C) + p_t(1|h,C) \geq p_t(0|h+1,C) + p_t(1|h+1,C)$ and $V^*_{t+1}(1) \geq V^*_{t+1}(2)$. The result follows when the same procedure is applied for $j = 2, \ldots, h$.

For inequality (4.5): Assumption (4.3.1) requires that

$$\sum_{j=h+1}^{H+1} p_t(j|h,C) \leq \sum_{j=h+1}^{H+1} p_t(j|h+1,C)$$

for $h = 0, \ldots, H$. Therefore,

$$\sum_{j=h+1}^{H+1} [p_t(j|h,C) - p_t(j|h+1,C)]V^*_{t+1}(j)$$

$$= [p_t(H+1|h,C) - p_t(H+1|h+1,C)]V^*_{t+1}(H+1) + \sum_{j=h+1}^{H} [p_t(j|h,C) - p_t(j|h+1,C)]V^*_{t+1}(j)$$

$$\geq [p_t(H+1|h,C) - p_t(H+1|h+1,C)]V^*_{t+1}(H) + \sum_{j=h+1}^{H} [p_t(j|h,C) - p_t(j|h+1,C)]V^*_{t+1}(j) \quad (4.8)$$

$$= [p_t(H+1|h,C) - p_t(H+1|h+1,C)]V^*_{t+1}(H) + [p_t(H|h,C) - p_t(H|h+1,C)]V^*_{t+1}(H)$$
$$+ \sum_{j=h+1}^{H-1} [p_t(j|h,C) - p_t(j|h+1,C)]V^*_{t+1}(j)$$

$$= [p_t(H+1|h,C) + p_t(H|h,C) - p_t(H+1|h+1,C) - p_t(H|h+1,C)]V^*_{t+1}(H)$$
$$+ \sum_{j=h+1}^{H-1} [p_t(j|h,C) - p_t(j|h+1,C)]V^*_{t+1}(j)$$

$$\geq [p_t(H+1|h,C) + p_t(H|h,C) - p_t(H+1|h+1,C) - p_t(H|h+1,C)]V^*_{t+1}(H-1)$$
$$+ \sum_{j=h+1}^{H-1} [p_t(j|h,C) - p_t(j|h+1,C)]V^*_{t+1}(j), \quad (4.9)$$

where (4.8) follows because $p_t(H+1|h,C) \leq p_t(H+1|h+1,C)$, by Assumption (4.3.1), and $V^*_{t+1}(H) \geq V^*_{t+1}(H+1)$, by the initial assumption on $V^*_{t+1}(h)$. Similarly, (4.9) holds because $p_t(H+1|h,C) + p_t(H|h,C) \leq p_t(H+1|h+1,C) + p_t(H|h+1,C)$ and $V^*_{t+1}(H-1) \geq V^*_{t+1}(H)$. The result follows when the same procedure is applied for $j = h+1, \ldots, H-1$. ∎

Lemma 4.3.5 *(Adapted from Lemma 2 in [3]) Given Assumption 4.3.1 and a function, $V^*_{t+1}(h)$, that is monotone decreasing in h for $t = 1, \ldots, N-1$, $\sum_{j=1}^{H+1} p_t(j|h,C)V^*_{t+1}(j) \geq \sum_{j=1}^{H+1} p_t(j|h+1,C)V^*_{t+1}(j)$ for $h = 1, \ldots, H$ and $t = 1, \ldots, N-1$.*

Proof Note that

$$\sum_{j=0}^{H+1} p_t(j|h,C)V_{t+1}^*(j) - \sum_{j=0}^{H+1} p_t(j|h+1,C)V_{t+1}^*(j)$$

$$= \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) + \sum_{j=h+1}^{H+1}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j)$$

$$\geq \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h) + \sum_{j=h+1}^{H+1}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h+1) \quad (4.10)$$

$$= \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h) + \left[\sum_{j=h+1}^{H+1} p_t(j|h,C) - \sum_{j=h+1}^{H+1} p_t(j|h+1,C)\right] V_{t+1}^*(h+1)$$

$$= \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h)$$

$$+ \left[\left(1 - \sum_{j=0}^{h} p_t(j|h,C)\right) - \left(1 - \sum_{j=0}^{h} p_t(j|h+1,C)\right)\right] V_{t+1}^*(h+1)$$

$$= \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h) - \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(h+1)$$

$$= [V_{t+1}^*(h) - V_{t+1}^*(h+1)] \sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)], \quad (4.11)$$

where (4.10) follows from Lemma 4.3.4. Following from the monotonicity assumption on $V_t^*(h)$ and Assumption (4.3.1), $V_{t+1}^*(h) - V_{t+1}^*(h+1)$ and $\sum_{j=0}^{h}[p_t(j|h,C) - p_t(j|h+1,C)]$ are nonnegative. Therefore, the quantity in (4.11) is also nonnegative and the desired result follows. ∎

Theorem 4.3.6 *Under Assumptions 4.3.1, 4.3.2, and 4.3.3 for* $h = 1, \ldots, H$, $V_t^*(h) \geq V_t^*(h+1)$ *for all* $t \in \mathcal{N}$.

Proof *(By induction)*

From Assumption 4.3.2 it follows that $V_N^*(h) \geq V_N^*(h+1)$ since $V_N^*(h) = r_N(h)$ for all $h \in \mathcal{H}$. Now suppose that $V_n^*(h) \geq V_n^*(h+1)$ for $h = 1, \ldots, H$ and for $n = t+1, \ldots, N-1$. It remains to show that $V_t^*(h) \geq V_t^*(h+1)$ for $h = 1, \ldots, H$. Note that

$$V_t^*(h) = \max\left\{r_t(h,D),\ r_t(h,C) + \sum_{j=1}^{H+1} p_t(j|h,C)V_{t+1}^*(j)\right\}, \text{ and} \quad (4.12)$$

$$V_t^*(h+1) = \max\left\{ r_t(h+1, D),\ r_t(h+1, C) + \sum_{j=1}^{H+1} p_t(j|h+1, C) V_{t+1}^*(j) \right\}. \tag{4.13}$$

If $V_t^*(h+1) = r_t(h+1, D)$, then by definition of $V_t^*(h)$ and Assumption 4.3.2, $V_t^*(h) \geq r_t(h, D) \geq r_t(h+1, D) = V_t^*(h+1)$ and the result follows. Otherwise,

$$V_t^*(h) - V_t^*(h+1) \geq r_t(h, C) - r_t(h+1, C) + \sum_{j=1}^{H+1} p_t(j|h, C) V_{t+1}^*(j) - \sum_{j=1}^{H+1} p_t(j|h+1, C) V_{t+1}^*(j)$$

$$\geq \sum_{j=1}^{H+1} p_t(j|h, C) V_{t+1}^*(j) - \sum_{j=1}^{H+1} p_t(j|h+1, C) V_{t+1}^*(j) \tag{4.14}$$

where (4.3.3) follows from the value functions (4.12) and (4.13) and the inequality (4.14) follows from Assumption 4.3.3. Following from the induction assumptions and Lemma 4.3.5, (4.14) is nonnegative and the desired result follows. ∎

This result demonstrates the intuitive conclusion that as a patient's health degrades, the patient's expected T-day survival does not improve.

In addition to showing that structure exists for the model value function, it is also desirable to extend these results by demonstrating structure for the resulting optimal solution. Of particular interest for this type of model is to demonstrate the existence of a control-limit policy.

Theorem 4.3.7 presents a general condition for the existence of a control-limit policy.

Theorem 4.3.7 *There exists an optimal non-stationary control-limit policy in h if*

$$r_t(h, D) - r_t(h+1, D) \geq r_t(h, C) - r_t(h+1, C) + \sum_{j=1}^{H+1} [p_t(j|h, C) - p_t(j|h+1, C)] V_{t+1}^*(j), \tag{4.15}$$

for $h = 1, \ldots, H$ and $t = 1, \ldots, N-1$. In other words, for each $t \in \mathcal{N}$ there exists a state h_t^* (the control limit) such that $a_t^*(1) = \cdots = a_t^*(h_t^* - 1) = D$ and $a_t^*(h_t^*) = a_t^*(h_t^* + 1) = \cdots = a_t^*(H+1) = C$.

Proof *(By contradiction)*

Recall that

$$V_t^*(h) = \max\left\{r_t(h,D), r_t(h,C) + \sum_{j=1}^{H+1} p_t(j|h,C)V_{t+1}^*(j)\right\}. \tag{4.16}$$

For a fixed value of t, assume that for some h, $a_t^*(h) = C$. Now suppose that $a_t^*(h+1) = D$. This implies that

$$r_t(h,D) \leq r_t(h,C) + \sum_{j=1}^{H+1} p_t(j|h,C)V_{t+1}^*(j) \tag{4.17}$$

and

$$r_t(h+1,D) > r_t(h+1,C) + \sum_{j=1}^{H+1} p_t(j|h+1,C)V_{t+1}^*(j).$$

Therefore,

$$r_t(h,D) - r_t(h+1,D) < r_t(h,C) - r_t(h+1,C) + \sum_{j=1}^{H+1}[p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j), \tag{4.18}$$

which contradicts Condition (4.15). Therefore, $a_t^*(h+1)$ must equal C, completing the proof. ∎

Condition (4.15) can be interpreted as follows: the marginal decrease in a patient's $(T-t)$-day expected survival in sequential health states must be no less than the marginal decrease in a patient's immediate reward received for remaining in the hospital for one more day plus the total expected reward for the remaining time the patient is in the hospital.

Note that Condition (4.15) implies that in order for a control-limit policy to exist, the function

$$w_t(h,a) = r_t(h,a) + \sum_{j=1}^{H+1} p_t(j|h,a)V_{t+1}^*(j),$$

must be superadditive for h and a, where h has the natural ordering $1, 2, \ldots, H+1$ and $C \geq D$. The traditional method of demonstrating that $w_t(h,a)$ is superadditive is to show that both $r_t(h,a)$ and $\sum_{j=1}^{H+1} p_t(j|h,a)V_{t+1}^*(j)$ are superadditive. Proposition 4.3.8 demonstrates that $r_t(h,a)$ is superadditive; unfortunately, Proposition 4.3.9 shows that $\sum_{j=1}^{H+1} p_t(j|h,a)V_{t+1}^*(j)$ is subadditive.

Proposition 4.3.8 *The function, $r_t(h, a)$, is superadditive.*

Proof By Assumption 4.3.2,

$$r_t(h, D) \geq r_t(h + 1, D).$$

As stated in the model formulation, $r_t(h, C) = r_t(h + 1, C) = 1$, therefore,

$$r_t(h, D) + r_t(h + 1, C) \geq r_t(h, C) + r_t(h + 1, D),$$

which is superadditive by Definition 4.3.1 and the assumed ordering of states and actions, thus completing the proof.

Proposition 4.3.9 *The function, $\sum_{j=1}^{H+1} p_t(j|h, a)V_{t+1}^*(j)$, is subadditive.*

Proof By Lemma 4.3.5,

$$\sum_{j=1}^{H+1} p_t(j|h + 1, C)V_{t+1}^*(j) \leq \sum_{j=1}^{H+1} p_t(j|h, C)V_{t+1}^*(j).$$

By definition of the model, the process terminates if the patient is discharged. In other words, $\sum_{j=1}^{H+1} p_t(j|h, D) = \sum_{j=1}^{H+1} p_t(j|h + 1, D) = 0$ for all $h \in \mathcal{H}$. Therefore,

$$\begin{aligned} &\sum_{j=1}^{H+1} p_t(j|h + 1, C)V_{t+1}^*(j) + \sum_{j=1}^{H+1} p_t(j|h, D)V_{t+1}^*(j) \\ \leq &\sum_{j=1}^{H+1} p_t(j|h, C)V_{t+1}^*(j) + \sum_{j=1}^{H+1} p_t(j|h + 1, D)V_{t+1}^*(j), \end{aligned}$$

which is subadditive by Definition 4.3.1 and the assumed ordering of states and actions, thus completing the proof.

Redefining the ordering of the states and/or actions would only reverse the superadditivity and subadditivity of $r_t(h,a)$ and $\sum_{j=1}^{H+1} p_t(j|h,a)V_{t+1}^*(j)$. Since the traditional proof will not work for this model, it is left to demonstrate a condition under which the superadditivity of $r_t(h,a)$ overcomes the subadditivity of $\sum_{j=1}^{H+1} p_t(j|h,a)V_{t+1}^*(j)$, to result in a superadditive $w_t(h,a)$. Theorem 4.3.10 revises Condition (4.15) by incorporating clinical knowledge to eliminate the dependence on $V_{t+1}^*(h)$. First, Δ_h is defined to be the maximum decrease between health states h and $h+1$ in a patient's total expected reward for the remaining time that the patient is in the hospital, where

$$\sum_{j=1}^{H+1} [p_t(j|h,C) - p_t(j|h+1,C)]V_{t+1}^*(j) \leq \Delta_h.$$

Using this information, Theorem 4.3.10 presents a sufficient condition for the existence of a control-limit policy that is independent of the value function, $V_{t+1}^*(h)$.

Theorem 4.3.10 *There exists an optimal non-stationary control-limit policy in h if*

$$r_t(h,D) - r_t(h,C) - [r_t(h+1,D) - r_t(h+1,C)] \geq \Delta_h, \tag{4.19}$$

for $h = 1,\ldots,H$ and $t = 1,\ldots,N-1$.

The proof of Theorem 4.3.10 is similar to the proof of Theorem 4.3.7 and is therefore omitted. Unfortunately, Condition (4.19) is too restrictive for reasonable values of Δ_h and does not consistently hold for the data tested in this dissertation. The exploration of less restrictive sufficient conditions is left to future research. The next section explores the existence of control-limit policies for this model through various computational experiments.

4.4 EXPLORING HOSPITAL DISCHARGE POLICIES USING PATIENT-BASED DATA

One of the aims of this dissertation is to explore the structure of hospital decision policies through the use of computational experiments for patients with sepsis with the hope of being able to suggest general strategies for patient discharge. Due to the complexity of the disease and the availability of data at this time, the model state space is defined by a single parameter, the total Sepsis-related Organ Failure Assessment (SOFA) score. This score is a reasonable approximation for patient health because, as will be discussed in Section 4.4.2, this score is calculated based on the complex interactions between multiple aspects of the patient's health, all of which were captured as part of the GenIMS trial. The model was solved using the standard backward induction algorithm [89] presented in Appendix C.

4.4.1 MDP data sources

The GenIMS trial data contains static and dynamic variables for 2320 patients. These patients were identified by the GenIMS investigators as potentially having community-acquired pneumonia (CAP). Of these patients, 2032 were admitted to the hospital and went on to develop varying degrees of sepsis. The computational experiments presented in this section utilize a sample of 2025 patients, with seven patients being excluded from the GenIMS inpatient cohort due to missing or irregular data. Static variables such as age and race are provided for each patient. Dynamic health variables are available on a daily basis, where missing data were estimated utilizing a clinically derived algorithm that combines last observation carried forward and other clinically based interpolation methods, as agreed upon by the GenIMS investigator team [45].

4.4.2 SOFA score

The patient's health state is represented by the total SOFA score, an integer value ranging between 0 and 24, where 24 corresponds to the sickest health state. The score was developed by the Working Group on Sepsis-related Problems of the European Society of Intensive

Care Medicine to describe quantitatively the degree of organ dysfunction/failure over time [112]. The correlation between organ dysfunction and mortality makes the SOFA score an appropriate descriptor of patient health in a model of severe sepsis and its use is supported by previous models of severe sepsis in the literature that have used the SOFA score to describe patient health [24, 40].

Total SOFA is calculated based on six component scores that evaluate different organ systems (respiratory, coagulation, liver, central nervous system, renal, and cardiac). Therefore, even though total SOFA is a single value, the score actually captures a wide range of patient health variables. The daily component SOFA scores and the resulting total SOFA scores were calculated by the GenIMS investigators for all patients in the GenIMS cohort. By capturing the time-varying nature of each patient's SOFA scores in the transition probabilities used as input to the model, the model captures the evolution of patient health through all stages of the disease and the patient's hospital stay.

Due to data sparseness, the 25 total SOFA score values are aggregated into four patient health states {0,1}, {2,3}, {4,5,6,7}, {8,...,24}, and defined as aggregated health states 1, 2, 3, and 4, respectively. This aggregation was chosen to capture changes in the SOFA score for those levels at which clinicians would consider the discharge decision. Scores of 8 or greater indicate a disease severity that would make the discharge decision improbable.

4.4.3 Problem instances considering various age/race cohorts

Based on conventions in the literature [22, 23, 24, 50, 62], the values $N = 30$ days and $T = 90$ days are used. Since age and race have been determined to be significant predictors of patient mortality [58, 62], these static variables are used to define the eleven problem instances described in Table 4.1. The age breaks (45, 65) used to describe the instances follow the conventions in [24]. Note that the instances are further stratified by race (Caucasian, non-Caucasian). Due to the small sample sizes associated with non-Caucasian patients under 65 years of age, not all combinations of age and race groups could be tested with the available data.

Table 4.1: Description of MDP Problem Instances

Instance	Sample Size	Age	Race	Stationary During Periods
1	468	< 65	Caucasian	1-3, 4-7, 8-29
2	1158	≥ 65	Caucasian	1-3, 4-7, 8-29
3	273	< 65	non-Caucasian	1-3, 4-7, 8-29
4	126	≥ 65	non-Caucasian	1-3, 4-7, 8-29
5	242	< 45	all	1, 2-9, 10-29
6	499	$[45, 65]$	all	1, 2-9, 10-29
7	1284	≥ 65	all	1, 2-9, 10-29
8	242	< 45	all	1-3, 4-7, 8-29
9	499	$[45, 65]$	all	1-3, 4-7, 8-29
10	741	< 65	all	1-3, 4-7, 8-29
11	1284	≥ 65	all	1-3, 4-7, 8-29

The far right column of Table 4.1 describes a third component of the problem instances. Due to data sparsity, the transition probabilities and rewards are assumed to be piece-wise constant over specific stages as defined in the far right column of Table 4.1, but are allowed to be time-varying between the grouped stages. For example, problem instance 2 considers Caucasian patients that are age 65 or older. By assumption, the transition probabilities and rewards for this problem instance are stationary during stages 1 through 3, stages 4 through 7, and stages 8 through 29, but can be time-varying between stages 3 and 4 and between stages 7 and 8. Note that there are two different groupings used to define different problem instances, (1-3, 4-7, 8-29) and (1, 2-9, 10-29). The former is based on the clinical expertise of the coauthors. The latter definition is similar to that used by Clermont et al. [24].

These instances provide valuable insights into the effect of hospital length of stay on the hospital discharge decision for patients of varying age and race. These results are described in more detail in the next section.

4.4.4 Hospital discharge policy results and clinical interpretation

Table 4.2 presents the optimal policy for problem instance 2 including the optimal value function value and the optimal action for each stage and state. The optimal value function value, $V_t^*(h)$, represents the $(90-t)$-day expected survival of a patient in state h at stage t given that the clinician chooses the optimal action in the current stage and in all stages moving forward. For example, for a patient in aggregated health state 2 on day 5, the optimal action is to Continue with an expected 85-day survival of 71.9 days given that the clinician chose to keep the patient in the hospital and then act optimally in all future stages.

The optimal action to take at each stage and for each state is interpreted for problem instance 2 as follows. During days 1, 2, and 3, it is optimal to discharge patients in aggregated health states 1 and 2 (corresponding to a SOFA score of 0, 1, 2, or 3). For patients in all other aggregated health states (corresponding to a SOFA score of 5 or greater), it is optimal to keep the patient in the hospital for one additional day. During days 4, 5, and 6, it is optimal to discharge patients in the healthiest aggregated state only (corresponding to a SOFA score of 0 or 1) and to keep all other patients (corresponding to a SOFA score of

2 or greater) in the hospital for one additional day. Finally, during days 7 through 29, it is optimal to discharge all but the sickest patients and keep the remaining patients in the hospital for one more day (corresponding to SOFA scores of 0 through 7 and then 8 through 24, respectively).

This optimal solution is a control-limit policy. It is interesting to note that the control limit for this instance is time varying. For example, a patient that has not been discharged by day 3 and that is in aggregated state 2 in day 4 would not be discharged under this policy. This means that the patient was in an aggregated health state of 3 or greater in all days prior to day 4 (or the patient would have been discharged previously). This policy is intuitive, because patients that are sicker may need to remain in the hospital for a longer period of time, even though they appear to improve over time. Recall that these policies are determined as a result of the time-varying transition probabilities and rewards used as input data to the model. These time-dependent inputs are an essential component of a robust model of sepsis progression [24].

For days 7 through 29, only patients in the sickest health state should be kept in the hospital, corresponding to similar results presented in the literature. For example, Halm et al. [50] found that the median time to overall clinical stability in patients with CAP was between 3 days for the most lenient definition of clinical stability and 7 days for the most conservative definition. Studies looking at intensive care unit (ICU) length of stay for sepsis patients found the median length of stay to be between 7 and 14 days [24, 93]. Given that the current trend in research is to find ways to reduce excessively long ICU and hospital stays, the results found through this analysis are quite promising. Therefore, while this type of policy does not hold exactly for all stages in all problem instances, it does suggest an easy-to-implement decision making strategy.

Table 4.2: Optimal Solution to Problem Instance 2 (D = Discharge, C = Continue)

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	85.367	83.606	63.617	31.023	D	D	C	C
2	84.408	82.667	61.567	28.743	D	D	C	C
3	83.449	81.727	58.123	25.187	D	D	C	C
4	76.947	72.695	59.095	17.014	D	C	C	C
5	76.053	71.632	57.770	16.271	D	C	C	C
6	75.158	70.561	56.167	14.861	D	C	C	C
7	74.263	69.541	54.269	11.550	D	D	D	C
8	67.739	57.245	45.556	6.415	D	D	D	C
9	66.913	56.547	45.000	6.353	D	D	D	C
10	66.087	55.849	44.444	6.291	D	D	D	C
11	65.261	55.151	43.889	6.229	D	D	D	C
12	64.435	54.453	43.333	6.167	D	D	D	C
13	63.609	53.755	42.778	6.105	D	D	D	C
14	62.783	53.057	42.222	6.043	D	D	D	C
15	61.957	52.358	41.667	5.980	D	D	D	C
16	61.130	51.660	41.111	5.918	D	D	D	C
17	60.304	50.962	40.556	5.856	D	D	D	C
18	59.478	50.264	40.000	5.794	D	D	D	C
19	58.652	49.566	39.444	5.732	D	D	D	C
20	57.826	48.868	38.889	5.670	D	D	D	C
21	57.000	48.170	38.333	5.608	D	D	D	C
22	56.174	47.472	37.778	5.545	D	D	D	C
23	55.348	46.774	37.222	5.483	D	D	D	C
24	54.522	46.075	36.667	5.418	D	D	D	C
25	53.696	45.377	36.111	5.349	D	D	D	C
26	52.870	44.679	35.556	5.261	D	D	D	C
27	52.043	43.981	35.000	5.112	D	D	D	C
28	51.217	43.283	34.444	4.749	D	D	D	C
29	50.391	42.585	33.889	3.648	D	D	D	C

Table 4.3 provides a summary of the optimal action by day and state for every problem instance. Control-limit policies are indicated in bold text. For example, control-limit policies exist for every stage and state for problem instances 6 and 9 in addition to problem instance 2. The optimal policies for most of the other problem instances are of control-limit type for the majority of states and stages. For example, problem instance 3 has a control-limit policy for all days except day 29 and problem instance 11 follows this type of policy for all days except days 5 and 6. Problem instances 5 and 8, however, vary further from the control-limit policy structure. For instance 5, the control-limit policy structure does not hold for day 9 on and for instance 8, a control-limit type policy exists in days 1, 2, 3, and 29 only.

Table 4.4 lists how the assumptions and conditions presented in Section 4.3 hold for each problem instance. For example, Assumption 4.3.1 does not hold for problem instance 4, while Assumption 4.3.3 does hold for this instance. It is interesting to note that a control-limit policy does not exist for those instances for which one or more assumptions does not hold.

Other possible reasons for the deviations from the non-stationary control-limit policy structure appear to be attributable to both cohort sample size and definition. For example, even though problem instance 7 has the largest sample size (1284) of all problem instances, the patients included in this sample are not stratified by race. However, when patients are separated by race (i.e., into problem instances 2 and 4), the new instances follow or closely follow the control-limit policy structure. Since cohort 4 only has a sample size of 126 patients, data sparseness may be the source of any policy deviations.

Table 4.3: Optimal Solutions (Problem Instances 1 Through 11)

Instance and Action:		Days										
		1	2	3	4	5-6	7	8	9	10-26	27-28	29
1	D	**1,2,3**	**1,2,3**	**1,2,3**	2,3	2,3	**1,2,3**	1,3	1,3	1,3	1,3	1,3
	C	**4**	**4**	**4**	1,4	1,4	**4**	2,4	2,4	2,4	2,4	2,4
2	D	**1,2**	**1,2**	**1,2**	**1**	**1**	**1,2,3**	**1,2,3**	**1,2,3**	**1,2,3**	**1,2,3**	**1,2,3**
	C	**3,4**	**3,4**	**3,4**	**2,3,4**	**2,3,4**	**4**	**4**	**4**	**4**	**4**	**4**
3	D	**1,2,3**	**1,2,3**	**1,2,3**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	1,3
	C	**4**	**4**	**4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	2,4
4	D	**1**	**1**	**1**	2,3	2,3	**1,2,3**	**1**	**1**	**1**	**1**	1,3
	C	**2,3,4**	**2,3,4**	**2,3,4**	1,4	1,4	**4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	2,4
5	D	**1,2,3**	**1,2,3**	2,3	2,3	2,3	2,3	2,3	3	3	3	3
	C	**4**	**4**	1,4	1,4	1,4	1,4	1,4	1,2,4	1,2,4	1,2,4	1,2,4
6	D	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1,2,3**	**1**	**1**	**1**
	C	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**4**	**2,3,4**	**2,3,4**	**2,3,4**
7	D	2	**1,2**	1,3	1,3	1,3	1,3	1,3	**1,2,3**	**1,2**	**1,2**	**1,2**
	C	1,3,4	**3,4**	2,4	2,4	2,4	2,4	2,4	**4**	**3,4**	**3,4**	**3,4**
8	D	**1,2,3**	**1,2,3**	**1,2,3**	3	3	3	3	3	3	3	**1,2,3**
	C	**4**	**4**	**4**	1,2,4	1,2,4	1,2,4	1,2,4	1,2,4	1,2,4	1,2,4	**4**
9	D	**1,2,3**	**1,2,3**	**1,2,3**	**1,2**	**1,2**	**1,2,3**	**1**	**1**	**1**	**1**	**1**
	C	**4**	**4**	**4**	**3,4**	**3,4**	**4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**	**2,3,4**
10	D	**1,2,3**	**1,2,3**	**1,2,3**	**1,2**	**1,2**	**1,2,3**	**1**	**1**	**1**	1,3	1,3
	C	**4**	**4**	**4**	**3,4**	**3,4**	**4**	**2,3,4**	**2,3,4**	**2,3,4**	2,4	2,4
11	D	**1,2**	**1,2**	**1,2**	**1**	1,3	**1,2,3**	**1,2,3**	**1,2,3**	**1,2,3**	**1,2,3**	**1,2,3**
	C	**3,4**	**3,4**	**3,4**	**2,3,4**	2,4	**4**	**4**	**4**	**4**	**4**	**4**

Table 4.4: Verification of Assumptions and Conditions from Section 4.2

Instance	Assumption 4.3.1 satisfied?	Assumption 4.3.2 satisfied?	Assumption 4.3.3 satisfied?	Condition 4.15 satisfied?	Control-limit Policy?
1	yes	yes	yes	no	no
2	yes	yes	yes	no	yes
3	yes	yes	yes	no	close
4	no	no	yes	no	no
5	no	no	yes	no	no
6	yes	yes	yes	no	yes
7	yes	yes	yes	no	no
8	yes	no	yes	no	no
9	yes	yes	yes	no	yes
10	yes	yes	yes	no	no
11	yes	yes	yes	no	close

4.5 CONCLUSIONS

Through mathematical analysis and computational experiments, study of this model found that optimal hospital discharge strategies for patients with pneumonia-related sepsis tend to follow a non-stationary control-limit type policy structure. These types of policies have an obvious advantage in that they are easy to understand and can be used to standardize an otherwise complicated and ad-hoc procedure. Introducing the medical community to this type of policy structure is the first step in standardizing the hospital discharge decision.

There are limitations, however, to describing patient health by a single dimension, such as total SOFA score. Future work will explore more complex state descriptions, such as those that include the component SOFA scores, which are necessary before such models can inform clinical practice. As more data become available, the model presented in this dissertation can be used to provide increasingly accurate values for the health-based non-stationary control limits. Clearly, additional data would help to resolve any issues with data sparseness and would allow for the testing of additional cohort stratifications.

For example, in addition to age and race, gender has also been cited as a key predictor of patient mortality among patients with severe sepsis. Additional data would allow for the testing of age, race, and gender cohorts.

The following chapter extends this model by incorporating testing decisions into the model. In this second model, the patient health state is no longer characterized by the SOFA score, rather the value of a single cytokine level is used as a predictor of patient health. While a more realistic model would use both SOFA and the cytokine level, data availability and computational complexity limit current capabilities. Chapter 5, however, will explore heuristic methods for combining the solutions from both models.

5.0 MODELING TESTING AND HOSPITAL DISCHARGE DECISIONS (A PARTIALLY OBSERVABLE MARKOV DECISION PROCESS APPROACH)

Extending the model presented in Chapter 4, this model considers the clinician's decision problem of when to test for cytokine information in addition to the decision problem of when to discharge the patient from the hospital in order to maximize the patient's expected survival over a finite observation horizon as measured from hospital admission. The model is formulated as a finite-horizon POMDP, where the patient's true health state can only be observed through an inaccurate testing procedure. As a simplified variant of the general model presented in Chapter 3, the patient health state is represented by a single variable that measures the value of a single cytokine level. It is assumed that throughout the patient's stay in the hospital, the patient is being treated according to standard methods of care. At each decision point, the clinician can decide to continue treating the patient in the hospital using standard care without testing (C), continue treating the patient in the hospital using standard care and also order a cytokine test (O), or discharge the patient from the hospital without testing (D). It is assumed that when a test is ordered, its result is not known until the beginning of the next time period. Therefore, if a cytokine test has been ordered for a patient, the patient will remain in the hospital for at least one more time period. The next stage's decision is made after the test result ordered in the previous stage is observed.

In this model, the patient's true health state is modeled as the value of a single cytokine level. This level can take on one of two values: low (L) or high (H). After a test is ordered, the clinician will observe either a L or H value, which relate probabilistically to the true underlying cytokine level through the observation probability matrix (i.e., the accuracy of the test). A low cytokine value is correlated with a high probability of patient survival [63]; however, it does not indicate patient survival with complete certainty. This uncertainty is

captured in the expected survival calculation used for the reward function. In the model it is assumed that the patient's cytokine level cannot be known without testing, testing is done only when action O is chosen, and test results are received at the beginning of the next time period.

5.1 POMDP NOTATION

The following notation is used:

- $\mathcal{N} = \{1, 2, \ldots, N\}$: discrete stages at which a decision must be made by the clinician, where N is the treatment horizon. If a patient has not died and has not been discharged by stage N, it is assumed that the patient is discharged at stage N. This dissertation defines a stage as one day; however, the model is flexible enough to consider smaller time intervals (hours, for example) as the data for solving such a model become available. Let t denote the current stage in the model.
- T: the observation time horizon used to measure patient survival from hospital admission.
- y_t: a scalar describing the true value of the patient's cytokine level. It is assumed that y_t can take on one of three values: low (L), high (H), or dead.
- o_t: a scalar describing the observed value of the patient's cytokine level. If the patient has died, then o_t = dead; otherwise, if a test is ordered, it is assumed that o_t can take on one of two values: low (L) or high (H). When no test result is received, let $o_t = \emptyset$.
- a_t: the action taken. Possible actions are to continue treating the patient in the hospital without ordering a cytokine test (C), to continue treating the patient in the hospital and order a cytokine test (O), or to discharge the patient from the hospital (D). It is assumed that when a cytokine test result is ordered, the test result is received at the beginning of the next period before the next decision is made. Therefore, the patient cannot be discharged before the next period. If the decision is made to discharge the patient from the hospital, the patient transitions out of the model.
- c: a scalar representing the cost of ordering a cytokine test (converted to patient life days using methods from cost-effectiveness analysis [47]).

- $f_t(y_t, D)$: the expected $(T-t)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage t with true health state y_t. Note that $f_t(\text{dead}, D) = 0$.
- $r_t(\pi_t, D)$: the expected $(T-t)$-day survival (in patient life days) of a patient that is discharged from the hospital at stage t with belief variable π_t, where $r_t(\pi_t, D) = f_t(L, D)\pi_t + f_t(H, D)(1-\pi_t)$.
- $f_t(y_t, O)$: the expected reward (in patient life days) received for keeping a patient with true health state y_t at time t in the hospital for one more stage and ordering a cytokine test. Note that $f_t(\text{dead}, O) = 0$.
- $r_t(\pi_t, O)$: the expected reward (in patient life days) received for keeping a patient with belief variable π_t at time t in the hospital for one more stage and ordering a cytokine test, where $r_t(\pi_t, O) = f_t(L, O)\pi_t + f_t(H, O)(1-\pi_t)$.
- $f_t(y_t, C)$: the expected reward (in patient life days) received for keeping a patient with true health state y_t at time t in the hospital for one more stage. Note that $f_t(\text{dead}, C) = 0$.
- $r_t(\pi_t, C)$: the expected reward (in patient life days) received for keeping a patient with belief variable π_t at time t in the hospital for one more stage, where $r_t(\pi_t, C) = f_t(L, C)\pi_t + f_t(H, C)(1-\pi_t)$.
- $f_N(y_t)$: the expected $(T-N)$-day survival of a patient that is discharged from the hospital at stage N with true health state y_t. Note that $f_N(\text{dead}) = 0$.
- $r_N(\pi_N)$: the expected $(T-N)$-day survival of a patient that is discharged from the hospital at stage N with belief variable π_N, where $r_N(\pi_N) = f_N(L)\pi_N + f_N(H)(1-\pi_N)$.
- $p_t(y_{t+1}|y_t, a_t)$: the probability that the patient's true health state is y_{t+1} at stage $t+1$ given that at stage t, the patient's true health state was y_t, and action a_t was chosen. Note that $p_t(\text{dead}|\text{dead}, \cdot) = 1$ and that the process will terminate with reward $f_t(\text{dead}, \cdot) = 0$ if the patient dies before the next time period.
- $z(o_{t+1}|y_{t+1}, a_t)$: the test accuracy, i.e., the probability of observing cytokine level o_{t+1} at stage $t+1$ when the patient's true cytokine level is y_{t+1} at stage $t+1$ and action a_t was chosen at stage t. Note that $z(\text{dead}|\text{dead}, \cdot) = 1$ and that $z(L|\text{dead}, \cdot) = z(H|\text{dead}, \cdot) = z(\emptyset|\text{dead}, \cdot) = 0$.

- π_t: the probability that the patient's true cytokine level is low ($y_t = L$) at stage t given that the patient is still alive at stage t.
- $\gamma_t(o_{t+1}|\pi_t, a_t)$: the probability of observing cytokine level o_{t+1} at time $t+1$ given that the belief variable was π_t and action a_t was taken at time t, where

$$\gamma_t(o_{t+1}|\pi_t, a_t) = \sum_{y_{t+1} \in \{L, H, \text{dead}\}} z(o_{t+1}|y_{t+1}, a_t)[p_t(y_{t+1}|L, a_t)\pi_t + p_t(y_{t+1}|H, a_t)(1-\pi_t)].$$

- $U(\pi_{t+1}|o_{t+1}, \pi_t, a_t)$: the updating function used to update the belief variable π_t to π_{t+1} based on o_{t+1}, the observation at stage $t+1$, π_t, the belief variable at time t, and a_t, the action taken at time t.
- $V_t(\pi_t)$: the value function used to calculate the total expected reward (in patient life days) at stage t when the patient is alive and in health state π_t, where $V_t^*(\pi_t)$ denotes the optimal value function value.
- $\mathcal{A}_t^*(\pi_t)$: the set of optimal actions at stage t when the system is in state π_t, where $a_t^*(\pi_t) \in \mathcal{A}_t^*(\pi_t)$ is an action that maximizes the value function $V_t(\pi_t)$.

After action a_t is taken at stage t, an immediate expected reward $r_t(\pi_t, a_t)$ is received. If $a_t = D$, the patient is discharged and receives an expected reward, $r_t(\pi_t, D)$. If $a_t = O$, the clinician orders a cytokine test, the patient receives an expected reward $r_t(\pi_t, C)$ (in patient life days), and the patient's core health state transitions to a new value, which includes the possibility of patient death. Finally, if $a_t = C$, no test is ordered, the patient receives an expected reward $r_t(\pi_t, C)$ (in patient life days), and the patient's core health state transitions to a new value, which includes the possibility of patient death. If a patient dies (i.e., $y_{t+1} = o_{t+1} = \text{dead}$), it is assumed that death occurs at the beginning of the next stage and that the patient exits the model and does not accumulate any future rewards. If the patient does not die, the belief variable, π_t, is updated from stage t to stage $t+1$ using the updating function, $U(\pi_{t+1}|o_{t+1}, \pi_t, a_t)$, which performs the update using the observation o_{t+1} from stage $t+1$, where

$$U(\pi_{t+1}|o_{t+1}, \pi_t, a_t) = \begin{cases} \frac{z(o_{t+1}|L)[p_t(L|L,a_t)\pi_t + p_t(L|H,a_t)(1-\pi_t)]}{\gamma_t(o_{t+1}|\pi_t, a_t)}, & \text{if } o_{t+1} \in \{H, L\}; \\ p_t(L|L, a_t)\pi_t + p_t(L|H, a_t)(1-\pi_t), & \text{if } o_{t+1} = \emptyset. \end{cases} \tag{5.1}$$

5.2 POMDP MODEL FORMULATION

Let the optimal value function, $V_t^*(\pi_t)$, be the total expected reward for a living patient with belief variable π_t for time t onward. $V_t^*(\pi_t)$ can then be defined recursively as follows.

$$V_N^*(\pi_N) = r_N(\pi_N), \text{ for all } \pi_N \in \Pi \tag{5.2}$$

$$V_t^*(\pi_t) = \max \begin{cases} r_t(\pi_t, D), \\ r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O)), \\ r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t, C)). \end{cases} \tag{5.3}$$

for all $\pi_t \in [0, 1]$ and $t = 1, \ldots, N-1$.

The structural results presented in Section 5.3 and the computational experiments described in Section 5.4 provide insight into the value of test accuracy and cost as they relate to prolonging the life years of a patient. Test accuracy refers not only to the result received through testing, but also its ability to inform correct clinical interpretation of the true underlying health state.

5.3 ANALYZING THE EFFECTS OF TEST COST AND ACCURACY ON CYTOKINE TESTING AND HOSPITAL DISCHARGE DECISIONS

This section presents structural results for the POMDP model pertaining to changes in test cost and accuracy. The purpose of this analysis is to provide insight into the robustness of the model results for a range of model parameters, as can be expected in practice. Experiments utilizing patient-based data are presented in the following section. Relevant proofs are provided in the Appendix. First, the following assumptions are made.

5.3.1 Additional assumptions for the POMDP model

Assumption 5.3.1 *It is assumed that* $f_t(y_t, C) = f_t(y_t, O) = 1$ *for all* $y_t \in \mathcal{Y}$ *and for all* $t \in \mathcal{N}$.

This assumption states that the immediate reward received for keeping the patient in the hospital, with or without ordering a cytokine test, is one more day of patient life. Future research could consider other reward functions, such as incorporating the cost of care.

Assumption 5.3.2 *The core patient health transition probability matrix* $P(t)$*, with entries* $[P(t)]_{yj} = p_t(j|y_t, C) = p_t(j|y_t, O)$*, is IFR for all* $t \in \mathcal{N}$.

From definition 4.3.1, this assumption implies that for two patients in health states y and $y+1$, respectively, the patient in health state y is more likely to transition to a health state worse than y in the next stage. In other words, sicker patients are more likely to progress to being even sicker than are healthier patients.

Assumption 5.3.3 *The test has accuracy* ρ*, i.e.,* $z(L|L) = z(H|H) = \rho$.

Through this assumption the model is restricted to consider only symmetric testing accuracies.

Assumption 5.3.4 *Tests can be inaccurate, i.e.,* $\rho \in (\frac{1}{2}, 1]$.

This assumption implies that the test results are more likely to be correct than incorrect.

5.3.2 Mathematical analysis of the effects of testing cost and accuracy

With these assumptions, this model is similar to the model presented in Monahan [81], but applied to a medical decision making application. This model also considers a time-dependent terminal reward, but this difference does not impact the structural results. Theorem 5.3.5 restates one of Monahan's results for the finite-horizon case, which demonstrates an ordering of the value function over π_t for fixed testing cost and test accuracy. As discussed in Section 2.3.1, a more general result was presented by Smallwood and Sondik [103].

Theorem 5.3.5 *[81, 103] For any fixed* $\rho \in (\frac{1}{2}, 1]$ *and IFR transition probability matrix,* $V_t^*(\pi_t)$ *is continuous, nondecreasing, and convex in* π_t*,* $\pi_t \in [0, 1]$.

Extending this result, an ordering on the value function with respect to testing cost is demonstrated in Theorem 5.3.6. Let $V_t^*(\pi_t, c)$ denote the optimal value function value at stage t when the testing cost is c and the belief variable is π_t.

Theorem 5.3.6 *$V_t^*(\pi_t, c') \leq V_t^*(\pi_t, c'')$ for $c' \geq c''$ and for all $t \in \mathcal{N}$.*

Proof *(By induction)* From (5.2) it is known that $V_N^*(\pi_N, c') = V_N^*(\pi_N, c'') = r_N(\pi_N)$. Assuming that $V_k^*(\pi_k, c') \leq V_k^*(\pi_k, c'')$ for $k = t+1, \ldots, N-1$, it suffices to show that $V_t^*(\pi_t, c') \leq V_t^*(\pi_t, c'')$. Following from (3.5), it follows that

$$V_t^*(\pi_t, c') = \max \begin{cases} r_t(\pi_t, D), \\ r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c'), \\ r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t, C), c'). \end{cases}$$

and

$$V_t^*(\pi_t, c'') = \max \begin{cases} r_t(\pi_t, D), \\ r_t(\pi_t, O) - c'' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c''), \\ r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t, C), c''). \end{cases}$$

The induction assumption implies that

$$V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c'') \geq V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c') \tag{5.4}$$

and

$$V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t, C), c'') \geq V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t, C), c'). \tag{5.5}$$

Suppose that $V_t^*(\pi_t, c') = r_t(\pi_t, D)$. By definition, $V_t^*(\pi_t, c'') \geq r_t(\pi_t, D) = V_t^*(\pi_t, c')$.

Next, suppose that

$$V_t^*(\pi_t, c') = r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c').$$

By definition,

$$\begin{aligned} V_t^*(\pi_t, c'') &\geq r_t(\pi_t, O) - c'' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c'') \\ &\geq r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c'') \end{aligned} \tag{5.6}$$

$$\geq r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), c') \tag{5.7}$$
$$= V^*_t(\pi_t, c'),$$

where (5.6) follows from the fact that $c' \geq c''$ and (5.7) follows from the induction assumption (5.4).

Finally, suppose that $V^*_t(\pi_t, c') = r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t, C), c')$. By definition,

$$V^*_t(\pi_t, c'') \geq r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t, C), c'')$$
$$\geq r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t, C), c') = V^*_t(\pi_t, c'), \tag{5.8}$$

where (5.8) follows from the induction assumption (5.5) and the desired result follows. ∎

The ordering shown in Theorem 5.3.6 leads to Corollary 5.3.7, which shows that as the testing cost decreases, the optimal testing region does not decrease.

Corollary 5.3.7 *If it is optimal to test in state π_t at stage t when the testing cost is c', then it is also optimal to test when the testing cost is $c'' \leq c'$.*

Proof Given π_t and testing cost c', $a^*_t = O$ and (3.5) imply that

$$r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}, O|\pi_t) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O)) \geq r_t(\pi_t, D)$$

and that

$$r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O))$$
$$\geq r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t, C)).$$

Note that $c'' \leq c'$ implies that

$$r_t(\pi_t, O) - c'' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O))$$
$$\geq r_t(\pi_t, O) - c' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O)).$$

Therefore,

$$r_t(\pi_t, O) - c'' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O)) \geq r_t(\pi_t, D)$$

and

$$r_t(\pi_t, O) - c'' + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O)) \\ \geq r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t, C)),$$

and the result follows. ∎

Corollary 5.3.7 demonstrates the intuitive result that when it is optimal to test for a given test cost c', then it is also be optimal to test for a lower test cost c'' given that all other model parameters remain the same.

Exploring the effects of test accuracy on the optimal value function value, Theorem 5.3.8 restates another result from Monahan [81] for the finite-horizon case. This result demonstrates an ordering on the value function with respect to test accuracy. Let $V^*_t(\pi_t, \rho)$ denote the optimal value function value at stage t when the test accuracy is ρ and the belief variable is π_t.

Theorem 5.3.8 *[81] $V^*_t(\pi_t, \rho') \leq V^*_t(\pi_t, \rho'')$ for $\rho' \leq \rho''$ and for all $t \in \mathcal{N}$.*

Corollary 5.3.9 expands this result, which shows that as testing accuracy increases, the optimal testing region does not decrease.

Corollary 5.3.9 *If it is optimal to test in state π_t at stage t when the test accuracy is ρ', then it is also optimal to test when the test accuracy is $\rho'' \geq \rho'$.*

Proof Given π_t and test accuracy ρ', $a^*_t = O$ and (3.5) imply that

$$r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), \rho') \geq r_t(\pi_t, D)$$

and that

$$r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), \rho') \\ \geq r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t, C)).$$

Note that following from Theorem 5.3.8, $\rho'' \geq \rho'$ implies that

$$r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), \rho'') \\ \geq r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), \rho').$$

Therefore,

$$r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), \rho'') \geq r_t(\pi_t, D)$$

and

$$\begin{aligned} r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t, O) V^*_{t+1}(U(\pi_{t+1}|o_{t+1}, \pi_t, O), \rho'') \\ \geq r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, C) V^*_{t+1}(U(\pi_{t+1}|\emptyset, \pi_t)), \end{aligned}$$

and the result follows. ∎

Corollary 5.3.9 demonstrates the intuitive result that when it is optimal to test for a given test accuracy ρ', then it is also be optimal to test when the test result and/or interpretation accuracy is improved.

The specific results described in Corollaries 5.3.7 and 5.3.9 are demonstrated through the computational experiments described in Section 5.4.

5.4 USING PATIENT DATA TO CALIBRATE MODELS OF CYTOKINE TESTING AND HOSPITAL DISCHARGE DECISIONS

The problem instances described in this section were solved using the modified backward induction algorithm detailed in Appendix C.

5.4.1 POMDP data sources

The GenIMS trial data contains static and dynamic variables for 2320 patients that were identified as potentially having community-acquired pneumonia (CAP). Of these patients, 2032 were admitted to the hospital and went on to develop varying degrees of sepsis. The computational experiments presented in this section utilize a sample of 1096 patients, with 936 patients being excluded from the GenIMS in-patient cohort because they did not have at least two consecutive days of cytokine test results. Static variables such as age and race are provided for each patient.

5.4.2 Interleukin-6

The following computational experiments utilize the cytokine interleukin-6 (IL-6) to represent the observed value of the patient's health state. IL-6 has been shown to be a predictor of severe sepsis and death through studies conducted as part of the GenIMS investigation [63]. In particular, it has been shown that elevated concentrations of this cytokine were higher for those patients that died following severe sepsis compared to those who survived [63]. Based on the results of this study, the problem instances tested in this dissertation considered an IL-6 level of 5.9 pg/mL or less to be low and all levels greater than this value to be high [63].

5.4.3 Problem instances considering various testing costs and accuracy levels

Based on conventions in the literature [22, 23, 24, 50, 62, 63, 65, 96], the values $N = 30$ days and $T = 90$ days are used. Since age and race have been determined to be significant predictors of patient mortality [58, 62], these static variables are used to define problem instances following the conventions used in [24], [65], and [96]. Due to the sparsity of cytokine data, problem instances were constructed for the cohort consisting of Caucasian patients age 65 and older. As additional data become available, the model can be used for other age and race cohorts.

Due to issues with data sparsity when attempting to develop non-stationary probability matrices, it is assumed that the transition and observation probabilities are stationary; however, the patient's $(T-t)$-day expected survival is calculated as a function of the patient's length of stay in the hospital. As additional controlled trials are conducted to study the role of inflammatory markers in predicting sepsis progression and survival, the availability of additional data will allow for the consideration of increasingly complex models, such as those with multiple non-stationary components.

Table 5.1 defines nine problem instances by their observation probabilities and testing costs. These instances were tested using the general model. As in Section 5.3, symmetric test accuracies are assumed, i.e., $z(L|L) = z(H|H)$. This dissertation focuses on how testing accuracy in general affects the optimal policies. Further investigation into changes in test specificity versus sensitivity is left to future research.

Table 5.1: Problem Instances Tested Using the POMDP Model

	$z(L\|L)$	c
A	1	0
B	1	0.5
C	1	1
D	0.95	0
E	0.95	0.5
F	0.95	1
G	0.90	0
H	0.90	0.5
I	0.90	1

5.4.4 Results for various testing costs and accuracy levels

Figure 5.1 displays the optimal discharge and testing policy regions for the nine problem instances described in Table 5.1. Recall that these problem instances were all solved for the cohort consisting of Caucasian patients age 65 and older. In Figure 5.1, the problem instances are presented such that that test accuracy increases from bottom to top and testing cost increases from left to right. These results are also presented in numerical form in Tables B1, B2, and B3 in Appendix B.

Three values of test accuracy are considered: 1 (completely accurate), 0.95, and 0.9. Three values of test cost (in life days) are considered: 0 (no cost), 0.5, and 1. Test cost is a conversion of dollars to life days using a willingness-to-pay threshold, as done in cost-effectiveness analysis [48]. For example, 0.5 life days would translate to $685 using a $500,000 per life year willingness-to-pay threshold value or to $1,370 using a $1,000,000 per life year willingness-to-pay threshold.

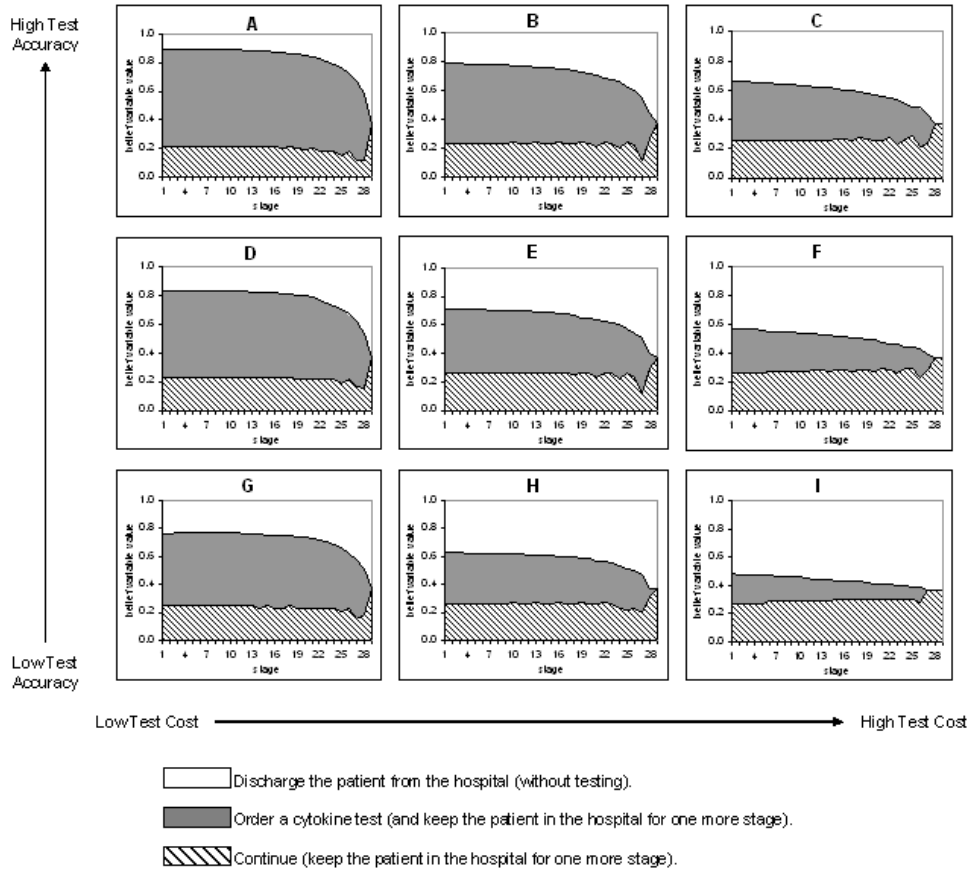

Figure 5.1: Optimal Policy Regions for Each POMDP Problem Instance

To interpret the optimal policy decision for a patient in a given health state (represented by the current value of the belief variable) at a particular stage, the clinician would look at the appropriate problem instance graph and first choose the current stage along the x-axis. The clinician would next find the current value of the belief variable (i.e., the clinician's belief that the patient's underlying health state is well) along the y-axis and then identify in which of the three possible action regions these x, y coordinates fall. The lower-most region, shaded in diagonal lines, represents the Continue region. If the coordinates fall within this region, then the optimal action is to keep the patient in the hospital for one more day without ordering a cytokine test. If the coordinates fall within the gray-shaded region, the testing region, then the optimal action would be to keep the patient in the hospital for one more day, but to also order a cytokine test. Recall that the test results will be received in the next stage before a decision is made. Finally, if the coordinates fall in the upper-most region, shaded white, then the optimal action is to discharge the patient from the hospital without ordering a cytokine test.

Note that the optimal policy on day 29 is the same for each problem instance. This policy can be interpreted as: If the belief variable is less than 0.37, keep the patient in the hospital for one more day and then discharge on day 30; otherwise, discharge the patient from the hospital on day 29. There is no testing region on day 29 since the model assumes that a patient that is still in the hospital on day 30 must be discharged from the hospital. A test result received on day 30 (for a test ordered on day 29) would not change this decision and is therefore not necessary. Since testing is not involved at the end of the horizon, the nine problem instances presented are the same on day 29 looking forward and therefore have the same optimal policy. For the previous stages, however, the varying test costs and accuracies impact the resulting policies, as can be seen in the differing optimal policies for each problem instance for days 1 through 28. The convergence of each policy to 0.37 on day 29 can be described as "end-of-horizon effects."

5.4.5 Discussion of POMDP results

To explain further, consider problem instance H, where $z(L|L) = 0.9$. Suppose that on day 9, the clinician believed that there was a 50% chance that the patient's IL-6 level was low. Based on the optimal policy illustrated in Figure 5.1, the clinician would choose to keep the patient in the hospital for one more day and order a cytokine test. When the test result is received at the beginning of day 10, the clinician observes the test result and makes another decision.

Suppose that the test result observed at the beginning of day 10 indicates that the patient's IL-6 level is low. With this new information about the patient and with $p_t(L|L) = 0.81, p_t(H|L) = 0.18, p_t(L|H) = 0.17$, and $p_t(H|H) = 0.82$, the clinician's new belief variable value becomes:

$$U(\pi_{t+1}|o_{t+1} = L, \pi_t = 0.5) = \frac{z(L|L)[p_t(L|L)\pi_t + p_t(L|H)(1-\pi_t)]}{\gamma_t(L|0.5)}$$

$$= \frac{0.9[0.81 * 0.5 + 0.17 * 0.5]}{0.491} = 0.898 \equiv \mathbf{90\%},$$

where

$$\gamma_t(L|0.5) = z(L|L)[p_t(L|L)\pi_t + p_t(L|H)(1-\pi_t)] + z(L|H)[p_t(H|L)\pi_t + p_t(H|H)(1-\pi_t)]$$

$$= 0.9[0.81 * 0.5 + 0.17 * 0.5] + 0.1[0.18 * 0.5 + 0.82 * 0.5] = 0.491.$$

Looking at the chart for instance H, day 10, and belief variable value 0.90, the optimal decision is to discharge the patient from the hospital.

If, on the other hand, the test results observed at the beginning of day 10 had indicated that the patient's IL-6 level was high, then the clinician's new belief variable value would have become:

$$U(\pi_{t+1}|o_{t+1} = H, \pi_t = 0.5) = \frac{z(H|L)[p_t(L|L)\pi_t + p_t(L|H)(1-\pi_t)]}{\gamma_t(H|0.5)}$$

$$= \frac{0.1[0.81 * 0.5 + 0.17 * 0.5]}{0.499} = 0.098 \equiv \mathbf{10\%},$$

where

$$\gamma_t(H|0.5) = z(H|L)[p_t(L|L)\pi_t + p_t(L|H)(1-\pi_t)] + z(H|H)[p_t(H|L)\pi_t + p_t(H|H)(1-\pi_t)]$$

$$= 0.1[0.81 * 0.5 + 0.17 * 0.5] + 0.9[0.18 * 0.5 + 0.82 * 0.5] = 0.499.$$

Looking at the chart for instance H, day 10, and belief variable value 0.1, the optimal decision is to keep the patient in the hospital without ordering a cytokine test. If no cytokine test is ordered on day 10, then the updated belief variable on day 11 becomes:

$$U(\pi_{11}|\emptyset, 0.098) = p_t(L|L)\pi_t + p_t(L|H)(1-\pi_t)$$

$$= 0.81 * 0.098 + 0.17 * 0.902 = 0.233 \equiv \mathbf{23\%}.$$

Note that the belief variable value increased from 0.1 to 0.23 and that the optimal decision on day 11 in this state is to order another cytokine test.

These regions can be further explained as follows. For very low values of the belief variable, the decision maker believes with a high probability that the patient's cytokine level is high and it is therefore not necessary to order a test for more information. This additional information would not change the decision to keep the patient in the hospital, so a test is not ordered. Similarly, for very high values of the belief variable, the decision maker is fairly certain that the patient is healthy enough to be discharged from the hospital, and it is therefore not necessary to order a test for more information. This additional information would most likely not change the decision to discharge the patient from the hospital, so a test is not ordered.

Notice that these observations are true even for the case when there is no testing cost and no test error. The Continue and Discharge regions increase in size as the associated testing cost increases, confirming the results of Corollary 5.3.7 in Section 5.3. These regions decrease in size as the associated testing accuracy increases, confirming the results of Corollary 5.3.9 in Section 5.3. The center testing region is highly dependent on both the cost of ordering a test and the accuracy of the test results. The testing region is largest for Problem Instance A in which there is no testing cost and the test results are completely accurate. For Problem Instance I, on the other hand, in which the testing cost is highest and the test accuracy

is lowest, the testing region is very narrow. The testing region appears to be less affected when only one parameter, cost or accuracy, is changed, but decreases dramatically as both become more unfavorable.

5.5 INTERPRETING OPTIMAL POLICIES FOR CLINICAL USE

One drawback of the POMDP model formulation and model output is the difficulty in translating the results to an applicable strategy. This stems mainly from the use of a belief variable in the model formulation. The belief variable is updated from one stage to the next using Bayesian updating, effectively capturing the history of patient health transitions and observed test results in a single value. Unfortunately, the clinician often considers the patient's history explicitly when making a treatment decision. The single belief variable value alone does not provide enough clinical information to the clinician to be immediately applicable in practice.

A finite horizon model that considers 30 patient days in the hospital and three possible test results (low, high, no test) will result in 3^{30} (over 200 trillion) possible combinations of test results. The belief variable allows for the use of solution techniques that avoid enumerating each possible solution. The following tables provide some insight into the belief variable calculation over a short time horizon of four days. In this example, it is assumed that tests are ordered for four consecutive days and that the first test result is observed on day 1. Assuming that the test interpretation accuracy is 0.95, Table 5.2 demonstrates how the belief variable is updated based on all combinations of four consecutive test results, starting with a low test result on day 1. Table 5.3 demonstrates how the belief variable is updated based on all combinations of four consecutive test results, starting with a high test result on day 1.

These tables demonstrate that a change in the test result received between two consecutive days significantly impacts the belief variable value. For example, even after a 3 low test results, a single high test result will reduce the belief variable value from 0.9874 to 0.1791. Similarly, after 3 high test results, a single low test result will raise the belief variable value from 0.0113 to 0.8039.

Table 5.2: Belief Variable Values (Starting With a Low Test Result on Day 1 and All Possible Results on Days 2, 3, and 4)

day 1 result	day 2 result	day 3 result	day 4 result
L 0.95	L 0.9855	L 0.9874	H 0.1791
L 0.95	L 0.9855	H 0.1780	L 0.8837
L 0.95	H 0.1587	L 0.8772	L 0.9814
L 0.95	L 0.9855	H 0.1780	H 0.02062
L 0.95	H 0.1587	L 0.8772	H 0.1276
L 0.95	H 0.1587	H 0.0194	L 0.8094
L 0.95	H 0.1587	H 0.0194	H 0.0116
L 0.95	L 0.9855	L 0.9874	L 0.9875

Table 5.3: Belief Variable Values (Starting With a High Test Result on Day 1 and All Possible Results on Days 2, 3, and 4)

day 1 result	day 2 result	day 3 result	day 4 result
H 0.05	L 0.8284	L 0.9784	H 0.1738
H 0.05	L 0.8284	H 0.1114	L 0.8588
H 0.05	H 0.0132	L 0.8052	L 0.9768
H 0.05	L 0.8284	H 0.1114	H 0.0166
H 0.05	H 0.0132	L 0.8052	H 0.1046
H 0.05	H 0.0132	H 0.0113	L 0.8039
H 0.05	H 0.0132	H 0.0113	H 0.0112
H 0.05	L 0.8284	L 0.9784	L 0.9870

These values are further put into perspective when compared to the testing and discharge control-limit values presented in Table 5.4. These testing and discharge control limits correspond to problem instances D, E, and F, which correspond to a test accuracy of 0.95 and test costs, 0, 0.5, and 1, respectively. When comparing the values in Tables 5.2 and 5.3 to Table 5.4, consider, for example, that two low test results in a row in either Table 5.2 or Table 5.3 will raise the belief variable value above the limit for discharging a patient for all problem instances. Similarly, even one high test result reduces the belief variable below the testing limit.

Table 5.4: POMDP Control Limits (Instances D, E, and F)

day	D		E		F	
	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$
1	0.23	0.83	0.26	0.71	0.26	0.57
2	0.23	0.83	0.26	0.71	0.26	0.57
3	0.23	0.83	0.26	0.71	0.26	0.57
4	0.23	0.83	0.26	0.71	0.26	0.57
5	0.23	0.83	0.26	0.71	0.26	0.56
6	0.23	0.83	0.26	0.71	0.27	0.55
7	0.23	0.83	0.26	0.7	0.27	0.55
8	0.23	0.83	0.26	0.7	0.27	0.55
9	0.23	0.83	0.26	0.7	0.27	0.55
10	0.23	0.83	0.26	0.7	0.27	0.54
11	0.23	0.83	0.26	0.7	0.27	0.54
12	0.23	0.83	0.26	0.7	0.29	0.54
13	0.23	0.82	0.26	0.69	0.29	0.53
14	0.23	0.82	0.26	0.69	0.27	0.53
15	0.23	0.82	0.26	0.69	0.29	0.52
16	0.23	0.82	0.26	0.68	0.29	0.52
17	0.23	0.81	0.26	0.68	0.27	0.51
18	0.23	0.81	0.25	0.67	0.29	0.51
19	0.22	0.8	0.26	0.65	0.29	0.5
20	0.22	0.8	0.26	0.65	0.27	0.5
21	0.22	0.79	0.23	0.64	0.3	0.48
22	0.22	0.77	0.26	0.62	0.29	0.47
23	0.22	0.75	0.26	0.61	0.26	0.46
24	0.22	0.73	0.22	0.6	0.3	0.44
25	0.19	0.71	0.26	0.57	0.3	0.44
26	0.22	0.67	0.23	0.54	0.23	0.43
27	0.16	0.62	0.12	0.51	0.29	0.39
28	0.15	0.54	0.29	0.4	0.37	0.37
29	0.37	0.37	0.37	0.37	0.37	0.37

Once a high test result is received and the belief variable value falls below the testing limit, then the optimal action is to keep the patient in the hospital for one more stage without ordering another test. The belief variable value is then updated according to the underlying core state transition probabilities. Only after the belief variable value has reached the testing limit, will another test be ordered for that patient. Figure 5.2 displays the testing and discharge control limits, by stage, for problem instance E, corresponding to a test accuracy of 0.95 and a test cost of 0.5. This figure also shows how the value of the belief variable is updated from one day to the next after a high test result on day one *and no further test results*. Notice that the belief variable value enters the testing region on day 3. In other words, if a high test result was received on day 1, then the patient would be kept in the hospital without testing on days 1 and 2. On day 3, the patient would be kept in the hospital and another cytokine test would be ordered.

Interpreting this information from a clinical perspective, it is clear that a high test result is a significant indication that the patient is very sick and should be kept in the hospital. However, if the patient survives for two days following the high test result, then the patient's health has likely improved, and another test result should be ordered.

To generalize this result, Table 5.5 demonstrates how various values of the belief variable π_t are updated from day t to $t+1$ after a high test result is received. Notice that π_{t+1} falls below the test threshold for all t and for all values of π_t, as shown in Table 5.5. It can therefore be concluded, that after a high test result is received, the clinic

Table 5.6 demonstrates how π_t is updated from day t to $t+1$ when no test result is received, for the updated belief variable values in Table 5.5. Note that only the largest values fall above the testing thresholds of the later days in the model, as indicated in Table 5.4.

Finally, Table 5.7 demonstrates the updating of π_{t+1} from day $t+1$ to day $t+2$ for the updated values in Table 5.6. These values all fall above the testing thresholds for all values of π_{t+1} and all days. Therefore, it is optimal to order another cytokine test when in any of these belief variable states.

Table 5.8 shows how various values of the belief variable π_t would be updated from day t to $t+1$ when a low test result is received. Notice that all updated belief variable values fall

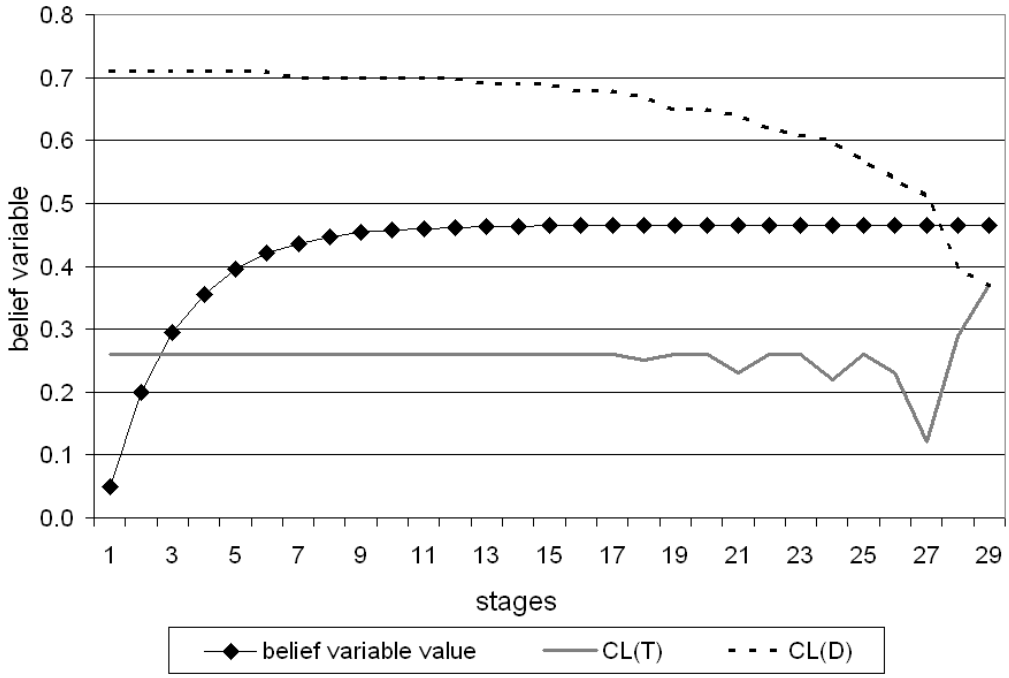

Figure 5.2: Belief Variable Value (After a High Test Result on Day 1 and No Further Test Results)

Table 5.5: Updating π_t from Day t to $t+1$ After a High Test Result

pi_t	pi_{t+1}
0.10	0.02
0.20	0.02
0.30	0.03
0.40	0.04
0.50	0.05
0.60	0.06
0.70	0.08
0.80	0.10
0.90	0.14

Table 5.6: Updating π_t from Day t to $t+1$ When No Test Result is Received

π_t	π_{t+1}
0.02	0.18
0.03	0.19
0.04	0.19
0.05	0.20
0.06	0.21
0.08	0.22
0.10	0.23
0.14	0.25

Table 5.7: Updating π_{t+1} from Day $t+1$ to $t+2$ When No Test Result is Received

π_{t+1}	π_{t+2}
0.18	0.28
0.19	0.29
0.20	0.29
0.21	0.30
0.22	0.31
0.23	0.32
0.25	0.33

above the discharge threshold for all days and all values of π_t. It can be concluded, therefore, that a patient should be discharged after a low test result is observed.

Exploring the results for problem instance E has resulted in decision rules that are easy to implement in practice. If a high test result is observed, wait one to two days, then test again. If a low test result is observed, discharge the patient.

This approach to interpreting and implementing the POMDP model solutions can be carried over into an interpretation of a combined SOFA and IL-6 policy based on the POMDP results discussed in this section and the SOFA results presented for problem instance 2 (Caucasian patients, age 65 years and older) in Chapter 4.

5.5.1 Combined SOFA and IL-6 policy

This section considers an approach for utilizing the SOFA-based MDP model results in addition to the IL-6-based POMDP model results to inform clinical practice. As additional data become available, the general model proposed in Section 3 can be used to generate optimal results combining completely observable elements (e.g., the SOFA score) and partially observable elements (e.g., the IL-6 level). In the absence of these data, however, key results

Table 5.8: Updating π_t from Day t to $t+1$ When a Low Test Result is Received

π_t	π_{t+1}
0.10	0.85
0.20	0.89
0.30	0.92
0.40	0.93
0.50	0.95
0.60	0.96
0.70	0.97
0.80	0.98
0.90	0.98

from Chapters 4 and 5 are combined to form a general strategy for patient care using the following heuristic approach.

An important consideration when discussing combining the SOFA score and IL-6 information is the fact that the SOFA score is directly observable by the clinician and will be the basis on which most decisions are made in practice. The SOFA score will be considered before other additional information, such as the patient's IL-6 level, when making the discharge decision. For example, if a patient has a very high SOFA score, then the patient will be kept in the hospital, even if the patient's IL-6 level is low. Therefore, the IL-6 level does not provide useful information unless the clinician is considering the decision to discharge the patient from the hospital.

The following heuristic approach to a combined SOFA and IL-6 policy is proposed for Caucasian patients age 65 and older (corresponding to MDP problem instance 2 and POMDP problem instance E): Observe the patient's SOFA score value following the MDP policy as described in Table A2 in Appendix A. At the point at which the decision is to discharge the patient, *do not* discharge the patient, but instead order a cytokine test. If the cytokine

test result is low in the next stage and the patient's SOFA score is still within the range to discharge the patient, then discharge the patient.

If the cytokine test result is high in the next stage, then keep the patient in the hospital for two more days. If on the second day, the patient's SOFA score is still in the discharge range, then order another cytokine test. Again, if the test result is low and the patient's SOFA score is still in the discharge range, then discharge the patient. Otherwise wait two more days, as before, and order another cytokine test. Continue this process until the patient either dies or is discharged from the hospital.

This policy builds off of current practice with elements of the optimal SOFA and IL-6 policies described in this dissertation. In the IL-6 model, the decision to test rather than discharge the patient resulted in an average $(90-t)$-day life expectancy improvement of more than 4 days, or an average increase of more than 6.7% (for the case of 95% test accuracy and 0.5 days for test cost). Again, as additional data become available the general model proposed in Section 3 can be used to measure this improvement for a combined SOFA and IL-6 model, but it is assumed that similar improvements in patient survival can be expected from implementing the suggested heuristic policy. Until addition data become available for more robust modeling, these approaches help to provide structure to the current process and begin to give insight into the interpretation of complex modeling techniques like POMDPs.

5.6 CONCLUSIONS

The results presented in this chapter demonstrate the need for inexpensive, accurate testing procedures as well as accurate interpretation of test results. More importantly, however, is the suggestion that even in light of completely accurate and cost-free tests, it is not optimal to test all the time. This at first seems to be counterintuitive until one considers that additional information is only needed if it will change the decision to discharge the patient.

Additionally, while this model considers an individual patient's perspective, one must also consider the amount of time that the clinician and other health care providers spend in administering, processing, and analyzing test results. Avoiding unnecessary tests will help to reduce health care costs from the system perspective.

Using the results of the IL-6-based model together with the SOFA-based policies from Chapter 4, a heuristic policy is developed utilizing testing as a qualification step for patient discharge. As additional data become available, the general model proposed in Chapter 3 can be used to develop optimal SOFA- and IL-6-based testing and hospital discharge policies. Until that time, the general strategies presented in this chapter will help to provide structured strategies for utilizing new cytokine testing procedures as part of patient treatment.

From this study it is clear that the POMDP framework can be used to solve medical decision making questions in which aspects of the patient's health state can only be observed through a costly or inaccurate testing procedure.

6.0 CONTRIBUTIONS AND DIRECTIONS FOR FUTURE RESEARCH

6.1 CONTRIBUTIONS TO PATIENT CARE

Expanding upon research to date that has sought to compare various treatment alternatives, this dissertation is the first study to address the optimization of decisions made in the management of severe sepsis. In addition, this is the first model of sepsis to consider the question of cytokine testing, specifically its impact on expected patient survival. The GenIMS trial is the first study of its kind to provide enough patient data to test the effects of cytokine testing on patient survival. Using the results of this trial in an optimization model greatly extends the impacts of this research in the medical community.

In this dissertation, novel models utilizing the MDP and POMDP methodologies that have only recently been introduced to the medical community are presented. Not only is the framework of these models shown to be useful for modeling medical decision making questions in the management of sepsis, but these models are also analyzed mathematically to uncover clinical conditions that ensure specific types of optimal solutions.

Most importantly, using data from the GenIMS study, these models are then solved for problem instances constructed from actual patient-based data. The results are interpreted from a clinical perspective to give recommendations on specific strategies that can be used to inform clinical practice moving forward.

Finally, the results of both models motivate the need for additional medical research and demonstrate that, with additional data, these modeling techniques can be used to solve complex problems that are otherwise too difficult to analyze with common techniques used by the medical decision making research community today.

6.2 CONTRIBUTIONS TO STOCHASTIC OPTIMIZATION

The interesting problems found in critical care medicine, such as the management of severe sepsis, provide the operations research community with unique applications that contribute to the development of MDP and POMDP theory.

For the MDP model, this dissertation demonstrates clinical conditions for the existence of control-limit policies using an approach that differs from the standard method described by Puterman [89]. Computational experiments further illustrate these types of policies for several problem instances.

For the POMDP models, this dissertation explores the effects of test cost and accuracy on the optimal testing region and resulting expected patient survival. The results demonstrate that as testing cost decreases and test accuracy increases, the optimal testing region does not decrease. In other words, these results give further evidence for the need for less expensive, more accurate tests. In addition, the results indicate that testing should be used as a qualification for hospital discharge, but is not necessary if the clinician has already decided not to discharge the patient based on completely observable factors. For example, even in the case of no cost, completely accurate tests, it is not always optimal to test. This is because the additional information from the test result would not change the clinician's decision to keep the patient in the hospital and is therefore unnecessary. It is also important to consider the unquantified savings experienced when not performing unnecessary tests, such as an improvement in quality of life for the patient and the additional time that the clinician can spend with the patient, with other patients, or on other activities.

From a research perspective, it is especially beneficial to realize the applicability of theoretical models in practical applications. It is also helpful to test these models to recognize their weaknesses and additional opportunities for improvement. While additional data are needed from the application perspective, the dependence on extremely large quantities of data from the modeling perspective is also an issue.

The artificial intelligence community is already well underway with exploring methods by which to capture disease progression by means of influence diagrams and neural networks so that complete transition probability networks are not needed. Unfortunately, the GenIMS

study was not far enough along to take advantage of these techniques in this dissertation. The next section explores several future opportunities for research in this area.

6.3 DIRECTIONS FOR FUTURE RESEARCH

To further validate and gain insights into these models, future research efforts could include the use of simulation to compare actual strategies to the policies proposed in this dissertation. For example, Saka et al. [96] presented a simulation model that captured changes in a patient's SOFA score over time. The end points in this model were either patient discharge or death. As a direction for future research, this simulation model could be used to test the impact of the optimal SOFA-based MDP model results on patient length of stay in the hospital. These results could then be compared to the actual length of stay in the GenIMS trial to validate the effectiveness of the model results. Similarly, the model could be updated to incorporate patient cytokine levels. The updated model could then be used to test the impact of the optimal IL-6-based POMDP model result on the patient length of stay and these results could be compared to the actual length of stay. Finally, the average length of stay resulting from the application of each policy could be compared to assess the value of the individual models as compared to using the proposed heuristic strategy.

As additional data become available from clinical studies like GenIMS, the models presented in this dissertation can be used to solve increasingly robust problems involving clinical decisions in the management of severe sepsis. At the same time, modeling techniques are being developed to capture the intricate relationships between disease parameters. These techniques, once refined, will be able to solve, in a reasonable amount of time, increasingly complex problems in medical decision making.

In addition to these broader directions for research, several immediate applications of the research from this dissertation are apparent. First, these models can be used to explore testing and discharge decisions for each cytokine collected in the GenIMS trial, IL-10 and TNF, for example. This dissertation focused mainly on the application of the model to a specific example, IL-6, but can also be used to explore these other variables.

Second, as the interdependencies between the cytokine values are better understood as a result of the statistical analysis currently being conducted by the GenIMS investigators, the general POMDP model can be used to consider models with multiple cytokine variables. This dissertation, due to the data requirements and availability, only considered a single cytokine variable. However, as more of the statistical relationships are understood, the data requirements will decrease and become more manageable.

Third, further exploration into the algorithms currently being developed in the artificial intelligence community may uncover additional ways to solve increasing complex POMDP models with multiple variables. These methods, combined with the increased understanding of sepsis resulting from the GenIMS study, will also allow for the development of a more complex model.

An investigation into conditions for the existence of control-limit policies for both the MDP and POMDP models would be beneficial. In addition to the investigation of several structural properties, computational results were mainly used in this dissertation to indicate that these types of policies exists for these models. However, robust mathematical proofs would advance the current state of research in these areas. Unfortunately, only very restrictive assumptions on the models, too restrictive for this type of application, are necessary. As was already demonstrated for the MDP model, the restrictive condition did not hold for the data tested, even though control limit type policies appeared to exist as demonstrated through the computational results.

Finally, the exploration of other factors that influence clinical decision making, such as the costs of care, can be incorporated into these models. Currently, the immediate reward received for keeping a patient in the hospital is a full day of life. Considering factors such as cost of care would reduce this reward and would likely increase the discharge region for many if not all problem instances.

It is clear that mathematical modeling techniques such as MDPs and POMDPs are useful to model questions in the management of sepsis. The investigation of structural properties and the results demonstrated through computational experiments not only validate the applicability of the modeling techniques, but also serve to provide clinical insight on management strategies. It is hoped that in the future, as additional data and solution techniques become

available, that the operations research and medical decision making research communities will be able to work together to develop clinically robust optimization models that can be used to inform daily decisions made by clinicians in sepsis management and in other areas of patient treatment as well.

APPENDIX A

MDP OPTIMAL SOLUTION OUTPUT

Table A1: Optimal Solution to MDP Problem Instance 1

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	88.444	89.000	89.000	60.936	D	D	D	C
2	87.370	88.000	88.000	59.567	D	D	D	C
3	86.281	87.000	87.000	57.687	D	D	D	C
4	85.177	86.000	86.000	54.907	C	D	D	C
5	84.016	85.000	85.000	51.148	C	D	D	C
6	82.823	84.000	84.000	44.295	C	D	D	C
7	81.593	83.000	83.000	31.257	D	D	D	C
8	78.720	73.919	60.529	18.376	D	C	D	C
9	77.760	73.022	59.804	18.170	D	C	D	C
10	76.800	72.124	59.077	17.963	D	C	D	C
11	75.840	71.225	58.347	17.756	D	C	D	C
12	74.880	70.325	57.614	17.547	D	C	D	C
13	73.920	69.423	56.877	17.337	D	C	D	C
14	72.960	68.519	56.135	17.125	D	C	D	C
15	72.000	67.612	55.387	16.911	D	C	D	C
16	71.040	66.702	54.630	16.693	D	C	D	C
17	70.080	65.787	53.862	16.471	D	C	D	C
18	69.120	64.865	53.080	16.245	D	C	D	C
19	68.160	63.935	52.282	16.013	D	C	D	C
20	67.200	62.995	51.462	15.773	D	C	D	C
21	66.240	62.039	50.616	15.525	D	C	D	C
22	65.280	61.060	49.739	15.268	D	C	D	C
23	64.320	60.049	48.825	14.999	D	C	D	C
24	63.360	58.986	47.873	14.718	D	C	D	C
25	62.400	57.838	46.889	14.419	D	C	D	C
26	61.440	56.544	45.903	14.081	D	C	D	C
27	60.480	54.986	45.000	13.625	D	C	D	C
28	59.520	52.951	44.286	12.824	D	C	D	C
29	58.560	50.103	43.571	11.186	D	C	D	C

Table A2: Optimal Solution to MDP Problem Instance 2

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	85.367	83.606	63.617	31.023	D	D	C	C
2	84.408	82.667	61.567	28.743	D	D	C	C
3	83.449	81.727	58.123	25.187	D	D	C	C
4	76.947	72.695	59.095	17.014	D	C	C	C
5	76.053	71.632	57.770	16.271	D	C	C	C
6	75.158	70.561	56.167	14.861	D	C	C	C
7	74.263	69.541	54.269	11.550	D	D	D	C
8	67.739	57.245	45.556	6.415	D	D	D	C
9	66.913	56.547	45.000	6.353	D	D	D	C
10	66.087	55.849	44.444	6.291	D	D	D	C
11	65.261	55.151	43.889	6.229	D	D	D	C
12	64.435	54.453	43.333	6.167	D	D	D	C
13	63.609	53.755	42.778	6.105	D	D	D	C
14	62.783	53.057	42.222	6.043	D	D	D	C
15	61.957	52.358	41.667	5.980	D	D	D	C
16	61.130	51.660	41.111	5.918	D	D	D	C
17	60.304	50.962	40.556	5.856	D	D	D	C
18	59.478	50.264	40.000	5.794	D	D	D	C
19	58.652	49.566	39.444	5.732	D	D	D	C
20	57.826	48.868	38.889	5.670	D	D	D	C
21	57.000	48.170	38.333	5.608	D	D	D	C
22	56.174	47.472	37.778	5.545	D	D	D	C
23	55.348	46.774	37.222	5.483	D	D	D	C
24	54.522	46.075	36.667	5.418	D	D	D	C
25	53.696	45.377	36.111	5.349	D	D	D	C
26	52.870	44.679	35.556	5.261	D	D	D	C
27	52.043	43.981	35.000	5.112	D	D	D	C
28	51.217	43.283	34.444	4.749	D	D	D	C
29	50.391	42.585	33.889	3.648	D	D	D	C

Table A3: Optimal Solution to MDP Problem Instance 3

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	89.000	89.000	89.000	78.496	D	D	D	C
2	88.000	88.000	88.000	74.406	D	D	D	C
3	87.000	87.000	87.000	69.408	D	D	D	C
4	86.000	82.480	76.807	65.938	D	C	C	C
5	85.000	80.926	74.468	62.215	D	C	C	C
6	84.000	79.265	71.930	58.154	D	C	C	C
7	83.000	77.467	69.167	53.722	D	C	C	C
8	77.176	74.693	66.802	48.699	D	C	C	C
9	76.235	73.742	65.865	47.967	D	C	C	C
10	75.294	72.787	64.910	47.215	D	C	C	C
11	74.353	71.825	63.936	46.439	D	C	C	C
12	73.412	70.857	62.940	45.636	D	C	C	C
13	72.471	69.881	61.920	44.802	D	C	C	C
14	71.529	68.897	60.872	43.933	D	C	C	C
15	70.588	67.904	59.795	43.023	D	C	C	C
16	69.647	66.901	58.684	42.065	D	C	C	C
17	68.706	65.888	57.537	41.050	D	C	C	C
18	67.765	64.866	56.350	39.967	D	C	C	C
19	66.823	63.834	55.122	38.803	D	C	C	C
20	65.882	62.793	53.850	37.537	D	C	C	C
21	64.941	61.746	52.537	36.143	D	C	C	C
22	64.000	60.696	51.185	34.584	D	C	C	C
23	63.059	59.646	49.808	32.805	D	C	C	C
24	62.118	58.596	48.429	30.727	D	C	C	C
25	61.176	57.536	47.095	28.233	D	C	C	C
26	60.235	56.422	45.895	25.142	D	C	C	C
27	59.294	55.123	45.000	21.169	D	C	C	C
28	58.353	53.409	44.286	15.963	D	C	C	C
29	57.412	50.867	43.571	9.095	D	C	D	C

Table A4: Optimal Solution to MDP Problem Instance 4

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	89.000	82.188	63.846	48.813	D	C	C	C
2	88.000	82.735	65.670	50.940	D	C	C	C
3	87.000	83.469	68.335	54.972	D	C	C	C
4	83.210	86.000	71.667	64.029	C	D	D	C
5	81.901	85.000	70.833	57.175	C	D	D	C
6	80.519	84.000	70.000	45.807	C	D	D	C
7	79.048	83.000	69.167	26.537	D	D	D	C
8	76.875	70.685	56.919	2.000	D	C	C	C
9	75.938	69.827	56.234	2.000	D	C	C	C
10	75.000	68.968	55.546	2.000	D	C	C	C
11	74.063	68.106	54.855	2.000	D	C	C	C
12	73.125	67.242	54.160	2.000	D	C	C	C
13	72.188	66.374	53.460	2.000	D	C	C	C
14	71.250	65.501	52.752	2.000	D	C	C	C
15	70.313	64.621	52.034	2.000	D	C	C	C
16	69.375	63.732	51.302	2.000	D	C	C	C
17	68.438	62.831	50.551	2.000	D	C	C	C
18	67.500	61.912	49.775	2.000	D	C	C	C
19	66.563	60.971	48.965	1.999	D	C	C	C
20	65.625	59.998	48.108	1.998	D	C	C	C
21	64.688	58.983	47.189	1.996	D	C	C	C
22	63.750	57.911	46.185	1.992	D	C	C	C
23	62.813	56.760	45.066	1.984	D	C	C	C
24	61.875	55.504	43.792	1.969	D	C	C	C
25	60.938	54.099	42.312	1.938	D	C	C	C
26	60.000	52.487	40.560	1.875	D	C	C	C
27	59.063	50.568	38.460	1.750	D	C	C	C
28	58.125	48.167	35.945	1.500	D	C	C	C
29	57.188	44.921	33.027	1.000	D	C	D	C

Table A5: Optimal Solution to MDP Problem Instance 5

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	89.000	89.000	89.000	68.074	D	D	D	C
2	87.857	88.000	88.000	52.147	D	D	D	C
3	86.865	87.000	87.000	51.490	C	D	D	C
4	85.874	86.000	86.000	50.770	C	D	D	C
5	84.884	85.000	85.000	49.932	C	D	D	C
6	83.897	84.000	84.000	48.876	C	D	D	C
7	82.913	83.000	83.000	47.415	C	D	D	C
8	81.934	82.000	82.000	45.201	C	D	D	C
9	80.961	81.000	81.000	41.589	C	C	D	C
10	80.000	80.000	80.000	35.380	C	C	D	C
11	79.000	79.000	79.000	34.951	C	C	D	C
12	78.000	78.000	78.000	34.522	C	C	D	C
13	77.000	77.000	77.000	34.094	C	C	D	C
14	76.000	76.000	76.000	33.665	C	C	D	C
15	75.000	75.000	75.000	33.236	C	C	D	C
16	74.000	74.000	74.000	32.807	C	C	D	C
17	73.000	73.000	73.000	32.378	C	C	D	C
18	72.000	72.000	72.000	31.947	C	C	D	C
19	71.000	71.000	71.000	31.515	C	C	D	C
20	70.000	70.000	70.000	31.077	C	C	D	C
21	69.000	69.000	69.000	30.631	C	C	D	C
22	68.000	68.000	68.000	30.164	C	C	D	C
23	67.000	67.000	67.000	29.656	C	C	D	C
24	66.000	66.000	66.000	29.061	C	C	D	C
25	65.000	65.000	65.000	28.285	C	C	D	C
26	64.000	64.000	64.000	27.128	C	C	D	C
27	63.000	63.000	63.000	25.173	C	C	D	C
28	62.000	62.000	62.000	21.550	C	C	D	C
29	61.000	61.000	61.000	14.433	C	C	D	C

Table A6: Optimal Solution to MDP Problem Instance 6

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	86.810	89.000	89.000	39.267	D	C	C	C
2	87.447	82.914	70.400	23.894	D	C	C	C
3	86.453	81.972	69.600	23.625	D	C	C	C
4	85.459	81.028	68.800	23.337	D	C	C	C
5	84.465	80.082	68.000	23.002	D	C	C	C
6	83.472	79.133	67.200	22.536	D	C	C	C
7	82.478	78.176	66.400	21.723	D	C	C	C
8	81.484	77.211	65.600	19.980	D	C	C	C
9	80.491	76.235	64.800	15.747	D	D	D	C
10	75.789	60.532	48.000	13.333	D	C	C	C
11	74.842	59.782	47.400	13.167	D	C	C	C
12	73.895	59.033	46.800	13.000	D	C	C	C
13	72.947	58.283	46.200	12.833	D	C	C	C
14	72.000	57.533	45.600	12.667	D	C	C	C
15	71.053	56.784	45.000	12.500	D	C	C	C
16	70.105	56.034	44.400	12.333	D	C	C	C
17	69.158	55.284	43.800	12.167	D	C	C	C
18	68.210	54.534	43.200	12.000	D	C	C	C
19	67.263	53.784	42.600	11.833	D	C	C	C
20	66.316	53.034	42.000	11.667	D	C	C	C
21	65.368	52.283	41.400	11.500	D	C	C	C
22	64.421	51.532	40.800	11.333	D	C	C	C
23	63.474	50.778	40.200	11.167	D	C	C	C
24	62.526	50.022	39.600	11.000	D	C	C	C
25	61.579	49.259	39.000	10.833	D	C	C	C
26	60.632	48.487	38.400	10.667	D	C	C	C
27	59.684	47.697	37.800	10.500	D	C	C	C
28	58.737	46.873	37.200	10.333	D	C	C	C
29	57.789	45.990	36.600	10.167	D	C	C	C

Table A7: Optimal Solution to MDP Problem Instance 7

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	82.154	79.876	69.406	31.069	C	D	C	C
2	79.666	79.810	66.449	20.423	D	D	C	C
3	78.760	78.897	65.694	19.739	D	C	D	C
4	77.855	77.968	64.939	18.828	D	C	D	C
5	76.950	77.001	64.184	17.582	D	C	D	C
6	76.045	75.925	63.429	15.839	D	C	D	C
7	75.139	74.554	62.673	13.358	D	C	D	C
8	74.234	72.380	61.918	9.783	D	C	D	C
9	73.329	68.015	61.163	4.586	D	D	D	C
10	64.167	51.765	31.304	1.000	D	D	C	C
11	63.365	51.118	30.913	1.000	D	D	C	C
12	62.562	50.471	30.522	1.000	D	D	C	C
13	61.760	49.824	30.130	1.000	D	D	C	C
14	60.958	49.176	29.739	1.000	D	D	C	C
15	60.156	48.529	29.348	1.000	D	D	C	C
16	59.354	47.882	28.956	1.000	D	D	C	C
17	58.552	47.235	28.565	1.000	D	D	C	C
18	57.750	46.588	28.174	1.000	D	D	C	C
19	56.948	45.941	27.783	1.000	D	D	C	C
20	56.146	45.294	27.391	1.000	D	D	C	C
21	55.344	44.647	27.000	1.000	D	D	C	C
22	54.542	44.000	26.609	1.000	D	D	C	C
23	53.740	43.353	26.217	1.000	D	D	C	C
24	52.937	42.706	25.826	1.000	D	D	C	C
25	52.135	42.059	25.435	1.000	D	D	C	C
26	51.333	41.412	25.043	1.000	D	D	C	C
27	50.531	40.765	24.652	1.000	D	D	C	C
28	49.729	40.118	24.261	1.000	D	D	C	C
29	48.927	39.471	23.870	1.000	D	D	C	C

Table A8: Optimal Solution to MDP Problem Instance 8

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	89.000	89.000	89.000	62.943	D	D	D	C
2	88.000	88.000	88.000	61.823	D	D	D	C
3	87.000	87.000	87.000	60.423	D	D	D	C
4	86.000	86.000	86.000	58.551	C	C	D	C
5	85.000	85.000	85.000	56.605	C	C	D	C
6	84.000	84.000	84.000	53.317	C	C	D	C
7	83.000	83.000	83.000	47.256	C	C	D	C
8	82.000	82.000	82.000	35.461	C	C	D	C
9	81.000	81.000	81.000	35.043	C	C	D	C
10	80.000	80.000	80.000	34.624	C	C	D	C
11	79.000	79.000	79.000	34.206	C	C	D	C
12	78.000	78.000	78.000	33.787	C	C	D	C
13	77.000	77.000	77.000	33.368	C	C	D	C
14	76.000	76.000	76.000	32.949	C	C	D	C
15	75.000	75.000	75.000	32.531	C	C	D	C
16	74.000	74.000	74.000	32.111	C	C	D	C
17	73.000	73.000	73.000	31.692	C	C	D	C
18	72.000	72.000	72.000	31.271	C	C	D	C
19	71.000	71.000	71.000	30.847	C	C	D	C
20	70.000	70.000	70.000	30.418	C	C	D	C
21	69.000	69.000	69.000	29.978	C	C	D	C
22	68.000	68.000	68.000	29.516	C	C	D	C
23	67.000	67.000	67.000	29.009	C	C	D	C
24	66.000	66.000	66.000	28.409	C	C	D	C
25	65.000	65.000	65.000	27.618	C	C	D	C
26	64.000	64.000	64.000	26.436	C	C	D	C
27	63.000	63.000	63.000	24.455	C	C	D	C
28	62.000	62.000	62.000	20.835	C	C	D	C
29	61.000	61.000	61.000	13.857	D	D	D	C

Table A9: Optimal Solution to MDP Problem Instance 9

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	87.749	89.000	89.000	42.929	D	D	D	C
2	86.738	88.000	88.000	39.206	D	D	D	C
3	85.721	87.000	87.000	31.115	D	D	D	C
4	85.196	83.795	68.800	21.923	D	D	C	C
5	84.206	82.821	68.000	21.146	D	D	C	C
6	83.215	81.846	67.200	19.466	D	D	C	C
7	82.224	80.872	66.400	15.361	D	D	D	C
8	79.311	63.490	47.953	13.667	D	C	C	C
9	78.344	62.726	47.386	13.500	D	C	C	C
10	77.377	61.962	46.818	13.333	D	C	C	C
11	76.410	61.198	46.250	13.167	D	C	C	C
12	75.443	60.434	45.681	13.000	D	C	C	C
13	74.475	59.668	45.111	12.833	D	C	C	C
14	73.508	58.902	44.540	12.667	D	C	C	C
15	72.541	58.134	43.966	12.500	D	C	C	C
16	71.574	57.364	43.389	12.333	D	C	C	C
17	70.607	56.591	42.809	12.167	D	C	C	C
18	69.639	55.815	42.223	12.000	D	C	C	C
19	68.672	55.034	41.631	11.833	D	C	C	C
20	67.705	54.245	41.028	11.667	D	C	C	C
21	66.738	53.447	40.413	11.500	D	C	C	C
22	65.770	52.636	39.781	11.333	D	C	C	C
23	64.803	51.805	39.125	11.167	D	C	C	C
24	63.836	50.945	38.442	11.000	D	C	C	C
25	62.869	50.041	37.724	10.833	D	C	C	C
26	61.902	49.066	36.970	10.667	D	C	C	C
27	60.934	47.969	36.193	10.500	D	C	C	C
28	59.967	46.648	35.443	10.333	D	C	C	C
29	59.000	44.891	34.857	10.167	D	C	C	C

Table A10: Optimal Solution to MDP Problem Instance 10

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	88.321	89.000	89.000	43.525	D	D	D	C
2	87.328	88.000	88.000	40.571	D	D	D	C
3	86.336	87.000	87.000	34.288	D	D	D	C
4	84.925	84.346	78.182	24.441	D	D	C	C
5	83.938	83.365	77.273	23.517	D	D	C	C
6	82.950	82.385	76.364	21.496	D	D	C	C
7	81.963	81.404	75.455	16.524	D	D	D	C
8	79.897	67.368	61.500	10.250	D	C	C	C
9	78.923	66.551	60.750	10.125	D	C	C	C
10	77.949	65.735	60.000	10.000	D	C	C	C
11	76.974	64.918	59.250	9.875	D	C	C	C
12	76.000	64.102	58.500	9.750	D	C	C	C
13	75.026	63.285	57.750	9.625	D	C	C	C
14	74.051	62.469	57.000	9.500	D	C	C	C
15	73.077	61.652	56.250	9.375	D	C	C	C
16	72.103	60.835	55.500	9.250	D	C	C	C
17	71.128	60.018	54.750	9.125	D	C	C	C
18	70.154	59.201	54.000	9.000	D	C	C	C
19	69.179	58.382	53.250	8.875	D	C	C	C
20	68.205	57.563	52.500	8.750	D	C	C	C
21	67.231	56.740	51.750	8.625	D	C	C	C
22	66.256	55.913	51.000	8.500	D	C	C	C
23	65.282	55.078	50.250	8.375	D	C	C	C
24	64.308	54.229	49.500	8.250	D	C	C	C
25	63.333	53.353	48.750	8.125	D	C	C	C
26	62.359	52.431	48.000	8.000	D	C	C	C
27	61.385	51.426	47.250	7.875	D	C	D	C
28	60.410	50.271	46.500	7.750	D	C	D	C
29	59.436	48.851	45.750	7.625	D	C	D	C

Table A11: Optimal Solution to MDP Problem Instance 11

t	$V_t^*(h)$				$A_t^*(h)$			
	$h=1$	$h=2$	$h=3$	$h=4$	$h=1$	$h=2$	$h=3$	$h=4$
1	85.764	81.472	69.388	44.093	D	D	C	C
2	84.800	79.617	66.126	36.006	D	D	C	C
3	83.836	77.333	62.826	26.446	D	D	C	C
4	77.502	76.750	59.939	14.833	D	C	C	C
5	76.601	75.470	59.242	12.595	D	C	D	C
6	75.700	73.598	58.545	9.542	D	C	D	C
7	74.798	70.231	57.849	5.348	D	D	D	C
8	68.163	56.871	44.514	1.000	D	D	D	C
9	67.331	56.177	43.971	1.000	D	D	D	C
10	66.500	55.484	43.429	1.000	D	D	D	C
11	65.669	54.790	42.886	1.000	D	D	D	C
12	64.838	54.097	42.343	1.000	D	D	D	C
13	64.006	53.403	41.800	1.000	D	D	D	C
14	63.175	52.710	41.257	1.000	D	D	D	C
15	62.344	52.016	40.714	1.000	D	D	D	C
16	61.513	51.323	40.171	1.000	D	D	D	C
17	60.681	50.629	39.629	1.000	D	D	D	C
18	59.850	49.935	39.086	1.000	D	D	D	C
19	59.019	49.242	38.543	1.000	D	D	D	C
20	58.188	48.548	38.000	1.000	D	D	D	C
21	57.356	47.855	37.457	1.000	D	D	D	C
22	56.525	47.161	36.914	1.000	D	D	D	C
23	55.694	46.468	36.371	1.000	D	D	D	C
24	54.863	45.774	35.829	1.000	D	D	D	C
25	54.031	45.081	35.286	1.000	D	D	D	C
26	53.200	44.387	34.743	1.000	D	D	D	C
27	52.369	43.694	34.200	1.000	D	D	D	C
28	51.538	43.000	33.657	1.000	D	D	D	C
29	50.706	42.306	33.114	1.000	D	D	D	C

APPENDIX B

POMDP OPTIMAL SOLUTION OUTPUT

Table B1: Optimal Solution to POMDP Problem Instances A, B, and C

t	A		B		C	
	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$
1	0.21	0.89	0.23	0.79	0.26	0.66
2	0.21	0.89	0.23	0.79	0.26	0.66
3	0.21	0.89	0.23	0.79	0.26	0.66
4	0.21	0.89	0.23	0.78	0.26	0.65
5	0.21	0.89	0.23	0.78	0.26	0.65
6	0.21	0.89	0.23	0.78	0.26	0.65
7	0.21	0.89	0.23	0.78	0.26	0.64
8	0.21	0.89	0.23	0.78	0.26	0.64
9	0.21	0.89	0.23	0.78	0.26	0.64
10	0.21	0.89	0.24	0.77	0.26	0.63
11	0.21	0.89	0.23	0.77	0.26	0.63
12	0.21	0.88	0.23	0.77	0.26	0.62
13	0.21	0.88	0.24	0.76	0.26	0.62
14	0.21	0.88	0.23	0.76	0.26	0.62
15	0.21	0.88	0.23	0.76	0.27	0.61
16	0.21	0.87	0.24	0.75	0.27	0.6
17	0.2	0.87	0.23	0.75	0.26	0.6
18	0.21	0.86	0.23	0.74	0.28	0.59
19	0.2	0.86	0.24	0.73	0.27	0.58
20	0.19	0.85	0.23	0.72	0.26	0.57
21	0.2	0.84	0.21	0.71	0.26	0.56
22	0.18	0.83	0.24	0.69	0.28	0.55
23	0.18	0.81	0.23	0.68	0.23	0.54
24	0.18	0.79	0.2	0.66	0.27	0.51
25	0.15	0.76	0.24	0.62	0.29	0.49
26	0.18	0.72	0.22	0.6	0.22	0.48
27	0.12	0.67	0.12	0.55	0.24	0.43
28	0.12	0.58	0.26	0.44	0.37	0.37
29	0.37	0.37	0.37	0.37	0.37	0.37

Table B2: Optimal Solution to POMDP Problem Instances D, E, and F

t	D		E		F	
	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$
1	0.23	0.83	0.26	0.71	0.26	0.57
2	0.23	0.83	0.26	0.71	0.26	0.57
3	0.23	0.83	0.26	0.71	0.26	0.57
4	0.23	0.83	0.26	0.71	0.26	0.57
5	0.23	0.83	0.26	0.71	0.26	0.56
6	0.23	0.83	0.26	0.71	0.27	0.55
7	0.23	0.83	0.26	0.7	0.27	0.55
8	0.23	0.83	0.26	0.7	0.27	0.55
9	0.23	0.83	0.26	0.7	0.27	0.55
10	0.23	0.83	0.26	0.7	0.27	0.54
11	0.23	0.83	0.26	0.7	0.27	0.54
12	0.23	0.83	0.26	0.7	0.29	0.54
13	0.23	0.82	0.26	0.69	0.29	0.53
14	0.23	0.82	0.26	0.69	0.27	0.53
15	0.23	0.82	0.26	0.69	0.29	0.52
16	0.23	0.82	0.26	0.68	0.29	0.52
17	0.23	0.81	0.26	0.68	0.27	0.51
18	0.23	0.81	0.25	0.67	0.29	0.51
19	0.22	0.8	0.26	0.65	0.29	0.5
20	0.22	0.8	0.26	0.65	0.27	0.5
21	0.22	0.79	0.23	0.64	0.3	0.48
22	0.22	0.77	0.26	0.62	0.29	0.47
23	0.22	0.75	0.26	0.61	0.26	0.46
24	0.22	0.73	0.22	0.6	0.3	0.44
25	0.19	0.71	0.26	0.57	0.3	0.44
26	0.22	0.67	0.23	0.54	0.23	0.43
27	0.16	0.62	0.12	0.51	0.29	0.39
28	0.15	0.54	0.29	0.4	0.37	0.37
29	0.37	0.37	0.37	0.37	0.37	0.37

Table B3: Optimal Solution to POMDP Problem Instances G, H, and I

t	G		H		I	
	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$	$CL(T)$	$CL(D)$
1	0.25	0.76	0.26	0.63	0.27	0.48
2	0.25	0.76	0.26	0.63	0.27	0.47
3	0.25	0.77	0.26	0.63	0.27	0.47
4	0.25	0.77	0.26	0.62	0.27	0.47
5	0.25	0.77	0.26	0.62	0.27	0.47
6	0.25	0.77	0.26	0.62	0.29	0.47
7	0.25	0.77	0.26	0.62	0.29	0.46
8	0.25	0.77	0.26	0.62	0.29	0.46
9	0.25	0.77	0.26	0.62	0.29	0.46
10	0.25	0.77	0.27	0.62	0.29	0.46
11	0.25	0.77	0.26	0.62	0.29	0.45
12	0.25	0.76	0.26	0.61	0.29	0.44
13	0.25	0.76	0.27	0.61	0.29	0.44
14	0.23	0.76	0.26	0.61	0.29	0.44
15	0.25	0.75	0.26	0.61	0.3	0.43
16	0.23	0.75	0.27	0.6	0.3	0.43
17	0.23	0.75	0.26	0.6	0.3	0.43
18	0.25	0.75	0.26	0.6	0.3	0.43
19	0.23	0.74	0.27	0.59	0.3	0.42
20	0.23	0.74	0.26	0.59	0.3	0.41
21	0.23	0.73	0.26	0.57	0.3	0.41
22	0.23	0.72	0.27	0.56	0.3	0.41
23	0.23	0.71	0.26	0.55	0.3	0.4
24	0.23	0.68	0.23	0.53	0.3	0.4
25	0.21	0.66	0.22	0.51	0.3	0.39
26	0.23	0.62	0.23	0.5	0.28	0.39
27	0.15	0.58	0.2	0.47	0.37	0.37
28	0.19	0.51	0.32	0.37	0.37	0.37
29	0.37	0.37	0.37	0.37	0.37	0.37

APPENDIX C

ALGORITHMS

C.1 BACKWARD INDUCTION ALGORITHM

1. Set $t = N$ and $V_t^*(h) = r_N(h_N)$ for all $h_N \in \mathcal{H}$.
2. Substitute $t-1$ for t. For all $h_t \in \mathcal{H}$, perform steps a and b. Then go to step 3.
   a. Set
   $$V_t^*(h_t) = \max \begin{cases} r_t(h_t, D), \\ r_t(h_t, C) + \sum_{j=1}^{H+1} p_t(j|h_t, C) V_{t+1}(j). \end{cases}$$
   b. Set
   $$\mathcal{A}_t^*(h_t) = \arg\max \begin{cases} r_t(h_t, D), \\ r_t(h_t, C) + \sum_{j=1}^{H+1} p_t(j|h_t, C) V_{t+1}(j). \end{cases}$$
3. If $t = 1$, stop; otherwise return to step 2.

The reader is referred to Puterman [89] for more information.

C.2 MODIFIED BACKWARD INDUCTION ALGORITHM

Discretize the possible belief variable values according to a specified precision, such as 0.01. Let Π denote the set of all possible values of π_t for all $t \in \mathcal{N}$.

1. Set $t = N$ and
   $$V_t^*(\pi_t) = r_N(\pi_N) \quad \text{for all } \pi_N \in \Pi.$$

2. Substitute $t-1$ for t. For all $\pi_t \in \Pi$, perform steps a and b. Then go to step 3.
   a. Calculate $\gamma_t(o_{t+1}|\pi_t)$ for all $o_{t+1} \in \{H, L\}$ given the current value of π_t.
   b. Calculate $U(\pi_{t+1}|o_{t+1}, \pi_t) = \pi_{t+1}$ for all $o_{t+1} \in \{H, L, \emptyset\}$ given the current value of π_t. Round the calculated value of π_{t+1} to the closest value contained in Π.
3. For all $\pi_t \in \Pi$, perform steps a and b. Then go to step 4.
   a. Set

$$V_t^*(\pi_t) = \max \begin{cases} r_t(\pi_t, D), \\ r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t)), \\ r_t(\pi_t, C) V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t)). \end{cases}$$

   b. Set

$$\mathcal{A}_t^*(\pi_t) = \arg\max \begin{cases} r_t(\pi_t, D), \\ r_t(\pi_t, O) - c + \sum_{o_{t+1} \in \{L,H\}} \gamma_t(o_{t+1}|\pi_t) V_{t+1}^*(U(\pi_{t+1}|o_{t+1}, \pi_t)), \\ r_t(\pi_t, C) + \gamma_t(\emptyset|\pi_t, a_t) V_{t+1}^*(U(\pi_{t+1}|\emptyset, \pi_t)). \end{cases}$$

4. If $t = 1$, stop; otherwise return to step 2.

The reader is referred to Monahan [82] for more information.

BIBLIOGRAPHY

[1] J.H. Ahn and J. C. Hornberger. Involving patients in the cadaveric kidney transplant allocation process: A decision-theoretic approach. *Management Science*, 42:629–641, 1996.

[2] O. Alagoz. *Optimal Policies for the Acceptance of Living-donor and Cadaveric-donor Livers*. PhD thesis, University of Pittsburgh, 2004.

[3] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. The optimal timing of living-donor liver transplantation. *Management Science*, 50(10):1420–1430, 2004.

[4] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research*, 55(1):24–36, 2007.

[5] E. Altman. Denumerable constrained Markov decision processes and finite approximations. *Mathematics of Operations Research*, 19(1):169–191, 1994.

[6] D.C. Angus. Study design issues in sepsis trials. *Sepsis*, 4:7–13, 2000.

[7] D.C. Angus, W.T. Linde-Zwirble, J. Lidicker, G. Clermont, J. Carcillo, and M.P. Pinsky. Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Critical Care Medicine*, 29(7):1303–1310, 2001.

[8] K.J. Arrow, D. Blackwell, and M.A. Girshick. Bayes and minimax solutions of sequential decision problems. *Econometrica*, 17:213–244, 1949.

[9] K.J. Astrom. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10:174–205, 1965.

[10] R.E. Barlow and F. Proscan. *Mathematical Theory of Reliability*. John Wiley and Sons, New York, 1965.

[11] R. Bauerle, A. Rucker, T.C. Schmandra, K. Holzer, and A. Encke. Markov cohort simulation study reveals evidence for sex-based risk differences in intensive care unit patients. *American Journal of Surgery*, 179:207–211, 2000.

[12] R. Bellman. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684, 1957.

[13] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Volumes 1 and 2*. Athena Scientific, Bemont, Massachusettes, 2001.

[14] G.R. Bitran and D. Tirupati. Hierarchical production planning. In S.C. Graves, A.H.G. Rinnooy Kan, and P.H. Zipkin, editors, *Handbooks on Operations Research and Management Science*, volume 4, pages 523–568. North-Holland, Amsterdam, 1993.

[15] R.C. Bone. Why sepsis trials fail. *Journal of the American Medical Association*, 276:565–566, 1996.

[16] H. Burchardi and H. Schneider. Economic aspects of severe sepsis: A review of intensive care unit costs, cost of illness and cost effectiveness of therapy. *Pharmacoeconomics*, 22(12):793–813, 2004.

[17] L. Burns and D. Wholey. The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Medical Care*, 29:225–271, 1991.

[18] A.R. Cassandra. *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*. PhD thesis, Brown University, 1998.

[19] A. Cayley. Mathematical questions with their solutions, no. 4528. *Educational Times*, 23:18, 1875.

[20] C.J.Wiedermann. The limitations of observational studies on the treatment of severe sepsis. *Critical Care*, 6(6):546–547, 2002.

[21] P.D. Cleary, S. Greenfield, A.G. Mulley, S.G. Pauker, S.A. Schroeder, L. Wexler, and B.J. McNeil. Variation in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. *Journal of the American Medical Association*, 266:73–89, 1991.

[22] G. Clermont, D.C. Angus, K.G. Kalassian, W.T. Linde-Zwirble, N. Ramakrishnan, P.K. Linden, and M.R. Pinsky. Reassessing the value of short-term mortality in sepsis: Comparing conventional approaches to modeling. *Critical Care Medicine*, 31(11):2627–2633, 2003.

[23] G. Clermont, D.C. Angus, W.T. Linde-Zwirble, M.F. Griffin, M.J. Fine, and M.R. Pinsky. Does acute organ dysfunction predict patient-centered outcomes? *Chest*, 121(6):1963–1971, 2002.

[24] G. Clermont, V. Kaplan, R. Moreno, J.L. Vincent, W.T. Linde-Zwirble, B. Van Hout, and D.C. Angus. Dynamic microsimulation to model multiple outcomes in cohorts of critically ill patients. *Intensive Care Medicine*, 30:2237–2255, 2004.

[25] J. Cohen. The immunopathogenesis of sepsis. *Nature*, 420(6917):885–891, 2002.

[26] I. David and U. Yechiali. A time-dependent stopping problem with application to live organ transplants. *Operations Research*, 33(2):491–504, 1985.

[27] A. Davies, S. Ridley, J. Hutton, C. Chinn, B. Barber, and D.C. Angus. Cost effectiveness of drotrecogin alfa (activated) for the treatment of severe sepsis in the United Kingdom. *Anesthesia*, 60(2):155–162, 2005.

[28] G.T. de Ghellinck and G.D. Eppen. Linear programming solutions for separable Markovian decision problems. *Management Science*, 13(6):371–394, 1967.

[29] F. D'Epenoux. A probabalistic production and inventory problem. *Management Science*, 10(1):98–108, 1963.

[30] C. Derman. On minimax surveillance schedules. *Naval Research Logistics Quarterly*, 8(4), 1961.

[31] C. Derman. *Finite State Markovian Decision Processes*. Academic Press, New York, 1970.

[32] C. Derman and M. Klein. Some remarks on finite horizon Markovian decision models. *Operations Research*, 13(2):272–278, 1965.

[33] M.E. Drummond, B. O'Brien, G.L. Stoddart, and G.W. Torrance. *Methods for the Evaluation of Health Care Programmes, second edition*. Oxford Medical Publications, 1997.

[34] E. Dynkin. The optimum choice for the instant for stopping a Markov process. *Doklady Akademii Nauk SSR*, 150:238–240, 1963.

[35] J.E. Eckles. *Optimum replacement of stochastically failing systems*. PhD thesis, Stanford University, 1966.

[36] J.E. Eckles. Optimum maintenance with incomplete information. *Operations Research*, 16:1058–1067, 1968.

[37] P. Eggimann and D. Pittet. Infection control in the ICU. *Chest*, 120(6):2059–2093, 2001.

[38] E.A. Feinberg and A. Shwartz. Markov decision models with weighted discounted criteria. *Mathematics of Operations Research*, 19(1):152–168, 1994.

[39] E.Z. Ferenstein and E.G. Enns. Optimal sequential selection from a known distribution with holding costs. *Journal of the Americal Statistical Association*, 83(402):382–386, 1988.

[40] F.L. Ferriera, D.P. Bota, A. Bross, C. Melot, and J.-L. Vincent. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *Journal of the American Medical Association*, 286(14):1754–1758, 2001.

[41] M.J. Fine, L.J. Hough, A.R. Medsger, Y.H. Li, E.M. Ricci, D.E. Singer, T.J. Marrie, C.M. Coley, M.B. Walsh, M. Karpf, K.C. Lahive, and W.N. Kapoor. The hospital discharge decision in patients with community-acquired pneumonia: Results from the pneumonia PORT cohort study. *Archives of Internal Medicine*, 157:47–56, 1997.

[42] M.J. Fine, D.E. Singer, A.L. Phelps, B.H. Hanusa, and W.N. Kapoor. Differences in length of hospital stay in patients with community-acquired pneumonia: A prospective four hospital study. *Medical Care*, 31:371–380, 1993.

[43] P.R. Freeman. The secretary problem and its extensions. *International Statistics Review*, 51:189–206, 1983.

[44] T.S. Furgeson. Who solved the secretary problem? *Statistical Science*, 4:282–296, 1989.

[45] GenIMS. http://genims.org/, 2007.

[46] J. Gilbert and F. Mosteller. Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, 61:35–73, 1966.

[47] H.S. Gold and R.C. Moellering. Antimicrobial-drug resistance. *New England Journal of Medicine*, 335:1445–1453, 1996.

[48] M.R. Gold, J.E. Siegel, L.B. Russel, and M.C. Weinstein. *Cost Effectiveness in Health and Medicine*. Oxford University Press, 1996.

[49] A. Gosavi, T.K. Das, and A. Sarkar. A simulation-based learning automata framework for solving semi-Markov decision problems under long-run average reward. *IIE Transactions*, 36:557–567, 2004.

[50] E.A. Halm, M.J. Fine, T.J. Marrie, C.M. Coley, W.N. Kapoor, D.S. Obrosky, and D.E. Singer. Time to clinical stability in patients hospitalized with community-acquired pneumonia. *Journal of the American Medical Association*, 279(18):1452–1457, 1998.

[51] M. Hauskrecht and H. Fraser. Planning treatment of ischemic heart disease with partially observable Markov decision processes. *Artificial Intelligence in Medicine*, 18:221–244, 2000.

[52] A. Hordijk and L.C.M. Kallenberg. Constrained undiscounted stochastic dynamic programming. *Mathematics of Operations Research*, 9(2):276–289, 1984.

[53] R.S. Hotchkiss and I.E. Karl. The pathology and treatment of sepsis. *The New England Journal of Medicine*, 348(2):138–150, 2003.

[54] D.L. Hoyert, M.P. Heron, S.L. Murphy, and H.-C. Kung. Deaths: Final data for 2003. *National Vital Statistics Reports*, 54(13), 2006.

[55] C. Hu, W.S. Lovejoy, and S.L. Shafer. Comparison of some suboptimal control policies in medical drug therapy. *Operations Research*, 44(5):696–709, 1996.

[56] D.L. Iglehart. Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. *Management Science*, 9(2):259–267, 1963.

[57] J. Ivy. A maintenance model for breast cancer detection and treatment. Submitted for publication, 2007.

[58] J.A. Johnston. Determinants of mortality in patients with severe sepsis. *Medical Decision Making*, JUL-AUG:374–386, 2005.

[59] J.S. Kakalik. Optimum policies for partially observable Markov systems. Technical Report 18, Operations Research Center, Massachusetts Institute of Technology, 1965.

[60] E.P.C. Kao. Modeling the movement of coronary patients within a hospital by semi-Markov processes. *Operations Research*, 22(4):683–699, 1974.

[61] S. Karlin. Stochastic models and optimal policies for selling an asset. In K. J. Arrow, S. Karlin, and H. Scarf, editors, *Studies in Applied Probability and Management Science*, pages 148–158. Stanford University Press, Palo Alto, CA, 1962.

[62] J. Kasal, Z. Jovanovic, G. Clermont, L.A. Weissfeld, V. Kaplan, R.S. Watson, and D.C. Angus. Comparison of Cox and Gray's survival models in severe sepsis. *Critical Care Medicine*, 32(3):700–707, 2004.

[63] J.A. Kellum, L. Kong, M.P. Fink, L.A. Weissfeld, D.M. Yealy, M.P. Pinsky, J. Fine, A. Krichevsky, R.L. Delude, and D.C. Angus. Understanding the inflammatory cytokine response in pneumonia and sepsis: Results of the GenIMS study. Technical report, University of Pittsburgh, 2007.

[64] M. Klein. Inspection-maintenance-replacement schedules under Markovian deterioration. *Management Science*, 9(1):25–32, 1962.

[65] J.E. Kreke, M.D. Bailey, A.J. Schaefer, D.C. Angus, and M.S. Roberts. Modeling hospital discharge policies for patients with pneumonia-related sepsis. Submitted for publication, 2007.

[66] J.E. Kreke, A.J. Schaefer, D.C. Angus, C.L. Bryce, and M.S. Roberts. Incorporating biology into simulation models of organ allocation. In Yucesan and Chen, editors, *Proceedings of the 2002 Winter Simulation Conference*, pages 532–536, 2002.

[67] J.E. Kreke, A.J. Schaefer, and M.S. Roberts. Simulation and critical care modeling. *Current Opinion in Critical Care*, 10(5):395–398, 2004.

[68] A.M. Law and W.D. Kelton. *Simulation Modeling and Analysis, 3rd Edition*. McGraw-Hill Higher Education, New York, 2000.

[69] C. Lefevre. Optimal control of a birth and death epidemic process. *Operations Research*, 29(5):971–982, 1981.

[70] M. Lipsitch and B.R. Levin. Population dynamics of tuberculosis treatment: Mathematical models of the roles of non-compliance and bacterial heterogeneity in the evolution of drug resistance. *International Journal of Tuberculosis and Lung Disease*, 2(3):187–199, 1998.

[71] W.S. Lovejoy. Computationally feasible bounds for partially observed Markov decision processes. *Operations Research*, 39(1):162–175, 1991.

[72] W.S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28:47–66, 1991.

[73] P. Magni and R. Bellazzi. DT-Planner: An environment for managing dynammic decision problems. *Computer Methods and Programs in Biomedicine*, 54:183–200, 1997.

[74] P. Magni, S. Quaglini, M. Marchetti, and G. Barosi. Deciding when to intervene: A Markov decision process approach. *International Journal of Medical Informatics*, 60:237–253, 2000.

[75] A.S. Manne. Linear programming and sequential decisions. *Operations Research*, 23:785–795, 1960.

[76] M. Marchetti, S. Quaglini, and G. Barosi. Prophylactic splenectomy and cholecystectomy in mild hereditary spherocytosis: Analyzing the decision in different clinical scenarios. *Journal of Internal Medicine*, 244:217–226, 1998.

[77] J.C. Marshall. Clinical trials of mediator-directed therapy in sepsis: What have we learned? *Intensive Care Medicine*, 26(Suppl.l):S75–83, 2000.

[78] D. McCormick, M.J. Fine, C.M. Coley, T.J. Marrie, J.R. Lave, D.S. Obrosky, W.N. Kapoor, and D.E. Singer. Variation in length of hospital stay in patients with community acquired pneumonia: Are shorter stays associated with worse medical outcomes? *Journal of General Internal Medicine*, 11(suppl)(72), 1996.

[79] G.E. Monahan. *On Optimal Stopping in a Partially Observable Markov Process with Costly Information*. PhD thesis, Northwestern University, 1977.

[80] G.E. Monahan. Optimal stopping in a partially observable, binary-valued Markov chain with costly perfect information. Working Paper MS-79-4, College of Industrial Management, Georgia Institute of Technology, 1979.

[81] G.E. Monahan. Optimal stopping in a partially observable Markov process with costly information. *Operations Research*, 28(6):1319–1334, 1980.

[82] G.E. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28:1–16, 1982.

[83] D. Naimark, M.D. Krahn, G. Naglie, D.A. Redelmeier, and A.S. Detsky. Primer on medical decision analysis: Part 5 - Working with Markov processes. *Medical Decision Making*, 17(2):913–919, 1997.

[84] C.H. Papadimitriou and J.N. Tsitsiklis. The complexity of Markov decision processes. *Mathematics of Operations Research*, 12(3):441–450, 1987.

[85] N.B. Peek. Explicit temporal models for decision-theoretic planning of clinical management. *Artificial Intelligence and Medicine*, 15(2):135–154, 1999.

[86] D.B. Petitti. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis*. Oxford University Press, 2000.

[87] J.D. Petruccelli. On a best choice problem with partial information. *Annals of Statistics*, 8:1171–1174, 1980.

[88] L.K. Platzman. *Finite-memory estimation and control of finite probabalistic systems*. PhD thesis, Massachusetts Institute of Technology, 1977.

[89] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York, 1994.

[90] A.A. Quartin, R.M. Schein, D.H. Kett, and P.N. Peduzzi. Magnitude and duration of the effect of sepsis on survival. Department of Veterans Affairs Systemic Sepsis Cooperative Studies Group. *Journal of the American Medical Association*, 277:1058–1063, 1997.

[91] M.S. Rangel-Frausto, D. Pittet, M. Costigan, T. Hwang, C.S. Davis, and R.P. Wenzel. The natural history of the systemic inflammatory response syndrome (SIRS): A prospective study. *Journal of the American Medical Association*, 273(2):117–123, 1995.

[92] M.S. Rangel-Frausto, D. Pittet, T. Hwang, R.F. Woolson, and R.P. Wenzel. The dynamics of disease progression in sepsis: Markov modeling describing the natural history and the likely impact of effective antisepsis agents. *Clinical Infectious Diseases*, 27:185–190, 1998.

[93] O. Roman-Marchant, C.E.A. Orellana-Jimenez, D. De Backer, C. Merlot, and J.-L. Vincent. Septic shock of early or late onset: Does it matter? *Chest*, 126(1):173–178, 2004.

[94] S.M. Ross. Quality control under Markovian deterioration. *Management Science*, 17(9):587–596, 1971.

[95] S.M Ross. *Introduction to Stochastic Dynammic Programming*. Academic Press, New York, 1984.

[96] G. Saka, J.E. Kreke, A.J. Schaefer, D.C. Angus, and M.S. Roberts. Predicting disease progression using dynamic microsimulation in pneumonia-related sepsis. To appear in *Critical Care*, 2007.

[97] S.M. Samuels. Minimax stopping rules when the uderlying distribution is uniform. *Journal of the American Statistical Association*, 76:188–197, 1981.

[98] A.J. Schaefer, M.D. Bailey, S.M. Shechter, and M.S. Roberts. Modeling medical treatment using Markov decision processes. In M. Brandeau, F. Sainfort, and W. Pierskalla, editors, *Handbook of Operations Research/Management Science Applications in Health Care*, pages 597–616. Kluwer Academic Publishers, 2004.

[99] R.D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, 1986.

[100] S. M. Shechter, M. D. Bailey, A. J. Schaefer, and M. S. Roberts. The optimal time to initiate HIV therapy under ordered health states. Technical report, University of Pittsburgh, 2007.

[101] S.M. Shechter, C.L. Bryce, O. Alagoz, J.E. Kreke, J.E. Stahl, A.J. Schaefer, D.C. Angus, and M.S. Roberts. A clinically based discrete event simulation of end-stage liver disease and the organ allocation process. *Medical Decision Making*, 25(2):199–209, 2005.

[102] S.M. Shechter, A.J. Schaefer, R.S. Braithwaite, and M.S. Roberts. Increasing the efficiency of monte carlo cohort simulations with variance reduction techniques. *Medical Decision Making*, 26(5):550–553, 2006.

[103] R.D. Smallwood and E.J. Sondik. The optimal control of partially observable Markov decision processes over a finite horizon. *Operations Research*, 21(5):1071–1088, 1973.

[104] S.M.Shechter, R.S. Braithwaite, A.J. Schaefer, and M.S. Roberts. Modeling the progression and treatment of HIV. In R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 953–959, 2004.

[105] E.J. Sondik. *The Optimal Control of Partially Observable Markov Processes*. PhD thesis, Stanford University, 1971.

[106] E.J. Sondik. The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Operations Research*, 26(2):282–304, 1978.

[107] F.A. Sonnenberg and J.R. Beck. Markov models in medical decision making: A practical guide. *Medical Decision Making*, 13(4):322–338, 1993.

[108] D.J. Spiegelhalter. Bayesian graphical modelling: A case-study in monitoring health otucomes. *Applied Statistics*, 47:115–133, 1998.

[109] T.J. Stewart. Optimal selection from a random sequence with learning of the underlying distribution. *Journal of the American Statistical Association*, 73:775–780, 1978.

[110] T.J. Stewart. The secretary problem with an unknown number of options. *Operations Research*, 29(1):130–145, 1981.

[111] X. Su, S.A. Zenios, and G.M. Chertow. Incorporating recipient choice in kidney transplantation. *Journal of the American Society of Nephrology*, 15:1656–1663, 2004.

[112] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Medonca, H. Bruining, C.K. Reinhart, R.M. Suter, and L.G. Thijs. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Medicine*, 22:707–710, 1996.

[113] A. Wald. *Sequential Analysis*. Wiley, 1947.

[114] A. Wald and J. Wolfowitz. Optimal character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19:326–339, 1948.

[115] E.C.Y. Wang, T.H. Grasela, and C.A. Walawander. Applying epidemiology-based outcomes research to clinical decision-making. *Pharmacoeconomics*, 15(4):385–393, 1999.

[116] D. Weycker, K.S. Akhras, J. Edelsberg, D.C. Angus, and G. Oster. Long-term mortality and medical care changes in patients with severe sepsis. *Critical Care Medicine*, 31(9):2316–2123, 2003.

[117] A.P. Wheeler and G.R. Bernard. Current concepts: Treating patients with severe sepsis. *The New England Journal of Medicine*, 340(3):207–214, 1999.

[118] C.C. White and W.T. Scherer. Solution procedures for partially observed Markov decision processes. *Operations Research*, 37(5):791–797, 1989.

[119] F. Zeni, B.F. Freeman, and C. Natanson. Anti-inflammatory therapies to treat sepsis and septic shock: A reassessment. *Critical Care Medicine*, 25:1095–1100, 1997.

[120] S.A. Zenios, G.M. Chertow, and L.M. Wein. Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research*, 48(4):549–569, 2000.

[121] S.A. Zenios, L.M. Wein, and G.M. Chertow. Evidence-based organ allocation. *The American Journal of Medicine*, 107:52–61, 1999.