Use of Markov Decision Process Models in Preventive Medicine

by

Gizem Sultan Nemutlu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Management Sciences

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Evrim Didem Gunes Assoc. Prof., Coll. of Business Admin. and Econ., Koc University
Supervisor:	Fatih Safa Erenay Prof., Dept. of Management Science, University of Waterloo
Internal Member:	James H. Bookbinder Prof., Dept. of Management Science, University of Waterloo
Internal Member:	Qi-Ming He Prof., Dept. of Management Science, University of Waterloo
Internal-External Member:	Ali Elkamel Prof., Dept. of Chemical Engineering, University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The biggest trade-off when proposing health care policies is about balancing the effectiveness and the practicality of the policies. The optimal policies providing benchmark performances can be driven through using operations research tools; however, they usually have complex structures that are necessary to sufficiently represent various aspects of the system being modeled. There are also policies either proposed in guidelines or followed in practice but they often vary with the system characteristics, i.e., preferences of the clinicians, available resources of the clinics, etc. Therefore, standardized, simple yet effective policies are needed for many healthcare applications, including preventive medicine. At this point, we study developing health care delivery policies that maximize the effect of the preventive interventions, while providing applicable policy structures that can be easily followed by health practitioners in practice. We focus on two applications of preventive medicine: childhood vaccine administration practices in developing countries; and colorectal cancer screening and surveillance.

Vaccine administration practices in developing countries suffer from open-vial wastage. Doses remaining from opened vials are disposed at the end of a day, due to lack of appropriate cold storage conditions. We propose administering vaccines *from different sizes of multi-dose vials* to address the open-vial wastage problem. We utilize a Markov decision process model to maximize the expected total number of doses administered via reducing vaccine wastage. The model dynamically decides which size of a multi-dose vial to open next, and when to terminate vaccination service for the day, given the time remaining in the replenishment cycle and available vaccine stocks. We show that the optimal policies are of control-limit type. Using data for routine pediatric vaccines, we show that the proposed optimal policies could cost-effectively reduce open-vial wastage and significantly improve the covered vaccine demand. We also analyze the initial vaccine inventory composition that specifies how many vials of each size should be kept in stock. We show that the optimal policy for the right vaccine inventory composition may improve the expected vaccine demand covered up to target levels without early termination of vaccination service while realizing reasonably small or no additional cost. Although the number of system

variables being tracked in our state space is manageable, the optimal policies still require significant effort to be adopted in practice. That is especially challenging in developing countries, where the resources, e.g., clinic staff, are limited. Therefore, we introduce *simple vaccine administration policies* that are developed with the guidance of the insights from our numerical and structural analyses. Our insights on the simple vaccine administration policies show that these policies can provide promising performance, in terms of costs and expected vaccine demand covered, compared to the optimal policies while requiring only a single system variable, i.e., time of a decision, to be monitored.

Colonoscopy screening prevents, and early-detects colorectal cancer (CRC), one of the most common and deadliest cancers in the world. Considering that the risk of developing CRC significantly increases after age 50, and that the North American population is aging, the colonoscopy screening and follow-up policies employed by gastroenterologists play a vital role in the well-being of the population. Existing clinical guidelines recommend colonoscopy screening policies that are shown to be cost-effective in CRC prevention and early detection. Nevertheless, almost half the practitioners do not follow these guidelines, indicating controversy around the best CRC screening practices. Several studies analyze alternative CRC screening policies using simulation and mathematical models. Especially, dynamic alternative policies, derived by a stochastic dynamic programming approach, can significantly increase health outcome improvements due to CRC screening and follow-up. However, under dynamic policies, colonoscopy screening and surveillance intervals significantly vary in factors such as age, gender, and personal history, which are harder to implement for clinicians. Our study on this second application aims at deriving efficient and simpler-to-implement colonoscopy screening and follow-up policies, but that perform closely to the optimal policies. We employ a patient-level discrete-event simulation model, built and validated using real data, to mimic CRC progression in asymptomatic and higher-risk individuals. We estimate the expected life-years, age-based risk of having CRC, CRC mortality, costs associated with CRC screening, and the number of required colonoscopies for a large set of screening policies. We evaluate the performances of all relevant simpler-to-implement colonoscopy policies, including the periodic screening policies currently used by practitioners, and all feasible periodic policies with n-period switch times (for $n = 0, 1, 2$).

Our analysis identifies under the parameter settings under which alternative and simpler policies are sufficient to provide close-to-optimal performance. These results provide insights on the types of policies on which to focus in future studies, for researchers from both medical and operational research fields.

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor, Professor Fatih Safa Erenay, for his invaluable guidance and advisory throughout my PhD. He has always been an in-act role model for me from who I can observe how to be a dedicated and determined researcher, and, at the same time, a gracious and modest colleague. Without his empathy, courage, and grace, I would not be able to overcome many of the hard times during my PhD. I am grateful that I have had a supervisor who is there for me every single time I needed his help.

The members of the Examining Committee have improved this thesis through their comments and suggestions. I am grateful for their time and effort spent in reading the manuscript, and providing their feedback.

The courses I have taken at the University of Waterloo have been instrumental in my development and in shaping this thesis. I would like to thank the instructors of these courses for making hard concepts easy to understand and introducing me to new topics and ideas: Professors James H. Bookbinder, Samir Elhedhli, Qi-ming He, Bon Koo, Levent Tuncel.

I would like to personally thank Professor James H. Bookbinder for many reasons. He has been a great source of help and motivation over the years. I truly enjoyed every piece of our countless conversations since my first year. His passion for teaching and research has always inspired me. I have learned a lot from him that I cannot possibly learn by myself spending years. I am sincerely grateful to him for his invaluable guidance and support.

I also would like to thank my friends in Waterloo, my time in PhD has been a memorable one thanks to them. Thanks to Zehra and Zeynep, I have always known that I also have a family here away from home. Many thanks to Ahmet, Mustafa, Taghi, Alpay, Ugur, Merve, Burak, Emre, Bahar, Forough, Abdelhalim, and many others for being true friends over the years.

I am indebted to my family for their motivation and unconditional support. I always feel the luckiest to not only have my parent's support but also have my aunt's continuous

and unconditional support over the years. Without her, I could not have achieved many of the things I have today. I am also grateful to have the greatest siblings, Burcu and Burak.

Last but not least, I would like to thank my husband Ozden Onur Dalgic who is my greatest support and closest friend in this journey. Being in the same building every day, working with the same supervisor, and getting through the same hard times are oftentimes meant hard to manage. He has been very successful managing all of these and making home a peaceful place away from any stress. I am very grateful for him and our family of Kemal Altan, Kepler, and Luna. I could not wish for more.

Dedication

As every other thing in my life, this was, is, and will be forever dedicated to my boy, Kemal Altan.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis focuses on developing health care delivery policies maximizing the effects of the preventive interventions while providing applicable policy structures that can be easily followed by health clinicians in practice. Specifically, two distinct application areas, childhood vaccine administration and colorectal cancer screening, are studied. We employ Markov Decision Process (MDP) models to derive the optimal policies for both problems. While the optimal policies provide significant improvements in health outcomes, they may sometimes be too complex to adopt in practice by health care professionals. Therefore, this thesis also considers the applicability and practicality of the proposed optimal policies, and aims at obtaining necessary insights from the problem analysis in order to use them when investigating simpler-to-implement policies.

In this chapter, Chapter 1, we present a brief review of the recent operations research and management sciences (OR/MS) applications in health care, and emphasize the importance of the preventive interventions that we study in this thesis.

1.1 Preface

The health care industry has grown exponentially, in line with the aging population of the world. Recent studies project that by 2050, the number of older persons in the world will exceed the number of young for the first time. The proportion of elderly, estimated to reach 21% by then, was around 10% in the early 2000s McNicoll (2002). Countries spend a great proportion of their gross domestic product (GDP) to provide effective and efficient health services. According to The World Bank, the health expenditure to cover the preventive and curative health services, family planning activities, nutrition activities, and emergency aid represented 17.1% and 10.4% of GDP in the US and Canada, respectively, between 2011-2015 (The World Bank (2016)). As health care has been pointed out as one of the biggest economic and social challenges, health care policy makers are expected to effectively and efficiently design their processes to deal with such a challenge.

Operations Research (OR) has been promoted for many decades to tackle the challenges in health care issues, since OR offers numerous methodologies and solution techniques. Some of the optimization issues tackled are related to health care delivery (e.g., capacity planning and resource scheduling, logistics, demand forecasting), health care finance (e.g., economic policy analysis, dynamic resource allocation), public policy analysis (e.g., disease management, vaccination, bio-terrorism, health technology assessment, policy evaluation and improvement), medical decision making (patient's and/or policy maker's action models), and pharmaceutical economics (e.g., economics of vaccine and drug development). Although many of the problems in the health care operations research literature may have similar analytic characteristics as the problems in other industries, measuring the performances of health care policies in terms of lives and utilities of individuals makes those problems unique. Health care OR problems mainly consider the possibilities of death or quality of remaining life, quality and value of the health care systems outcomes, the shared decisions among several decision makers (policy administrators such as governments and insurance companies, physicians, nurses), the utilization of resources for diagnosis and treatments, and the concept of health care access as a right of citizens in society (i.e., equity) Pierskalla and Brailer (1994). While considering those, both deterministic and

stochastic approaches have been commonly used in health care decision making. More thorough reviews of the related literature (OR applications in health care) can be found in Kahraman and Topcu (2018); Ayer et al. (2014); Dobrzykowski et al. (2014); Lakshmi and Iyer (2013); Hulshof et al. (2012); Rais and Viana (2011); Zhang et al. (2011); Denton et al. (2011); Brailsford et al. (2009); Gupta and Denton (2008); Cayirli and Veral (2003).

Mathematical Programming (linear programming, integer programming, etc.) is one of the commonly used OR tools for dealing with health care problems. Examples of such problems generally originate from operational problems such as hospital admissions scheduling, staff assignments, operations room scheduling, selection of service locations, budget allocation, and inventory management decisions (Burke et al., 2004; Cheang et al., 2003; Daskin and Dean, 2005; Ernst et al., 2004; Nicholson et al., 2004). Providing effective and efficient solutions to those problems significantly improves the quality of health care services, while reducing the resulting costs. For example, according to standards in US, ambulances should reach their destinations within ten minutes, 95% of the time. Although increasing fleet sizes and locating ambulances at broader sites may provide a strategy satisfying the service requirements, it is definitely not a cost-efficient solution. However, Repede and Bernardo (1994) proposes a maximal-expected-coverage location model (integer programming model), and integrates this model into a decision support system to relocate ambulances in order to maximize the expected total demand that can be served within the time limit without increasing fleet sizes. The challenges of such problems incorporate the uncertainty of applications into decisions. That is, for the ambulance location model, the demand pattern changes over time and the fleet size is not constant, since the dispatched ambulances are not available for demands observed during the time that they are away for service. Hence, utilizing OR tools in health care enables decision makers to derive strategies that incorporate such uncertainties and propose more realistic solutions.

Queueing models have been also used in the OR health care literature to analyze patient flows, waiting times, patient delays, utilization of resources, system design, and appointment systems. Review of related literature is presented in Fomundam and Herrmann (2007); Lakshmi and Iyer (2013); Nosek and Wilson (2001); Preater (2002). The key component of high-quality health care is to provide timely access to care with limited resources.

However, patient delays are inevitable for many of the health care services, as capacities and resources are limited. To illustrate, escalation of overcrowding in an emergency department (ED) may affect quality and access of health care. Green et al. (2006) studies the effectiveness of ED provider staffing using queueing models (M/M/s) to reduce the fraction of patients who leave the ED without being served. Using queueing models to analyze the current system provides important insights, as queueing models identify the bottlenecks of the current system and determine the areas for improvements. In this particular study, the results show that the proportion of patients who left the ED without being served can be decreased by 23% if the provider hours are increased 3.1% per week. This improvement illustrates the effectiveness of queueing system in reducing waiting times and delays.

Using simulation models (discrete event simulation, Monte Carlo simulation, system dynamics, and agent-based modeling) is a promising alternative for applications of OR in health care, especially when queueing models are too complex to be theoretically analyzed without simplifying assumptions. A review of applications of simulation in health care can be found in Brailsford et al. (2009); Günal and Pidd (2010); Jun et al. (1999); Katsaliaki and Mustafee (2011). In addition to providing a comprehensive analysis of the current system similar to what queueing models provide, simulation models also visualize how particular changes (decisions) may affect the whole system. However, simulation models require a quality of data which is difficult to be obtained from such complex systems. Moreover, the activities usually involve many different activities (e.g., physicians may serve several patients at once). Thus, determining service times is also one challenge for such systems.

Markov Decision Processes (MDP) models are also used considerably in health care, especially in medical decision making. Güneş and Örmeci (2018); Denton et al. (2011); Schaefer et al. (2005) present a thorough review of the studies using MDP models in medical decision making and a discussion to reveal the appropriate application areas of MDPs within the health care context. The specific characteristics of problems in medical decision making that need to be captured by MDP models are: the continuous risks associated with decisions over the time (e.g., risk of post-treatment complications if the decision is *do treatment*, or risk of disease progress if the decision is *do nothing*); the timing of events (e.g., developing colorectal cancer at age 50 versus at age 80); the timing of decisions

(e.g., scree every year versus every 5 years); the probability of observing the same event more than once (Sonnenberg and Beck, 1993). Hence, such characteristics require making sequential decisions under uncertainty, that can be properly modeled using MDPs. Since the optimal policies in this thesis are derived using MDP models, Section 1.2 is devoted to review MDP models in health care, specifically the models for treatment and prevention of diseases.

1.2 Use of Markov Decision Processes Models in Disease Prevention

In this section, a brief introduction to MDP models is given. Additionally, a review of health care problems using MDP models in disease prevention is presented. Although there are many studies using MDP models in health care, providing a comprehensive review of related literature is beyond the scope of this thesis. An extensive analysis of application of OR in prevention, detection, and treatment of disease is presented in Güneş and Örmeci (2018); Zhang et al. (2011).

Making decisions under uncertainty is a challenge in health care. Most of the time, decision makers include physicians, patients, and third-party-payers. Each of them has different constraints, objectives, and preferences that directly affect the decisions and the policy performances. Moreover, availability of resources also affects decisions, in the majority of cases, limiting the performances of the policies. All those bring uncertainty to the system; thus, decisions should be dynamically evaluated over the time to handle the uncertainty. MDP is a strong methodology to evaluate such dynamic and complex decisions. Stochastic modeling techniques such as discrete-event simulation or Markov models, generally examine one particular policy at a time for a fully specified stochastic model but they do not provide an optimal policy. On the other hand, MDP models implicitly evaluate every feasible policy to provide the policy that optimizes a particular objective (e.g., maximizing expected total reward) while assuming that decisions are made at various time epochs. Although, MDP is a strong methodology to analyze complex systems,

it has also some drawbacks. The number of feasible solutions may rapidly increase with the size of the problem, thus, solving MDPs exactly becomes challenging. Furthermore, MDP models require quality data to better define system characteristics, specifically, to estimate transition probabilities for the next epoch. However, in medical decision making, collecting quality data is also challenging.

We now briefly describe a Markov decision process model. Puterman (2014) defines an MDP model as a particular sequential decision model which consists of five elements: decision epochs (t) from a set of either discrete or continuum time points, states (s) and actions (a) from finite or countably infinite sets (discrete), transition probabilities ($p_t(j|s,a)$) determining the state at the next decision epoch, and a real valued function of rewards ($r_t(s,a)$). Suppose that a decision maker considers a probabilistic system which evolves through time. After observing (or partially observing) the current state, the action is chosen at the decision epoch, and a reward (or cost) is received. The actions chosen affect the system and the state of the system may change in accordance with these actions.

The objective of MDP models is to choose a sequence of actions that optimizes some predetermined criterion. MDP models have the Markovian property, that is; the state of the system prior to any decision, the rewards, and the future transitions depend only on the current decision (independent of the past). However, MDPs allow decision makers to bind previous, current, and future system decisions by properly defining the system states (Schaefer et al., 2005). We can categorize MDPs into many different groups based on the system characteristics: finite-horizon; infinite-horizon; discrete-time; continuous-time; fully observable; partially observable; semi-Markov decision processes. For a complete coverage of MDPs, Bellman and Dreyfus (2015); Bertsekas et al. (1995); Puterman (2014) can be referred.

The focus of our study is the use of MDP models in preventive medicine. The following sections discuss the vaccination practices and the cancer screening as they are the most prominent examples of preventive medicine applications in the OR literature.

1.2.0.1 Vaccination Practices

Organization et al. (2008) estimated that the numbers of deaths among children under 5-year old, from diseases that are preventable by vaccination in 2008 are 199,000 for HIB, 195,000 for pertussis, 118,000 for measles, 59,000 for neonatal tetanus, 2,000 for tetanus (non-neonatal), 476,000 for pneumococcal disease, and 453,000 for rotavirus. Moreover, seasonal epidemics, which can also be prevented via proper vaccination, cause a devastating burden. That is, worldwide, annual epidemics result in approximately 3 to 5 million cases of severe illness, and 250,000 to 500,000 deaths (Fauci, 2006). Therefore, vaccine practices should be designed effectively and efficiently to decrease the burden of vaccine-preventable diseases.

OR tools can be employed to tackle many of the challenges associated with the utilization of vaccines that prevent a number of diseases. One challenge is to incorporate specific requirements of vaccines into decisions, such as multi-dose vaccine formats must be administered within a minimum or maximum time window. Another challenge is conflicting vaccines: certain vaccines interact with others, thus, when decisions are made, vaccine specifics should be considered. Moreover, recently new vaccines have been introduced that are combinations of different vaccines (multi-valent) that can cover multiple diseases. Decisions on their administration should be evaluated separately. Also prior to introduction of new vaccines, determining optimal design of vaccines is required. Decisions associated with the vaccine supply chains are also quite challenging: yields are uncertain; demand levels are difficult to estimate; designing logistics, determining storage location of vaccine supplies, quantity of supplies and storage capacities (especially for the vaccine formats requiring cold chains) reveal complex decisions.

There are many studies in the OR literature visiting the challenges listed above for childhood (pediatric) and influenza vaccinations as well as vaccination for bio-defense (Dalgıç et al., 2017; Chick et al., 2008; Jacobson et al., 1999, 2003; Mamani et al., 2013; Medlock and Galvani, 2009; Mofrad et al., 2014, 2016; Ozaltin et al., 2011; Yaesoubi and Cohen, 2011a). However, to the best of our knowledge, there are only three studies (Mofrad et al., 2014, 2016; Yaesoubi and Cohen, 2011a) using MDP models to analyze vaccination prac-

tices. Yaesoubi and Cohen (2011a) describes an epidemic influenza with an SIR model and develops a MDP model that dynamically determines the optimal health policies maximizing the overall population's health during the epidemic. The model provides adaptive decisions as the system characteristics are updated, based on real-time data available during the epidemic. Mofrad et al. (2014) and Mofrad et al. (2016) will be addressed in detail in Chapter 2.

1.2.0.2 Cancer Screening

The American Cancer Society reports that the number of new cancer cases expected to be diagnosed is approximately $1,735,350$ and the number of people expected to die of cancer is $609,640$ in the United States in 2018 (Siegel et al., 2016; ACS, 2018). The Canadian Cancer Society estimates those numbers, respectively, as $206,200$ and $80,800$ for Canada in 2015 (CCS, 2018). Cancer is the second most common cause of death, accounting for nearly 25% of all deaths in the US and the leading cause of death in Canada, accounting for nearly 30% of all deaths (CDC, 2018). The estimated direct medical costs of cancer (total health care expenditures) were $74.8 billion (USD) in the US in 2013, and $3.8 billion (USD) in Canada representing the 7^{th} most costly illness or injury in 2008.

Screening is the primary method used to detect and/or to prevent cancers in the early stages, when treatment is more likely to be successful. For colorectal and cervical cancers, screening also allows removal of precancerous lesions by detecting them. Therefore, screening is suggested to be used to reduce mortality for several types of cancers including breast, prostate, colon, rectum, and lung (ACS, 2018). Although, guidelines suggest that individuals undergo screening at regular time intervals (e.g., every 5-year), screening of a population is expensive; the screened population may experience disutility resulted from screening (e.g., complication risks associated with the screening method, the false-positive test results, etc.). To determine the screening policy that optimizes a particular objective (e.g., minimizing expected costs, maximizing quality adjusted life years (QALYs), etc.) MDP models are proposed in several existing studies in the literature.

Chhatwal et al. (2010) investigates the optimal breast biopsy decisions using a finite-

horizon discrete-time MDP. This study considers the mammographic features, and demographic factors of patients to propose personalized screening policies that maximize QALYs. A comparison of the current practices of radiologists and the proposed model, which takes the patients' ages into account, is presented. It is shown that the optimal policy of the MDP model outperforms the current practice, based on the clinical data. Burnside et al. (2012) also addresses breast biopsy decisions. An MDP model maximizing total QALYs of a patient is proposed to determine the optimal threshold of breast cancer risk that requires that the patient undergoes a breast biopsy. Alagoz et al. (2013) also incorporates short interval follow up exams into an MDP model, and presents the optimal policies for postmammography diagnostic decisions that maximize a patient's QALYs. Cantor et al. (1995) addresses an MDP model to determine prostate cancer screening decisions that optimize QALYs. In addition to determining the optimal policies for a particular screening modality, Vijan et al. (2001) presents a comparison of screening costs, effectiveness and compliance of different screening modalities. Comprehensive reviews of studies determining cancer screening policies using OR models can be found in Alagoz et al. (2011) and Pierskalla and Brailer (1994).

1.3 Thesis Outline

This thesis includes two main studies, each focusing on different practices of preventive medicine; childhood vaccine administration practices and colorectal cancer screening and surveillance policies.

The first study examines operational decisions of health clinics that administer childhood vaccines during a clinic day in developing countries. We consider the practice setting of holding the inventory in different sizes of multi-dose vials for a particular vaccine type, and administering doses from those vials dynamically over time to cover the vaccine demand. The objective is to dynamically decide which size of a vial to use next given the available vaccine inventory and the number of remaining days until next inventory replenishment to maximize the demand coverage. We develop a discrete-time finite horizon MDP

model to evaluate administration policies and determine the one maximizing the demand coverage. The optimal policies determine whether to terminate services for the remainder of a clinic day (if applicable), or open a new vial of a particular size among various vial size options. We perform a comprehensive numerical analysis to investigate the dynamics of the optimal policies, along with the structural analysis. Using the insights from both analyses, simple yet effective (i.e., providing policy performances close to the performances of the optimal policies) vaccine administration policies that require tracking a single time threshold for switching actions are derived. To further improve the performances of the optimal policies, a Pareto-efficiency analysis on the expected total cost and the expected demand coverage spectrum is presented to better design initial inventory replenishment decisions.

The second study is on optimizing colorectal cancer screening and surveillance policies, considering screening and treatment costs, and has two fold objectives. The first objective is again to derive simple yet effective CRC screening policies that can be preferred by health care practitioners over the optimal policies of a POMDP model. We conduct an extensive numerical analysis by using a microsimulation model which mimics the CRC progression of low- and high-risk individuals using clinical data. The numerical analysis focuses on finding the Pareto-efficient simple policies designed by considering various factors which may affect the performances of the policies. Using the insights from this analysis, we then aim at deriving the optimal screening policies, to maximize expected QALYs and minimize expected total cost or number of required colonoscopies, from the multi-objective optimization point of view. We develop a bi-criteria constrained MDP model framework. Our initial analyses indicate that there are screening policies that accrue most of the QALYs improvements without increasing the expected total cost, but requiring slightly more frequent colonoscopy screening than the guidelines.

Both studies aim to determine how to design a health policy to improve efficiency from both the social and economic perspectives. Moreover, the health policies are aimed to be simpler-to-implement in practice. We can put both studies under the same objective umbrella of designing effective yet simple policies, using the insights from the optimal MDP policies.

In the remainder of the thesis, we devote one chapter for each application area, Chapter 2 for the study analyzing childhood vaccine administration policies, and Chapter 3 for colorectal cancer screening and surveillance policies. Each chapter presents the motivation, related literature, and proposed model of the problems separately. For the vaccine administration problem, we study structural properties and design a comprehensive numerical analysis of the problem. Together these both educate us when developing the simple policies for the problem. We also perform a Pareto-efficiency analysis to better design the initial inventory replenishment decisions. On the other hand, we study more on numerical experiments for the CRC problem. Thus, Chapter 3 presents an extensive numerical experiment on the simpler-to-implement policy derivation. We also provide an idea of an analytic model, a bi-criteria constrained MDP model.

Chapter 2

Designing Effective Childhood Vaccine Administration Practices

In this chapter, we introduce the problem of designing effective childhood vaccine administration practices. The proposed model introduces the concept of using different sizes of multi-dose vials during a replenishment cycle, and aims to determine a vaccine administration policy that dynamically decides which vial-size to use next during a clinic-day. The following sections present the motivation of this study, related literature review, proposed MDP model, and the structural and numerical analysis. We also provide a heuristic policy that uses the intuition from the problem analysis, and develop vaccine administration policies that are simple enough to follow in practice while achieving similar coverages as the optimal policies. We then discuss our findings in detail at the end of Chapter 2.

2.1 Introduction

Infectious diseases are among the leading causes of death, i.e., approximately 20% of worldwide deaths in 2012 were caused by infectious diseases (WHO, 2012). The burden of infectious diseases is quite significant, especially in developing countries (Clemens et al., 2010).

Although they are mostly preventable or treatable, each year diseases such as diarrhea, lower respiratory infections of children, HIV/AIDS, hepatitis B, tuberculosis, malaria, and measles cause more than 13.7 million deaths worldwide, accounting for nearly 90% of deaths recorded in the poorest countries (Moxon et al., 2011).

Infectious diseases are especially devastating for children, with approximately 10 million pediatric deaths per year (Clemens et al., 2010). Pediatric vaccination is the primary preventive measure against infectious diseases among children GAVI (2012); Ozawa et al. (2011); Stack et al. (2011), and effectively reduces pediatric deaths (WHO et al., 2015; Shefer et al., 1999). Therefore, the World Health Organization (WHO) runs the Expanded Program on Immunization (EPI) and recommends routine pediatric vaccination with Bacille Calmette-Guérin (BCG), diphtheria-tetanus-pertussis (DTP3), polio, and measles vaccines.

There are ongoing efforts towards improving routine childhood vaccination coverage; EPI has increased coverage of these four pediatric vaccines from 5% to 84% since 1974. In addition, the Global Vaccine Action Plan (GVAP) targets improving coverage rates for all vaccines including $DTP3$ (a major performance indicator of national immunization programs) to 90% in every country and 80% in every district (WHO et al., 2015). Recent studies report that increasing the coverage of routine pediatric vaccines in the world's poorest countries to 90% between 2011-2020 (as GVAP targets) would avert 6.4 million pediatric deaths, $63,000$ disabilities, and 426 million illnesses. This would save \$6.2 billion in disease treatment costs and \$145 billion in productivity losses (Ozawa et al., 2011; Stack et al., 2011).

Despite the benefits of vaccination, its full potential has not been achieved, as current vaccination practices have failed in achieving the targeted coverage rates. The estimates of WHO and the United Nations Children's Fund (UNICEF) show that although the coverage rate of $DTP3$ vaccine increased from 20% to 84% between 1980 and 2013, 21.8 million eligible children could not complete $DTP3$ vaccine series in 2013 (UNICEF and WHO, 2014). Moreover, some developing countries are significantly below these global averages, e.g., the $DTP3$ coverage is still less than 50% in 6 countries (UNICEF and WHO, 2014).

Although it is desirable to improve vaccination coverage by increasing accessibility to vaccination services around the world, this is hard to achieve because many countries have less accessible regions lacking the necessary storage, transportation, and administration conditions. Improving the accessibility in such countries may take years. Due to this and other similar factors, global coverage rates did not improve between 2010-2013 (UNICEF and WHO, 2014). Another way to increase vaccination coverage is improving efficiency of vaccination practices via reducing vaccine wastage which is defined as *"loss by use, decay, erosion, or leakage or through wastefulness"*. It is reported that vaccine wastage accounts for 50% of vaccines distributed around the world in 2005 (WHO, 2005). Recently, countries put more effort in reducing vaccine wastage due to increasing vaccine costs during the last two years (WHO, 2005).

In this context, novel methods for efficient management and planning of vaccine administration practices are needed to reduce vaccine wastage and further increase vaccine coverages. There are basically two categories of vaccine wastage: i) wasting unopened vials due to expiration (decay, erosion) or leakage, ii) wasting unused portion of opened multi-dose vials due to ineffective vaccine administration strategies as well as inappropriate handling practices and contamination (Setia et al., 2002). Because the latter category contributes more to the total vaccine wastage and is associated with more variability compared to the prior one, we focus on vaccine wastage in opened multi-dose vials (open-vial wastage) in this study.

Although the use of preloaded single-dose syringes and availability of proper cold-storage minimize vaccine wastage in North America, the practices in developing countries suffer from open-vial wastage due to the use of multi-dose vials and lack of appropriate storage/handling conditions. When multi-dose vaccine vials cannot be kept under ideal storage conditions in clinics lacking proper equipment or during outreach vaccination sessions, the remaining doses from the opened vials must be discarded at the end of the day (WHO, 2014a). The high open-vial vaccine wastage rate in developing countries is one of the main obstacles in achieving GVAP's targets on global vaccine coverage rate (WHO et al., 2015; UNICEF and WHO, 2014). The success of a pediatric vaccination program is often associated with the strength of the vaccine delivery systems within a country (Mc-

Coy et al., 2009). Therefore, developing countries recently put more effort in improving their vaccine delivery systems and the effectiveness of their immunization programs (WHO, 2005). However, innovative short-term and long-term strategies are still needed to reduce pediatric vaccine wastage and improve vaccine coverage (Moxon et al., 2011; Briss et al., 2000). Administering pediatric vaccines from multi-dose vials of different sizes efficiently has been promoted as a potential short-term solution to this problem Parmar et al. (2010); Drain et al. (2003), while accessibility to remote districts and the availability of proper cold storage are improved to solve this problem permanently in the long-run.

Using single-dose vials minimizes the open-vial wastage (and maximize demand coverage), it is a very costly practice. Recent studies show that single-dose vials may not be cost-effective (if not infeasible) compared to using a multi-dose vial for routine pediatric vaccination in particular developing countries (Lee et al., 2010, 2011; Assi et al., 2011). Administration of the same-size multi-dose vaccine vials by dynamically deciding whether to open a new vial or to terminate vaccination service for the current day has been addressed in the literature (Mofrad et al., 2014, 2016). The concern with this practice is that vaccine wastage might still occur and a significant number of patients might be denied service due to early termination of vaccination service. A promising direction to alleviate these concerns is holding a vaccine inventory of multi-dose vials in different sizes, and dynamically using them to further reduce vaccine wastage and the number of patients denied service.

In this context, we propose a novel discrete-time finite-horizon Markov decision process model to determine i) which vial size should be used after the depletion of the previously opened vial and ii) when to terminate the vaccination service during a clinic day. We take society's perspective, and maximize the expected number of doses administered (representing the benefit to the society) from the available pediatric vaccine inventory of multi-dose vials in different sizes. We may use the number of doses administered and *benefit* interchangeably hereafter. Our numerical analysis specifically focuses on measles, DTP-HepB-Hib (Pentavalent), yellow fever (YF), and BCG vaccination in developing countries. These vaccine types are included in most routine childhood vaccination programs; thus, more data is available for these vaccines in the literature. In addition, these vaccine types

are available in various vial sizes including 2-, 5-, 10-, and 20-dose vials in developing countries; therefore, this provides us ample problem instances to show the effect of vial sizes on the performance of the proposed approach. We compile data about the routine pediatric vaccination practices in developing countries from clinical literature. Details about these vaccines are presented in Appendices B. In developing countries, vaccine administration practices may be carried at drop-in clinics, i.e., fixed sites for vaccination (e.g., hospitals, health clinics, etc.) or through outreach sessions providing mobile vaccination. Our study focuses on the vaccination practices at drop-in clinics.

Using the proposed model, we show that the optimal policies are of control-limit type. In addition, we numerically illustrate that vaccine coverage may be improved to desired levels by keeping vaccine doses in two different vial sizes (i.e., small and large vials) and administering them dynamically even without early termination of vaccination service. Note that there is a trade-off between reducing open-vial wastage by using smaller vials and facing higher total cost, which is important to consider for developing countries. The proposed model also helps determining the best inventory levels for each vial size to balance this trade-off. Our numerical results from a Pareto-frontier analysis based on the MDP model indicate that vaccine administration practices can be improved to increase expected demand covered cost-efficiently (i.e., without increasing the total cost significantly or at all) by determining the ideal levels of the initial vaccine inventory held in small and large vials.

2.2 Literature

Operations research (OR) applications in the vaccination literature primarily focus on (i) effective allocation of vaccine resources to mitigate infectious diseases Dalgıç et al. (2017); Yarmand et al. (2014); Medlock and Galvani (2009); Brandeau (2005); Weniger et al. (1998); (ii) determining the optimal biological composition of vaccinesOzaltin et al. (2011); Cho (2010); Kornish and Keeney (2008); Lim and Lee (2008); (iii) planning vaccine procurement, pricing, and distribution mostly from the supply chain management perspective

Dai et al. (2016); Robbins and Jacobson (2015); Ramirez-Nafarrate et al. (2015); Fleischhacker et al. (2015); Proano et al. (2012); Robbins and Jacobson (2011); Jacobson et al. (2006), (iv) pediatric vaccine coverage and wastage in developing countries and the impact of vial size selection on them.

We position this study within a few studies in (iv), and yet differentiate our work from them because this study: (1) proposes using several vial sizes interchangeably during a clinic day; (2) presents an analytic model to determine the optimal vaccine administration decisions; (3) suggests an alternative administration practice to improve the current vaccine coverage through reducing open-vial wastage in developing countries; (4) finds initial vial inventory levels for each vial size that improves the current practice, without significantly increasing the total cost.

Most studies in (iv) analyze the effect of multi-dose vial size on open-vial wastage and total cost by statistically estimating vaccine wastage rates using empirical data, or explicitly modeling the demand arrival and vaccine wastage processes. Guichard et al. (2010) estimate vaccine wastage rates in Bangladesh in 2004 by analyzing data from vaccine distribution and stock registries with cluster sampling. They show that the average vaccine wastage from opened vials is significantly larger than that from unopened vials for various pediatric vaccines (e.g., 84.4% and 0.5% for BCG vaccine, respectively).

Parmar et al. (2010) use empirical data from WHO to estimate the ranges of pneumococcal conjugate vaccine (PCV) wastage rates ($25^{th} - 75^{th}$ percentile) for 1-, 2-, 5- and 10-dose vials, and determine associated cold chain, vaccine procurement, and wastage costs for several developing countries based on a set of hypothetical cost functions. They empirically determine the best vial size option with the lowest average total cost per vaccinated child for each country. Parmar et al. (2010) show that although multi-dose vials may be the best for inexpensive vaccines such as BCG and DTP, higher wastage rates can negate the savings realized from lower storage and purchasing costs per dose of multi-dose vials for expensive vaccines such as PCV. Similar observations are made by others (e.g., Drain et al. (2003)).

Lee et al. (2010) and Yang et al. (2014) analyze the cost-effectiveness of various multi-

dose vial sizes for several pediatric vaccines with limited open-vial shelf-life by explicitly modeling vaccine wastage. Lee et al. (2010) propose an spreadsheet model to estimate the expected vaccine utilization and costs associated with vaccine wastage, vial disposal, and vial storage for each vial size based on a Poisson arrival process representing vaccine demand realization. Using this approach, Lee et al. (2010) determine the most cost-saving vial size (among all feasible options) for different demand levels. Yang et al. (2014)) analyze the session size data of WHO for four countries to derive the best-fit statistical arrival processes for vaccine demand realization, and then use the approach proposed in Lee et al. (2010) to compare 5-dose and 10-dose vial options for each country. They report that using 5-dose vials reduces open-vial wastage while increasing the expected total cost. Dhamodharan and Proano (2012) propose a binary integer programming model integrated into a Monte Carlo simulation (specifying a sample of vaccine demand realizations) to determine the vial size and inventory reordering level combinations that minimize the total cost of vaccine procurement and wastage for specific vaccine types. They validate their approach by comparing the optimal vial sizes they derive with those from (Lee et al., 2010).

A few studies in (iv) also analyze the effect of vial size on the performance of a vaccination program while considering the capacity limitations in storage and transportation of smaller dose vials. Lee et al. (2011) and Assi et al. (2011) study the effect of using different vial sizes for a set of vaccines on the overall supply chain performance, open-vial wastage, and costs, using dynamic discrete-event simulation models that keep track of vaccine inventory levels and their daily utilization in facilities distributed over regions of Niger and Thailand, respectively. Their numerical results illustrate that even though using smaller vial sizes may result in lower open-vial wastage, they have larger per dose volumes which may reduce the vaccine availability because of increased storage and transportation capacity needs.

Although there are several papers analyzing the sizes of vials in a vaccine inventory, to the best of our knowledge, there is only one analytic model in the literature, Mofrad et al. (2014), that studies how vaccine doses from same-size vials should be administered to maximize the demand coverage. Mofrad et al. (2014) determine the optimal time to

terminate vaccination service in a clinic for the rest of the day, by using a finite-horizon discrete-time Markov decision process (MDP) model with the objective of maximizing the expected total number of doses administered, based on the available vaccine inventory level, and the number of days until the next inventory replenishment. They state that termination of vaccination service for the rest of the day, rather than opening a new vial for the next patient, may help in avoiding open-vial wastage. Mofrad et al. (2014) prove the existence of the optimal threshold-type policies for daily termination of vaccination service. In a follow-up study, Mofrad et al. (2016) extend their numerical analyses to provide guidance to practitioners about the best selection of vaccination clinic operating hours and session frequency. Using a simulation model, they conduct sensitivity analyses on model parameters and analyze the robustness of the decisions derived from their MDP model to random vial yield.

The aforementioned studies generally agree that selection of the ideal vial size depends on several factors including vaccine characteristics, wastage rates, required coverage levels, targeted vaccination program efficiency, available storage/transportation capacities, costs, and demand levels. As a result, while one study highlights that the savings from using larger vials may not compensate for the increase in wastage cost Parmar et al. (2010), another one reports that the extra procurement cost associated with smaller vials can negate the savings from wastage cost (Yang et al., 2014). Such conflicting observations in the literature illustrate the complexity of the best vaccine administration policy. More comprehensive and innovative policies are still in need since the current practices suffer from high open-vial wastage rates even with the actively promoted options in use (e.g., outreach sessions, scheduled appointments, etc.) (Guichard et al., 2010).

All of the studies mentioned above consider either keeping all vaccine stocks in smaller same-size vials or dynamically terminating vaccine administration during a clinic day. Although both approaches can increase vaccine coverage and reduce vaccine wastage, adopting the former alone is associated with higher vaccine procurement costs, whereas the latter increases the number of patients denied service. Previous studies highlight the potential of using a mixture of vial sizes in vaccine administration to further improve vaccination practices (Parmar et al., 2010; Drain et al., 2003). A vaccine administration strategy that

dynamically uses a mixture of vial sizes (e.g., small and large vials) may be less expensive than always using small vials, and may terminate vaccine administration services as infrequently as possible to improve service rate. The feasibility of keeping different sizes of vaccine vials in inventory, and using them dynamically over time to cover the vaccine demand is implied by the previous studies promoting it as an alternative to existing vaccination practices Parmar et al. (2010); Drain et al. (2003), and reports of WHO stating that pediatric vaccines such as measles, pentavalent, YF, and BCG vaccines may be available in several vial sizes in developing countries (GAVI, 2015a). However, the benefits of such pediatric vaccine administration policies are not evaluated in the literature.

In order to fill the preciding gap, our MDP model incorporates the option of keeping vaccine doses in vials of different sizes, and analyzes the decisions of selecting the size of the vial to open next and the time of service termination. We generalize the structural results in Mofrad et al. (2014) to the case of non-homogeneous vial sizes, and provide new analytical results specifying the effect of inventory level for each vial size on the optimal solution. Through analytical and numeric results, we show the existence of the optimal thresholds for switching from smaller to larger vials and terminating vaccination services, which change monotonically in terms of time and inventory level. These dynamic optimal policies may require managing a more complex inventory system, and cause variance in day-to-day vaccination practices. Therefore, based on the monotone behaviour of the optimal solutions to our MDP model, we derive easier-to-implement vaccine administration policies to promote the applicability of our findings in practice.

Moreover, through a Pareto-frontier analysis, we identify a set of ideal initial inventory levels for different vial sizes in the spectrum of expected demand covered and total cost. We show that the expected demand covered can be increased to a desired level by keeping a relatively small portion of the vaccines (e.g., $< 18\%$ for pentavalent) in small vials without early termination of vaccination services. Thus, the proposed policies may improve the pediatric vaccine coverage to the targeted levels in exchange for a reasonably small (if at all) increase in cost compared to administering vaccines from large vials. The insights reported in this thesis are unique contributions to the literature because Mofrad et al. (2014) and Mofrad et al. (2016) do not consider the option of using different size vials,

and they provide a limited analysis of the effect of initial pediatric vaccine inventory on demand coverage and total cost.

2.3 MDP Formulation

We determine the optimal vaccine administration policy using an MDP model that tracks the number of available vials of each size and the time until the next inventory replenishment. When a patient arrives to get vaccinated after the previously opened vial is depleted, the MDP model considers two decisions; terminating the vaccination service for the rest of the day, or continuing the service by selecting which size of vial to open next in order to maximize the benefit to society. The main assumptions of our model are listed as below.

(A1) The daily vaccine demand follows a Poisson process.

(A2) The initial inventory level for each vial size is known and vaccine inventory is periodically replenished to this level.

(A3) The unused doses from an open-vial are discarded at the end of each clinic day.

(A4) Any demand after service termination is lost.

While (A1), (A2), and (A3) are generally valid in the current vaccination practice, (A4) may hold only for particular healthcare settings. (A1) is justifiable since several studies show that a Poisson process may represent vaccine demand arrival in some of the developing countries reasonably well (Yang et al., 2014; Rajgopal et al., 2011). Note that this is a common assumption among the studies proposing efficient vaccine administration practices (Mofrad et al., 2014; Yang et al., 2014; Rajgopal et al., 2011; Lee et al., 2010). (A2) also complies with the current practice as many studies in the literature assume periodic replenishment of vaccine inventory based on the information and data that they collect from several developing countries Lee et al. (2011); Assi et al. (2011); Guichard et al. (2010). In addition, considering that WHO recommends discarding unused vaccine doses

6-8 hours after opening a vial in the absence of proper cold-chain storage facilities (WHO, 2016b), (A3) is in line with the current practice guidelines. Although there can be practices for which (A4) is not valid, considering accessibility issues in developing countries, we can assume that a significant portion of the denied patients may not visit the clinic the next day. Furthermore, the proposed model aims at delaying early termination by using small vials, which limits the expected number patients denied service.

Table 2.1 presents the notation used throughout the thesis. The time horizon, i.e., the time between two vaccine inventory replenishments, is divided into clinic days, each consisting of η timeslots with equal length τ. We set τ (η) short (large) enough to ensure that the probability of having more than one demand in each timeslot is negligible. Therefore, the Poisson demand arrival within a clinic day is approximated by a discrete-time arrival process with independent and identically distributed (*IID*) geometric inter-arrival times ($p = \mu/\eta$). The probability that the next arrival occurs x timeslots later is $p_X(x) = p(1-p)^{(x-1)},\ x = \{1, .., \eta\}$, and $p_X(\eta+1) = (1-p)^{\eta}$.

We denote the state of our MDP model by $s = (t, \mathbf{q}, h)$, where $t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$, and $h \in \mathbb{H}$ refer to the number of days until the next replenishment (including the current day), the number of vials on-hand, and the current timeslot for a service decision, respectively. The decisions are always made at the beginning of a timeslot. In other words, an action is taken at timeslot h if an arrival occurs after the depletion of the previously opened vial. While considering termination of vaccination service, we assume that either termination is allowed at any time, starting from the beginning of each day (terminated service), or no termination is allowed until the end of each day (continuous service). Let h_c be the earliest timeslot at which daily vaccination services can be terminated. Then, these service scenarios are denoted by $h_c = 0$ and $h_c = \eta + 1$, respectively. Hence, the possible actions for state s, $\mathbb{A}_s$, include *opening a type i vial* (i) and *termination of vaccination service* (N)

Table 2.1: List of notation

Notation	Desciption
State indices and parameters	
T	The total number of days in a replenishment cycle
$t \in \mathbb{T} \equiv \{1, .., T\}$	The number of days remaining until the next inventory replenishment
$i \in \mathbb{I} \equiv \{1, .., n\}$	The vial type among n different size options
z_i	The number of doses per type i vial, $\forall i \in \mathbb{I}$ s.t. $z_i < z_{i+1}$.
Q_i	The initial inventory level for type i vials, $\forall i \in \mathbb{I}$
$\mathbf{Q} = (Q_1, Q_2, .., Q_n)$	The vector representing Q_i values, $\forall i \in \mathbb{I}$
$q_i \in \mathbb{Q}_{\mathbb{i}} \equiv \{0, 1, .., Q_i\}$	The number of type i vials on-hand, $\forall i \in \mathbb{I}$
$\mathbf{q} = (q_1, q_2, .., q_n) \in \mathbb{Q}$	The vector representing q_i values, $\forall i \in \mathbb{I}$ where $\mathbb{Q} \equiv \mathbb{Q}_{\mathbb{1}} \times \ldots \times \mathbb{Q}_{\mathbb{n}}$
η	The total number of timeslots in a day
τ	The length of each timeslot, i.e., $\tau = 1/\eta$ day
$h \in \mathbb{H} \equiv \{1, .., \eta\}$	The timeslot of the first arrival after the previously opened vial is depleted
Probability distributions and parameters	
μ	The daily demand rate for vaccination service
$p = \mu/\eta$	The probability that a demand for vaccination service arrives at the beginning of a timeslot
$p_X(x)$	The probability that the next arrival occurs x timeslots later
D_h	The binomial random variable denoting the total demand occurred between timeslots h and η
Y_h^i	The negative binomial random variable referring to the timeslot at which the first arrival occurs after depleting the type i vial opened at timeslot h
Secondary performance metrics	
$N_i(t, \mathbf{q}, h)$	The expected number of type i vials opened starting from state s until the next replenishment under the optimal policy
$N_i(t, \mathbf{q})$	The expected number of type i vials opened starting from the beginning of day t until the next replenishment under the optimal policy
$W_i(T, \mathbf{Q})$	The expected open-vial wastage from type i vials under the optimal policy
$W(T, \mathbf{Q})$	The expected open-vial wastage of the optimal policy during the replenishment cycle
$\omega_i(T, Q)$	The expected percentage of open-vial wastage from type i vials under the optimal policy
$\omega(T, Q)$	The expected percentage of open-vial wastage under the optimal policy
$\phi_i(T, Q)$	The expected percentage of demand covered by the doses administered from type i vials under the optimal policy
$\phi(T, Q)$	The expected percentage of demand covered under the optimal policy
$\pi(T, Q)$	The expected total cost of the optimal policy within a replenishment cycle

List of notation continued

Notation	Desciption
Model components and parameters	
$s = (t, \mathbf{q}, h) \in \mathbb{S}$	The current state of the system where $\mathbb{S} \equiv \mathbb{T} \times \mathbb{Q}_{\mathbb{1}} \times .. \times \mathbb{Q}_{\mathbb{n}} \times \mathbb{H}$
$A_s \in \mathbb{A}_s \subset \{1, 2, .., n, N\}$	The action in state s
$A_s^* \in \mathbb{A}_s$	The optimal action taken in state s
$h_c \in \{0, \eta + 1\}$	The earliest timeslot at which vaccination service can be terminated
$I_{h_c} \in \{0, 1\}$	The indicator function defining the service scenario, i.e., $I_{h_c} = 1$ if it is a *terminated service*, $I_{h_c} = 0$ if it is a *continuous service*
$V(t, \mathbf{q}, h)$	The maximum expected number of doses administered until the next replenishment starting from state s
$\nu(t, \mathbf{q})$	The expected number of doses administered starting from the beginning of day t to the next replenishment given the inventory of $\mathbf{q}$
$g_i(t, \mathbf{q}, h)$	The expected number of doses administered until the next replenishment when the decision is opening a type i vial in state s
$\mathbf{e_i}$	The binary row vector showing whether a type i vial is opened, i.e., $\mathbf{e_{i,j}} = 1$ if $j = i$, $\mathbf{e_{i,j}} = 0$ otherwise, $\forall i, j \in \mathbb{I}$, $\forall s \in \mathbb{S}$
$V_i(t, \mathbf{q}, h)$	The expected number of doses administered from type i vials starting from state s to the next replenishment cycle under the optimal policy
$\nu_i(t, \mathbf{q})$	The expected number of doses administered from type i starting from the beginning of day t to the next replenishment vials under the optimal policy
I_s^i	The indicator function showing whether the optimal action in state s is opening a type i vial
k_s	The index variable denoting the optimal action in state s, i.e., $k_s = i$ when $A_s^* = i$, $k_s = n + 1$ when $A_s^* = i$
$\mathbf{e_s}$	The binary row vector showing the type of the vial opened in state s, i.e., $\mathbf{e_{s,i}} = 1$ if $I_s^i = 1$, $\mathbf{e_{s,i}} = 0$ otherwise, $\forall i \in \mathbb{I}$, $\forall s \in \mathbb{S}$
$h_i^*(t, \mathbf{q})$	The timeslot when the optimal action is using a type i vial for the first time on day t given the vaccine inventory of $\mathbf{q}$
$h_N^*(t, \mathbf{q})$	The timeslot when the optimal action is terminating daily vaccination services for the first time on day t given the vaccine inventory of $\mathbf{q}$
α_i	The proportion of doses held in type i vials at the beginning of replenishment cycle
Cost parameters	
c_i	The procurement cost per dose of a type i vial, $\forall i \in \mathbb{I}$
w_i	The weight of an empty type i vial , $\forall i \in \mathbb{I}$
w_d	The weight per dose
w_{syr}	The weight of an empty syringe
c_d	The cost per kg of medical waste disposed
v_i	The volume per dose of a type i vial, $\forall i \in \mathbb{I}$
c_s	The unit storage cost

as stated by the following definition.

$$A_s \in \mathbb{A}_s \equiv \begin{cases} \{1,2,..,n,N\} \setminus \bigcup\limits_{\{i:\, q_i=0\}} \{i\}\,, & \text{if } h_c = 0 \\ \{1,2,..,n\} \setminus \bigcup\limits_{\{i:\, q_i=0\}} \{i\}\,, & \text{if } h_c = \eta + 1 \end{cases}$$

Note that $h_c = \eta + 1$ represents the traditional administration policy of providing full-time service as long as vaccine stocks are available (as in Yang et al. (2014), Lee et al. (2011), Lee et al. (2010)), and $h_c = 0$ allows termination of vaccination service at any time as in Mofrad et al. (2014). Indeed, any minimum working-hour constraint can be easily incorporated into the model; however, we exclude this option for brevity of the analyses. Early termination of vaccination services requires denying patients a vaccination, which reduces access to medical resources and equity in resource allocation. We aim to derive optimal policies with continuous service, whose performance is as good as those with service termination, to eliminate the cost/burden of denying patients the service. That is, we use optimal policies with service termination for benchmarking. Figure 2.1 visualizes the state components and progress of time in our MDP model when two different vial sizes are available. A detailed discussion of the model parameters and their specifics is presented.

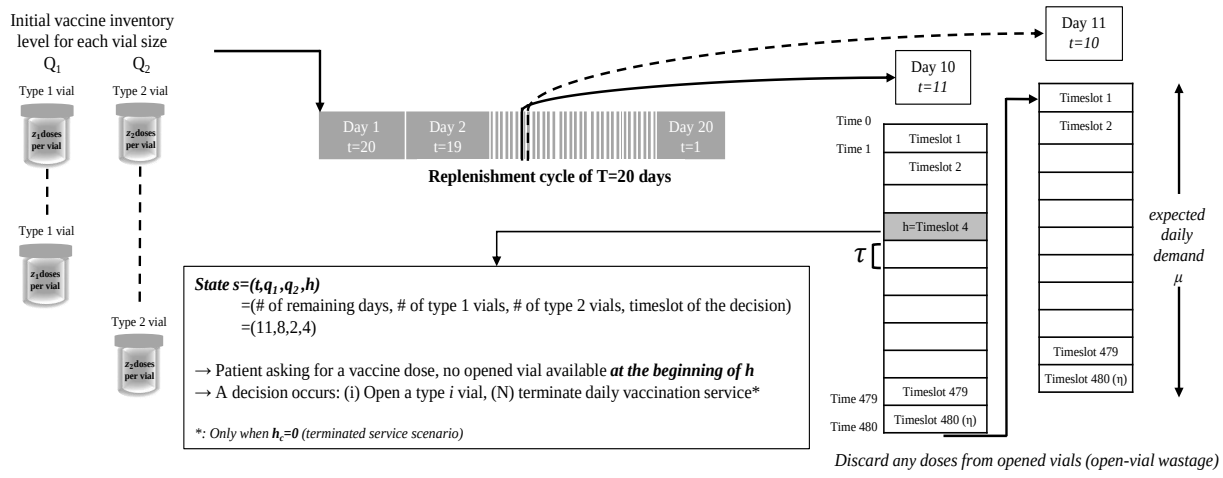


Figure 2.1: A representation of the MDP model

The benefit of opening a new vial depends on the remaining service time on a given day, i.e., $(\eta - h)$. The total demand by the end of the clinic day, D_h, is binomially distributed with parameters $(\eta - h, p)$ as given in Equation 2.1. The possible values of D_h start from 1 due to the definition of h; i.e., the decision of opening a vial can occur only upon an arrival. In addition, Y_h^i refers to the timeslot of the first arrival after the depletion of a z_i-dose vial opened at timeslot h. Basically, Y_h^i has a truncated negative binomial distribution with parameters z_i and p as stated in Equation 2.2.

$$p_{D_h}(d) = \begin{cases} \binom{\eta-h}{d-1} p^{(d-1)}(1-p)^{(\eta-h-(d-1))}, & d \in \{1, .., \eta - h + 1\}, \quad h \in \mathbb{H} \setminus \{\eta + 1\} \\ 0, & otherwise \end{cases} \tag{2.1}$$

$$p_{Y_h^i}(y) = \begin{cases} 0, & y < h + z_i \\ \binom{y-h-1}{z_i-1} p^{z_i}(1-p)^{(y-h-z_i)}, & y \in \{h + z_i, .., \eta\}, \quad h \in \mathbb{H} \setminus \{\eta + 1\} \\ \sum_{k=1}^{z_i} p_{D_h}(k), & y = \eta + 1 \end{cases} \tag{2.2}$$

2.3.1 Optimality Equation

We denote the optimal value function of the MDP model by $V(t, \mathbf{q}, h)$, which refers to the maximum expected number of doses administered until the next replenishment starting from state $s = (t, \mathbf{q}, h) \in \mathbb{S}$. The following Bellman optimality equation characterizes $V(t, \mathbf{q}, h)$ for both service scenarios with the use of an indicator function (I_{h_c}), which is equal to 1 and 0 if $h_c = 0$ and $h_c = \eta + 1$, respectively. Hence, for all $t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$, $\mathbf{q} > 0$, $h \in \mathbb{H}$, and $i \in \mathbb{I}$:

$$V(t, \mathbf{q}, h) = \max \left\{ I_{h_c} \nu(t-1, \mathbf{q}), \max_{i \in \{1,..,n\}} \{g_i(t, \mathbf{q}, h)\} \right\}, \tag{2.3}$$

where

$$\nu(t, \mathbf{q}) = \sum_{h=1}^{\eta+1} V(t, \mathbf{q}, h) p_X(h), \tag{2.4}$$

$$g_i(t, \mathbf{q}, h) = \sum_{y=h+z_i}^{\eta} \Big[z_i + V(t, \mathbf{q} - \mathbf{e_i}, y) \Big] p_{Y_h^i}(y) + \sum_{d=1}^{z_i} \Big[d + \nu(t-1, \mathbf{q} - \mathbf{e_i}) \Big] p_{D_h}(d), \tag{2.5}$$

$$I_{h_c} = \begin{cases} 1, & h_c = 0 \\ 0, & h_c = \eta + 1 \end{cases}, \quad \text{and} \quad \mathbf{e_i} = \begin{array}{c} 1 \quad \dots \quad i-1 \quad i \quad i+1 \quad \dots \quad n \\ \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix} \end{array}$$

In Equation 2.4, $\nu(t, \mathbf{q})$ denotes the maximum expected number of doses administered from the beginning of day t to the next inventory replenishment, given the current vaccine inventory levels for each vial size $\mathbf{q}$. In Equation 2.5, $g_i(t, \mathbf{q}, h)$ represents the expected number of doses administered after opening a type i vial in state $(t, \mathbf{q}, h)$ and following the optimal vaccine administration policy afterwards. The first term in the equation incorporates the possibility of depleting the opened vial before the end of the day and observing a demand, prompting a new decision between timeslots $h + z_i$ and η. The second term represents the case where the vial opened at timeslot h is not completely depleted during the rest of the clinic day t.

When the vaccination service scenario allows termination, $I_{h_c} = 1$ in Equation 2.3. The objective function of this scenario accrues $\nu(t, \mathbf{q})$ by terminating vaccination service for the rest of the current clinic day if the maximum expected benefit of saving the current vaccine inventory for the next day is greater than that of opening a new vial of any size at timeslot h of day t, i.e., $\nu(t, \mathbf{q}) > \max_{i \in \{1,..,n\}} \{g_i(t, \mathbf{q}, h)\}$. On the other hand, in case of continuous vaccination service scenario (i.e., $h_c = \eta + 1$), the model only decides which vial size to open based on the expected demand level during the remaining portion of the day. In this formulation, we do not consider any terminal rewards; $V(0, \mathbf{q}, h) = 0$, $\forall\ \mathbf{q} \in \mathbb{Q}$, $h \in \mathbb{H}$. In addition, we define $V(t, \mathbf{q}, \eta + 1)$ as the number of doses that can be administered after day t is over ($h = \eta + 1$ on day t). Since the clinic is closed starting from this time, i.e., timeslot $\eta + 1$ on day t, until the beginning of the next day, i.e., timeslot 1 on day $t - 1$,

we assume $V(t, \mathbf{q}, \eta + 1) = \nu(t-1, \mathbf{q})$, $\forall\ t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$. Lastly, we set $V(t, \mathbf{0}, h) = 0$, for all $t \in \mathbb{T}$, $h \in \mathbb{H}$ where $\mathbf{0}$ refers to a vector of zeros.

The following equations define the expected number of doses administered from each vial type under the optimal policy which is important to track for the derivation of secondary performance metrics. Let $V_i(t, \mathbf{q}, h)$ specify this amount for type i vial and $A_s^* \in \mathbb{A}_s$ be the optimal action taken in state $s = (t, \mathbf{q}, h)$. Hence, for all $t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$, $h \in \mathbb{H}$, and $i \in \mathbb{I}$:

$$V_i(t, \mathbf{q}, h) = \left(\max_{j \in \{1,..,n\}} I_s^j \right) \left[\sum_{y=h+z_{k_s}}^{\eta} \left(I_s^i z_i + V_i\left(t, \mathbf{q} - \mathbf{e_s}\ , y\right) \right) p_{Y_h^{k_s}}(y) \right.$$
$$\left. + \sum_{d=1}^{z_{k_s}} \left(I_s^i d + \nu_i\left(t-1, \mathbf{q} - \mathbf{e_s}\ \right) \right) p_{D_h}(d) \right] + \left(1 - \max_{j \in \{1,..,n\}} I_s^j \right) \nu_i(t-1, \mathbf{q}),$$

$$\nu_i(t, \mathbf{q}) = \sum_{h=1}^{\eta+1} V_i(t, \mathbf{q}, h) p_X(h), \quad \text{where} \quad \forall s = (t, \mathbf{q}, h) \in \mathbb{S}$$

$$I_s^i = \begin{cases} 1, & \text{if } A_s^* = i \\ 0, & \text{otherwise} \end{cases}, \qquad k_s = \begin{cases} i, & \text{if } A_s^* = i \\ n+1, & \text{if } A_s^* = N \end{cases}, \qquad \mathbf{e_s} = \begin{pmatrix} I_s^1 & I_s^2 & \dots & I_s^n \end{pmatrix}$$

2.3.2 Secondary Performance Metrics

Although we set the objective of the proposed MDP model as maximizing the expected number of vaccine doses administered to represent a societal perspective, secondary performance measures such as expected total cost, expected number of vials used, and expected number of doses wasted are also important when evaluating the economic efficiency of proposed policies. Assessing whether the proposed policies may cost-efficiently (not increasing costs significantly or at all) improve the performance of pediatric vaccine administration is crucial for developing countries due to scarcity of resources.

We define the expected total number of type i vials opened until the next inventory replenishment under the optimal policy as $N_i(t, \mathbf{q})$ given the vaccine inventory levels for

each vial size $\mathbf{q} = (q_1, q_2, .., q_n)$ at the beginning of day t. We derive this metric as follows:

$$N_i(t, \mathbf{q}) = \sum_{h=1}^{\eta+1} N_i(t, \mathbf{q}, h) p_X(h), \quad \forall t \in \mathbb{T}, \forall \mathbf{q} \in \mathbb{Q}, \forall i \in \mathbb{I}$$

where $N_i(t, \mathbf{q}, h)$ denotes the expected number of type i vials opened starting from state $(t, \mathbf{q}, h)$ until the next inventory replenishment.

$$\begin{aligned} N_i(t, \mathbf{q}, h) = (\max_{j \in \{1,..,n\}} I_s^j) \Bigg[& \sum_{y=h+z_{k_s}}^{\eta} \left(I_s^i + N_i\left(t, \mathbf{q} - \mathbf{e_s}, y\right)\right) p_{Y_h^{k_s}}(y) \\ & + \sum_{d=1}^{z_{k_s}} \left(I_s^i + N_i\left(t-1, \mathbf{q} - \mathbf{e_s}\right)\right) p_{D_h}(d) \Bigg] \\ & + \left(1 - \max_{j \in \{1,..,n\}} I_s^j\right) N_i(t-1, \mathbf{q}), \quad \forall t \in \mathbb{T}, \forall \mathbf{q} \in \mathbb{Q}, \forall h \in \mathbb{H}, \forall i \in \mathbb{I} \end{aligned}$$

If the optimal action is to open a new vial at timeslot h, then this vial and the expected number of vials to be opened during the rest of day t are included in the calculation of $N_i(t, \mathbf{q}, h)$. If the optimal action is to terminate vaccination service for the day (N), then the expected number of vials to be opened is calculated starting from the beginning of the next day, i.e., $N_i(t, \mathbf{q}, h) = N_i(t-1, \mathbf{q})$. Naturally, the following hold for each vial type i, $i \in \mathbb{I}$: $N_i(t, \mathbf{0}, h) = 0 \ \forall t \in \mathbb{T}$, $h \in \mathbb{H}$; $N_i(0, \mathbf{q}, h) = 0 \ \forall \mathbf{q} \in \mathbb{Q}$, $h \in \mathbb{H}$; and $N_i(t, \mathbf{q}, \eta+1) = N_i(t-1, \mathbf{q}) \ \forall t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$.

Let $W(T, \mathbf{Q})$ be the expected open-vial wastage of the optimal policy during the replenishment cycle and $W_i(T, \mathbf{Q})$ be the expected open-vial wastage from type i vial for a given $(T, \mathbf{Q})$ pair. We calculate $W_i(T, \mathbf{Q})$ as the difference between the expected total number of doses in the opened type i vials and the expected number of doses administered from type i vials.

$$W(T, \mathbf{Q}) = \sum_{i=1}^{n} W_i(T, \mathbf{Q}), \qquad \text{where } \ W_i(T, \mathbf{Q}) = z_i N_i(T, \mathbf{Q}) - \nu_i(T, \mathbf{Q}) \ \ \forall i \in \mathbb{I}$$

When we compare the optimal vaccine administration policies for different parameters, we consider the expected percentage of demand covered, i.e., $\phi(T, \mathbf{Q})$, and the expected percentage of open-vial wastage, i.e., $\omega(T, \mathbf{Q})$. The definitions of these two metrics are similar to those in Mofrad et al. (2014). Let $\phi_i(T, \mathbf{Q})$ be the proportion of the expected demand vaccinated via administering doses from type i vial over the expected total demand, and $\omega_i(T, \mathbf{Q})$ be the proportion of doses wasted from type i vial over the expected total number of doses in the opened vials, respectively. Hence, for all $i \in \mathbb{I}$, these metrics are calculated as follows:

$$\phi(T, \mathbf{Q}) = \sum_{i=1}^{n} \phi_i(T, \mathbf{Q}), \quad \text{where } \phi_i(T, \mathbf{Q}) = \frac{\nu_i(T, \mathbf{Q})}{\mu T}$$

$$\omega(T, \mathbf{Q}) = \sum_{i=1}^{n} \omega_i(T, \mathbf{Q}), \quad \text{where } \omega_i(T, \mathbf{Q}) = \frac{W_i(T, \mathbf{Q})}{\nu(T, \mathbf{Q}) + W(T, \mathbf{Q})}$$

We also calculate the expected total cost of the optimal policy within a replenishment cycle based on Equation 2.6. The expected total cost ($\pi(T, \mathbf{Q})$) comprises the costs of (i) administered and wasted doses; (ii) disposing empty vials, syringes and wasted doses; and (iii) storage. While (i) captures the increased vaccine procurement cost and the reduced cost of wastage, (ii) and (iii) reflect the increased expenditure for disposing of more empty vials and needing more storage space, due to holding a portion of doses in smaller vials in the initial vaccine inventory, $\mathbf{Q}$. Having non-homogeneous-size vials in the inventory may complicate both inventory handling and disposal processes, and increase the likelihood of errors. Therefore, the actual cost of disposal and storage may be higher than our estimates, which can be addressed with additional sensitivity analyses. One can also incorporate the societal cost of vaccine wastage (resulting from delayed immunization and increased susceptibility to infectious diseases), and/or the cost of denying services to patients due to service termination decisions, into the total cost function. However, these costs are not readily available in the literature, and require additional modeling to derive them. Furthermore, considering these costs would only improve the performance of the proposed

optimal policies with continuous service, and strengthen our conclusions.

$$\pi(T,\mathbf{Q}) = \sum_{i=1}^{n}\Bigg[\Big[(\nu_i(T,\mathbf{Q}) + W_i(T,\mathbf{Q})) \times c_i\Big] + \Big[N_i(T,\mathbf{Q}) \times (w_i + w_{syr} \times (z_i+1)) \times c_d\Big] + \Big[W_i(T,\mathbf{Q}) \times w_d \times c_d\Big] + \Big[\frac{1}{2} \times Q_i \times z_i \times v_i \times c_s\Big]\Bigg] \tag{2.6}$$

2.4 Structural Properties

We now present a set of structural properties of the optimal value function and the optimal vaccine administration decisions in terms of time and vaccine inventory on-hand. We assume that set $\mathbb{I}$ represents the order of vial sizes, i.e., $z_i < z_{i+1}\ \forall i \in \mathbb{I}\ n$. The proofs of the structural properties are available in Appendices A. Note that, $h_i^*(t,\mathbf{q})\ \forall i \in \mathbb{I}$, and $h_N^*(t,\mathbf{q})$ refer to the timeslots when the optimal action for a given t and $\mathbf{q}$ is using a type i vial and terminating daily vaccination service for the first time on that day, respectively. That is, $h_i^*(t,\mathbf{q}) = \min\{h \in \mathbb{H} : A^*_{(t,\mathbf{q},h)} = i\}$ and $h_N^*(t,\mathbf{q}) = \min\{h \in \mathbb{H} : A^*_{(t,\mathbf{q},h)} = N\}$. Naturally, $h_N^*(t,\mathbf{q}) = \eta+1$ when $h_c = \eta+1$ for all $\mathbf{q} \in \mathbb{Q}\backslash\mathbf{0}$ and $h_N^*(t,\mathbf{0}) = 1$ for all $t \in \mathbb{T}$.

Proposition 1 states that the optimal value function, $V(t,\mathbf{q},h)$, is non-increasing in h under the terminated vaccination service scenario ($h_c = 0$). However, this property does not hold for the continuous service scenario ($h_c = \eta+1$) because, when there are fewer remaining timeslots in a day, the demand realization in the rest of the current day may leave more vaccine vials to be used in the next day leading to an increase in $V(t,\mathbf{q},h)$ (See Example 1).

Proposition 1. *When $h_c = 0$, the maximum expected number of doses administered until the next inventory replenishment for a given state $s = (t,\boldsymbol{q},h)$ is non-increasing in h. That is, if $1 \le h_1 \le h_2 \le \eta$, then $V(t,\boldsymbol{q},h_1) \ge V(t,\boldsymbol{q},h_2)$ for all $t \in \mathbb{T}$, $\boldsymbol{q} \in \mathbb{Q}$.*

Example 1. *Suppose $h_c = \eta+1$, $n = 2$, $T = 20$, $Q_1 = 14$, $z_1 = 5$, $Q_2 = 15$, $z_2 = 10$, $\eta = 480$, and $\mu = 24$ (i.e., $p = 0.05$). When $t = 2$, $q_1 = 1$, and $q_2 = 1$, $V(t,q_1,q_2,h)$ is increasing between timeslots $h = 386$ and $h = 437$, e.g., $V(2,1,1,386) = 10.6125$, ...,*

$V(2,1,1,437) = 12.5270$. *This example implies that* $V(t, \boldsymbol{q}, h)$ *may increase in* h *under the continuous service scenario when* p *is large enough.*

Proposition 1 is sufficient to show that the optimal policies are of control-limit type in terms of service termination decisions. That is, when $h_c = 0$, the optimal action is to terminate vaccination service at and after a particular threshold timeslot $h_N^*(t, \mathbf{q})$, for each combination of day $t \in \mathbb{T}$ and inventory level $\mathbf{q} \in \mathbb{Q}$ as stated by Theorem 1. The existence of the optimal control-limit type policies simplifies the solution process for the proposed model, which could be very useful in solving the problem instances with large state space or in infinite-horizon. We observe that the optimal service termination decisions are also of control-limit type when $h_c = \eta + 1$ for all the problem instances in our extensive numerical experiments. Our numerical results also illustrate that such optimal threshold timeslots exist for switching from using larger vials to smaller vials, i.e., $A^*_{(t,\mathbf{q},h)} \in \{1, \ldots, i, N\}$ for all $h \geq h_i^*(t, \mathbf{q}) = \min\{h \in \mathbb{H} : A^*_{(t,\mathbf{q},h)} = i\}$.

Theorem 1. *When* $h_c = 0$*, the optimal decisions for terminating vaccination service are of control-limit type in terms of timeslot* h*. That is; there exists a threshold timeslot* $h_N^*(t, \boldsymbol{q}) \in \mathbb{H}$ *for each combination of* $t \in \mathbb{T}$ *and* $\boldsymbol{q} \in \mathbb{Q}$ *s.t.* $A^*_{(t,\boldsymbol{q},h)} = N$ *for all* $h \geq h_N^*(t, \boldsymbol{q})$.

For simplicity, we assume that vial size $i + 1$ is divisible to vial size i $\forall i \in \mathbb{I}$ n (i.e., $z_{i+1} = n_i z_i$ for an integer n_i) in the rest of the paper. Propositions 2 and 3.a state that $V(t, \mathbf{q}, h)$ is a non-decreasing function of the number of remaining days until the next inventory replenishment (t) and the vaccine inventory on-hand ($\mathbf{q}$) for both service scenarios, respectively. These properties provide justification for some trends we observe in our numerical experiments. For instance, $h_N^*(t, \mathbf{q})$ is non-decreasing when t decreases and $\mathbf{q}$ increases for all problem instances analyzed in Section 2.5. In addition, Proposition 3.b illustrates that the optimal value function is non-decreasing as the proportion of the smaller vials in the inventory increases for a given total number of vaccine doses.

Proposition 2. $V(t, \boldsymbol{q}, h)$ *is non-decreasing in* t*, i.e.,* $V(t_1, \boldsymbol{q}, h) \leq V(t_2, \boldsymbol{q}, h)$ *when* $1 \leq t_1 \leq t_2 \leq T$ *for any given* $\boldsymbol{q} \in \mathbb{Q}$*, and* $h \in \mathbb{H}$.

Proposition 3. *$V(t, \boldsymbol{q}, h) \geq V(t, \boldsymbol{q} - \boldsymbol{r}, h)$ for all $t \in \mathbb{T}$, $\boldsymbol{q} \in \mathbb{Q}$, $h \in \mathbb{H}$, and $\boldsymbol{r} = (r_1 \dots r_n)$ if one of the following conditions holds;*

a) $q_i \geq r_i \geq 0$ $\forall i \in \mathbb{I}$ or,

b) Recall that $\frac{z_i}{z_j}$ is integer, then, $\boldsymbol{r} = \boldsymbol{e}_i - \frac{z_i}{z_j}\boldsymbol{e}_j$ when $n \geq i \geq j \geq 1$, $q_i > 0$.

The structural properties and observations reported above imply that when policy makers design alternative pediatric vaccine administration policies that administer doses from non-homogeneous size vials, they should only specify time thresholds to switch from using a smaller size vial to a larger size one, and when to terminate vaccination services for each day and vaccine inventory level. These thresholds increase or decrease in monotone patterns, generally in the form of smooth curves (See Section 2.5.1), depending on the number of days remaining in the replenishment cycle and the available vaccine stocks. On one hand, this smooth monotone behavior suggests that implementing the proposed optimal policies may not cause extreme variations in day-to-day practice. On the other hand, implementing the proposed dynamic optimal policies would still cause variation in day-to-day practices, and thus could be challenging for medical practitioners to apply in real time. However, using the insights mentioned above, policy makers can design partially dynamic policies by categorizing the remaining days and available inventory levels into subgroups and assigning monotone time thresholds to these subgroups. The feasibility of deriving such more-stable and easier-to-implement alternatives to the optimal policies is illustrated in Section 2.5.

2.5 Numerical Results

We solve Equations 2.3–2.5 to derive the optimal administration policies for a set of childhood vaccines: pentavalent, measles, YF, and BCG. For ease of presentation, we consider only two vial sizes ($n = 2$) for each vaccine type. Table 2.2 presents the data on features of the vaccines used in our numerical analyses. We design a base case scenario that represents the pediatric vaccine administration practices in developing countries. These

practices usually replenish the clinic inventory on a monthly basis and provide service on each weekday in order to eliminate excessive missed demand (Lee et al., 2011; Assi et al., 2011; Guichard et al., 2010). Thus, we set the number of clinic days between two replenishments to $T = 20$. We divide each day into $\eta = 480$ timeslots so that $\tau = 1$ minute can satisfy Poisson arrival assumption (see Appendices C for justification). We set the *base case* demand rate to 11 patients per day ($\mu = 11$) which complies with the reported demand rates at drop-in clinics in developing countries (Yang et al., 2014; Chinnakali et al., 2012; Guichard et al., 2010). Considering the fact that the initial inventory levels are often determined based on the previous month's vaccine demand Assi et al. (2011), we fix the total number of doses available at the beginning of the time horizon to the expected total demand ($Q = 11 \times 20 = 220$ doses). Note that we use the same base case vaccine demand arrival rate for all vaccine types in our analysis because they all are on the routine pediatric immunization schedule recommended by Organization et al. (2017). The base-case demand rate (11/day) is within the range of daily pediatric vaccine demand rate reported in the literature for all considered vaccine types (Lee et al., 2010; Guichard et al., 2010). The parameter values are consistent with those in Mofrad et al. (2014). We examine the base case under both service scenarios (i.e., $h_c = \eta + 1$ and $h_c = 0$) and conduct detailed sensitivity analyses measuring the effects of length of time horizon (T), initial inventory level (Q), and vial sizes (z_1, z_2) on the performance of the proposed optimal policies.

In order to analyze the effect of the initial vaccine inventory levels for each vial size on the performance of the optimal policies, we define α_i as the proportion of vaccine doses held in type i vials on day 1 ($t = T$) in Equation 2.7. Note that the vaccine administration policy for $\alpha_1 = 0\%$, $h_c = \eta + 1$ reflects the current medical practice of providing continuous service using a single vial type while the optimal policies for $\alpha_1 = 0\%$, $h_c = 0$ are the same as those proposed in Mofrad et al. (2014). On the other hand, the optimal policies for $\alpha_1 > 0\%$ under both service scenarios represent the alternative practices proposed by our analysis.

$$\alpha_i = 100\% \frac{\text{number of type } i \text{ vials} \times \text{number of doses in a type } i \text{ vial}}{\text{total number of doses}} = 100\% \frac{Q_i \times z_i}{Q}, \quad \forall i \in \mathbb{I} \tag{2.7}$$

2.5.1 Optimal Vaccine Administration Decisions

We analyze the structures of the optimal policies for the base case parameters ($T = 20$, $Q = 220$, $\eta = 480$, and $\mu = 11$) and verify the existence of the properties presented in Section 2.4. Figure 2.2 shows the optimal actions during all timeslots on particular days ($t \in \{5, 10, 15, 20\}$) for various inventory levels (q_1, q_2) and vaccine types under the terminated service scenario. This figure confirms that the optimal timeslot thresholds for switching vial type and terminating vaccination service exist for all numbers of remaining days, inventory levels, and vaccine type combinations. That is, once the optimal decision is to open a small vial at the beginning of a timeslot for the first time on a given day, the optimal decision is either to open a small vial or to terminate vaccination service (if $h_c = 0$) for the remaining timeslots on that day. In all cases, vaccines are administered from small vials before terminating vaccination service for the day. We have verified that these observations hold for all parameter combinations in our numerical experiment setting.

Figure 2.2 also illustrates that the optimal timeslot thresholds typically increase as the number of remaining days until the next replenishment decreases and the number of large vials on-hand (q_2) increases. The optimal policies administer doses less conservatively for any given pair of (q_1, q_2) as t decreases by using large vials for a longer duration of time and terminating vaccination service later. Similarly, the optimal thresholds $h_1^*(t, \mathbf{q})$ and $h_N^*(t, \mathbf{q})$ increase as q_2 increases for all pairs of (t, q_1) in our numerical experiments. These two observations are intuitive as the benefit of saving available vaccines for the next day is smaller when there are more vaccines on-hand to cover the demand for fewer days.

In the majority of the cases, $h_1^*(t, \mathbf{q})$ decreases and $h_N^*(t, \mathbf{q})$ increases when q_1 increases for a given pair of (t, q_2). This pattern is reasonable as having more vaccines on-hand allows later service termination and having more small vials leads to less conservative use of them to cover the demand. Counterintuitively, this pattern may not always hold. For example, $h_1^*(15, 5, 14) = 130$ and $h_1^*(15, 6, 14) = 131$ for measles vaccine (Figure 2.2f). In this case, the model starts using small vials later when an additional small vial is available to keep more small vials for the next days and better control the vaccine wastage in the rest of the replenishment cycle. Moreover, the optimal policies terminate vaccination service later on

a day when either z_1 or z_2 (sizes of small and large vials) decrease due to the additional flexibility provided by smaller vials.

The above observations on the monotonic behaviors of the threshold timeslots $h_1^*(t, \mathbf{q})$ and $h_N^*(t, \mathbf{q})$ are important, as they provide insights to practitioners on how the current vaccine administration practices can be adjusted to improve vaccine coverage. Note that when the available inventory is significantly greater than the expected total demand in the remaining t days, the optimal policy may become indifferent between next opening a small or a large vial. Our MDP model selects opening small vials in such cases to minimize open-vial wastage. Therefore, in Figure 2.2, $h_1^*(t, \mathbf{q})$ may drop to zero as $q_1 + q_2$ increases. For example, when $t = 5$ and $q_2 = 12$, $g_1(t, q_1, q_2, h) = g_2(t, q_1, q_2, h)$ for all $q_1 \in \mathbb{Q}_{\mathbb{1}}$, $h \in \mathbb{H}$ for pentavalent vaccine (See Figure 2.2d). Such cases do not violate the monotonicity of the optimal thresholds, as both large and small vials can be used at any timeslot. Also note that the optimal timeslot thresholds for switching to small vials follow similar monotonic trends under the continuous service scenario (see Appendices D).

2.5.2 Base Case Performance of the Optimal Policies

Figure 2.3 illustrates the expected base case performance of the optimal policies for all vaccine types under various service and α_1 value (the proportion of doses held in small vials) scenarios. As α_1 increases, the expected base case system performance improves in terms of expected percentage demand covered ($\phi(T, \mathbf{Q})$) and open-vial wastage ($\omega(T, \mathbf{Q})$) in expense of increased expected total cost ($\pi(T, \mathbf{Q})$). The optimal policies using only small vials ($\alpha_1 = 100\%$) significantly reduce $\omega(T, \mathbf{Q})$ compared to those using only large vials ($\alpha_1 = 0\%$) for all vaccine types. However, most of the performance improvements can be achieved by only keeping a portion of vaccine stocks in small vials which costs significantly less than using only small vials. For example, when $h_c = \eta + 1$, the optimal policy for pentavalent vaccine with $\alpha_1 = 100\%$ reduces $\omega(T, \mathbf{Q})$ from 28% to 4% while increasing the expected cost by 75% compared to having $\alpha_1 = 0\%$. The reduction in expected vaccine wastage is from 12% to 4% for the same additional cost when $h_c = 0$. On the other hand, setting $\alpha_1 = 18\%$ reduces $\omega(T, \mathbf{Q})$ to 10% and 7% compared to $\alpha_1 = 0\%$ under continuous

Table 2.2: Input data

	Mean	Low	High	Source
Weight per dose (w_d), g				UNICEF (2015)
Pentavalent vaccine	2.840	1.120	4.827	
Measles vaccine	1.314	1.1478	1.481	
YF vaccine	1.340	0.448	2.610	
BCG vaccine	0.104	0.096	0.112	
Weight of vaccine vial disposals (w_i), g				Lee et al. (2010)
2-dose vial	1.914	1.723	2.106	
5-dose vial*	2.612	2.391	2.833	
10-dose vial	3.522	3.169	3.874	
20-dose vial	5.531	5.431	5.631	
Weight of reconstitution syringe disposal (w_{syr}), g	6.625	5.967	7.293	Lee et al. (2010)
Cost of medical waste disposal per kg (c_d), $	5.737	1.730	9.070	Lee et al. (2010)
Cost of vaccine storage per cm^3 (c_s), $	0.030	0.027	0.033	Lee et al. (2010)
Volume per dose (v_i), cm^3				Lee et al. (2010)
Pentavalent vaccine				
2-dose vial	11.000	9.900	12.100	
10-dose vial	4.400	3.960	4.840	
Measles vaccine				
5-dose vial	6.720	6.050	7.390	
10-dose vial	3.500	3.150	3.850	
YF vaccine				
5-dose vial	6.500	5.850	7.150	
20-dose vial	1.500	1.350	1.650	
BCG vaccine				
10-dose vial	2.200	1.980	2.420	
20-dose vial	1.200	1.080	1.320	
Cost per dose (c_i), $				Lee et al. (2010)
Pentavalent vaccine				
2-dose vial	3.500	3.250	3.750	
10-dose vial	2.000	1.800	2.200	
Measles vaccine				
5-dose vial*	0.633	0.570	0.696	
10-dose vial	0.246	0.221	0.271	
YF vaccine				
5-dose vial	0.728	0.655	0.801	
20-dose vial	0.645	0.581	0.710	
BCG vaccine				
10-dose vial	0.191	0.172	0.211	
20-dose vial	0.104	0.094	0.114	

Low and high parameter values provide a range for the input data. Our numerical analyses in Section 2.5 are performed with mean values while low and high values are used for sensitivity analyses (see Online Supplement H). The values obtained from interpolation are referred by *.

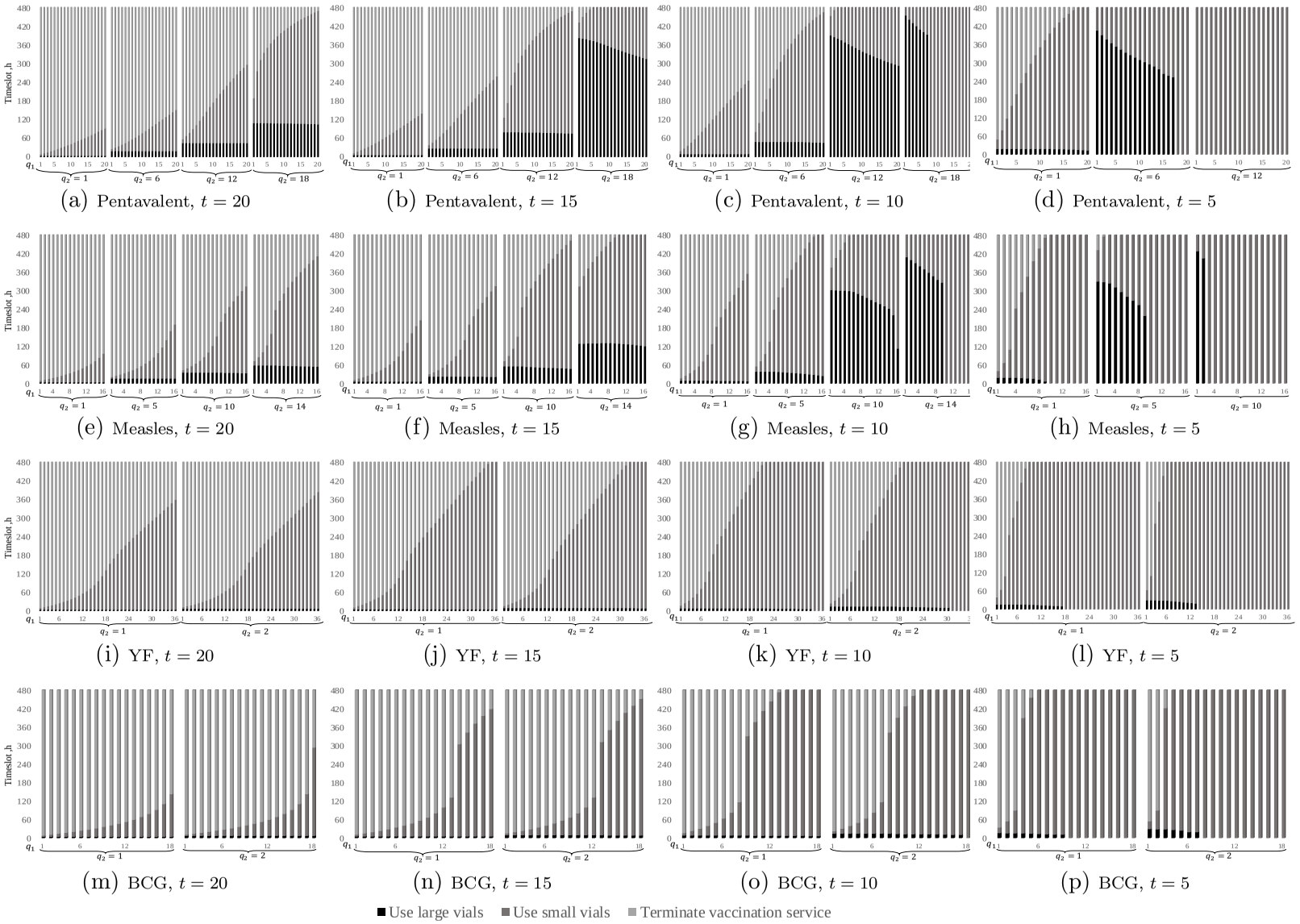


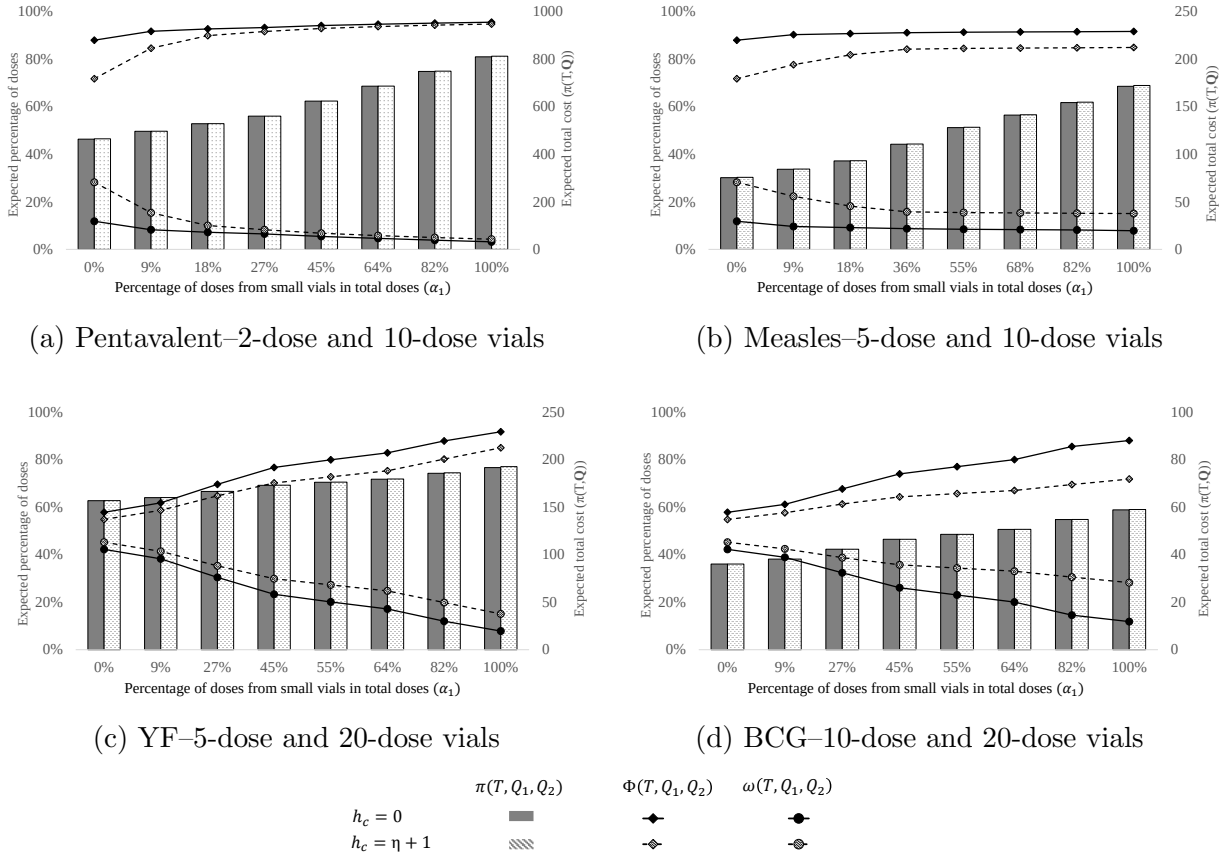
Figure 2.2: The optimal vaccine administration decisions for various (t, q_1, q_2) combinations under base case parameter scenario

and terminated service scenarios, respectively, in the expense of 14% increase in expected total cost.

We define $\bar{\alpha}_1$ as the minimum α_1 value for which the optimal policy can attain more than 95% of the improvement in $\phi(T, \mathbf{Q})$ achieved by the optimal policy with $\alpha_1 = 100\%$. The required $\bar{\alpha}_1$ values are 18%, 18%, 82%, 82% respectively for pentavalent, measles, YF, and BCG vaccines (see Appendix E for details). Note that when the size of large vials decreases, the corresponding $\bar{\alpha}_1$ value also decreases, e.g. $\bar{\alpha_1}$ is 18% for measles vaccines held in 5-dose and 10-dose vials; whereas, $\bar{\alpha_1}$ is 82% for YF vaccines held in 5-dose and 20-dose vials under the base case scenario.

Figure 2.3 also shows that when the sizes of both vial types are lower than the daily demand rate (e.g., for pentavalent and measles vaccines $z_1, z_2 \leq 11/day$), the number of doses administered increases and open-vial wastage decreases very rapidly as α_1 increases. Thus, the room for improvement diminishes before α_1 reaches 100%. Additionally, when $\alpha_1 \geq \bar{\alpha}_1$, the optimal policies with continuous service ($h_c = \eta + 1$) perform close to those with terminated service ($h_c = 0$) when z_1 is small enough. This is because vaccine wastage decreases as z_1 decreases, and greater vaccine availability may limit the additional benefits of terminating vaccination service. However, when the size of large vials is greater than the daily demand rate (e.g., $z_2 \geq 11/day$ for YF or BCG), opening a large vial even at the beginning of the day is associated with high risk of open-vial wastage. That is, when $\mu = 11$, the likelihood of having daily YF or BCG vaccine demand less than $z_2 = 20$ is around 0.98. Therefore, the marginal effect of α_1 on performance metrics may not diminish fast when $z_2 \geq \mu$ as seen in Figures 2.3c and 2.3d.

In Figure 2.3, the difference between the expected total costs of the optimal policies for continuous ($h_c = \eta + 1$) and terminated ($h_c = 0$) service is not very high because almost all available vials in inventory are opened due to having limited number of doses available in the base case, e.g., $Q = \mu \times T$. It is also interesting that the expected total costs of the optimal policies for YF vaccine do not vary much as α_1 increases, e.g., the expected total cost is \$157 when $\alpha_1 = 0\%$ and \$192 when $\alpha_1 = 100\%$ for $h_c = 0$ scenario. This result shows that, in some cases, the cost saved due to reduced vaccine wastage by using small



(a) Pentavalent–2-dose and 10-dose vials

(b) Measles–5-dose and 10-dose vials

(c) YF–5-dose and 20-dose vials

(d) BCG–10-dose and 20-dose vials

Figure 2.3: Base case performance of the optimal policies with continuous and terminated service scenarios

vials may compensate for the increase in vaccine procurement cost.

Our base case analysis shows that administering available vaccines by using only large vials and allowing early termination of vaccination service, as proposed in Mofrad et al. (2014), i.e., $\alpha_1 = 0\%$, $h_c = 0$, performs well in terms of expected demand covered for measles vaccine (See Figure 2.3b). However, alternative policies using both small and large vials under each service scenario (e.g., the optimal policies for $\alpha_1 = \bar{\alpha}_1$, $h_c = 0$ and $\alpha_1 = \bar{\alpha}_1$, $h_c = \eta + 1$) outperform the current practice and the optimal policies proposed in Mofrad et al. (2014) for the other three vaccine types. In case of pentavalent vaccine, the optimal policy for $\alpha_1 = \bar{\alpha}_1$ (18%), $h_c = \eta + 1$ covers more expected vaccine demand

than the optimal policy in Mofrad et al. (2014) without early termination of vaccination service. This observation verifies that practitioners can improve vaccine coverage and avoid negative effects of early service termination by keeping vaccine stocks in multi-dose vials of different sizes, while incurring a reasonably small increase in total cost.

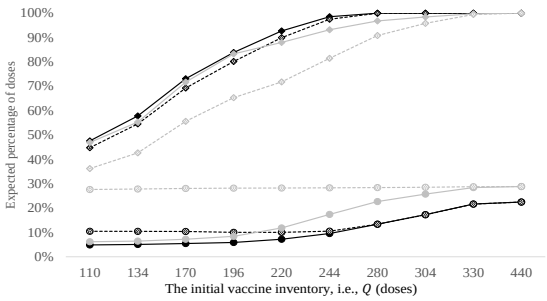
2.5.3 Sensitivity Analysis

Factors such as demand for vaccination service (μ), size of the initial vaccine inventory (Q), number of clinic days between inventory replenishments (T) may vary among healthcare centers. Therefore, we conduct sensitivity analyses measuring the effect of Q and T on the performance of the optimal vaccine administration policies. The sensitivity analysis on T also examines the effect of demand rate (μ) on the results.

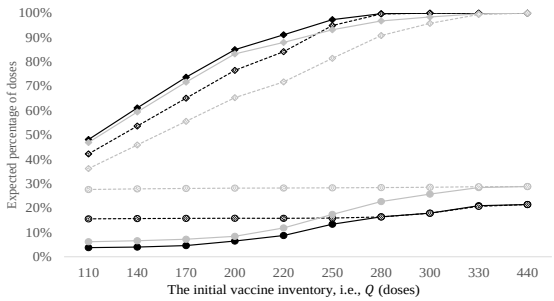
2.5.3.1 Sensitivity Analysis on Initial Inventory Level (Q)

Figure 2.4 compares the performances of the optimal policies with different Q values under various service and α_1 value scenarios for all vaccine types given the base case μ, T, and η values. In these comparisons, we set α_1 equal to $\bar{\alpha}_1$ to represent the scenarios using both small and large vials ($\alpha_1 > 0\%$). Figure 2.4 provides important insights about the combined effect of Q, h_c, α_1, and vial size combination (z_1, z_2) on the expected percentage demand covered and open-vial wastage.

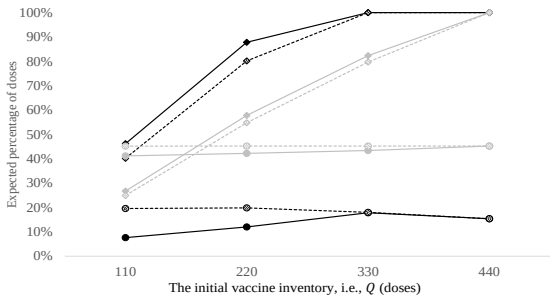
The expected percentage demand covered ($\phi(T, \mathbf{Q})$) is always non-decreasing in Q for all service and α_1 scenarios. On the other hand, the expected percentage of vaccine wastage ($\omega(T, \mathbf{Q})$) shows varying patterns. When vial sizes are smaller than the expected daily demand ($z_2 < \mu$), $\omega(T, \mathbf{Q})$ is non-decreasing in Q (See Figures 2.4a, 2.4b). This is reasonable because, as there are more vaccine vials in the stock, the optimal policies may use them less conservatively and waste more doses. However, $\omega(T, \mathbf{Q})$ may decrease as Q increases when $z_2 > \mu$ (See Figures 2.4c, 2.4d). This is mainly because, increasing Q means more small vials in the stock and the marginal benefit of an additional small vial in reducing vaccine wastage does not decrease fast when $z_2 > \mu$ (see Figures 2.3c and 2.3d).



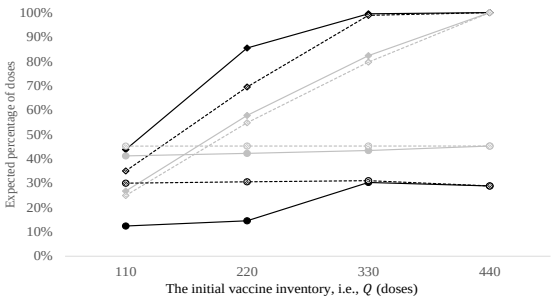
(a) Pentavalent - 2-dose and 10-dose vials



(b) Measles - 5-dose and 10-dose vials



(c) YF - 5-dose and 20-dose vials



(d) BCG - 10-dose and 20-dose vials

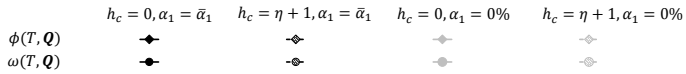


Figure 2.4: Expected percentage of demand covered and vaccine wastage corresponding to the optimal vaccine administration policies for various Q levels given the base case μ, T, and η values

The relative performances of the optimal policies with different service and α_1 scenarios merit a detailed discussion. Figure 2.4 demonstrates that, when the initial vaccine inventory is sufficiently large (e.g., $Q > 220$), the optimal policies utilizing both small and large vials with $\alpha_1 = \bar{\alpha}_1$ under both service scenarios perform better than the current practice ($\alpha_1 = 0\%$, $h_c = \eta + 1$) and the policy proposed in Mofrad et al. (2014) ($\alpha_1 = 0\%$, $h_c = 0$). The performance gaps between the models utilizing both vial types and those utilizing only large vials are more significant for YF and BCG, vaccines as the absolute differences between the vial sizes are greater in these cases.

The good performance of the optimal policies using only large vials and allowing termination, as proposed in Mofrad et al. (2014) for measles and pentavalent vaccines, is also noteworthy. The optimal policies for $\alpha_1 = 0\%$, $h_c = 0$ perform close to $\alpha_1 = \bar{\alpha}_1$, $h_c = 0$ and attain most of the achievable performance improvements when there are limited measles and pentavalent vaccines in stock (e.g., $Q \leq 220$). This is because when vaccine stocks are insufficient to cover the mean demand and $z_2 < \mu$, the size of stocks matters more than the levels of inventory for each vial size. This observation is important. The optimal policies for $\alpha_1 = 0\%$, $h_c = 0$ may be easier to implement, as they do not require specifying when to switch from using large to small vials.

Finally, Figure 2.4 shows that the optimal policies using both vial sizes with continuous service ($\alpha_1 = \bar{\alpha}_1$, $h_c = \eta + 1$) perform better than or as well as those using only large vials with terminated service ($\alpha_1 = 0\%$, $h_c = 0$) when vaccine stocks are sufficiently large. For instance, the performance improvements achieved by administering vaccines as proposed in Mofrad et al. (2014), by using only large vials and allowing early termination of vaccination service, can also be attained by using both small and large vials without any termination when Q is equal to or larger than the mean total demand (220 doses) for measles and pentavalent vaccines. This is one of the most important results in our study because the optimal policies for $\alpha_1 = \bar{\alpha}_1$, $h_c = \eta + 1$ may be desirable alternatives for practitioners preferring continuous vaccination service. Policy makers may prefer avoid early termination of vaccine services, as patients denied service may not be able to revisit the clinic later, due to accessibility issues in developing countries.

Given the above result and the higher procurement cost of smaller vaccine vials, it is desirable to determine the minimum α_1 value for which the performance of the optimal policies with $\alpha_1 > 0\%$, $h_c = \eta + 1$ match the performance of those with $\alpha_1 = 0\%$, $h_c = 0$. We refer to this threshold α_1 value as α_1^*. Table 2.3 presents α_1^* for various Q values in the cases of measles and pentavalent vaccines. As Q increases, the required α_1^* decreases. In some cases, α_1^* can be very small, e.g., practitioners can avoid early termination of vaccination service by keeping only 8% of the pentavalent vaccine doses in small vials while still improving the the expected vaccine demand covered when $Q = 240$. Such a policy would cost slightly higher than keeping all pentavalent vaccines in large vials.

Figure 2.4 and Table 2.3 imply that the performance of the optimal policies depends on Q and α_1 selection in terms of expected demand covered, wastage, and total cost. Thus, the optimal initial vaccine inventory level for each vial size is an important design problem, which we address in Section 2.5.4.

Table 2.3: The minimum α_1 values (α_1^*) for which the expected demand covered under the optimal policy with $\alpha_1 = \alpha_1^*$, $h_c = \eta + 1$ matches that under the optimal policy with $\alpha_1 = 0\%$, $h_c = 0$

	Terminated Service ($h_c = 0$)				Continuous Service ($h_c = \eta + 1$)					
	Measles & Pentavalent				Measles			Pentavalent		
Q	α_1	(Q_1, Q_2)	$\omega(T, \mathbf{Q})$	$\phi(T, \mathbf{Q})$	α_1^*	(Q_1, Q_2)	$\omega(T, \mathbf{Q})$	α_1^*	(Q_1, Q_2)	$\omega(T, \mathbf{Q})$
110	0%	(0,11)	6.17%	47%	NA	NA	NA	36%	(20,17)	7.46%
200	0%	(0,20)	8.19%	83%	NA	NA	NA	25%	(25,15)	8.55%
220	0%	(0,22)	11.83%	88%	NA	NA	NA	14%	(15,19)	12.11%
240	0%	(0,24)	15.55%	92%	33%	(16,16)	15.93%	8%	(10,22)	16.14%
250	0%	(0,25)	17.40%	93%	20%	(10,20)	17.75%	8%	(10,23)	16.60%
260	0%	(0,26)	19.22%	95%	15%	(8,22)	19.47%	8%	(10,24)	17.12%
270	0%	(0,27)	20.99%	96%	11%	(6,24)	21.52%	4%	(5,26)	21.59%
280	0%	(0,28)	22.68%	97%	11%	(6,25)	21.90%	4%	(5,27)	21.95%
300	0%	(0,30)	25.74%	99%	7%	(4,28)	24.33%	3%	(5,29)	22.75%
330	0%	(0,33)	28.41%	100%	0%	(0,33)	28.75%	0%	(0,33)	28.75%
440	0%	(0,44)	28.83%	100%	0%	(0,44)	28.83%	0%	(0,33)	28.83%

When $\alpha_1 = 0\%$, measles and pentavalent vaccines are administered from only 10-dose vials. Thus, the resulting $\omega(T, \mathbf{Q})$ and $\phi(T, \mathbf{Q})$ are the same for both vaccine types. When $Q \leq 220$, the optimal policies with continuous service may not outperform those with terminated service for any α_1 value for measles vaccine.

2.5.3.2 Sensitivity Analysis on the Total Number of Clinic Days During a Replenishment Cycle (T)

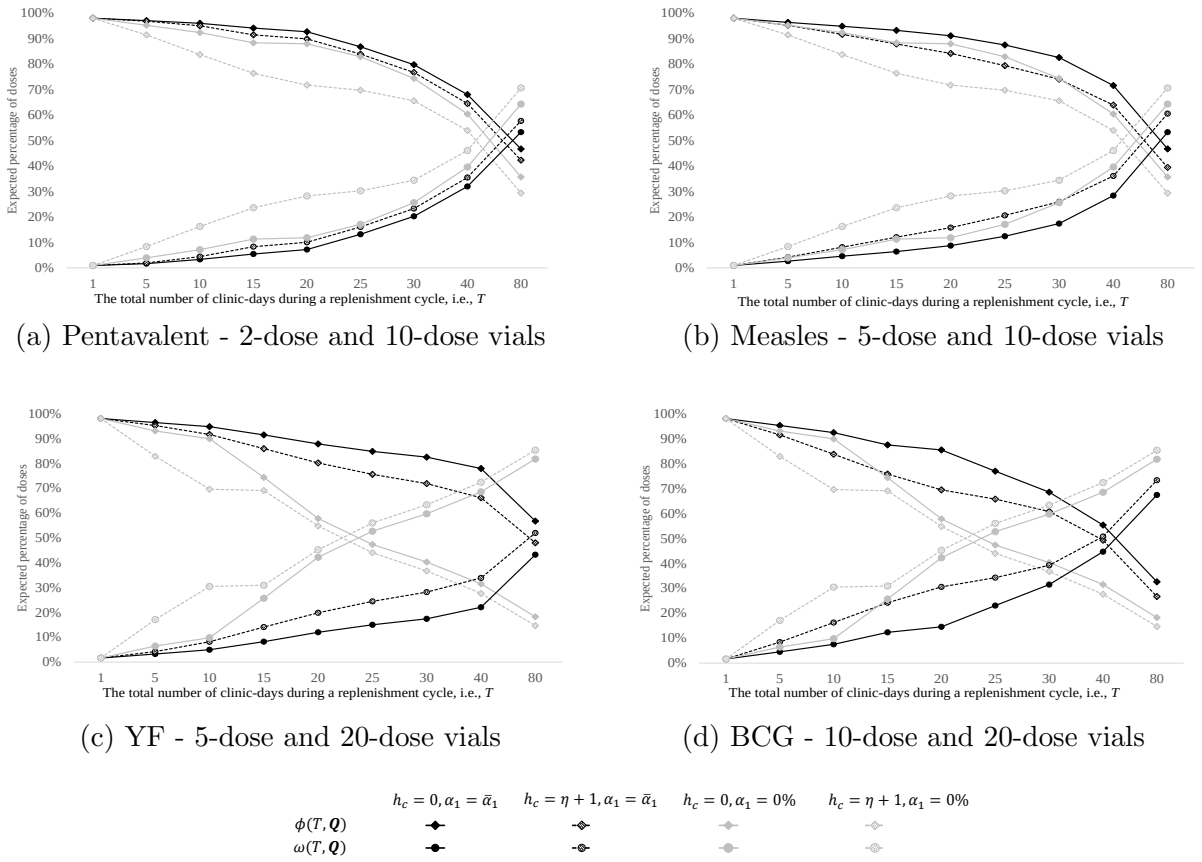
We know that the expected demand covered monotonically decreases as T increases for given μ and Q values (Proposition 2). In addition, our initial experiments show that the patterns observed as μ increases for fixed T and Q are very similar to those observed when Q decreases for given μ and T values (as depicted in Figure 2.4). Therefore, we omit these one-way sensitivity analyses, and focus on varying T and μ simultaneously so that the expected total demand during the replenishment cycle remains the same (i.e., $\mu \times T = 220$). This way, we investigate the effects of providing vaccination service on a limited number of clinic days over the time horizon, and pooling the demand on these days. Announcing a number of specific days for vaccine administration service to pool the demand

is a medical practice applied in developing countries. We observe that, as T decreases, the expected percentage of demand covered increases and the expected percentage of vaccine wastage decreases, due to the effect of pooling demand on fewer days. When T increases, the number of days with wastage also increases; therefore, it becomes more difficult to control the expected open-vial wastage without using small vials in addition to large vials. Especially, when $z_1 \geq \mu$, a significant amount of doses is wasted for both the $h_c = \eta+1$ and $h_c = 0$ scenarios when $\alpha_1 = 0\%$, compared to $\alpha_1 = \bar{\alpha}_1$. Figure 2.5 also implies that if the demand can be pooled to provide vaccination service on a few clinic days, vaccine demand coverage and wastage can be improved almost to the full extent by using only large vials and allowing service termination as recommended in Mofrad et al. (2014). However, if T is higher than a particular value, system performance can be further improved by using both vial sizes. In particular cases like pentavalent vaccine, both vials sizes can be used with continuous service without losing significant potential improvement in the expected demand covered. These results show the potential improvements that can be achieved by administering vaccines from both small and large vials in the setting of outreach vaccination sessions. A more accurate analysis requires an extension of our model and additional data; thus, left for future research.

2.5.3.3 Performance of the Optimal Policies under Different Cost Scenarios

Table 2.2 presents the input data used in total cost calculations for our numerical analyses. The numerical results in Section 2.5 are conducted using the mean values of these cost related parameters. We analyze the effect of the cost related parameters on the performances of the optimal policies by repeating some of the analyses in Section 2.5 with low and high values of these input parameters. We observe that the numerical results for the low and high cost parameters follow the same trends as those for the mean cost parameters.

Figure 2.6 shows the expected total cost of the optimal policies for the base-case T, μ, and Q under low, mean, and high cost-related parameter scenarios. As α_1 increases, the expected total costs of the optimal policies increase in similar fashions under all cost scenarios. Therefore, the percentage increases in the expected total costs associated with



(a) Pentavalent - 2-dose and 10-dose vials

(b) Measles - 5-dose and 10-dose vials

(c) YF - 5-dose and 20-dose vials

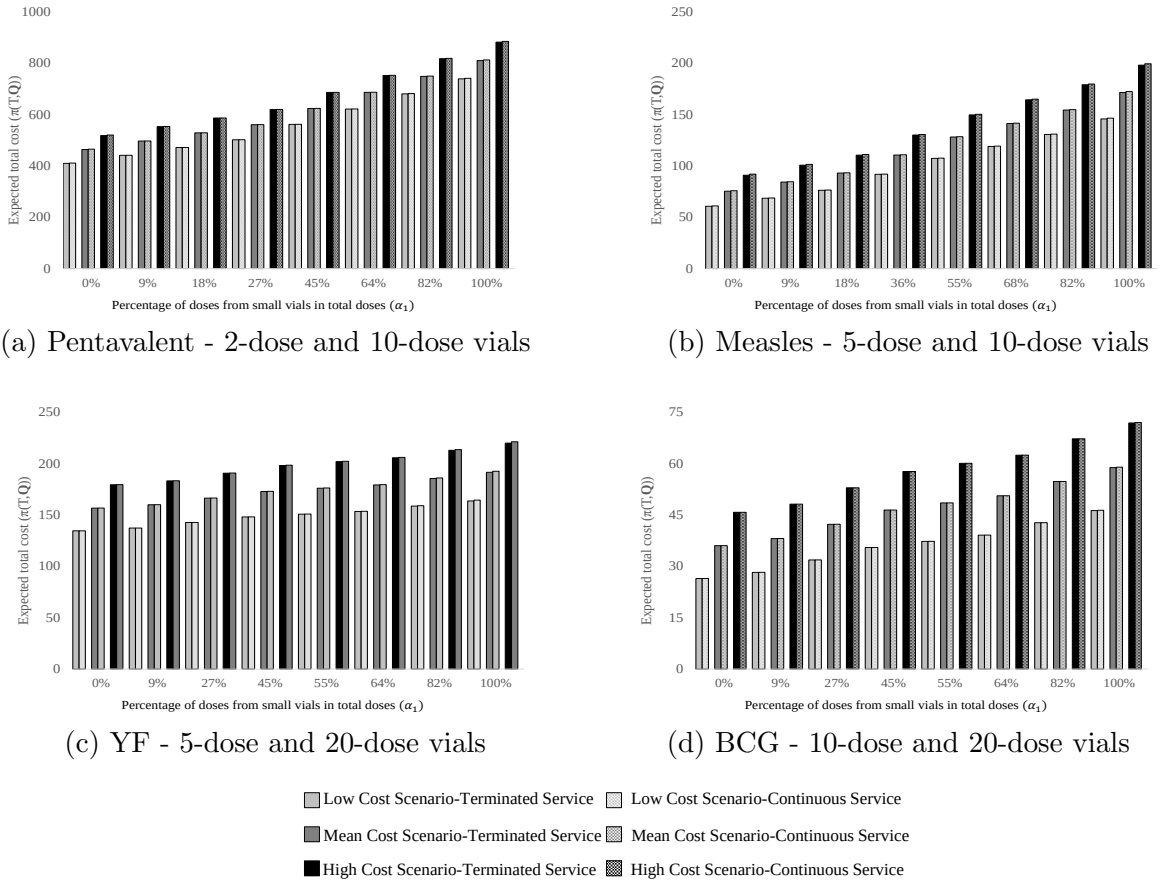
(d) BCG - 10-dose and 20-dose vials

Figure 2.5: Expected percentage of demand covered and vaccine wastage corresponding to the optimal vaccine administration policies for various T levels given the base case μ, Q, and η values

$\bar{\alpha}_1$ under low and high cost scenarios (defined in Section 5.2) are similar to those under mean cost scenario.

2.5.4 Implications on Vaccine Inventory Management

Section 2.5.3 implies that initial vaccine stock levels significantly affect performances of the immunization programs. The use of consumption data is promoted as a simple and effective method to find efficient stock levels (PATH, 2012). However, the number of vaccines consumed is highly correlated with the vaccine administration policy adopted by



(a) Pentavalent - 2-dose and 10-dose vials

(b) Measles - 5-dose and 10-dose vials

(c) YF - 5-dose and 20-dose vials

(d) BCG - 10-dose and 20-dose vials

Figure 2.6: The expected total costs of the optimal policies for low, mean, and high cost-related parameter values under various service and α_1 value scenarios

the immunization programs. Due to the recursive relationship between them, the optimal vaccine administration decisions and effective vaccine stock levels should be determined together.

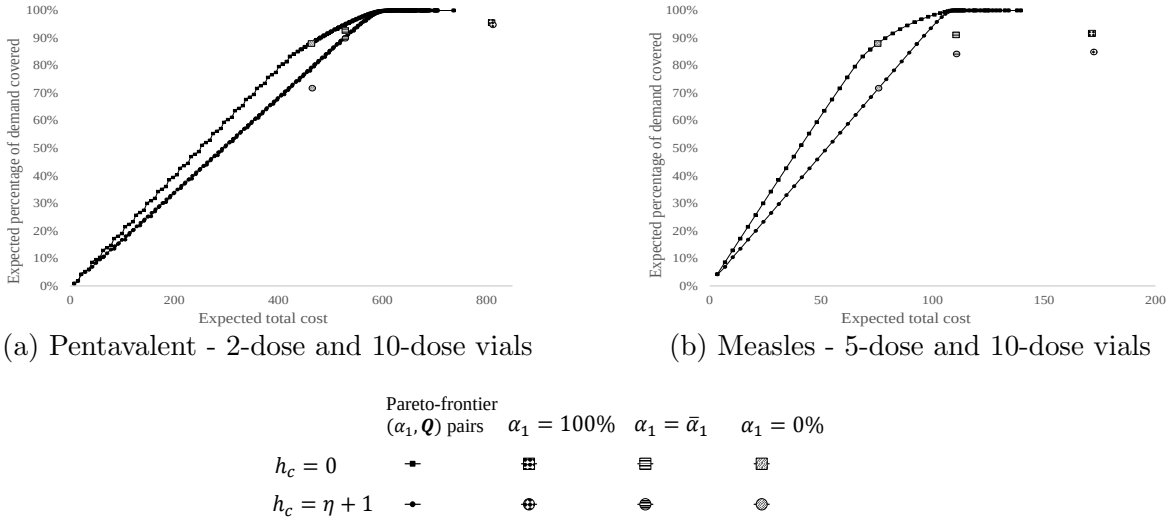
For this purpose, we use our MDP model to determine the effective vaccine stock level parameters, α_1 and Q. We derive the set of Pareto-efficient (α_1, Q) pairs that improve the vaccine administration practices with respect to expected percentage demand covered and total cost. We assume that each initial vaccine stock level (α_1, Q) considered is used to maximize the expected number of doses administered. In order to derive the Pareto-

efficient (α_1, Q) pairs and the corresponding optimal vaccine administration policies, we solve our MDP model for a large set of Q_1 and Q_2 combinations until reaching a sufficiently large $Q = Q_1 + Q_2$, for which having an additional vial does not increase expected demand covered. This approach is reasonable since the optimal vaccine demand covered is non-decreasing in Q and α_1 (Proposition 3). Then, we calculate the expected total costs using Equation 2.6, and eliminate dominated (α_1, Q) pairs. Note that we do not follow a complete multi-criteria optimization approach in this analysis, as the MDP model does not consider the associated total cost when deriving the optimal vaccine administration policies.

Figure 2.7 illustrates the performances of the optimal pentavalent and measles vaccine administration policies for the Pareto-frontier (α_1, Q) pairs under both service scenarios with the base case parameters. Note that the number of Pareto-efficient pairs is quite limited compared to the number of feasible (α_1, Q) pairs for each vaccine type (see Appendices F). The performances of the base case optimal policies with $\alpha_1 = 0\%$, $\alpha_1 = \bar{\alpha}_1$, and $\alpha_1 = 100\%$ are also highlighted in the figure for comparison.

In Figure 2.7a, there are 234 (289) Pareto-efficient initial inventory levels for pentavalent vaccine, and 87% (96%) of them utilize both small and large vials under terminated (continuous) service scenario. On the other hand, there are 62 Pareto-efficient (α_1, Q) pairs in Figure 2.7b for measles vaccine, and the majority of them (78%) uses only large vials under both service scenarios. The (α_1, Q) pairs utilizing small vials lie on the far right of the Pareto-frontier in Figure 2.7b; therefore, they are associated with incremental improvements towards 100% expected demand covered. These differences are possibly because the ratio of z_2/z_1 and that between procurement costs of small and large vials (c_1/c_2) are lower for pentavalent vaccine compared to measles vaccine. This result confirms that Mofrad et al. (2014)'s approach is sufficient to develop better vaccine administration practices for particular vaccine types (e.g., measles); however, the approach proposed in this paper can help further improve vaccine administration practices for other vaccine types such as pentavalent vaccine.

Although they may perform close to the Pareto-frontier, the optimal base case policies do not appear on the Pareto-frontier curves in most cases. That is, policy makers can



(a) Pentavalent - 2-dose and 10-dose vials

(b) Measles - 5-dose and 10-dose vials

Figure 2.7: The performance of the optimal vaccine administration policies for the Pareto-frontier (α_1, Q) pairs compared to the optimal base case vaccine administration policies with $\alpha_1 = 100\%$, $\alpha_1 = \bar{\alpha}_1$, and $\alpha_1 = 0\%$ under both terminated and continuous service scenarios given the base case values of T, μ, and η

improve vaccine coverage in a cost-effective, if not a cost-saving, way by smartly selecting the initial vaccine inventory levels for each vial size. To clarify this notion, Figure 2.8 shows the efficient (α_1, Q) pairs for pentavalent vaccine on the Pareto-frontier in more detail, and highlights some of the promising pairs. For instance, administering pentavalent vaccines using 22 large vials (220 doses) under continuous service (i.e., $\alpha_1 = 0\%, h_c = \eta + 1$), which reflects the current medical practice, is associated with 72% expected demand covered and \$465 expected cost per inventory replenishment cycle. Policy makers can improve the covered expected demand to 80% without additional cost by using only 208 doses, 9% of which are kept in small vials (see inventory Solution A), under the continuous service scenario. However, at this cost level, policy makers still have a significant incentive to allow early termination, and further improve expected demand covered to 88% using 216 doses with 3% vaccines kept in small vials (see Solution B).

The incentive to allow early termination of vaccination service decreases as the desired level of expected demand covered or/and total cost increase on the Pareto-frontier curves.

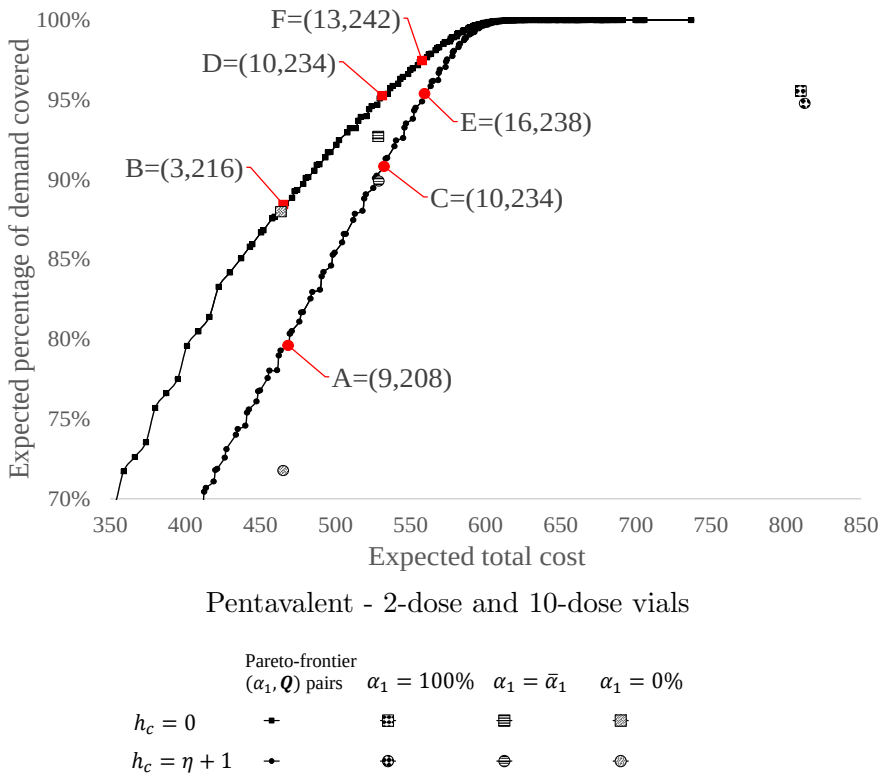


Figure 2.8: The detailed view of Pareto-frontier (α_1, Q) pairs for pentavalent vaccine under both terminated and continuous service scenarios given the base case values of T, μ, η

Therefore, if the policy makers can afford a reasonably small increase in cost, they can use alternative vaccine stock levels that are associated with higher expected demand covered, and thus a lower incentive to allow early termination of vaccination service. For example, Solutions C and E cost only 14% and 20% more than the current practice and increase the expected demand covered to 91% and 95% levels, respectively. At these solutions, policy makers can increase the expected demand covered to 95% and 97%, respectively, by switching to another vaccine inventory level with the same cost (i.e., Solutions D and F) and allowing early termination. Also, note that Solutions E and D achieve almost the same level of expected demand covered, but cost much less than keeping all vaccines in small vials with a base case vaccine supply (i.e., 220 doses in 110 small vials).

This Pareto-frontier analysis is also useful in specifying how the current vaccine administration practices and initial vaccine inventory levels can be improved cost-effectively to achieve target levels of expected demand covered. Such an analysis is relevant for organizations like the WHO and Global Alliance for Vaccines and Immunization (GAVI), which continuously strive to improve vaccine coverage beyond targeted levels in developing countries with a limited budget. Table 2.4 presents the (α_1, Q) pairs for which the optimal policies with $\alpha_1 = 0\%, h_c = \eta + 1$; $\alpha_1 = 0\%, h_c = 0$; $\alpha_1 > 0\%, h_c = \eta + 1$; and $\alpha_1 > 0\%, h_c = 0$ achieve 90%, 95% and 99% expected demand covered.

Table 2.4: The initial vaccine inventory levels that achieve the targeted levels of expected demand covered with minimal cost given the base case values of T, μ, and η

	(α_1, Q), $\pi(T, \mathbf{Q})$			
Target Demand	$\alpha_1 = 0\%, h_c = \eta + 1$	$\alpha_1 = 0\%, h_c = 0$	$\alpha_1 > 0\%, h_c = \eta + 1$	$\alpha_1 > 0\%, h_c = 0$
90%	(0,28), \$590	(0,23), \$484	**(11,21)**, \$525	**(2,22)**, \$478
95%	(0,30), \$624	(0,26), \$544	**(18,20)**, \$553	**(11,21)**, \$524
99%	(0,32), \$645	(0,30), \$617	**(18,22)**, \$582	**(16,22)**, \$573

The Pareto-efficient initial vaccine levels are highlighted with bold font.

Table 2.4 shows that, by using the proposed policies utilizing both small and large vials with the right vaccine stock level, the target levels of expected demand covered can be achieved with significantly less cost than using only large vials with (as in Mofrad et al. (2014)) or without (as in the current practice) early termination of vaccination service. Table 2.4 also illustrates that the proposed optimal vaccine allocation policies using both vial types under continuous service can achieve the target levels of expected demand covered with minimal or no cost incentive to allow termination (compare the columns for $\alpha_1 = 0\%, h_c = 0$ and $\alpha_1 > 0\%, h_c = \eta + 1$ in Table 2.4). That is, by using both small and large vials with the right inventory level, policy makers can avoid early termination of vaccination service and incur either negligible or no additional cost to improve the expected demand covered beyond 90%.

Global policy makers such as WHO and GAVI may benefit from the proposed model

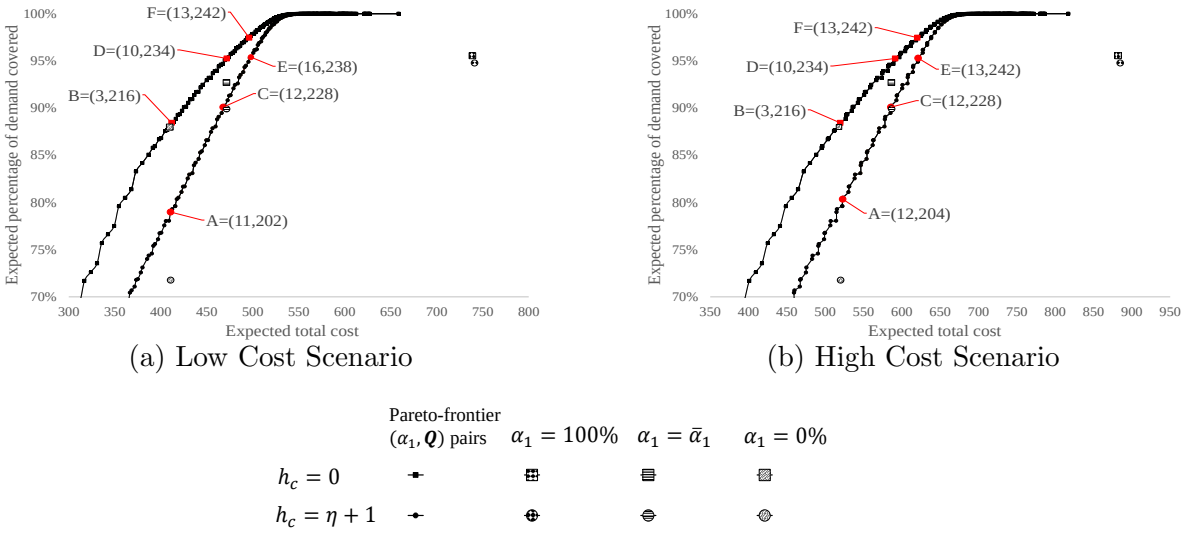
and the results provided in this section to update their future improvement targets for vaccine coverage, and to improve vaccination practices and guidelines. Using the proposed model, policy makers can determine to what level vaccine coverage may be improved by using multi-dose vials of different sizes, given limited vaccination resources. For instance, Table 2.4 shows whether the proportion of the expected demand covered can be increased to the 90-99% level in clinics for pentavalent vaccine using both small and large vials for a small increase in cost. Depending on the feasibility of the associated costs in Table 2.4 compared to the cost of current practices, WHO and GAVI may decide to update their current and future improvement targets for vaccine coverage levels.

Policy makers may also use the results of the proposed model to update their recommendations or develop alternative guidelines for effective vaccine administration as well as efficient selection of multi-dose vial-sizes and periodic vaccine stock levels. For instance, WHO recommends developing countries to monthly replenish pentavalent vaccine inventories, held in 10-dose vials, to the level of 1.25 times the expected total demand (e.g., 25% buffer stock) (WHO, 2002, 2013). In our base case, this recommendation requires keeping 28 10-dose vials whose performance is presented in the first column and the first row of Table 2.4. The third column of the same table (see rows 1 and 2) shows that holding a buffer stock of 5-8% of the expected demand may achieve a higher expected demand covered for less cost given that 9-15% of the pentavalent vaccine stocks are held in 2-dose vials and the rest in 10-dose vials. Such alternative recommendations may also help reducing vaccine buffer stocks which is a desirable outcome as reported in the literature (Zaffran et al., 2013).

In order to achieve the potential performance improvements associated them, such alternative vaccine stock level recommendations should be supported with implementable guidelines specifying when to switch from using large vials to small vials. This brings us to the final set of results we provide in Section 2.5.

2.5.4.1 Pareto-efficient Initial Vaccine Inventory Levels for Each Vial Size under Different Cost Scenarios

Figures 2.9a and 2.9b show the Pareto-efficient (α_1, Q) pairs derived using the low and high cost-related parameter values, respectively, for pentavalent vaccine, and highlight some of the promising pairs. Recall that the promising Pareto-frontier (α_1, Q) pairs for mean cost-related parameter values (see Figure F.1a of Section 2.5.4) are Solutions A=$(9, 208)$, B=$(3, 216)$, C=$(10, 234)$, D=$(10, 234)$, E=$(16, 238)$, and F=$(13, 242)$. The promising inventory solutions under terminated service scenario, i.e., Solutions B, D, and F, in Figure F.1a are the same as those in Figures 2.9a and 2.9b. However, the promising solutions for low and high cost scenarios are slightly different than those for mean cost scenario under continuous service. This implies that the performances of initial vaccine inventory levels for each vial size are more sensitive to variations in cost parameters under continuous service scenario.



(a) Low Cost Scenario (b) High Cost Scenario

Figure 2.9: The detailed view of Pareto-frontier (α_1, Q) pairs for pentavalent vaccine under both terminated and continuous service scenarios given the base case values of T, μ, η

Table 2.5 presents the (α_1, Q) pairs attaining 90%, 95% and 99% expected demand covered under various service and α_1 scenarios with low and high cost-related parameter

values. The initial inventory levels for each vial size presented in this table are the same as those presented in Table 2.4 of Section 2.5.4.

Table 2.5: The initial vaccine inventory levels for each vial size that achieve the targeted levels of expected demand covered with minimal cost given the base case values of T, μ, and η

	(α_1, Q), $\pi(T, \mathbf{Q})$			
Target Demand	$\alpha_1 = 0\%, h_c = \eta + 1$	$\alpha_1 = 0\%, h_c = 0$	$\alpha_1 > 0\%, h_c = \eta + 1$	$\alpha_1 > 0\%, h_c = 0$
90%	(0,28), \$521*, \$661**	(0,23), \$428*, \$542**	**(11,21)**, \$466*, \$585**	**(2,22)**, \$423*, \$535**
95%	(0,30), \$551*, \$699**	(0,26), \$481*, \$609**	**(18,20)**, \$493*, \$615**	**(11,21)**, \$465*, \$584**
99%	(0,32), \$569*, \$723**	(0,30), \$545*, \$691**	**(18,22)**, \$518*, \$647**	**(16,22)**, \$518*, \$637**

The Pareto-efficient initial vaccine inventory levels are highlighted with bold font. The expected total costs for low and high cost scenarios are denoted with * and **, respectively.

2.5.5 Simple Vaccine Administration Policies

In order to achieve the high expected demand coverage rates discussed in the previous section, the compliance of the healthcare practitioners to the proposed vaccine administration policies is critical. However, it may not be practical for healthcare practitioners to keep track of the optimal actions which depend on the number of days remaining until the next vaccine inventory replenishment, the number of vials on-hand, and the timeslot of the day. Nevertheless, the complexities regarding the practical implications of the proposed policies can be significantly reduced by deriving easier-to-implement policies based on the information and patterns provided by the optimal solution of the MDP model.

In this section, we show that such easier-to-implement policies with good performances, hereafter called as the *simple policies*, can be derived using our modeling framework. For this purpose, we focus on the pentavalent vaccine case and derive the simple vaccine administration policies using two different vial sizes (i.e., $z_1 = 2$, $z_2 = 10$) under the continuous service scenario (i.e., $\alpha_1 > 0\%$, $h_c = \mu + 1$). We divide the replenishment cycle into sub-horizons of equal length and determine a single switching timeslot threshold for each sub-horizon which is independent of the inventory level on-hand. The simple policy basically recommends practitioners to switch from using large vials to small vials after the

same timeslot on each day of a sub-horizon. We refer readers to Appendices G for the details about the derivation process of the simple policy.

We evaluate the performance of the simple policy by slightly adjusting the proposed model. Let j denote the number of sub-horizons remaining until next replenishment and h_j be the switching timeslot threshold for sub-horizon j where $j \in \{1, \ldots, J\}$. The expected number of vaccines administered under the simple policy can be obtained using Equations 2.8-2.11.

$$\tilde{V}(t, q_1, q_2, h) = \begin{cases} \tilde{g}_1(t, q_1, q_2, h), h \geq h_j \\ \tilde{g}_2(t, q_1, q_2, h), h < h_j \end{cases} \quad , \forall j \in \{1, \ldots, J\}, t \in [(j-1)(T/J), j(T/J)], \mathbf{q} \in \mathbb{Q}, h \in \mathbb{H}, \tag{2.8}$$

$$where \quad \tilde{\nu}(t, q_1, q_2) = \sum_{h=1}^{\eta+1} \tilde{V}(t, q_1, q_2, h) p_X(h), \tag{2.9}$$

$$\tilde{g}_1(t, q_1, q_2, h) = \sum_{y=h+z_1}^{\eta} \left[z_1 + \tilde{V}(t, q_1 - 1, q_2, y) \right] p_{Y_h^1}(y) + \sum_{d=1}^{z_1} \left[d + \tilde{\nu}(t-1, q_1 - 1, q_2) \right] p_{D_h}(d), \tag{2.10}$$

$$\tilde{g}_2(t, q_1, q_2, h) = \sum_{y=h+z_2}^{\eta} \left[z_2 + \tilde{V}(t, q_1, q_2 - 1, y) \right] p_{Y_h^2}(y) + \sum_{d=1}^{z_2} \left[d + \tilde{\nu}(t-1, q_1, q_2 - 1) \right] p_{D_h}(d), \tag{2.11}$$

We use our base-case scenario parameters ($T = 20$ days, $\mu = 11$ patients/day, $Q = 220$ doses, $\eta = 480$) for pentavalent vaccine ($z_1 = 2$-dose, $z_2 = 10$-dose) and assume $\alpha_1 = \bar{\alpha}_1 = 18\%$ to illustrate the performance of the proposed simple policies. We divide the monthly replenishment cycle of the base case (i.e., 20 days) into 5-day periods (i.e., $J = 4$) corresponding to one-week-long sub-horizons. We derived the switching timeslots as $h_4 = 108$, $h_3 = 286$, $h_2 = 292$, and $h_1 = 317$ using the simple policy development procedure in Appendices G. Table 2.6 shows that the simple policy performs almost as well as the optimal policy with $\alpha_1 = 18\%, h_c = \eta + 1$ (i.e., no early service termination, 18% of vaccines held in small vials) in terms of the expected percentage demand covered and total cost. The performance of such simple policies could be further improved by incorporating weekly-specified service termination times; however, the magnitude of these improvements would probably be minor (See the first and last columns of Table 2.6).

Table 2.6: The performance of the simple policy compared to the optimal policies for the base case pentavalent vaccine scenario

	The Simple Policy	The Optimal Policies			
	$h_c = \eta + 1$	$h_c = \eta + 1$		$h_c = 0$	
	$\alpha_1 = 18\%$	$\alpha_1 = 0\%$	$\alpha_1 = 18\%$	$\alpha_1 = 0\%$	$\alpha_1 = 18\%$
$\phi(T, \mathbf{Q})$	89.54%	71.77%	89.91%	87.99%	92.71%
$\omega(T, \mathbf{Q})$	10.42%	28.23%	10.04%	11.83%	7.19%
$\pi(T, \mathbf{Q})$	\$528.88	\$465.17	\$528.81	\$463.65	\$528.55

2.6 Discussion and Conclusion

Our analysis provides particular key insights that may help policy makers design more effective pediatric vaccine administration practices. Using an MDP model, we show that expected percentage of demand covered can be significantly improved by holding pediatric vaccine stocks in multidose-vials of different sizes and dynamically switching from using large vials to small vials when administering pediatric vaccines in a given clinic day. We also show that, by smartly selecting an initial pediatric vaccine stock level, target demand coverage levels can be achieved with no or limited additional costs and without early terminating vaccination services.

These results suggest that vaccine administration and inventory management decisions for routine pediatric vaccines should be considered simultaneously to improve performance of the immunization programs in a cost-effective way. The landscapes of national immunization programs change due to the introduction of many new and expensive vaccines, and changing population dynamics. Thus, the immunization supply chain and logistics systems that were designed in the 1980s may not be effective anymore (WHO, 2014b). A recent study reports that more than 75% of 57 GAVI-eligible countries operated below the standards for the vaccine supply chain management practices including vaccine stock management and distribution Colrain (2013) which may result in stock-outs and avoidable wastage. The proposed model may help policy makers design integrated pediatric vaccine administration and vaccine stock management policies to reduce wastage further.

We derive the Pareto-efficient initial vaccine inventory levels and the associated optimal vaccine administration policies in the trade-off spectrum between expected demand covered and cost. This way we show that, by choosing the right initial inventory level on the Pareto-frontier, 1) the performance of the optimal policies can be further improved (cost-effectively) and 2) the optimal policies using both vial types with continuous service can achieve desired level of expected demand covered (e.g., $> 90\%$) with no additional cost compared to the current practice and with minimal incentive to allow early termination.

The good performance of policies with two vial sizes and continuous service in our numerical experiments is noteworthy because: i) these policies are simpler to follow in real-life practice compared to those allowing both service termination and utilization of small and large vials; ii) providing an uninterrupted vaccination service could be desirable for policy makers since it increases patient convenience and access to vaccination services which are shown to improve vaccine coverage significantly in different health care provision settings (Bach and Goad, 2015; Papastergiou et al., 2014). This is especially important for developing countries where accessibility to medical resources is a major problem: patients denied vaccination service due to an early service termination may not be able to revisit a clinic soon or at all (Heaton et al., 2017; Torun and Bakırcı, 2006) and are therefore, in the meantime, exposed to higher risk of infection (Akmatov et al., 2008). In addition, early termination of vaccination service may also cause equity/fairness problems as it may limit accessibility to vaccination service for many patients who need to visit clinics towards the end of the day (e.g., full-time working parents) (Degli-Atti et al., 2004; Goad et al., 2013).

The limitations of our analysis should be understood well to appreciate its practical importance. Although our results suggest that the optimal threshold timeslots for switching to small vials and terminating vaccination service follows monotone patterns, these thresholds vary significantly depending on the remaining days in the replenishment cycle and available vaccine stocks. Therefore, following the optimal vaccine administration policies may lead to significant variation in day-to-day practices which could be challenging for medical practitioners. However, we show that we can derived more stable and easier-to-implement policies with very good performance, by using the optimal policies and their structural properties.

Our results on vaccine stock management should be also viewed cautiously. We only focus on improving the performance of the vaccine inventory and administration in the practice setting component of the vaccine supply and delivery chain. For a more accurate analysis, which is beyond the scope of our work, the overall supply chain performance needs to be optimized by considering the decisions in other echelons of the chain including selection of suppliers and national order quantities based on discount schemes, transportation/distribution costs, and lead times. Furthermore, the proposed policies, which administer pediatric vaccines from vials of different sizes, may cause complexities in inventory handling and waste disposal. For example, the proposed optimal policies may require more storage space and disposal of more medical waste. However, the proposed policies keeps only a small portion of vaccines in small vials; therefore, they are still feasible in terms of facility capacities reported in the literature (Lee et al., 2011). Our preliminary sensitivity analysis on storage and disposal costs shows that the effects of such complexities on the cost performance of the optimal policies are limited. Moreover, we assume that vaccine doses in opened vials expire at the end of a clinic day, which is reasonable for many clinics and vaccine types. Although one could still incorporate expiration of the open-vials before the end of a day into our model, this would only increase the vaccine wastage under each policy and magnify the effect of using vials of different sizes.

Our analysis and results focus on pediatric vaccination in walk-in clinics in developing countries. However, the proposed approach and main findings may be applicable to the administration of other vaccine types (such as flu vaccines) and setting of developed countries. Concerns regarding vaccine wastage have also been raised in developed countries as well. As immunization authority expands to include additional healthcare providers, vaccine has now started to be distributed through non-traditional channels such as community pharmacies (PIWG, 2015). Although this provides a convenient alternative for immunizing patients, estimating quantities and predicting demand in the retail setting can be complicated. Thus, strategies to combat vaccine stockpiling and wastage of publicly funded vaccine have been discussed (HPPA, 2016). Moreover, our modeling approach may be applied to high-priced and slow-moving multi-dose pharmaceutical products with limited shelf-life such as oral agents used to treat HIV, hepatitis, and cancer. When patients

present a prescription for unusual quantities of these types of drugs, pharmacists may either dispense the requested amount and risk having the remaining stock expire or dispense the entire vial/bottle. Ultimately, the patient or a third party payer will incur the cost of the unused medication. Note that, for these applications, one may need to extend our MDP model to incorporate a non-stationary arrival process into (e.g., vaccine demand fluctuates during the flu season). Finally, conducting a comprehensive cost-effectiveness analysis to evaluate the actual performance of the proposed policies for a given country is desirable. However, for brevity, we left such an analysis for future studies.

Chapter 3

Deriving Effective and Simpler-to-implement Colonoscopy Screening Policies for Preventive and Follow-up CRC Screenings

In this chapter, we present our two-fold work on deriving effective and simpler-to-implement colonoscopy screening policies. The main motivation of this research is driven from the findings of Erenay et al. (2014). That is, the performances of colonoscopy screening policies suggested by guidelines can be significantly improved by making personalized screening decisions; however, the optimal policies are complex to be followed in practice which may hinder the potential of the optimal policies. Using a microsimulation model, we evaluate all relative alternative periodic policies under different scenario settings that also takes risk-level, age, and gender into account like the optimal POMDP policies but in a much simpler manner. The following sections present the motivation, the literature review, the simulation model and a detailed numerical experiments on the alternative policies based on a Pareto-efficiency analysis. We also present the idea of a novel bi-criteria constrained MDP model which can be solved to obtain the true Pareto-frontier that may further help

improving the performances of the alternative policies.

3.1 Introduction

Malignant neoplasm or commonly known as cancer is one of the most important diseases that the world is experiencing for the last decades. The American Cancer Society expects, approximately, $1,735,350$ new cancer cases and $609,640$ cancer deaths in the US in 2018 (Siegel et al., 2016; ACS, 2018). Similarly, Canadian Cancer Society reports $206,200$ expected new cancer cases and $80,800$ expected deaths from cancer for Canada in 2017 (CCS, 2018). Cancer is the second most common cause of death accounting for nearly 25% of all deaths in the US and the leading cause of death in Canada accounting for nearly 30% of all deaths (CDC, 2018).

The relative survival rate can be used as a criterion for measuring the improvements in diagnosis and treatment of cancers. This rate is to determine the likelihood of surviving for a designated duration of time (e.g., 5 years) after being diagnosed with a cancer compared to the survival of comparable people without any cancers based on normal life expectancy. The 5-year relative survival rate is estimated 70% during 2007-2013 in the US (ACS, 2018). In Canada, it is 63% during 2006-2008 based on the Canadian Cancer Statistics published in 2017 (CCS, 2018). Recent incentives and improvements in preventing and treating cancers help increasing the survival rate; however, the economic burden of cancer is still devastating. The estimated direct medical costs of cancer (total of health care expenditures) were \$74.8 billion in the US in 2013 and \$3.8 billion in Canada representing the 7^{th} most costly illness or injury in 2008. The most commonly diagnosed cancer in both men and women is lung and bronchus cancer followed by breast and colon and rectum (colorectal) cancers in the US (ACS, 2018). However, when sexes are considered separately the most frequently diagnosed cancers are breast and prostate cancers in women and men, respectively. Since the focus of this study is colorectal cancer, for further reading of cancer facts we refer readers to Cancer Facts and Figures in the US in 2018 presented by ACS (2018).

Colorectal cancer (CRC) is the third and second most commonly diagnosed cancer in both men and women that is expected to results in approximately $140,250$ and $26,800$ new cases; and $50,630$ and $9,400$ moralities, respectively, in 2018 in the US and in 2017 in Canada (ACS, 2018; CCS, 2018; Siegel et al., 2017). The incidence and mortality rates for CRC have been declining since mid-1980s in both the US and Canada (American Cancer Society, 2018; CCS, 2018). This declining trend is the indicator of how CRC screenings may help detection and/or prevention of CRC at early stages when treatment is more likely to be successful. Besides its benefits, there are also harms associated with CRC screening such as economic costs and screening disutility. Yabroff et al. (2008) estimate that the total health expenditures associated with CRC in the US will increase from \$7.5 billion to \$14 billion between 2000-2020. Therefore, optimally designing CRC screening policies is still a challenge that should be addressed.

Colorectal cancer is the development of a malignant tumor from the inner lining of the colon or the rectum which are parts of large intestine. Most cancerous tumors start as a noncancerous growth often referred as benign tumor or precancerous lesion in the form of a *polyp*. Adenomatous polyps (adenomas) are the most common type of polyps which are usually considered as the first sign of CRC development. The development from adenomas to invasive cancer may take time, e.g., in average more than 10 years. This slow course of development provides the opportunity for prevention and early detection of CRC through screening. CRC screening differs from the screening tests for other cancers by allowing removal of polyps at the same time when they are detected. Therefore, screening is very important to prevent CRC before its development as well as detecting cancer at early stages when it is more likely to be cured via less extensive treatments and recovery after treatment is faster. For more information about CRC, its development, and the facts and statistics, readers can refer to the annual reports of American Cancer Society (2018) and CCS (2018).

Table 3.1 presents several CRC screening tests, the associated accuracy levels, complexity (involving patient preparation, facilities and equipment needed, and inconvenience), false negative rates, and disutility. Among CRC screening tests, colonoscopy provides the highest accuracy since the entire colon can be examined. It also can biopsy and remove

polyps before becoming malignant. However, it requires the highest complexity (i.e., full bowel preparation, and sedation) and incurs the highest disutility (i.e., highest risk of bowel tears or infections compared with other tests). Sigmoidoscopy has a high performance as well but only for rectum and lower one-third of the colon. Its complexity is intermediate, it requires only minimal bowel preparation and no sedation or no specialist. Since sigmoidoscopy cannot remove large polyps, in case of detecting any abnormalities colonoscopy is needed. CT (computed tomographic) colonography provides a high performance when polyps are large. The entire colon can be examined by CT colonography, however; polyps cannot be biopsied or be removed. Thus, it is required to perform colonoscopy if abnormalities are detected. Its complexity is intermediate, however; a full bowel preparation is still needed. Barium enema is another effective screening test to detect large polyps, however, its accuracy is relatively low, i.e., it may produce false positive test results. Although, the disutility and complexity associated with Stool DNA test and Fecal Occult Blood test are very low, polyps are missed most of the time when those tests are performed. However, they may provide a better performance (still intermediate) only in case of having cancer. In this study, we mainly focus on colonoscopy screening as it is the most commonly recommended screening and follow-up test (Krist et al., 2007; Winawer, 2007; Zapka et al., 2012). A thorough analyses and comparison of CRC screening tests can be found in American Cancer Society (2018).

Table 3.1: CRC screening tests

Screening Test	**Test Type**	**Accuracy**	**Complexity**	**False Positive Rate**	**Disutility**
Colonoscopy	Invasive/Optical	Highest	Highest	None	Highest
Sigmoidoscopy	Invasive/Optical	High	Intermediate	Low	High
CT Colonography	X-ray	High	Intermediate	Low	Low
Barium Enema	X-ray	Low	High	High	Low
Stool DNA Test	Stool	Intermediate	Low	Intermediate	Lower
Fecal Occult Blood Test	Stool	Lower	Low	High	Lower

Source: American Cancer Society (2018); Erenay et al. (2014); NCI (2014)

The American Cancer Society, the American College of Radiology and the US Multi-

Society Task Force on Colorectal Cancer (a consortium representing the American College of Gastroenterology, the American Society of Gastrointestinal Endoscopy, the American Gastroenterological Association, and representation from the American College of Physicians) were collaborated to publish the most recent CRC screening guideline in 2008. Table 3.2 presents the recommendations of the most recent CRC screening guidelines for low-risk asymptomatic people. The recommendations for people with increased or high risk of CRC based on specific risk factors (history of adenomatous polyp and its size, personal history of CRC, and family history of CRC or polyps) differ from the ones presented in Table 3.2 and can be found in American Cancer Society (2018); Winawer et al. (2003). The recommendations are usually more intensive for increased or high risk people. For example, high-risk people with adenomatous polyp are suggested to undergo follow-up colonoscopy three years after the removal of the polyp (Winawer et al., 2003).

Table 3.2: The most recent screening guideline for the early detection of CRC in low-risk asymptomatic people, American Cancer Society - 2008

Screening Test	**Frequency**	**Recommendation**[1]
Fecal occult blood test (FOBT) with at least 50% test sensitivity, or fecal immunochemical test (FIT) with at least 50% test sensitivity,	Every year	Starting at age 50. Collection techniques and number of samples required is recommended by clinicians. FOBT with the single stool sample during a clinician office visit is not recommended. Guaiac-based toilet bowl FOBT tests are not recommended. Immunochemical tests are usually chosen over guaiac-based tests since they are more patient-friendly while providing equal or better sensitivity and specificity. There is no justification for repeating FOBT in response to an initial positive finding.
Stool DNA test	Every 3 years	Starting at age 50.
Flexible sigmoidoscopy (FSIG)[3]	Every 5 years	Starting at age 50. FSIG can be performed alone, or with a highly sensitive gFOBT or FIT performed annually.
Double-contrast barium enema (DCBE)	Every 5 years	Starting at age 50.
Colonoscopy	Every 10 years,	Starting at age 50.
CT Colonography [3]	Every 5 years,	Starting at age 50.

Source: ACS (2018)

[1] All positive tests (other than colonoscopy) should be followed up with colonoscopy.

The reduction in CRC incidence rates and mortality rates due to the CRC screening guideline is an indicator showing that regular screening can prevent developing CRC or deaths from CRC. Regular flexible sigmoidoscopy results in 21% and 26% reduction in CRC incidence and mortality, respectively (Schoen et al., 2012). Zauber et al. (2012) shows that patients with the history of adenomatous polyp removed during colonoscopy are 53% less likely to die from CRC compared to general population. Regular FOBT screening reduces the risk of death from CRC by 32% after 30 years of follow-up (Shaukat et al., 2013). Moreover, FOBT decreases CRC incidence rate by 20% through detecting large adenomatous polyps (Mandel et al., 2000). The cost-effectiveness of CRC screening methods is also important when designing screening guidelines. Economic evaluation of several CRC screening methods can be found in Cruzado et al. (2013).

In spite of the benefits of regular screening, American Cancer Society reports that only 59% of the US population older than 49-year old follow the recommended screening. The prevalence of absence from recommended screening is especially higher among adults younger than 65-year old (American Cancer Society, 2018). Also it is reported that there is a significant difference between screening frequencies suggested by the guideline and by clinicians (Boolchand et al., 2006; Klabunde et al., 2009; Nodora et al., 2011; Yabroff et al., 2011b). As presented in Table 3.2, current guidelines recommend asymptomatic low-risk patients undergoing colonoscopy every 10 years for CRC prevention and early detection (USPSTF, 2008). In addition, guidelines recommend high-risk patients with history of adenomatous polyp undergoing follow-up colonoscopy 3 years after the polyp removal (Winawer et al., 2003). If first follow-up colonoscopy finds no abnormality, later follow-ups are done in every 5 years. Two independent surveys report that around 43% of the clinicians do not comply with colonoscopy screening guidelines by recommending asymptomatic low-risk patients to follow alternative screening policies such as preventive colonoscopy screening of every 1-2, 2-4, 3, 4, 5, or 6-9 years rather than every 10 years (Klabunde et al., 2009; Nodora et al., 2011). Similarly, some other surveys report that more than half of the clinicians recommend alternative policies such as undergoing a follow-up colonoscopy every 3 years or sooner after a polypectomy rather than every 3-5 years (Mysliwiec et al., 2004; Boolchand et al., 2006). In addition, colonoscopy screening is

recommended more frequently than the US guidelines in some countries including Germany and Greece (Haidinger et al., 2008). We refer these practiced alternative policies not complying to current guidelines as *practiced static policies*.

The incompatibility of clinical practices with the CRC screening guidelines reveals that designing effective colonoscopy screening policies to balance the benefits and harms of colonoscopy is still an important question that should be addressed. The main driver of this incompatibility is projected as the difference between the number of CRC screening that the guidelines and the practitioners suggest. This results also in a difference between the costs associated with the suggested CRC screening policies as well as between the other performance metrics of them (risk of CRC, mortality, etc). Although one way to resolve the controversy on the best colonoscopy screening practice is to derive the optimal policies utilizing a POMDP model as discussed in Erenay et al. (2014), the optimal policies proposed in Erenay et al. (2014) are dynamic in time and may be hard to implement in real-life.

In this context, the first part of this chapter, namely, the first fold, addresses i) whether the gap between current colonoscopy screening practices and the guidelines are justifiable in terms of health improvements achieved by the practiced static policies compared to guidelines, and ii) if the gap is justifiable, how to improve current guidelines practically to close this gap. For (ii), we evaluate performances of all simpler-to-implement colonoscopy screening policies, i.e., a large set of static-periodic and age-dependent dynamic periodic policies, described in detail in Section 3.4, using a discrete-event patient-level microsimulation model.

Our simulation model mimics the progression of CRC in asymptomatic low-risk, high-risk, and post-CRC patients. Note that, the CRC progression in the proposed simulation model is equivalent to the progression defined in the POMDP model of Erenay et al. (2014). The considered performance measures in this simulation analysis include the expected total life-years, total quality adjusted life-years (TQALYs: the difference between expected lifetime and disutility), age-based CRC risk and mortality, total costs, average screening interval, and the number of required colonoscopies. Our results illustrate that if

the colonoscopy disutility is not extremely higher than our estimates, some of the incompliant colonoscopy policies (i.e., practiced static policies) with shorter screening intervals than the guidelines may outperform the guidelines in terms of health outcomes as well as being cost-effective. Moreover, our analysis with the simpler-to-implement policies illustrates that the some of the age-dependent dynamic periodic policies may perform very close to optimal policies under certain parameter conditions. The simulation is the first one developed using clinical data for detailed performance evaluation of alternative policies compared to the optimal CRC screening protocols derived from an analytical model by Erenay et al. (2014). We utilize the simulation method because it is hard to estimate some of the statistics (e.g., average screening interval) from the analytical model without significant modification.

In the second fold, we introduce an idea for a modeling framework that can be used to optimize multiple objectives. We propose the idea of a bi-criteria constraint POMDP model optimizing the objective under resource constraints such as number of colonoscopies performed or total cost. Although it is noted in the literature that solving a constraint POMDP model exactly is quite challenging and utilizing approximate approaches are praiseworthy (Cevik et al., 2018), the specific nature of colonoscopy screening allows us to reduce this constraint POMDP model to a constraint MDP. That is, since the colonoscopy screening not only diagnoses the abnormal lesions but also removes and biopsies them, the risk of developing a cancerous lesion is reset after every polypectomy. Therefore, keeping only a partial history of events in track is sufficient to perform this reduction from POMDP to MDP. We believe that this approach can provide screening policies that accrue most of TQALYs improvements without increasing the expected total cost and requiring slightly more frequent colonoscopy screening than the guidelines.

3.2 Literature Review

Cancer screening has been noticeably addressed using OR techniques in the literature (Güneş and Örmeci, 2018; Saville et al., 2018; Alagoz et al., 2011; Ivy, 2009; Heidenberger,

1996; Pierskalla and Brailer, 1994; Stevenson, 1995). Many of the studies in this literature use discrete-event simulation and Markov modeling techniques as they provide practical modeling of uncertainty and complexity associated with the cancer screening practices under fewer simplifying assumptions that can be validated (Alagoz, 2011; Roberts, 2011; Knudsen et al., 2007; Schaefer et al., 2005). In the cancer screening domain, there are several studies specifically focusing on colorectal cancer screening. Using natural history models that represent the progression of CRC, screening decisions are analyzed under different assumptions about patient characteristics (e.g., risk-level, gender, age, etc.) and screening frequencies (every 5-, 10-year, etc.) to improve the healthoutcomes for individuals or society (Meester et al., 2015; Schreuders et al., 2015; USPSTF, 2008). As noted before, Chapter 3 proposes a discrete-event simulation model and introduce an idea of a constrained MDP model. Therefore, this section reviews the most relevant studies from the CRC literature that also use either of the approaches in their analysis.

3.2.1 Discrete-Event Simulation

The studies using discrete-event simulation as a tool aims at informing the policy makers for effective CRC screening by facilitating analysis of CRC progression and performance evaluation of alternative screening protocols for cost-effectiveness analyses (van Ballegooijen et al., 2011). Among the several simulation models proposed to mimic CRC progression and test different screening strategies in the literature (Neilson and Whynes, 1995; Roberts et al., 2007; Subramanian et al., 2009; Campbell et al., 2017), there are three micro-simulation models developed by interdisciplinary research teams: (1) Cancer Intervention and Surveillance Modeling Network (CISNET) consortium: Micro-simulation Screening Analysis (MISCAN) (Loeve et al., 1999), (2) Simulation Model of Colorectal Cancer (SimCRC) (CISNET, 2012), and (3) Colorectal Cancer Simulated Population model for Incidence and Natural History (CRC-SPIN) (Rutter and Savarino, 2010). These micro-simulations mimic CRC incidences and mortalities in a particular population by explicitly incorporating the CRC progression for each individual. They are clinically-validated and very detailed models with extensive state space definitions covering all clinically relevant

conditions. The models use different data sources and assumptions to model the CRC progression. Detailed comparison of these micro-simulations is available in Kuntz et al. (2011) and van Ballegooijen et al. (2011).

Clinical researchers have been using these micro-simulation models to address several controversial issues. For example, using MISCAN Simulation, Loeve et al. (2000) show that the savings from prevented and early detected CRC cases through CRC screening may compensate the associated costs under particular conditions. Vogelaar et al. (2006) also use MISCAN Simulation model to estimate the extend of reduction in CRC mortality associated with a set of CRC screening strategies under different scenarios of risk factors, compliance rates, and treatment options. Using CISNET micro-simulation models, several studies evaluate the cost-effectiveness of new screening modalities such as DNA in stool and CT Colonography, and the conditions under which these new screening methods can replace the existing screening modalities such as colonoscopy (Ahlquist et al., 2008; Lansdorp-Vogelaar et al., 2009a; Knudsen et al., 2010). Lansdorp-Vogelaar et al. (2009b) analyze the effect of individualizing the CRC screening according to patient's age and gender by evaluating performances of a small set of different screening frequency and screening modality combinations using MISCAN. In another study, Lansdorp-Vogelaar et al. (2009c) use MISCAN model to analyze the effect of several factors such as increasing chemotherapy costs on the cost-saving behavior of CRC screening.

Ramsey et al. (2010) also use MISCAN model to evaluate the cost-effectiveness of current CRC screening guidelines for the patients with different family history while Wilschut et al. (2011c) evaluate a set of different periodic colonoscopy screening policies to determine how frequently the patients with different family history should undergo colonoscopy. Wilschut et al. (2011a) also use MISCAN model to determine how FOBT screening should be adopted in order to comply with the capacity limits for colonoscopy surveillance after a positive FOBT result. In addition, Wilschut et al. (2011b) and Goede et al. (2012) compare the performances of different types of FOBT tests for different screening intervals and screening age ranges using MISCAN model. Vanness et al. (2011) use all of the three CISNET CRC micro-simulations in order to compare the heath outcomes associated with every 5- and 10-year CT Colonography with some of the screening strategies recommended

by the guidelines such as every 10-year colonoscopy for asymptomatic low-risk patients. Rutter et al. (2011) evaluate the assumptions about the potential effects of risk factors on the natural history of CRC using CRC-SPIN model. Finally, Haug et al. (2012) use SimCRC model to determine the best screening protocol followed for an individual with prior false-positive FOBT result.

Among the applications of CISNET CRC micro-simulations, that of Zauber et al. (2008) is praiseworthy for showing the resourcefulness of discrete-event simulation in informing policy makers. Zauber et al. (2008) use MISCAN and SimCRC to evaluate different CRC screening termination ages to determine when to stop CRC screening for elderly patients. Recent US Preventive Services Task Force (USPSTF) guidelines promote their findings as evidence based on which USPSTF now recommends stopping CRC screening after age-75 for asymptomatic low-risk patients (USPSTF, 2008).

Recently, MISCAN model is adjusted to capture the effect of some factors on the screening performances and the guideline policies are re-evaluated using this adjusted MISCAN model to inform the 2018 American Cancer Society CRC screening guideline. Peterse et al. (2018) consider the fact that the number of young adults has been consistently increasing when adjusting the MISCAN model. Their adjusted simulation model evaluates the alternative policies that are designed using different screening start ages (e.g., 40, 45, 50), stopping ages (e.g., 75, 80, 85), and intervals (e.g., 5-, 10-, 15-year for colonoscopy, 1-, 2-, 3-year for stool-based tests, etc.) for several screening modalities (e.g., colonoscopy, FOBT, CT colonography, etc.) for a total number of strategies evaluated of 145. Using clinical data, they determine the efficient and near-efficient screening strategies from a Pareto-efficiency analysis. On the other hand, Meester et al. (2018) takes race and sex into account with different assumptions on CRC risk trend over time when conducting very similar analysis to Peterse et al. (2018). They use the same policy alternative set of 145 and evaluate them using two discrete-event simulation models, MISCAN and SimCRC. Pareto-efficiency analysis is utilized when generating the efficient and near-efficient screening strategies. Note that these very recent updates support our claim that the guideline policies can be significantly improved by considering several factors when determining screening frequencies (Knudsen et al., 2016; Van Hees et al., 2015).

Besides CISNET, several other clinical groups also develop discrete-event simulation models to address challenges in CRC screening. Ness et al. (2000) develop a discrete-event simulation model and use it to analyze the costs and utility associated with one-time colonoscopy screening at different age intervals. Subramanian et al. (2009) develop an agent-based simulation model to measure the effect of the compliance on the performance of CRC screening. Later, they use their model to compare the performances of CRC screening using colonoscopy and FOBT tests under budget constraints (Subramanian et al., 2010).

There are also CRC discrete-event simulation models developed within the OR community. For example, Roberts et al. (2007) develop a detailed discrete-event simulation model that mimics the progression of CRC. Using this simulation model, Tafazzoli et al. (2009) compare the cost-effectiveness of some selected CRC screening policies proposed by the existing guidelines. Moreover, Yaesoubi and Roberts also use this simulation model in order to measure the willingness-to-pay for CRC screening in both insurer's (Yaesoubi and Roberts, 2008) and patient's (Yaesoubi and Roberts, 2010) perspectives. In addition, Pilgrim et al. (2008) develop a discrete-event simulation model that mimics the complete pathways of CRC patients including screening, diagnosis, referral, treatment, metastasis, and end-of-life care. They use this simulation model to analyze the costs and benefits of several development options across CRC pathway in England. They also use this model for estimating the direct cost of CRC care services provided by National Heath Services in England (Bending et al., 2010). Finally, Erenay et al. (2011) develop a discrete-event simulation model to analyze the progression of metachronous CRC among CRC survivors. They use this simulation model to estimate the unknown parameters of metachronous CRC progression using an inverse estimation approach.

Our proposed simulation model is different than the models reviewed above due to the following reasons.

1) Many of the simulation applications are very detail-oriented and require a significant amount of computational resources to evaluate large sets of screening policies. Our model is more compact than the existing simulation models and requires less computational efforts. However, it can still estimate the key performance measures

with sufficient accuracy. Thus proposed framework is appropriate for simulation optimization purposes and assessing the optimal policies from analytic models which requires large number of simulation replication as such models consider all possible action and screening result pathways. This way we analyze a large set of alternative policies without limiting our analysis to only periodic policies considered in previous studies in the literature.

2) The existing simulation models generally focus on only asymptomatic low-risk patients while our model proposes a holistic approach and allows patients to change their risk-level during the simulation based on their screening results (e.g., low-risk to high-risk after a polyp is found).

3) Unlike many of the simulation studies, we use clinical data for our numerical experiments rather than completely relying on published data in clinical literature.

4) Using the proposed simulation model, we shed light to some questions which have not been addressed in the literature before.

   – "*Are the practices of clinicians in-compliant to current guidelines justifiable?*"

   – "*What form of easier-to-implement and dynamic periodic policy would perform close to the benchmark policy?*"

   – " *To what extent, can the current practice be improved without significantly increasing the number of colonoscopies recommended?*"

3.2.2 Markov Decision Process Models

Although the discrete-event simulation is employed more by the studies in this literature, Markov models and fully or partially observable MDPs are also utilized to model and analyze CRC progression. Leshno et al. (2003) develop a Hidden Markov Model to evaluate the cost-effectiveness of several CRC screening policies including one-time colonoscopy screening, colonoscopy followed by a 10-year interval of follow-up, annual FOBT, annual

FOBT and sigmoidoscopy in a 5-year interval, and annual DNA stool test for the patients with average risk of CRC. The model mimics the progression of CRC from non-cancerous lesions to adenomatous polyps to cancerous lesions. Patients are categorized into seven different states representing their health states based on the size of polyp, the stage of cancer, and death. Moreover, Wu et al. (2006) focus on analyzing the cost-effectiveness of DNA stool test and provide a comparison with several CRC screening tests (FOBT, flexible sigmoidoscopy, and colonoscopy). The comparison is derived from a Markov model which depicts the natural progression of CRC by representing health states as in Leshno et al. (2003). He et al. (2017) also utilize a Markov model to improve the colonoscopy screening practices by determining the bottlenecks. Moreover, determining the capacity of colonoscopy services is addressed in Güneş et al. (2015) to minimize CRC mortality through solving a dynamic compartmental model. Örmeci et al. (2015) utilize MDPs in queueing control systems and develop an event-based dynamic programming approach to model the endogenous relationship between colonoscopy screening and diagnostic services that arises from the shared utilization of the capacity. Erenay et al. (2014) also develop a POMDP model that determines colonoscopy screening decisions maximizing QALYs.

To the best of our knowledge, there is no study addressing CRC screening and surveillance practices using a constraint MDP or POMDP model in the literature. However, there are two studies using constraint MDP and POMDP modeling approach for breast cancer screening, respectively, Ayvaci et al. (2012) and Cevik et al. (2018). Ayvaci et al. (2012) introduce a finite-horizon discrete-time constraint MDP model to maximize the TQALYs through optimizing the diagnostic decisions after mammography under budgetary restrictions. The proposed idea for a constraint POMDP model will refer to an equivalent colonoscopy history-baed MDP model which can be modeled as mixed-integer program (MIP) and solved under additional to budget constraints. The proposed model of (Ayvaci et al., 2012) is similar to our idea given in Section 3.5 in the sense that the optimal solution is derived considering a resource constraint. We also aim at providing policies that optimize health outcomes and utilized resources (e.g., number of colonoscopies performed, expected costs, etc.) simultaneously. However, these models are still different since the states in Ayvaci et al. (2012) are completely observable whereas we still incorporate the

partially observable states into our decision process. Cevik et al. (2018), on the other hand, develop a constraint POMDP model that maximizes TQALYs under a constraint on the number of mammography screenings performed. This study is similar to our study by assuming that the health states are not completely observable as they depend on the screening results which may not be completely accurate (i.e., sensitivities of both mammography and colonoscopy are not 100%). The constraint POMDP model in our study is aimed to be solved exactly whereas the model in Cevik et al. (2018) is solved using an approximate method. Therefore, the solution procedure for these studies are different. Last but not least, the modeling idea given in this study also differs from others by incorporating not only specific measures into the objective function but many of them, e.g., number of colonoscopy screenings performed vs. TQALYS, costs vs. TQALYs, costs vs. CRC risks.

3.3 Simulation Model

In this section, we first describe the CRC progression and discuss the effects of colonoscopy screening on the progression in order to better explain how the simulation model works. We then introduce our simulation model and give the details about the assumptions, run settings, and input parameters. After briefly describing how the simulation outputs, namely, the policy performances, calculated, we explain the simulation calibration process.

3.3.1 CRC Progression and Colonoscopy Screening

Developing an *adenomatous polyp*, i.e., a noncancerous growth often referred as benign tumor or precancerous lesion, is the fist sign of CRC development. As commonly assumed in the literature (Loeve et al., 1999; Frazier et al., 2000; Roberts et al., 2007), we assume that CRC progression starts with developing an adenomatous polyp which then turns into adenocarcinoma, cancerous lesion. In other words, the progression follows polyp-to-adenocarcinoma sequence for each patient-type (e.g., low-risk). Figure 3.1 illustrates the annual CRC progression in all risk-levels. We assume no colonoscopy screening,

colonoscopy screening with diagnosis of a polyp, and colonoscopy screening with diagnosis of a cancer in Figures 3.1a, 3.1b, and 3.1c, respectively. The pointed arrows show the annual health state transitions. We define three pre-clinical (undiagnosed) states for each patient group in our model: "*No lesion*", "*Polyp*" and "*CRC*". We also define a death (D) and under cancer treatment (UCT) states. To make the Figure 3.1 tractable for the readers, we give the list of notation in the figure by Table 3.3.

Table 3.3: Reference table for Figure 3.1

Notation	Description
p_{onset}	The adenomatous polyp onset probability
p_{P-C}	The probability of adenomatous polyp-to-CRC progression
p_{d_i}	The probability of mortality, respectively, in the *No lesion*, *Polyp*, and *CRC* states for $i \in \{0, 1, 2\}$ before developing any CRC
p_{P-P}	The probability of developing a new polyp after a polypectomy developed at the beginning of the same year
$p_{d_{Co}}$	The probability of mortality immediately after a colonoscopy with polypectomy due to fatal complications
$\bar{p}_{C-N}$	The probability of developing no lesion after immediate treatment
$\bar{p}_{C-P}$	The probability of developing a polyp after immediate treatment
$\bar{p}_{C-C}$	The probability of local recurrence of CRC
p_{UCT-N}	The probability of developing no lesion after continuous treatment
p_{UCT-P}	The probability of developing a polyp after continuous treatment
p_{UCT-C}	The probability of developing CRC after continuous treatment
$p_{UCT-UCT}$	The probability that the continuous treatment fails
$\bar{p}_{d_i}$	The probability of mortality, respectively, in the *No lesion*, *Polyp*, and *CRC* states for $i \in \{0, 1, 2\}$ after developing CRC
p_{SD}	The probability of self-diagnosing of CRC
p_{ST}	The probability of successfully completing the immediate treatment

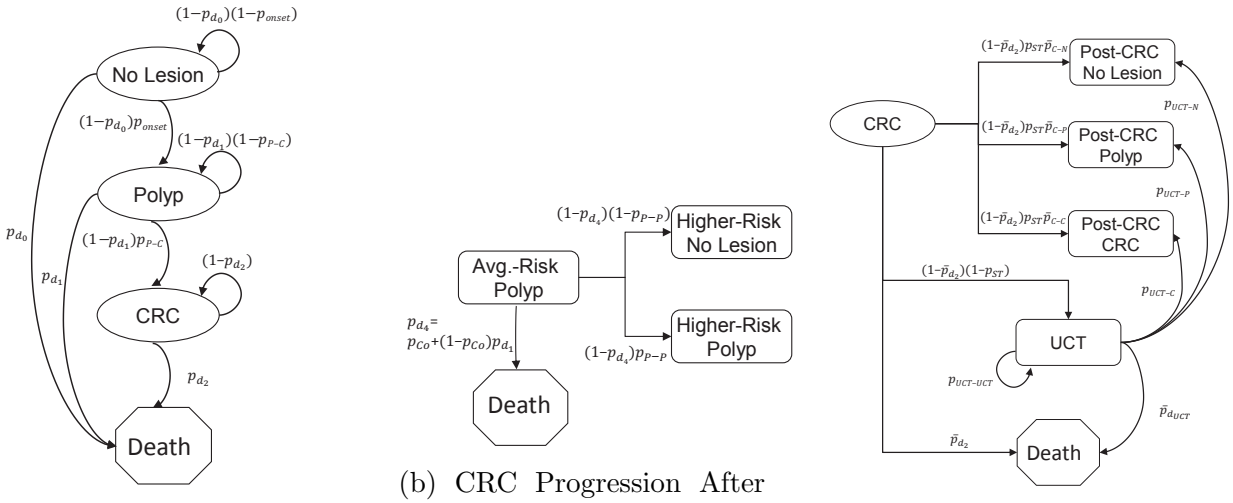
Immediate treatment refers to the first treatment after detecting CRC which takes less than a year

Continuous treatment refers to the treatment following the immediate treatment

Without preventive screening or self-diagnosis, colorectal cancer lesions randomly progress towards the next stage as shown in Figure 3.1a for all patient-types. The progression assumes that if a patient has no lesion at the beginning of year t, s/he may either develop an

adenomatous polyp with probability p_{onset} during the year and transit into the *Polyp* state at the beginning of next year, $t+1$, or may develop no polyp and stay in the "*No Lesion*" state with probability $1-p_{onset}$ given that s/he survives during year t with probability $1-p_{d_0}$. Similarly, if a patient with existing adenomatous polyp survives during year t with probability $1-p_{d_1}$, the adenomatous polyp may either progress to CRC with probability p_{P-C} or remain as an adenomatous polyp. Therefore, the patient who is in *Polyp* state in year t can be in the *Polyp* or *CRC* states in year $t+1$. If a patient has CRC and it is not self-diagnosed in year t, then s/he may either die from CRC or other natural causes during the year with probability p_{d_2} or continue to the next year with an undiagnosed CRC.

The progression illustrated in Figure 3.1a usually takes time, e.g., adenomatous polyps turn into invasive cancer in more than 10 years on average. This slow course of progression provides the opportunity for prevention and early detection of CRC through screening. Undergoing preventive or diagnostic CRC screening breaks down the polyp-to-adenocarcinoma sequence leading to different progression paths based on the screening result as in Figures 3.1b and 3.1c. Figure 3.1b mimics the CRC progression after a positive result for an ade-



(a) Progression Without Preventive Screening

(b) CRC Progression After Adenomatous Polyp Diagnosis

(c) CRC Progression After Cancer Diagnosis

Figure 3.1: CRC progression flow-charts: The probabilities on these figures vary by age. The probabilities on Figure 3.1 vary by patient-type

nomatous polyp from preventive colonoscopy screening. If a patient is diagnosed with an adenomatous polyp at the beginning of year t and survives during the year, he may develop a new adenomatous polyp with probability p_{P-P} and be in the "*Polyp*" state in year $t+1$ as a high-risk patient. The other possibility is staying polyp-free during year t, thus, being in the "*No Lesion*" state at the beginning of year $t+1$ as a high-risk level patient. Note that, in both cases, the CRC progression in the later years takes place according to Figure 3.1a with the progression parameters of high-risk patients until the next polyp or CRC diagnosis due to having history of adenomatous polyp.

Figure 3.1c illustrates the progression after CRC diagnosis. Whenever a patient is diagnosed with CRC s/he undergoes CRC treatment. In this study, we consider two different CRC treatment types based on the treatment success. The *immediate treatment* is the treatment that patients undergo right after the diagnosis of CRC and recovers from CRC in less than a year. On the other hand, the *continuous treatment* is followed if the treatment fails at the first attempt (i.e., unsuccessful immediate treatment) and the patient needs to continue undergoing treatment during the following year. If a patient survives an immediate CRC treatment at year t with probability $1-\overline{p}_{d_2}$, the cancer can be successfully treated with probability p_{ST}. In that case, the patient becomes a post-CRC patient in *No Lesion*, *Polyp*, or *CRC* states in the next year with probabilities $\bar{p}_{C-N}$, $\bar{p}_{C-P}$, and $\bar{p}_{C-C}$, respectively. On the other hand, if the immediate treatment fails and the patient survives during year t, her/his treatment continues during year t as continuous treatment and the patient stays in the "UCT" state in year $t+1$. If a patient is in the UCT state, then the treatment may fail and the patient may die (with probability $p_{d_{UCT}}$) or be in the UCT state in the next year (with probability $p_{UCT-UCT}$). If the treatment is successful, then the patient becomes a post-CRC patient in *No Lesion*, *Polyp*, or *CRC* states in the next year with probabilities p_{UCT-N}, p_{UCT-P}, and p_{UCT-C}, respectively. Note that, regardless of the treatment type reqired, transition to *CRC* state refers to the local recurrence of cancer. We refer the readers to the state transition matrices in Erenay et al. (2014) for more details about the CRC progression probabilities in the proposed simulation model.

Colonoscopy screening, the screening mode considered in this study, differs from the screening tests for other cancers by allowing removal of polyps at the same time when

they are detected. Therefore, colonoscopy screening is very effective to prevent CRC before its development as well as detecting cancer at early stages when it is more likely to be cured via less extensive treatments and recovery after treatment is faster (American Cancer Society, 2018). Colonoscopy screening may diagnose the existing lesions with a probability equal to the test sensitivity, i.e., probability of true positive, when a patient undergoes screening and survives against any colonoscopic complications. The sensitivity of colonoscopy for adenomatous polyps and CRC are reported to be around 80-90% and 85-95%, respectively (Frazier et al., 2000; Vijan et al., 2001). During colonoscopy, other abnormalities that are not associated with CRC (hyperplastic polyps, inflamation, etc.) can be also detected (Lieberman et al., 2000). Because all detected abnormalities are removed and sent to pathology which reveals their true nature, the specificity of colonoscopy (probability of true negative) is 100%. However, we consider the possibility of detecting non-CRC related abnormalities in calculating $p_{d_{Co}}$ because biopsy of such abnormalities increases the likelihood of fatal colonoscopy complications. Figure 3.2 illustrates how the screening results affect future CRC progression where τ_1 and τ_2 refer to colonoscopy sensitivity for adenomatous polyps and cancer, respectively. If a patient has no lesion, then the screening result is negative with probability 1 because the specificity of colonoscopy is 100%. On the other hand, if a patient has a polyp then the colonoscopy screening may detect and remove the polyp with the probability that equals the sensitivity of colonoscopy for polyps, τ_1, or miss the polyp (with $1-\tau_1$ probability). In case of having CRC, either the colonoscopy screening may detect the cancerous lesions with probability τ_2, or patients may self-detect the cancer if it is not detected by the colonoscopy (with probability $(1-\tau_2)p_{SD}$). The CRC progression follows the path in (i) Figure 3.1a if no lesion is detected, (ii) Figure 3.1b if an adenomatous polyp is detected, (iii) Figure 3.1c if a cancer is detected through the colonoscopy screening or self-detection (only in case of CRC).

3.3.2 Simulation Flow and Run Settings

The simulation model mimics the CRC progression at individual level for a particular population and synthesizes the individual outcomes to determine the health outcomes

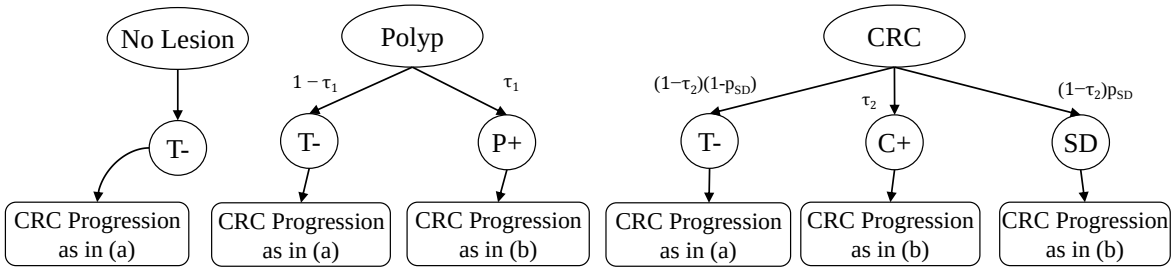


Figure 3.2: Screening results for particular health states.

for the population. For example, in order to estimate the heath outcomes of a particular colonoscopy screening policy among low-risk males with ages in 50-60, our model generates individuals in this risk-level one-by-one at random ages between 50 and 60 according to 2010 US Census population data (USCB, 2012). Figure 3.3 illustrates the simulation process flow. After initializing the input parameters (see Section 3.3.3 for the details), we generate the run scenario which specifies the population characteristics and the screening policies being evaluated. The population characteristics are risk level (e.g., high-risk), existing colorectal lesion (e.g., undetected polyp), gender (e.g., females), and age (e.g., aged between 50-65) of the patients according to their distributions among the size of the population that is given at parameter initializing step (e.g., 100,000 patients). The screening policies that are considered in this study are AGA guideline policies, the optimal POMDP policies of Erenay et al. (2014), the practice policies commonly followed by the clinicians, the alternative policies that we propose. The details of these policies are given in Section 2.5.

We consider three dynamic risk-levels, i.e., low-risk, high-risk, and post-CRC, and two lesion types, i.e., adenomatous polyp and CRC lesion. After patient generation, each year, the model determines whether a patient should undergo colonoscopy screening according to the screening policy being evaluated. For the simplicity, we assume that colonoscopy screening takes place at the beginning of the year. Based on the screening decision (i.e., being performed in that year or not), the simulation model follows the CRC progression described in Section 3.3.1 until the patients are 100 years old. If performed that year, colonoscopy

screening detects the existing colorectal lesions (CRC, adenomatous polyp, or others) with a probability equal to the sensitivity of colonoscopy. When the colonoscopy finds a lesion, the lesion is removed or treated. A patient may die from fatal colonoscopy complications such as perforation during colonoscopy with a particular probability. Polypectomy and biopsy of non-CRC related lesions increases the fatal complication probability. Note that, a patient with undiagnosed CRC may experience severe symptoms and undergo a diagnostic colonoscopy which reveals her/his cancer. We refer to this event as self-diagnosis and assumed that CRC self-diagnosis may happen if the simulated patient do not undergo a colonoscopy during the corresponding year.

Following the screening or self-diagnosis, the statistics about the screening results are updated and if a lesion is found then the risk-level of the simulated patient may change accordingly. After updating the statistics, the next screening age is determined. The patient moves to the lesion progression module where s/he ages one year during which either new lesions may develop, or existing lesions may progress to the next stage, or the patient may die from natural reasons, undetected CRC, or CRC treatment. When the lesion progression is completed, the statistics about the existing lesions and age are updated. The patient continues aging in the same manner until the simulation excludes her/him in the analysis because either s/he is 100-year-old or dead. When a patient is excluded, all of the statistics about him/her are recorded and another patient is generated. When simulation of a scenario is complete ($N_p >$Population size), the simulation continues with the next one until all scenarios are evaluated.

The simulation model is coded in Java as a microsimulation model. We run the simulation for $100,000$ patients for each gender, risk-level, and age-group (when applicable) for 50 years. The consistency in calculations are ensured by using the same seed in each run (16807).

3.3.3 Input Parameters and Validation

Table 3.4 presents the input parameters used for the numerical experiments. The parameter values are mainly based on the estimations of Erenay (2010); Erenay et al. (2011) and

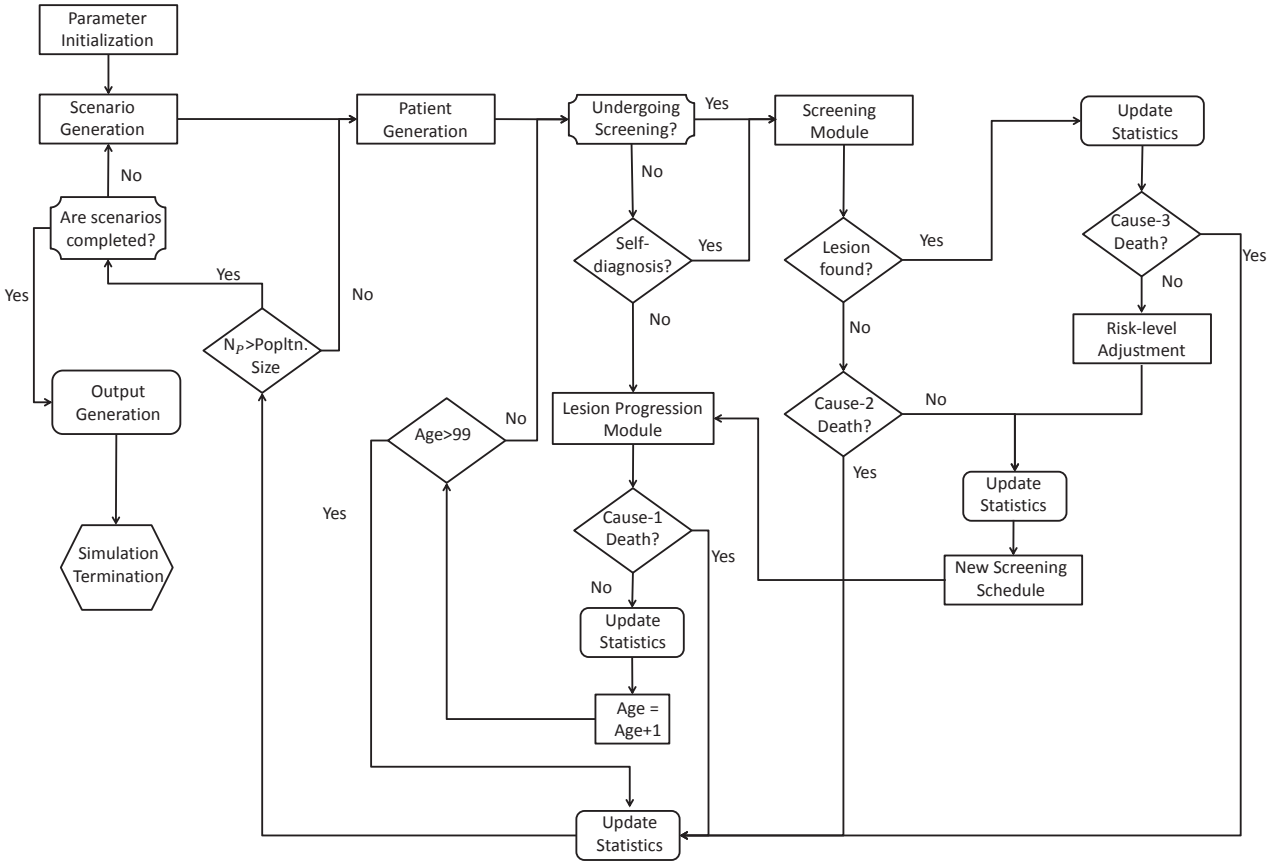


Figure 3.3: CRC simulation flow chart

N_p refers to the number of patients generated.
Cause-I death refers to all-cause deaths during a given year including death from undiagnosed CRC and cancer treatment.
Cause-II and Cause-III deaths refer to deaths due to colonoscopy complications without and with polypectomy.

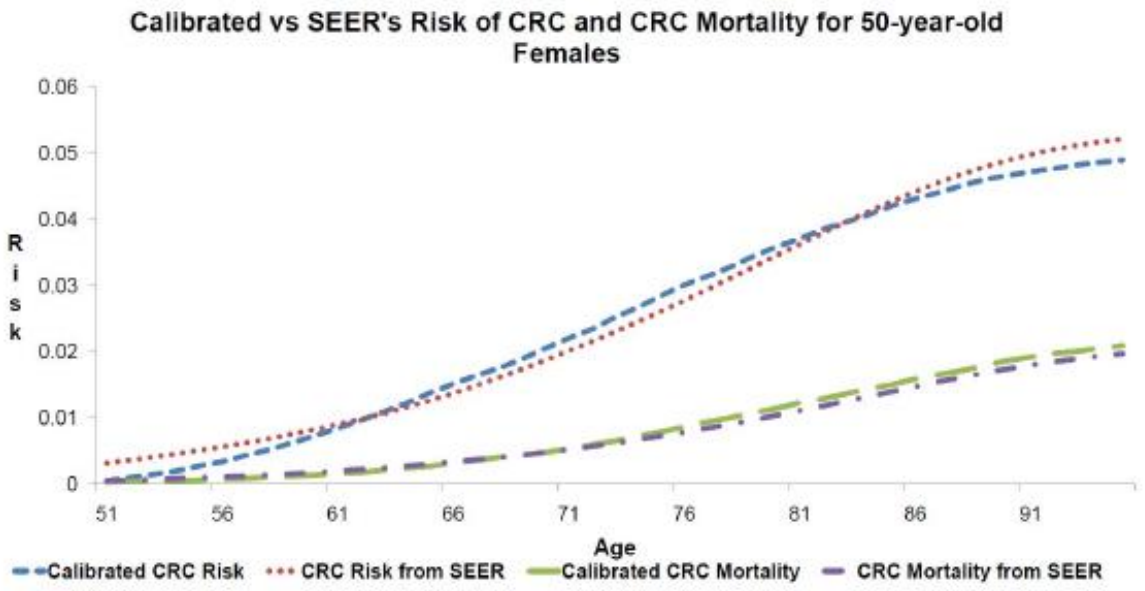
several clinical studies mimicking the progression of CRC in the recent literature. Although there are updates on some data in the recent literature (e.g., compliance rates), we use the data estimations of Erenay (2010) to be able to provide a complete comparison. The parameter estimation in Erenay (2010) is performed using a metachronous CRC natural history (MCRC-NH) simulation model with clinical data from MAYO Clinic, Rochester. This estimation along with the data from SEER database provides the transition probabil-

ities for the patients whose CRC lesions have been removed through colonoscopy screening (i.e., post-CRC patients). On the other hand, we obtain the polyp-onset probabilities for low-risk and high-risk patients from the National Polyp Study (Loeve et al., 2004). Mortality probabilities for CRC-free patients and patients with clinical CRC are based on Erenay (2010)'s estimates which are from US life tables (Arias, 2007) and SEER database (Howlader et al., 2012). The input parameters such as characteristics of colonoscopy screening and discount factors are derived directly from clinical literature whereas some of the other input parameter such as disutility of colonoscopy and expected TQALYs after age-100 are directly from Erenay (2010)'s meta-analyses. There is limited data for reliable estimation of annual probability of polyp-to-CRC progression and mortality from undetected CRC because detected polyps and CRC cases are immediately removed or treated. Therefore, we again refer to Erenay (2010) for estimates of these probabilities which they obtain through calibrating their model to replicate the cumulative age-based CRC risk and CRC mortality from SEER data for general population (see Figures 3.4a and 3.4b). We derive most of the cost components from Lansdorp-Vogelaar et al. (2009a) because theirs is the only study that estimated CRC treatment and continuing care costs separately. The details of the meta-analyses and calibration process are available in Erenay (2010) and the online supplement of Erenay et al. (2014).

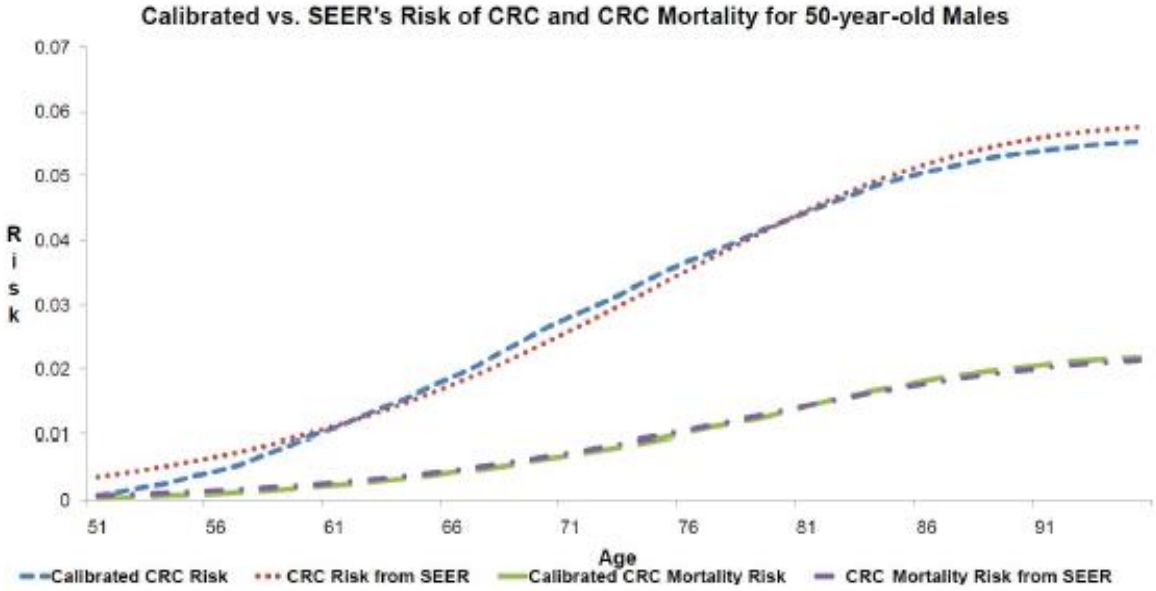
We assume a simulation environment where at age-50 1% of the population have personal or family history of CRC, i.e., high-risk patients, and 0.94% of the population have only asymptomatic risk towards developing CRC, i.e., low-risk patients. These patients are assumed to undergo colonoscopy screening with 100% compliance starting at age-50 until age-75 and -85, respectively, for low- and high-risk levels, and the simulation runs to calculate performances either they die or reach age-100 (i.e., decision horizon of 50 years).

3.3.4 Simulation Outputs: The Performance Measures

We use the proposed simulation model to estimate several performance measures reflecting the harms and benefits of colonoscopy screening. The harms include increased complication risk, disutility associated with undergoing colonoscopy, and clinical resource con-



(a) Female



(b) Male

Figure 3.4: Comparison of age-based CRC risk and CRC mortality from SEER database and calibrated simulation model of Erenay et al. (2014) (Source: Erenay (2010))

Parameter			
Annual colorectal lesion progression parameters (A.S.w.M[1])	**Low-risk**	**High-risk**	**Post-CRC**
Polyp-onset probability	0.021	0.052	0.12
Polyp-to-CRC progression probability		0.01	
Polyp-to-MCRC progression probability		0.348	
Probability of local CRC recurrence after treatment		0.038	
Probability of metastatic recurrence of CRC after treatment		0.061	
Colonoscopy Screening Characteristics (S[2], R[3])	**Low**	**Base**	**High**
Sensitivity for polyps^{S,R}	0.80	0.85	0.90
Sensitivity for cancer^{S,R}	0.85	0.90	0.95
Specificity	-	1	-
Probability of self-detection of CRC	-	0.49	-
Probability of major complication with polypectomy/biopsy	1	2	3
Probability of major complication without polypectomy/biopsy	1	2	3
Mortality related Probabilities[4] (A.S.w.M)			
Mortality with polypectomy/biopsy		1.635×10^{-4}	
Mortality without polypectomy/biopsy		0.892×10^{-4}	
Mortality after complete CRC treatment		0.256	
Mortality after incomplete CRC treatment		0.602	
Mortality from undetected CRC		0.356	
Mortality from other causes		0.070	
Disutility (S, R)	**Low**	**Base**	**High**
Colonoscopy Screening^{S,R} (in weeks)	1	2	3
Undetected CRC (in months)	1	1.5	2, 3
CRC treatment (in months)	4	4.5	5
Costs, $ (R)	**Low**	**Base**	**High**
Colonoscopy screening	370	740	1,110
Polypectomy	200	400	600
Removing a non-CRC lesion	200	400	600
Complications	2,900	5,800	8,700
CRC/MCRC Treatment	24,550	49,100	73,650
Life after CRC treatment	31,640	63,280	94,920
End of life	9,925	19,850	29,775

[1]: *Age Specific with Mean.* The given values are used to obtain the age specific values.
[2]: The parameters with low- and high-values considered in our sensitivity analysis.
[3]: The parameters with low- and high-values considered in our robustness analysis.
[4]: These probabilities differ based on the patients' health transitions. The given values represent the base values that are used to calculate the case-specific probabilities.

Table 3.4: Simulation input parameters

sumption and they are captured by number of required colonoscopies (NC) and average screening/follow-up interval (L). On the other hand, the increase in life-years (LY) and TQALYs (QY), and reduction in cumulative CRC risk (C_T) and all-cause mortality (M_T) reflect the benefits of the considered policies. Total cost partially reflects both the harms and benefits, e.g., total cost includes the increasing cost of colonoscopy tests and decreasing cost of CRC treatment when frequency of colonoscopy screening increases. The disutility of colonoscopy and complication risks are also considered in calculation of TQALYs.

The following equations represent the derivation of the aforementioned performance measures from the simulation results while Table 3.5 lists the notation we use for the formulation below.

$$NC = \sum_{n=1}^{N} \frac{NC_n}{N}, \quad \text{where} \tag{3.1}$$

$$NC_n = \sum_{t=1}^{T} I_t^n, \text{ and } I_t^n = \begin{cases} 1, & \text{if n}^{th} \text{ patient had a colonoscopy in year } t \\ 0, & \text{otherwise} \end{cases}$$

$$L = \frac{\sum_{n=1}^{N} L_n}{\sum_{n=1}^{N} I_E(n)} \quad \text{where } I_E(n) = \begin{cases} 1, & \text{if } NC_n > 1, n \in \{1, 2, \dots N\} \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

$$L_n = \begin{cases} \frac{\sum_{i=1}^{NC_n - 1} CoT_{i+1}^n - CoT_i^n}{NC_n - 1}, & \text{if } NC_n > 1, n \in \{1, 2, \dots N\}, \text{ given that} \\ & CoT_1^n = 0,\ CoT_i^n = \min\left(t : I_t^n = 1,\ t > CoT_{i-1}^n\right) \text{ for } i > 2 \\ 0, & \text{othewise} \end{cases} \tag{3.3}$$

(3.4)

Notation	Description
$t \in \{0, 1, \ldots, T\}$	The set of years after age of 50 years
$n \in \{0, 1, \ldots, N\}$	The set of patient identifiers
$s \in \mathbf{S}$	Current state where $\mathbf{S}$ consists of the health states
$a \in \mathbf{A} \equiv \{Co, DN\}$	Current action (undergo colonoscopy or do nothing)
$o \in \mathbf{O} \equiv \{T-, P+, C+, SD\}$	Current test results (Test negative, a polyp is found, CRC is found, and CRC is self-diagnosed, respectively)
LRi, HRi, PCi, $i \in \{0, 1, 2\}$	Health states for low-risk, high-risk, and post-CRC patients where $i = 0$, 1, and 2 refers to without lesion, polyp and CRC states, respectively.
NC_n	The number of colonocopies the n^{th} patient underwent during the simulation
L_n	The average colonoscopy interval for the n^{th} patient
I_t^n	Indicator function showing whether the n^{th} patient underwent a colonoscopy at the beginning of year t
$I_E(n)$	A function indicating whether the n^{th} patient undergoes more than one colonoscopy
CoT_i^n	The year of the i^{th} colonoscopy for the n^{th} patient
CR_t and MR_t	Cancer and mortality risk between age 50 and $50 + t$
CR_t^n and MR_t^n	Indicator functions specifying whether the n^{th} patient develops CRC or dies at age $50 + t$, respectively
$l(s, s')$	The expected life years for a patient who moves from state s to state s' at year t
$q_t(s, a, o, s')$	The expected QALYs at year t for a patient who moves from state s to state s' given action a and test result o
$c_t^n(s, a, o, s')$	The expected cost if the n^{th} patient moves from state s to state s' at year t given action a and test result o
$c_{Co}, c_{P/B}$	Cost of colonoscopy and polypectomy/biopsy, respectively
c_{cmp},c_{ccc}	Costs of treatment for colonoscopy complications and continuing cancer care, respectively
c_{MCT}, c_{ct}	Cost of metastatic CRC and CRC treatment, respectively
c_{tc}, c_{tcc}	Cost of terminal care for a patient without and with CRC, respectively
λ_h, λ_c	Discount factors for health outcomes and costs, respectively.

Table 3.5: Model notation

$$C_T = \sum_{n=1}^{N} \frac{\max_{0<t'<T}(C_{t'}^n)}{N}, \text{ and } C_t^n = \begin{cases} 1, & \text{if the } n^{th} \text{ patient has CRC in year t} \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

$$M_T = \sum_{n=1}^{N} \frac{\max_{0<t'<T}(M_{t'}^n)}{N}, \text{ and } M_t^n = \begin{cases} 1, & \text{if the } n^{th} \text{ patient dies in year t} \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

Let $I_t^n(s, s')$ and $I_t^n(s, a, o, s')$ be the indicator functions that, respectively, show whether the n^{th} patient was in state s in year t, moved to state s' in the next year and did not die at the beginning of year t due to colonoscopy complication or CRC treatment and the n^{th} patient was in state s in year t, moved to state s' in the next year after action a is taken and test result o is observed. Then;

$$LY = \sum_{n=1}^{N} \left(\sum_{t=0}^{T} \sum_{s' \in \mathbf{S}} \sum_{s \in \mathbf{S}} \frac{(\lambda_h)^t l(s, s') \times I_t^n(s, s')}{N} \right), \; l_t(s, s') = \begin{cases} 1, & \text{if} s, s' \in \mathbf{S} \backslash D \\ 0.5, & \text{if} s \in \mathbf{S} \backslash D \; and \; s' = D \\ 0, & otherwise \end{cases} \tag{3.7}$$

$$QY = \sum_{n=1}^{N} \left(\sum_{t=0}^{T} \sum_{s' \in \mathbf{S}} \sum_{s \in \mathbf{S}} \sum_{a \in \mathbf{A}} \sum_{o \in \mathbf{O}} \frac{(\lambda_h)^t q_t(s, a, o, s') I_t^n(s, a, o, s')}{N} \right) \tag{3.8}$$

$$TC = \sum_{n=1}^{N} \left(\sum_{t=0}^{T} \sum_{s' \in \mathbf{S}} \sum_{s \in \mathbf{S}} \sum_{a \in \mathbf{A}} \sum_{o \in \mathbf{O}} \frac{(\lambda_c)^t c_t^n(s, a, o, s') I_t^n(s, a, o, s')}{N} \right) \tag{3.9}$$

$$c_t^n(s,a,o,s') = \begin{cases} I_t^n c_{Co} + B_t^n c_{P/B} + Icm_t^n c_{cmp}, & a \in \mathbf{A},\ o = T-,\ s \in \mathbf{S}\backslash\{PC0, PC1, PC2, UCT, D\}, \\ & s' \in \mathbf{S}\backslash\{UCT, D\}, \\ I_t^n c_{Co} + B_t^n c_{P/B} + Icm_t^n c_{cmp} + c_{ccc} & a \in \mathbf{A},\ o = T-,\ s, s' \in \{PC0, PC1, PC2\}, \\ c_{Co} + c_{P/B} + Icm_t^n c_{cmp} & a = Co,\ o = P+,\ s \in \{LR1, HR1\}, s' \in \{HR0, HR1\}, \\ c_{Co} + c_{P/B} + Icm_t^n c_{cmp} + c_{ccc} & a = Co,\ o = P+,\ s \in \{PC1\}, s' \in \{PC0, PC1\}, \\ c_{MCT} & o = T- , s = UCT,\ s' \in \mathbf{S}\backslash D,\ a = DN, \\ c_{Co} + c_{P/B} + Icm_t^n c_{cmp} + c_{ct} & o \in \{C+, SD\},\ s \in \{LR2, HR2, PC2\},\ s' \in \mathbf{S}\backslash D,\ a \in \mathbf{A}, \\ I_t^n c_{Co} + B_t^n c_{P/B} + Icm_t^n c_{cmp} + c_{tc} & a \in \mathbf{A}, o = T-,\ s \in \mathbf{S}\backslash\{LR2, HR2, PC2, UCT, D\},\ \ s' = D, \\ c_{Co} + c_{P/B} + Icm_t^n c_{cmp} + c_{tc} & o = P+,\ s \in \{LR1, HR1, PC1\},\ s' = D,\ a = Co, \\ I_t^n c_{Co} + B_t^n c_{P/B} + Icm_t^n c_{cmp} + c_{tcc} & o = T-,\ s \in \{LR2, HR2, PC2\},\ s' = D,\ a \in \mathbf{A}, \\ c_{MCT} + c_{tcc} & o = T-,\ s = UCT,\ s' = D,\ a = DN, \\ c_{Co} + c_{P/B} + Icm_t^n c_{cmp} + c_{ct} + c_{tcc} & o \in \{C+, SD\},\ s \in \{LR2, HR2, PC2\},\ s' = D,\ a \in \mathbf{A}, \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

In the formulation of QY, $q_t(s,a,o,s')$ refers to the difference between expected life years ($l(s,s')$) and expected disutility which is a combination of disutilities of colonoscopy, complications, CRC treatment, and undiagnosed CRC. The formulation of $q_t(s,a,o,s')$ is available in Erenay et al. (2014). The detailed definition of $c_t^n(s,a,o,s')$ is as follows where B_t^n and Icm_t^n are binary variables representing whether the n^{th} patient has a polypectomy/biopsy and serious complication at age $50+t$, respectively. Note that, cost of complication treatment refers to the costs associated with hospitalization due to serious colonoscopy complications, while, cost of continuing care denotes the costs associated with medication after the initial CRC treatment for post-CRC patients. Unless it is metastasized, CRC treatments are completed within a year, thus, cost of being in the UCT state (c_{MCT}) is taken as the cost of metastatic CRC treatment.

3.4 Numerical Experiments

This section presents extensive numerical experiments for evaluating the performances of the periodic colonoscopy screening policies recommended by practitioners and all feasible

simpler-to-implement colonoscopy screening policies. The insights derived from our numerical results are then used to determine the characteristics of the promising alternative screening policies that can provide close-to-optimal performances compared to the optimal personalized screening policies found by the POMDP model proposed in Erenay et al. (2014).

Simpler-to-implement policies include *static periodic policies* and *dynamic periodic policies* with n period switch times ($n = 1, 2$) at the beginning of given ages. A brief representation of simpler-to-implement policies for *low-risk* patients is given in Figure 3.5. The static-periodic policy follows a single screening frequency of every 10-year and starting from age-50, the colonoscopy screenings are performed in every 10-year until the stopping age of age-80 for low-risk patients. The age-dependent periodic policy with $n = 1$, considers different screening frequencies over each of two age intervals, every 8-year screening in the age interval 50-60, and every 5-year screening age-60 onward. When there are three age intervals, equivalently 2 switch times, the screening policy with $n = 2$ suggests screening patients in every 8-, 5-, and 7-year, respectively, in the age intervals 50-60, 60-70, 70-80. This scheme is different for hig-risk patients since the test result affects the screening decisions for them, e.g., every 3-year screening after polypectomy and 5-year screening after no positive finding. Details are discussed in the remainder of Section 3.4.

3.4.1 The Rationale Behind the Proposed Policy Structures

In our numerical analysis, we focus on two different policy structures to define the simpler-to-implement policies; static-periodic, and age-dependent dynamic-periodic colonoscopy screening policies, which are referred to as dynamic-periodic policies hereafter. A *static-periodic screening policy* recommends patients undergo colonoscopy screening periodically with a single frequency depending on the risk-level and gender. On the other hand, an *age-dependent dynamic-periodic screening policy* recommends undergoing periodic screening with distinct frequencies in particular predefined age intervals.

We consider static-periodic policies because i) they are the simplest form of screening policies; and ii) current CRC screening guidelines follow this structure (USPSTF, 2008),

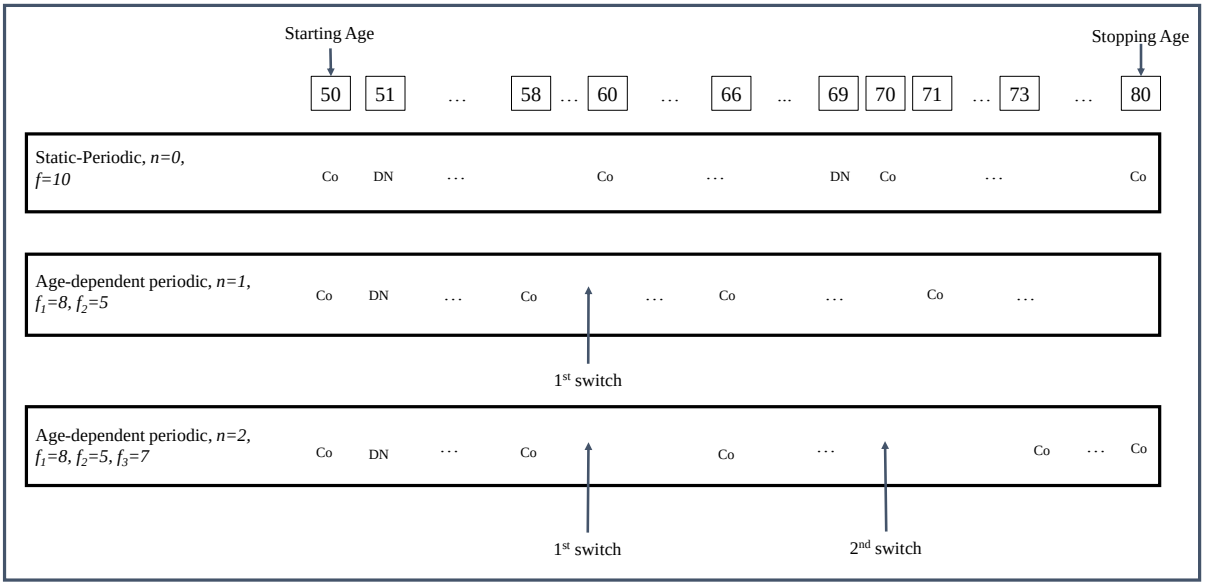


Figure 3.5: A representation of the simpler-to-implement policies for low-risk patients

e.g., every 10-year colonoscopy from age 50 to 75 for low-risk patients, and screening in the third year after a polypectomy and every 5-year screening thereafter until a positive colonoscopy for high-risk patients. We consider dynamic-periodic policies to further improve the performances of the proposed simpler-to-implement colonoscopy screening policies. Note that age-dependent dynamic-periodic policies are practiced for mammographic breast cancer screening based on the guidelines of American Cancer Society (Oeffinger et al., 2015).

Several studies reports that slightly more than half of the clinicians follow the CRC screening guidelines (Yabroff et al., 2011a; Klabunde et al., 2009, 2005; Cabana et al., 1999). Those who do not follow them recommend more frequent colonoscopy screening than the current guidelines again in the form of periodic screening for low-risk and high-risk patients. Mysliwiec et al. (2004), Yabroff et al. (2011a), and Wilschut et al. (2011c) report the use of the more frequent static-periodic policies listed in Table 3.6. As a result, existing clinical studies primarily focus on analyzing the cost-effectiveness of the static-periodic policies [reference]. These observations from CRC literature imply that clinicians

would be more likely to adopt static-periodic policies compared to dynamic policies, such as dynamic-periodic policies or those derived by POMDP models.

In the cases, for which static-periodic policies may not satisfactorily close the performance gap between the current colonoscopy screening guidelines and the optimal policies from the POMDP model, we aim to identify ideal simpler-to-implement policies among the dynamic-periodic policies. By considering dynamic-periodic policies, we seek to establish an interaction between the patients' ages and recommended screening intervals. This is important as i) many parameters depend on the age including the probability of colorectal lesion, and probability of mortality, ii) capturing this interaction is likely to improve CRC screening performance since the screening frequencies of the optimal policies from the POMDP model vary significantly in age. To tackle integration of age factor into the simple policies, we propose dividing the age range into n sub-intervals of varying lengths (n-1 frequency switches), and assigning a different screening frequency to each of these intervals (see Table 3.9).

Although there is a general consensus on initiating CRC screening at age 50 (USPSTF, 2008), determining the optimal screening stopping age, beyond which periodic screening is terminated except for colonoscopies prompted by self-detection of CRC, is a controversial issue. It is reasonable to terminate CRC screening at an age, beyond which the savings in TQALYs from early-detecting or removing a colorectal lesion may not compensate the disutility due to an invasive colonoscopy operation. Guidelines recently propose a stopping age for colonoscopy screening based on the results of recent studies which analyze the effects of stopping ages in the ranges of age 75 to 80 (Long and Sands, 2018; Knudsen et al., 2016; Wilson, 2010; Zauber et al., 2009, 2008; Maheshwari et al., 2008; Stevens and Burke, 2003). Therefore, in our numerical experiments on both static and age-dependent periodic screening policies, we investigate the effect of stopping age within this age interval, i.e., we consider stopping age options of 75, 80, and 85 for low-risk patients and 80, 85, and 90 for the high-risk patients.

For our numerical experiments, we first run the simulation for the current CRC guidelines and the other static-periodic policies recommended by the clinicians with different

parameter settings (See Table 3.6). We compare these policies with the optimal policy from the POMDP model proposed in Erenay et al. (2014) in order to see whether the currently practiced static-periodic policies may perform close to optimal under any parameter setting. We conduct this analysis because no study reports a performance comparison specifically between the guidelines and the other practiced static-periodic policies; therefore, how closely these practiced static-periodic policies may perform to the best possible practice is important to reveal. Next, we evaluate all possible static policies for both for low-risk and high-risk patients, determine the static policies performing closest to the optimal POMDP policy for each gender. Finally, in order to find simpler-to-implement policies with further reduced optimality gap, we run our simulation model for all possible dynamic-periodic policies with $n-1$ frequency switches given the reasonable n-interval partitions of the age range, $n = 2, 3$,for colonoscopy screening. Given a set of parameters, we derive promising simpler-to-implement policies (i.e., performing close to the optimal policies) for each gender, as earlier studies indicate gender as a factor to consider when analyzing CRC progression and determining the best colonoscopy screening practices (Erenay et al., 2014, 2011). We then perform sensitivity analysis on several parameters to see how those parameters may affect the policy performances and the selection of the promising simple-to-implement policies.

Recall that the main objective of this study is to improve the CRC screening guidelines proposing simpler-to-implement policies that are convenient to be adopted by clinicians in practice. Therefore, it is also important to propose limited number of alternative policies to the clinicians while conducting a comprehensive policy analysis. To present a concise evaluation, we perform a Pareto-efficiency analysis on the cost and TQALYs spectrum. That is, we enumerate all reasonable policies but only present the ones that cannot be dominated by another policy in terms of TQALY and cost measures. Furthermore, among the Pareto-frontier simpler-to-implement policies, we only highlight the promising ones that i) perform close to the optimal policies, 2) do not increase required number of colonoscopies significantly, and have favorable incremental cost-effectiveness ratios (ICER) given in Equation 3.11, i.e., ICER$< \$50,000$ or $\$100,000$. We then perform robustness analyses on the selected promising simple-to-implement policies for low- and high-risk patients to

see how their performances change under different input parameter scenarios.

$$ICER = \frac{C_1 - C_0}{\text{TQALYs}_1 - \text{TQALYs}_0}, \tag{3.11}$$

where 0 is the reference policy, 1 is the candidate policy, and C_i and TQALYs_0 denote the cost and expected TQALYs associated with a policy.

3.4.2 Commonly Practiced Policies

Klabunde et al. (2009) provide information about a survey analyzing the screening recommendations of clinicians in 2006 and 2007 that are commonly used for low-risk (i.e., asymptomatic average risk) patients in the practice. The screening intervals are given for several CRC screening modalities, and every 3-, 4-, 5-year and a randomized screening interval between 6 to 9 years (6:9-year) are highlighted for colonoscopy screening. Figure 3.6 presents the performances of these screening frequencies which are referred as *commonly practiced policies*, henceforth.

The policy suggested by AGA guidelines, every 10-year, performs the worst followed by the policies suggesting every 6:9-year and 3-year screening, for both genders. The best performing commonly practiced policy is every 5-year screening for females and every 4-year screening for males as seen in Figures 3.6a and 3.6b, respectively. We have observed that the best performing commonly practiced policy for males is associated with higher screening frequency than that for females. This is because females develop CRC lesions with a slightly lower rates than males do.

Moreover, the gap between the best commonly practiced policy and the optimal POMDP solution is greater for females than males in terms of TQALYs; however, the policy for females perform better than the one for males in terms of total costs, i.e., 0.04% vs. 0.03% gap between the corresponding POMDP in TQALYs and 2.08% vs. 1.26% less costs than the corresponding POMDP, respectively, for females and males. More frequent screenings

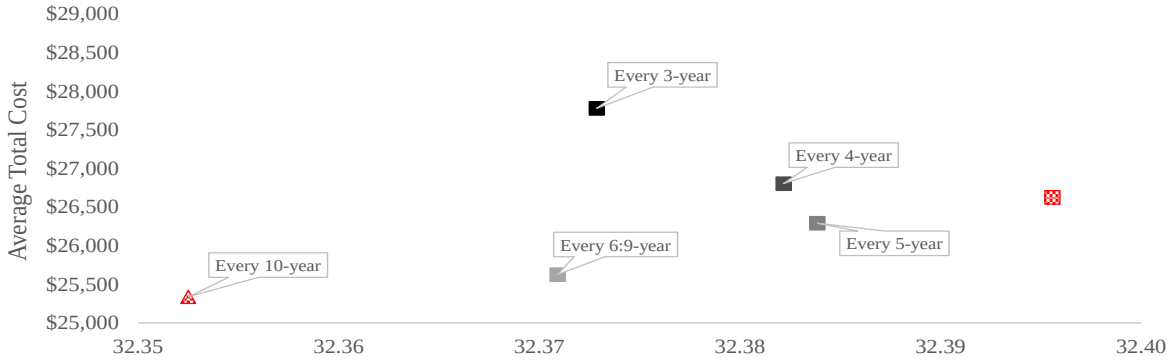
considerably work better for males than they work for females as every 3-year also performs closer to the optimal POMDP for males than that does for females. This observation reveals the subtle effect of total expected life years remaining between females and males. Since females have longer life expectancy, every 3-year screening results in greater number of colonoscopies scheduled in total (8.42 vs. 7.91), thus, resulting higher disutility, namely, lower TQALYs, and greater costs.

Our insights from the evaluation of commonly practiced policies strongly indicate that there is opportunity to improve the performance of the current guideline screening policy suggesting every 10-year screening even following the clinicians uneducated proposals on the screening frequencies. Therefore, we analyze a greater set of periodic frequencies under the assumptions that screening decisions are made with and without the consideration of patient's age.

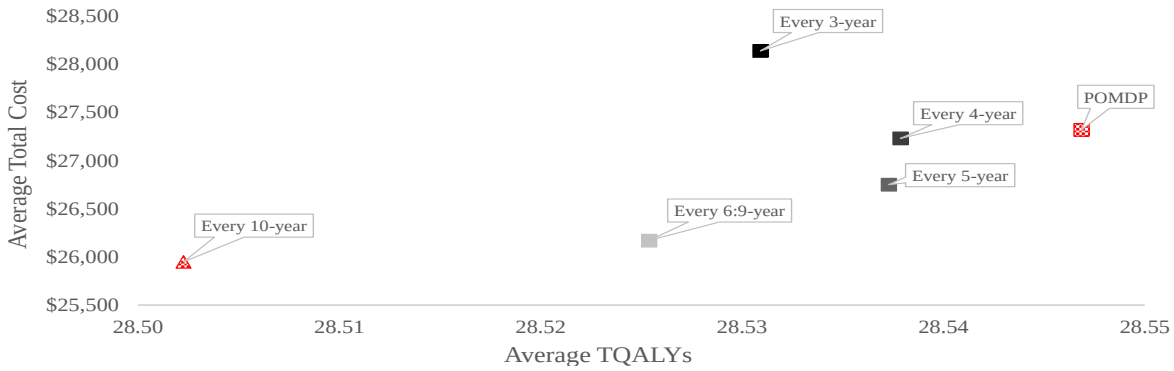
3.4.3 Static-periodic Policies

In this section, we evaluate the performances of the *static-periodic* colonoscopy screening policies, interchangeably referred as *static policies* hereafter, to find the best performing simpler-to-implement static-periodic policy among the Pareto-efficient policies while taking gender and risk-level of the patients into account. The effect of stopping age is also incorporated in our evaluation by comparing the policy performances for the stopping ages of 75, 80, 85 for low-risk patients and 80, 85, 90 for high-risk patients. The screening frequencies that we consider in our analysis for static policies are given in Table 3.6. The screening frequency is every-f year for low-risk patients whereas it is every-f^1 years after the polypectomy, and every-f^2 years afterwards when there is nothing significant found in the screening for high-risk patients. Note that we omit including the randomized policy of screen every 6:9-year policy into the proposed policies since our analysis is to inform the guidelines with a one-type screening policy which is followed by all clinicians based on the patients' characteristics.

Besides analyzing the static policies with base-case parameters, we also show how the changes in some of the system parameters affect their performances through presenting a



(a) Female



(b) Male

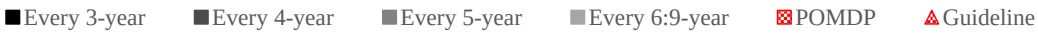


Figure 3.6: *Commonly practiced policies* reported in Klabunde et al. (2009) with *base-case* parameters for low-risk patients

Table 3.6: The screening frequencies for static policies

Low-risk patients	$f \in \{3, 4, 5, 6, 7, 8, 9, 10\}$
High-risk patients	$f^1, f^2 \in \{1, 2, 3, 4, 5\} \otimes \{1, 2, 3, 4, 5\}$, s.t. $f^1 \leq f^2$

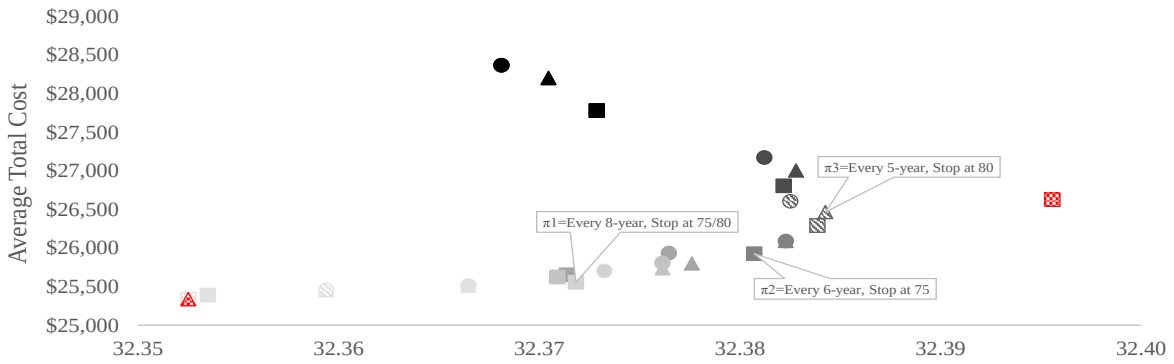
comparison with the guideline and the optimal POMDP policies, in Sections 3.4.3.1 and 3.4.3.2, respectively, for low- and high-risk patients.

3.4.3.1 The Best Performing Static Policies for Low-Risk Patients

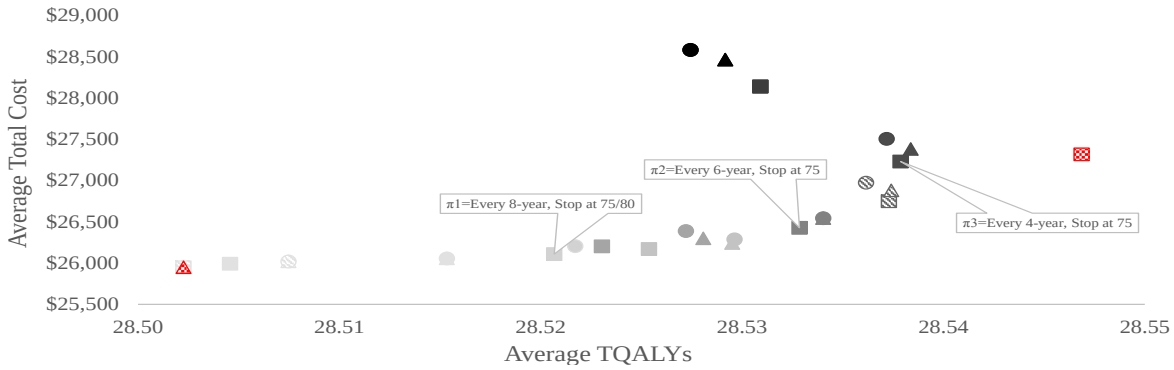
Figure 3.7 shows the performances (i.e., TQALYs and costs) of the static policies with *base-case* parameters for low-risk patients given stopping ages of 75, 80, 85. The number of different screening policies analyzed is 27, i.e., 9 different screening policies for each stopping age. Note that the figures in this section have a generic format to provide tractable representation, that is, the screening frequency increases when the color of the markers gets darker, and the stopping age changes with the shape of the marker, as shown in the legend of Figure 3.7.

The performances of the static policies are in a similar fashion for both female and male patients as shown in Figures 3.7a and 3.7b. TQALYs increase with the screening frequency for the policies that suggest performing colonoscopy screening at most in every 5-year for females and in every 4-year for males and decrease for more frequent screenings than those. We also observe that TQALYs further decrease for these undesirably frequent screenings when the stopping age increases. These are mainly because of the fact that colonoscopy procedure is associated with some disutility; thus, any additional colonoscopy screening performed results in more harms than the prevention benefits after a particular screening frequency.

The policy that improves the current guideline's TQALYs significantly with a reasonably small costs while realizing an ICER less than $\$50,000$ suggests every 8-year screening for both genders with a stopping age of either 75 or 80. The best-performing static policy under the base-case scenario, π, suggests every 5-year screening with a stopping age



(a) Female



(b) Male

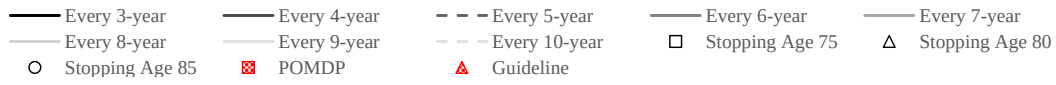


Figure 3.7: *Static screening policies* with *base-case* parameters for low-risk patients given stopping ages of 75, 80, 85

of 80 for females and every 4-year screening with a stopping age of 75 for males. These two policies perform closely to the optimal POMDP while requiring slightly less cost. We can further improve TQALYs for males by following the same static policy as for females (i.e., every 5-year screening); however, this requires slightly more cost than the optimal POMDP. Note that the average cost in this study only includes the costs of screening and treatment. If the cost of implementing a policy in practice is also included in the analysis, it is likely that following the same policy for both genders would be more favorable by the clinicians. However, this kind of cost is difficult to calculate due to lack of data.

Table 3.7 presents a comparison among the guideline policy of every 10-year, the policies suggesting every 8- (π_1), 6-(π_2), and 5- year, the best performing static-periodic policy (π_3), the static-periodic policies suggesting one and two additional colonoscopy screenings than the guideline, and the optimal POMDP policy of Erenay et al. (2014). The stopping age is assumed to be 75 for the guideline policy as suggested in USPSTF (2008). The static policies with one additional colonoscopy than the guidelines suggest every 8-year screening with the stopping age of 85. The ones suggesting two additional colonoscopy than the guidelines adopt every 5-year screening with a stopping age of 75. The stopping age for the guideline policies and every 8-, 6-, and 5-year screening policies is assumed to be 75 following the recommendations in USPSTF (2008).

The performances of the static policies in Table 3.7 are promising as we can achieve almost half of the improvement in TQALYs that the optimal POMDP attains with a reasonably small cost by following a considerably less complex policy structure. Although the static policies start performing closely in terms of TQALYs when the screening frequency increases, the improvement in terms of CRC risk and mortality continues improving by increasing frequency. This is noteworthy as these measures are also important while quantifying the effectiveness of a screening policy. Increasing the screening frequency alone is not a sufficient strategy as there is still a gap between the best performing static-periodic policies, π, and the corresponding POMDP policies for both genders. Therefore, we conduct further analysis considering the relationship between the screening age and frequency to fill this gap without compromising the simple structures of periodic policies. We aim to inform the current guidelines for a potential update using the insights from our analysis.

Note that the analysis in this section strongly indicates the need for updating the current guidelines as even the practice policies suggesting every 5-year screening only require two additional colonoscopy screenings with the same stopping age of 75 and perform very close to the optimal POMDP.

Table 3.7: Comparison of the US guidelines, the optimal POMDP policies of Erenay et al. (2014), and the promising static-periodic policies (i.e., π_1, π_2, π_3) for low-risk patients

		Guideline	π_1	π_2	5-year	π_3	+1 Col	+2 Col	POMDP*
Females									
	Total QALYs	32.353	32.372	32.381	32.384	32.385	32.373		32.396
	No. of colonoscopies	4.00	4.78	5.56	6.23	6.58	5.12		7.05
	Lifetime CRC risk	3.16%	2.68%	2.44%	2.26%	2.08%	2.45%	5-year	1.95%
	CRC mortality	1.24%	1.02%	0.93%	0.86%	0.76%	0.87%		0.65%
	Total Cost ($)	25,338	25,558	25,926	26,292	26,467	25,703		26,627
	Screening Interval	8.37	6.87	5.46	4.75	4.74	6.83		4.52
Males									
	Total QALYs	28.502	28.521	28.533	28.537	28.538	28.522		28.547
	No. of colonoscopies	3.72	4.42	5.17	5.80	6.57	4.68		6.96
	Lifetime CRC risk	3.41%	2.91%	2.61%	2.40%	2.27%	2.71%	5-year	2.01%
	CRC mortality	1.24%	1.01%	0.92%	0.84%	0.81%	0.86%		0.67%
	Total Cost ($)	25,949	26,109	26,428	26,751	27,231	26,204		27,319
	Screening Interval	8.45	6.94	5.49	4.76	4.02	6.90		4.12

* Source: Erenay et al. (2014)

Note that the guideline policies assume the stopping age of 75 whereas the stopping age may vary for the other policies.

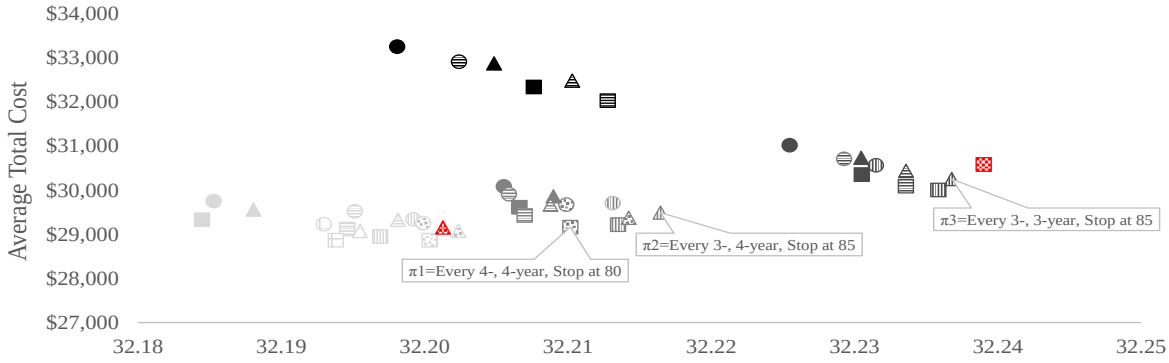
3.4.3.2 The Best Performing Static Policies for High-Risk Patients

The performance evaluation of the static policies with *base-case* parameters for high-risk patients is presented in Figure 3.8 for the stopping ages of 80, 85, and 90. We consider 15 different screening policy for each stopping age, thus, analyze 45 screening policy in total. The screening frequency sets for these policies is given in Table 3.6. We again design easily tractable representation, that is, the screening frequency after getting negative result, f^2, increases when the color of the markers gets darker, and the fill type of marker changes with

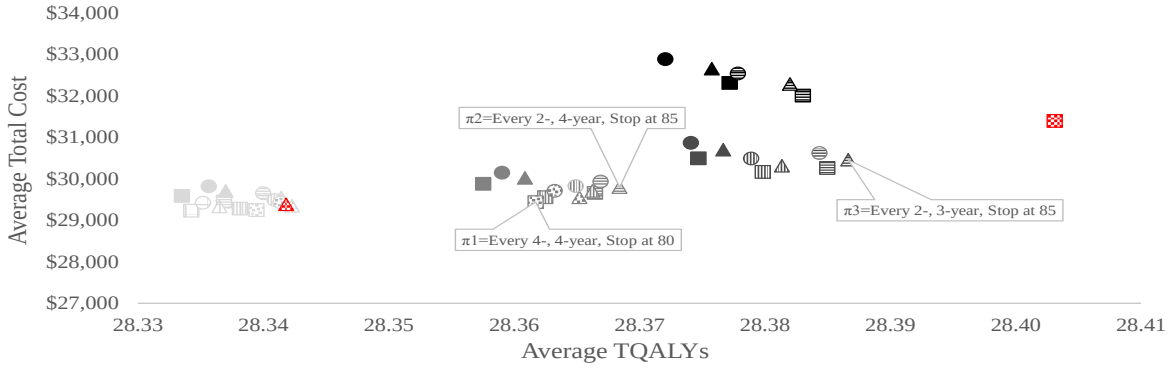
the screening frequency after polypectomy, f^1, for each f^2. Similar to all other graphs in this thesis, the stopping age changes with the shape of the marker, as shown in the legend of Figure 3.8.

Although the behaviors of the static policies are not noticeably different for low-risk female and male patients (see Figures 3.7a and 3.7b), the policies plot in a different fashion for both genders of high-risk patients and differences in the performances become significant (see Figures 3.8a and 3.8b). Firstly, the TQALYs vary significantly based on f^1, the screening frequency after polypectomy, for female patients whereas they perform very closely independent from their f^1 for male patients for a given f^2, the screening frequency after a negative result. This observation especially becomes noticeable for lower screening frequencies (e.g., the points are closer for all f^1 values when $f^2 = 5$ than $f^2 = 4$). On the other hand, the effect of f^1 is more prominent for male patients, e.g., the best performing policy suggests every 2-, 3-year screening for males whereas it is every 3-, 3-year screening for females. This is mainly because male patients develop more CRC lesions than females, thus, undergoing colonoscopy screening frequently after polypectomy is very important for early detection and prevention of CRC. Moreover, the proportion of patients having CRC among male patients is greater than that among female patients in simulation. Therefore, the changes in f^1 is more pronounced when calculating the average TQALYs as a simulation output. Secondly, increasing the screening frequency after a negative test result, f^1, too much decreases TQALYs more for females than that does for males ($f^1 < 3$). This is due to longer life expectancy of females, in other words, having more colonoscopy screenings scheduled for female patients over a longer horizon. Lastly, the greatest attainable improvement in TQALYs is more for female patients; the best-performing policy results in a similar TQALYs as the optimal POMDP with less costs. This observation is consistent with the findings of Erenay et al. (2014) that the improvements in TQALYs for female patients is more pronounced due to longer life expectancy.

Table 3.8 lists the performances of the guideline policy of every 3-, 5-year screening, the guideline policy with the stopping ages of 90 and 80, respectively, for females and males, the promising static-periodic policies (π_1, π_2, π_3), the static-periodic policies suggesting one and two additional colonoscopy screenings than the guideline, and the optimal POMDP



(a) Female



(b) Male

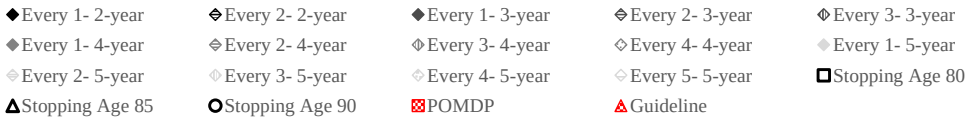


Figure 3.8: *Static screening policies* with *base-case* parameters for high-risk patients given stopping ages of 80, 85, 90

policy of Erenay et al. (2014). Note that we assume a stopping age of 85 for the guideline policy as in USPSTF (2008). We consider different stopping ages for the guideline policy of every 3-, 5-year screening in our analysis motivated by our findings from Section 3.4.3.1. That is, females have longer life expectancy, thus, screenings may continue further in age for them (e.g. until age-90) whereas males may get affected more by terminating screening at a younger age (e.g. at age-80).

We see that the guideline policy performs slightly better if the screening is terminated at age-90 rather than 85 for female patients and the stopping age-85 is the age that the guideline policy performs the best for males. The static policy suggesting one more colonoscopy screening than the guideline is among the promising points that we select, π_2, for both genders and performs reasonably well at a very similar costs as the guideline policy. This finding is very important as it shows the potential that following a simpler-to-implement policy for both genders may improve the TQALYs significantly with a reasonably small increase in costs. The static policy suggesting two additional colonoscopy than the guidelines is every 2-, 3-year screening with the stopping age of 80 for both females and males. The static policies perform reasonably well compared to the guideline for males and perform very closely to the optimal POMDP for females. However, there is still room for improving the static-periodic policy for male patients. Therefore, we analyze age-dependent policies also for high-risk patients in order to fill this gap for males and observe the changes for females in Section 3.4.4.

Table 3.8: Comparison of the US guidelines, the optimal POMDP policies of Erenay et al. (2014), and the promising static-periodic policies (π_1, π_2, π_3) for high-risk patients

	Guideline	π_1	π_2	3-, 5-year[1]	π_3	+1 Col	+2 Col	POMDP*
Females								
Total QALYs	32.191	32.210	32.217	32.200	32.237		32.233	32.239
No. of colonoscopies	7.00	7.11	7.99	7.38	9.41		9.05	10.18
Lifetime CRC risk	3.56%	3.67%	3.13%	3.35%	2.78%	π_2	3.09%	2.66%
CRC mortality	1.16%	1.27%	0.97%	0.99%	0.85%		1.08%	0.76%
Total Cost ($)	29,198	29,166	29,492	29,342	30,253		30,091	30,589
Screening Interval	4.50	4.01	3.91	4.51	3.28		3.16	2.97
Males								
High-risk patients								
Total QALYs	28.348	28.362	28.368	28.338	28.387		28.385	28.403
No. of colonoscopies	6.43	6.66	7.51	6.03	9.02		8.48	10.58
Lifetime CRC risk	3.93%	3.84%	3.34%	4.18%	2.81%	π_4	3.20	2.59%
CRC mortality	1.17%	1.28%	1.00%	1.37%	0.83%		1.06%	0.75%
Total Cost ($)	29,513	29,439	29,797	29,282	30,456		30,261	31,397
Screening Interval	4.50	4.00	3.78	4.49	3.16		3.13	2.54

* Source: Erenay et al. (2014)

[1] Stopping ages are 90 for females and 80 for males.

3.4.3.3 Sensitivity Analysis for Static Policies

Sensitivity on Disutility of Colonoscopy

The first sensitivity analysis that we conduct to understand the behaviors of the static-periodic policies is on the disutility of colonoscopy. We assume 2-week of colonoscopy disutility for the base-case and consider ±50% more disutility (i.e., 1- and 3-week) for the sensitivity analysis. As Figures 3.9 and 3.10 mimic, when the disutility of colonoscopy increases, TQALYs of all policies (i.e., the static, guideline, and optimal POMDP) decrease for both genders, as expected. The gap between the optimal POMDP policy and the best performing static-policy also closes when the disutility increases. This is mainly because, the optimal POMDP policy determines the actions (do nothing or undergo colonoscopy screening) to maximize the TQALYs, thus, scheduling fewer colonoscopy screenings at higher disutility levels is the strategy that the optimal policy follows. When the decision is

do nothing for most of the ages, the personalized screening policy does not provide as much improvement as it does with the base-case parameters. Moreover, the gap closes faster for male patients than female patients. This finding is in line with the findings of Erenay et al. (2014) that the optimal policy provides greater improvements over the guidelines for male patients, namely, following a personalized screening is more important for male patients in terms of improving TQALYs. Therefore, lack of personalized screening helps closing the gap faster for male patients.

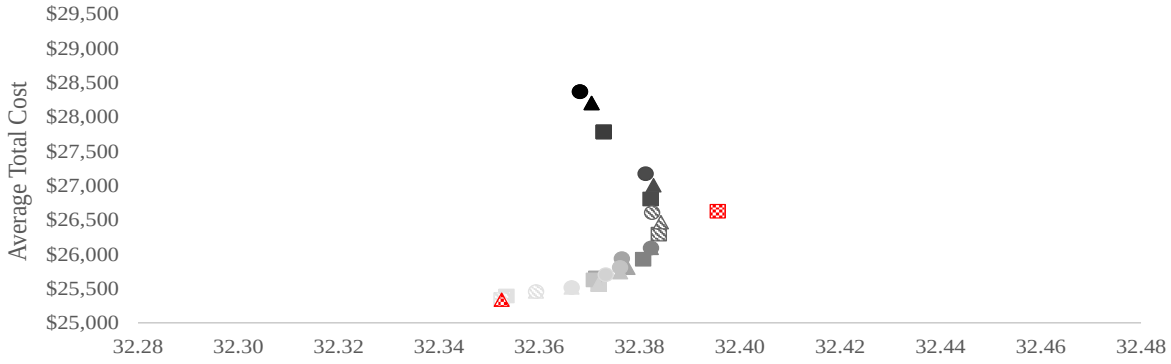
Sensitivity on Colonoscopy Sensitivity for Polyps and Cancer

We also conduct sensitivity analysis to see how the performances change with colonoscopy sensitivity for polyps and cancer as the screening decisions are prompt to change based on the effectiveness of the screening method. Figures 3.13-3.16 imply that the performances of the static policies and the gap are more sensitive to the changes in sensitivity of polyps than sensitivity of cancer. Moreover, the gap decreases more significantly for high-risk patients when the sensitivity of colonoscopy increases, as expected. It is also important to note that the order of the alternative policies in terms of TQALYs does not noticeably change with the change of either of the sensitivities. This stability in the order shows that the effects of the stopping age and the screening frequency on the performance measures are independent from the screening sensitivity. This is a strong conclusion that says determining the best-performing periodic policy with a given colonospocy sensitivity is sufficient to inform guidelines without further analysis with respect to any updates on the sensitivity of colonoscopy screening.

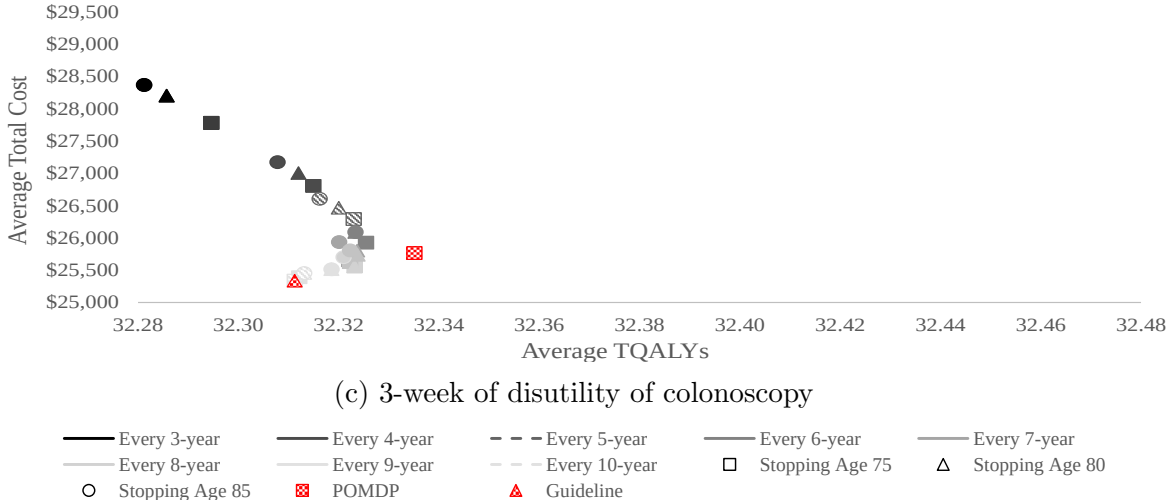
Sensitivity on Discount Rates for TQALYs and Costs

Our analysis on discount rates show how the performances change when the effects of costs and TQALYs change. Note that this change has a significant effect on the POMDP policy structure since the optimal policy determines the actions to maximize TQALYs. As Figures 3.17a-3.20e mimic, when the discount on TQALYs increases, the policies vary more in terms of their TQALYs, namely, TQALYs become more sensitive to the effects of screening frequency. On the other hand, when the discount on the costs increases the policies do not scatter noticeably different on the Pareto spectrum, because the effect of screening frequency on costs is proportional.

(a) 1-week of disutility of colonoscopy

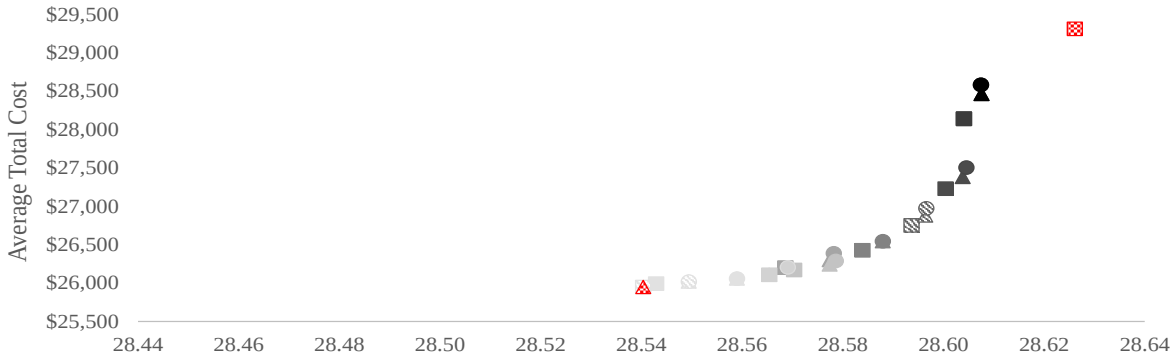


(b) 2-week of disutility of colonoscopy (*The base-case*)

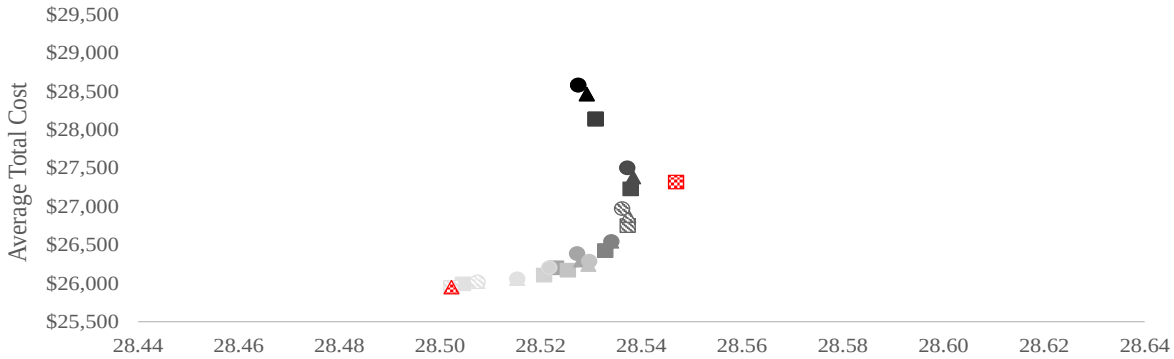


(c) 3-week of disutility of colonoscopy

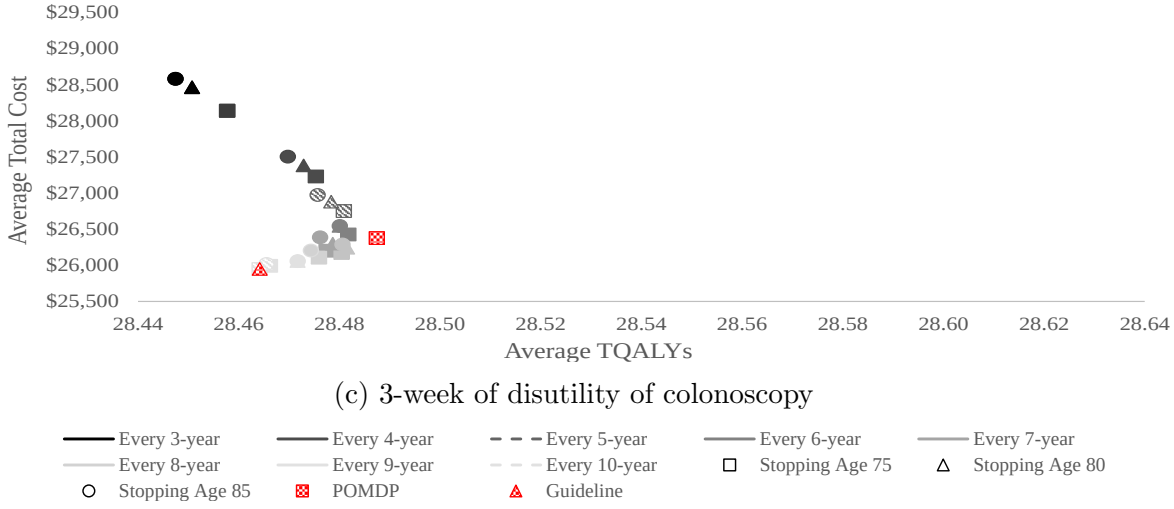
Figure 3.9: *Static screening policies* with varying disutility levels of colonoscopy screening for low-risk female patients given stopping ages of 75, 80, 85



(a) 1-week of disutility of colonoscopy

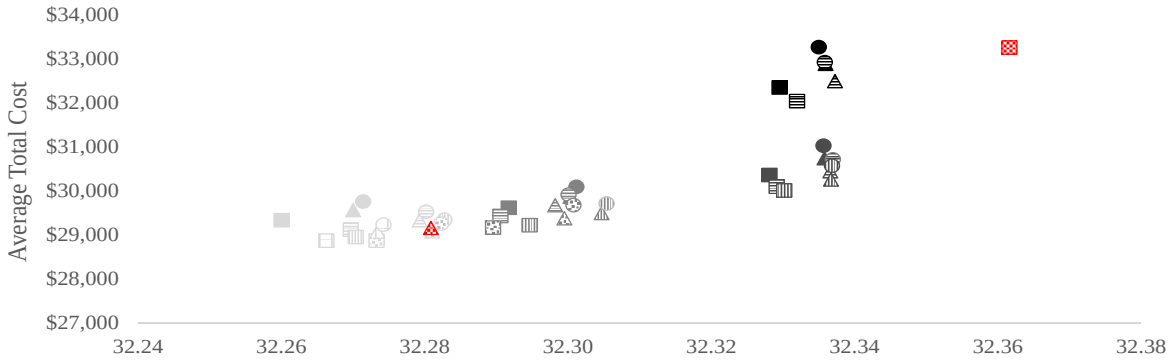


(b) 2-week of disutility of colonoscopy (*The base-case*)

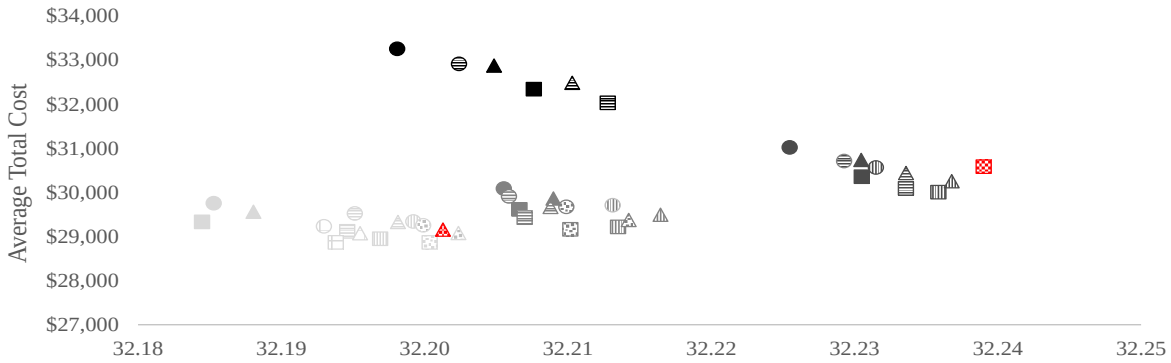


(c) 3-week of disutility of colonoscopy

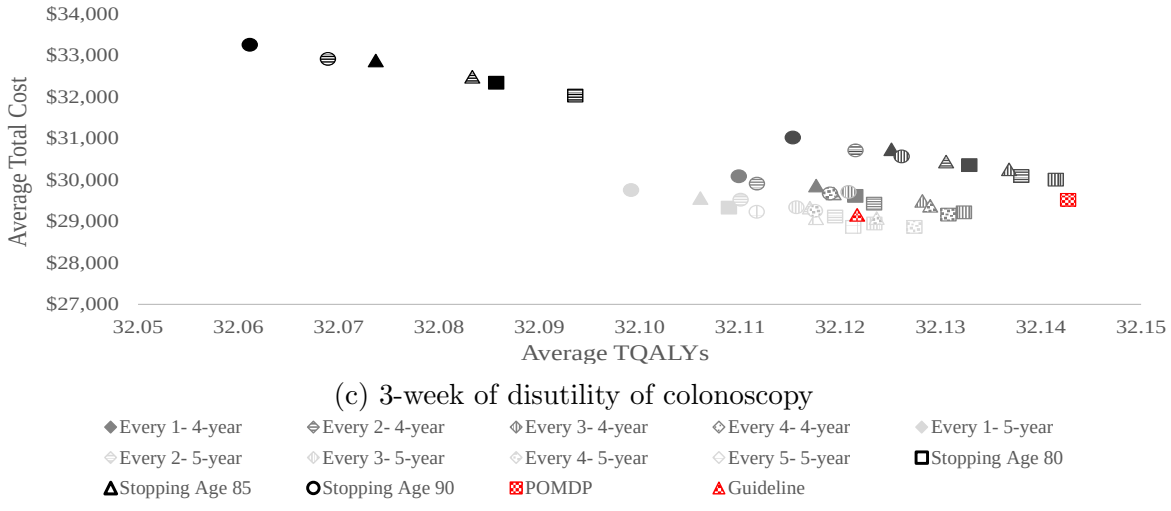
Figure 3.10: *Static screening policies* with varying disutility levels of colonoscopy screening for low-risk male patients given stopping ages of 75, 80, 85



(a) 1-week of disutility of colonoscopy

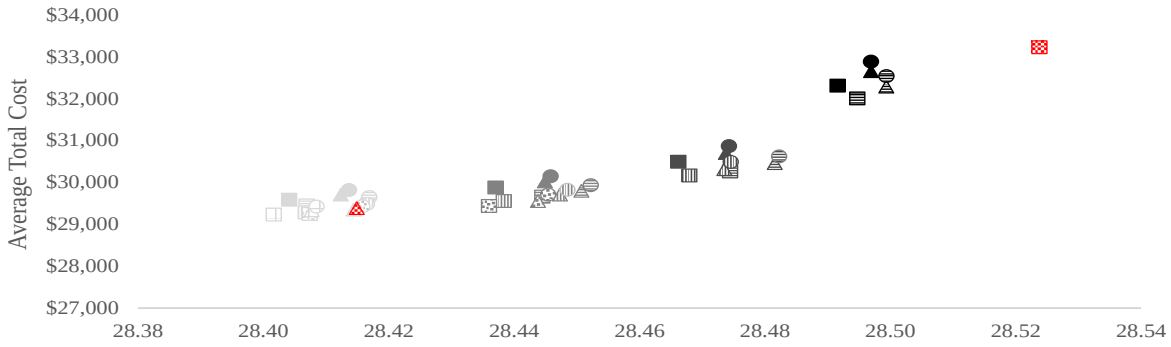


(b) 2-week of disutility of colonoscopy (*The base-case*)

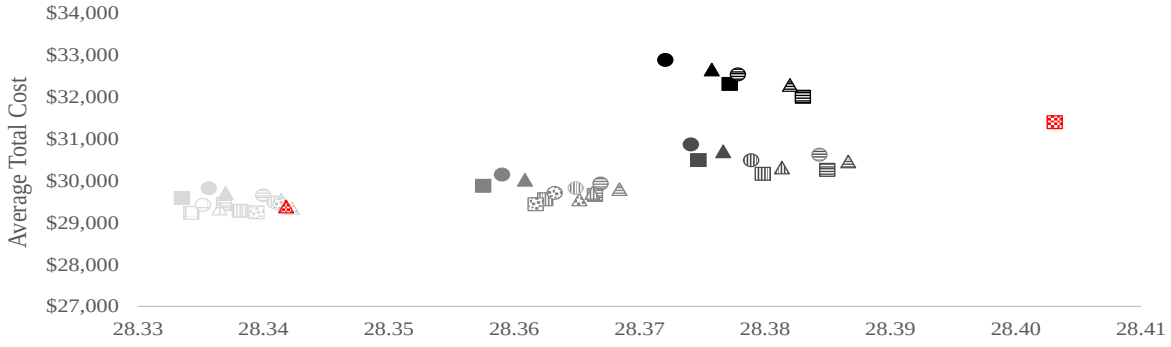


(c) 3-week of disutility of colonoscopy

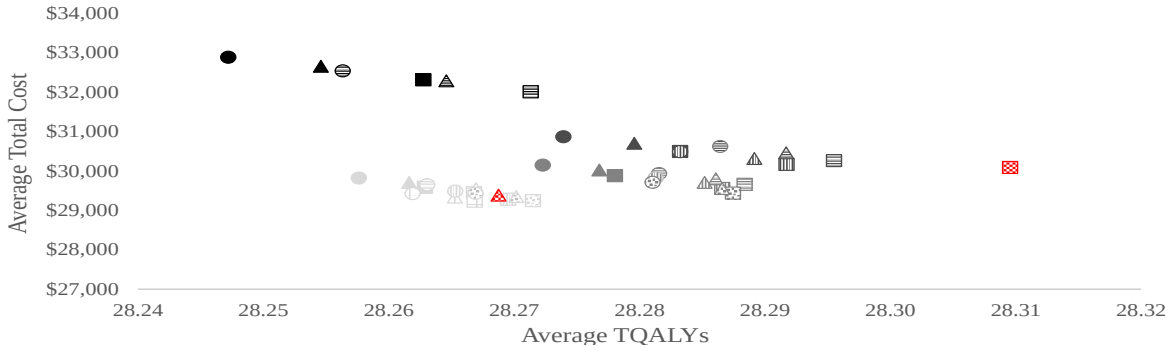
Figure 3.11: *Static screening policies* with varying disutility levels of colonoscopy screening for high-risk female patients given stopping ages of 75, 80, 85



(a) 1-week of disutility of colonoscopy



(b) 2-week of disutility of colonoscopy (*The base-case*)



(c) 3-week of disutility of colonoscopy

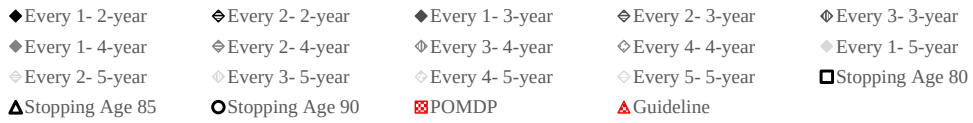
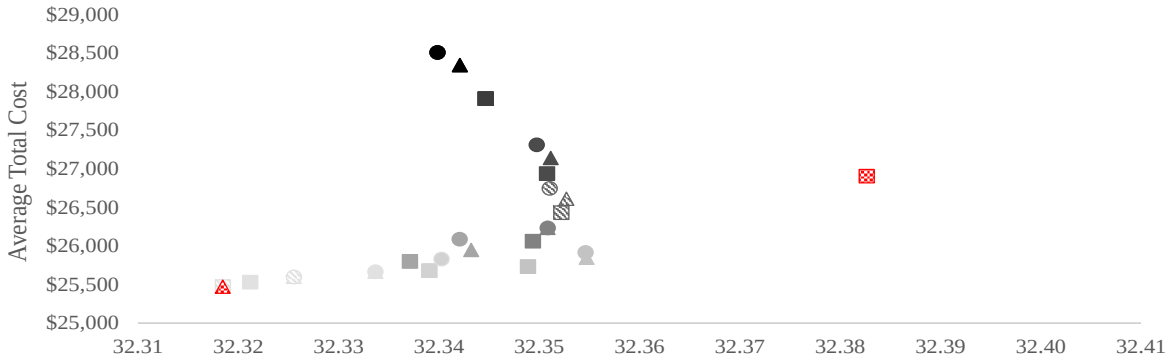
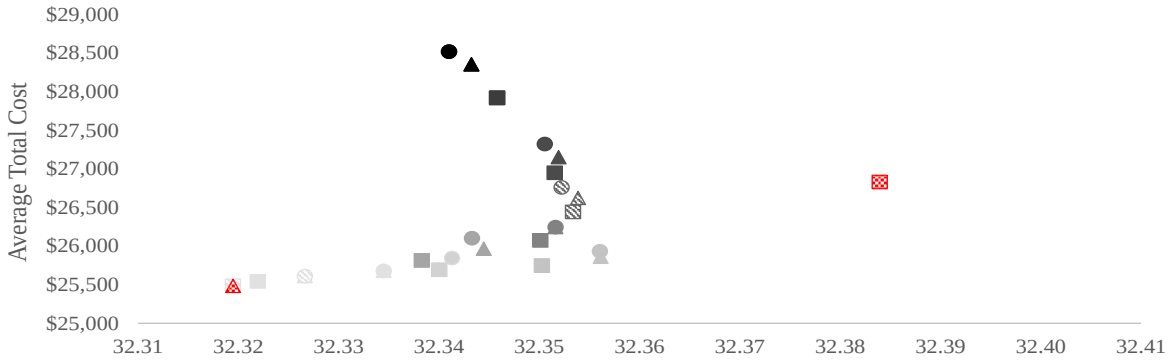


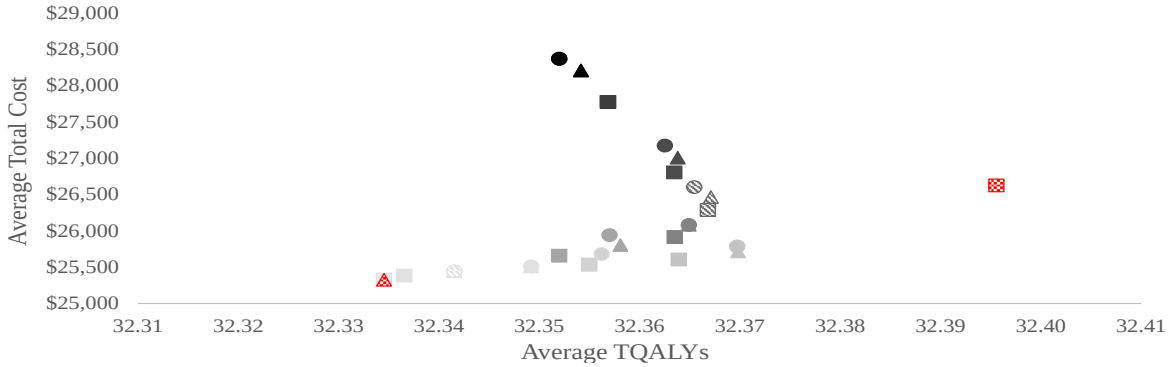
Figure 3.12: *Static screening policies* with varying disutility levels of colonoscopy screening for high-risk male patients given stopping ages of 75, 80, 85



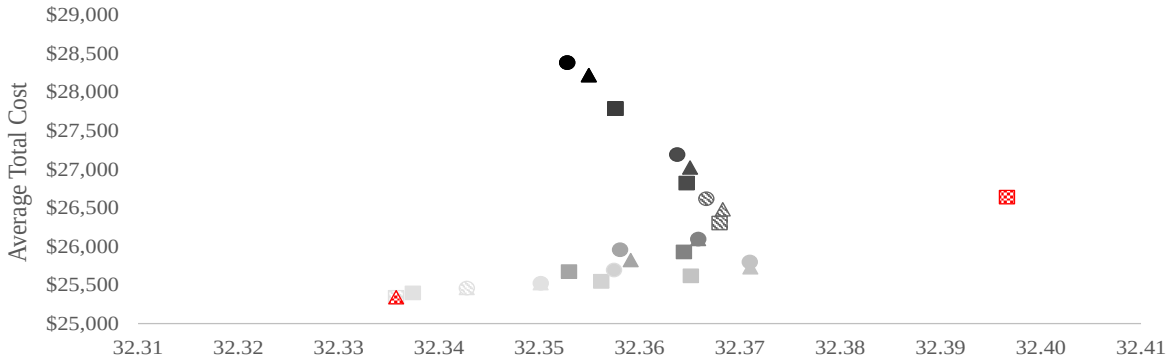
(a) Sensitivity of 80% for polyps and 85% for cancer



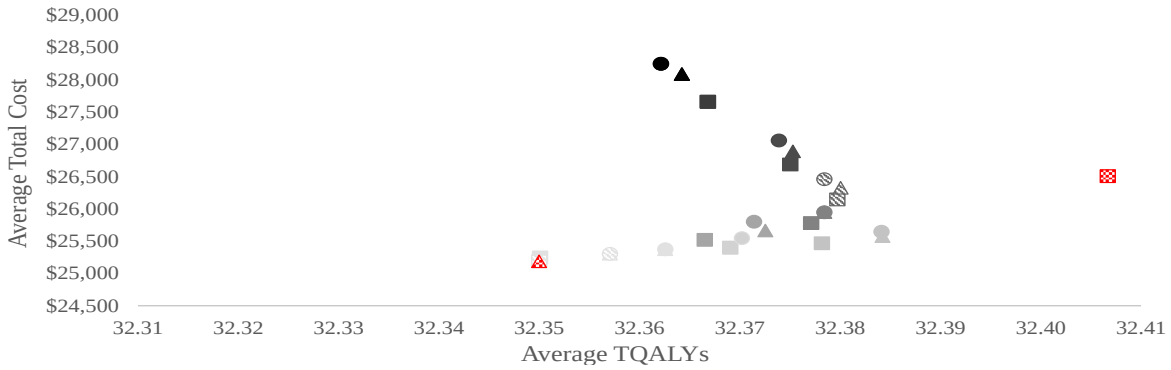
(b) Sensitivity of 80% for polyps and 90% for cancer



(c) Sensitivity of 85% for polyps and 90% for cancer (*The base-case*)



(d) Sensitivity of 85% for polyps and 95% for cancer



(e) Sensitivity of 90% for polyps and 95% for cancer

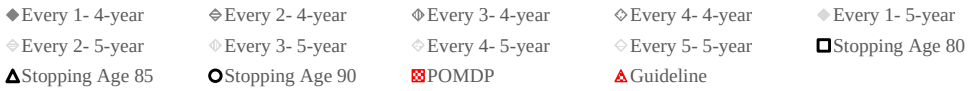
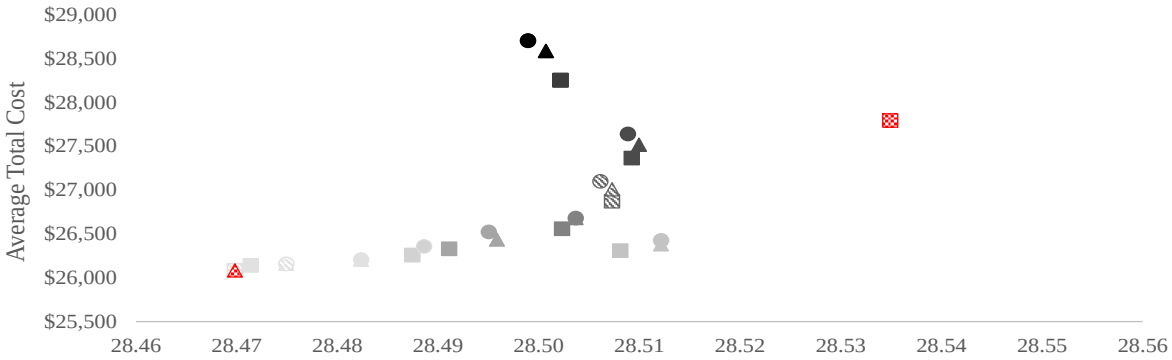
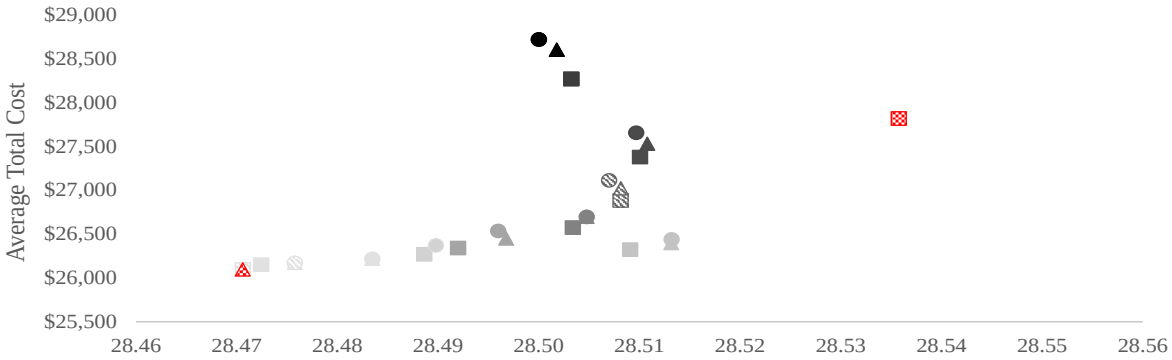


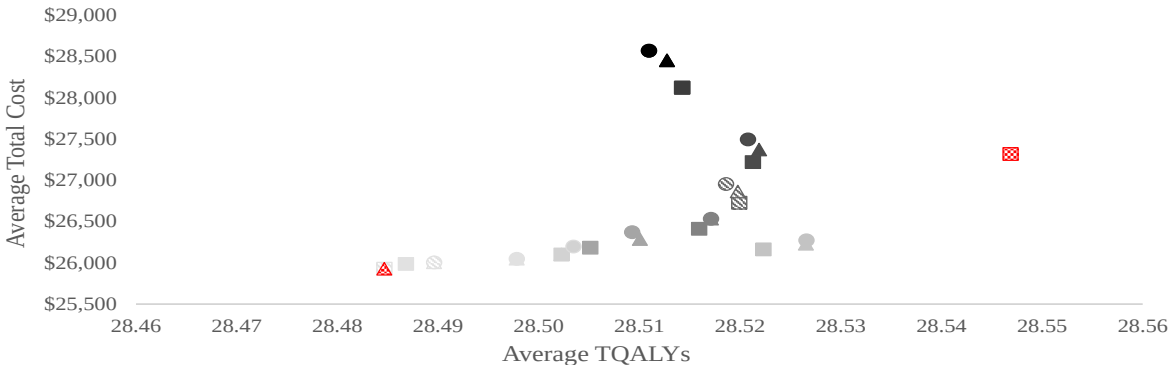
Figure 3.13: *Static screening policies* with varying colonoscopy screening sensitivities for adenomatous polyps and cancer for low-risk female patients given stopping ages of 75, 80, 85



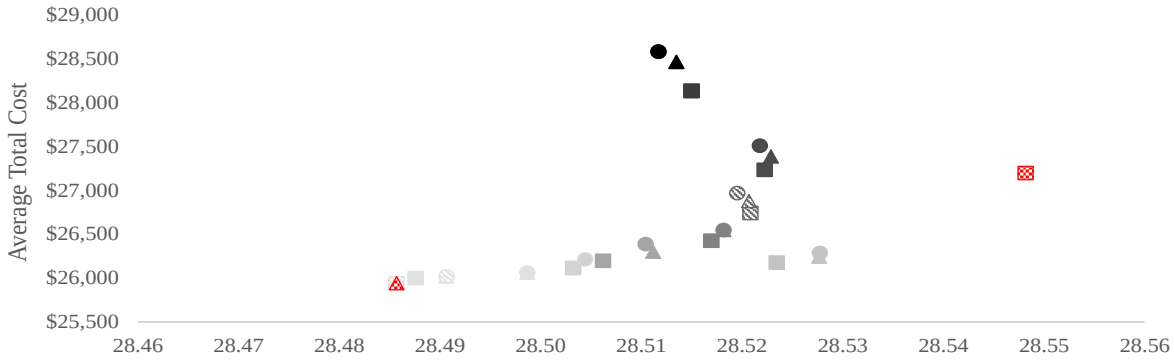
(a) Sensitivity of 80% for polyps and 85% for cancer



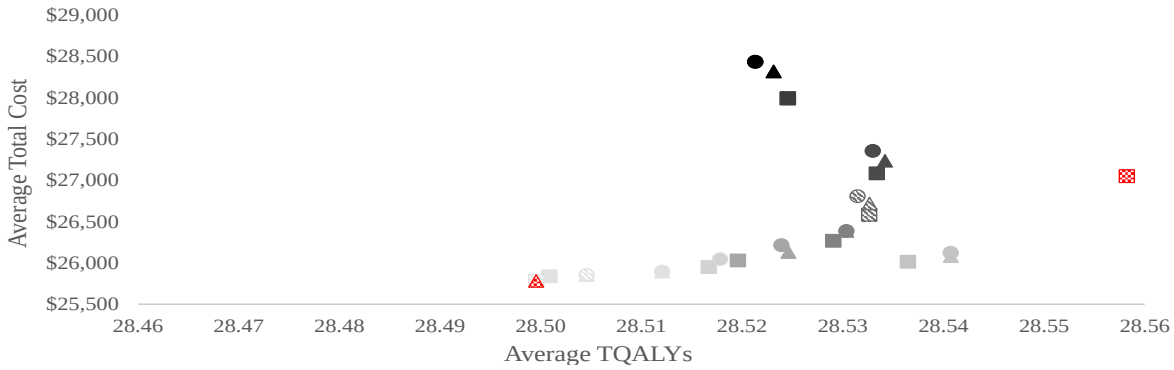
(b) Sensitivity of 80% for polyps and 90% for cancer



(c) Sensitivity of 85% for polyps and 90% for cancer (*The base-case*)



(d) Sensitivity of 85% for polyps and 95% for cancer



(e) Sensitivity of 90% for polyps and 95% for cancer

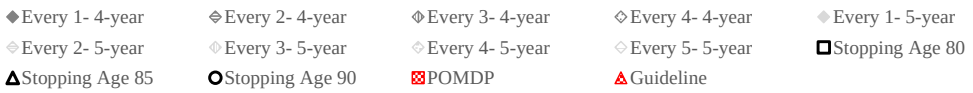
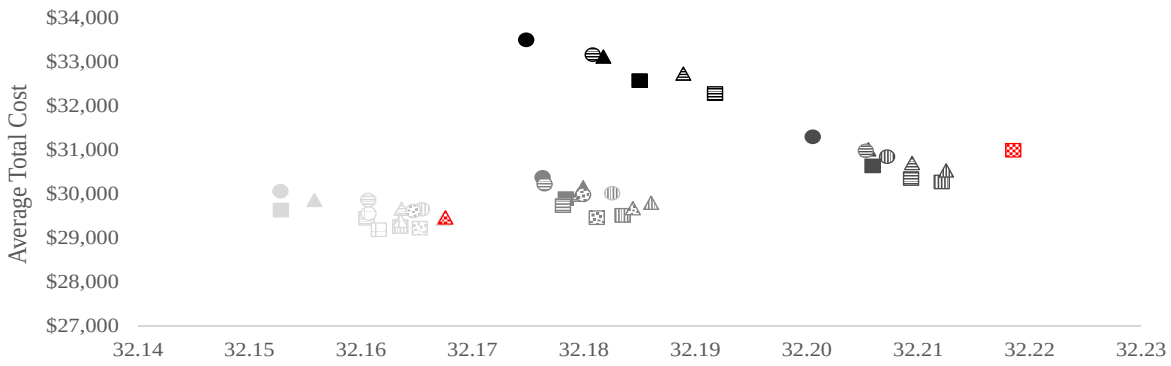
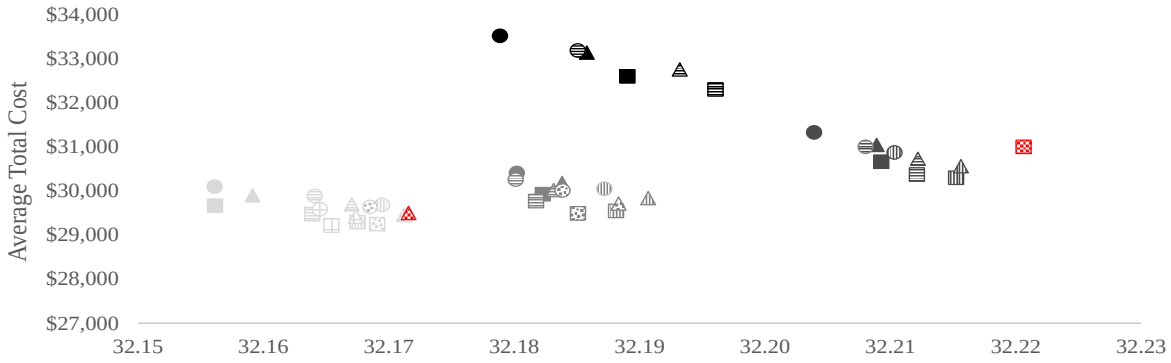


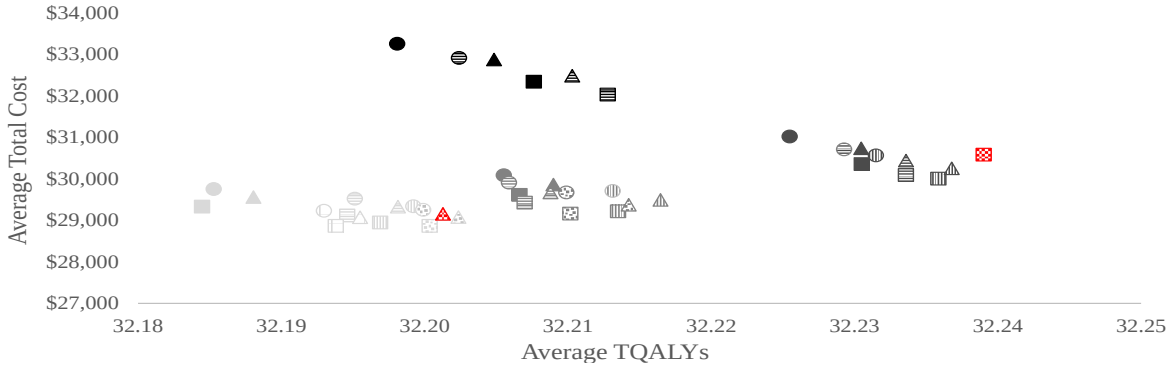
Figure 3.14: *Static screening policies* with varying colonoscopy screening sensitivities for adenomatous polyps and cancer for low-risk male patients given stopping ages of 75, 80, 85



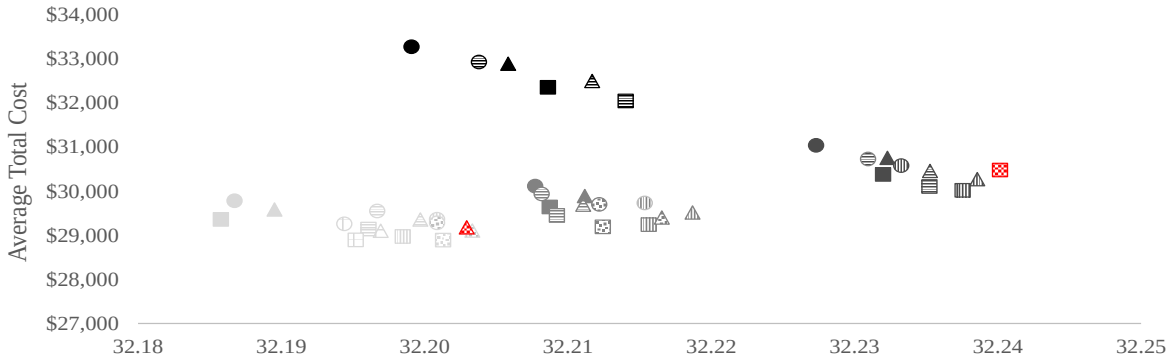
(a) Sensitivity of 80% for polyps and 85% for cancer



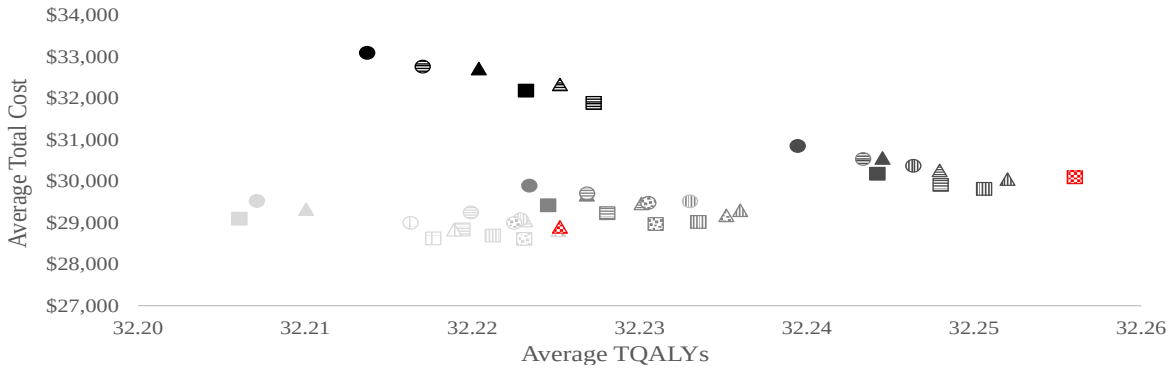
(b) Sensitivity of 80% for polyps and 90% for cancer



(c) Sensitivity of 85% for polyps and 90% for cancer (*The base-case*)



(d) Sensitivity of 85% for polyps and 95% for cancer



(e) Sensitivity of 90% for polyps and 95% for cancer

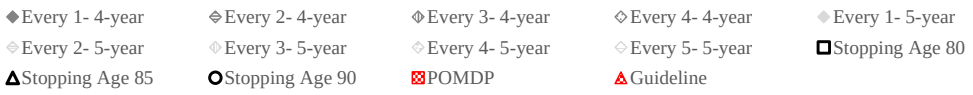
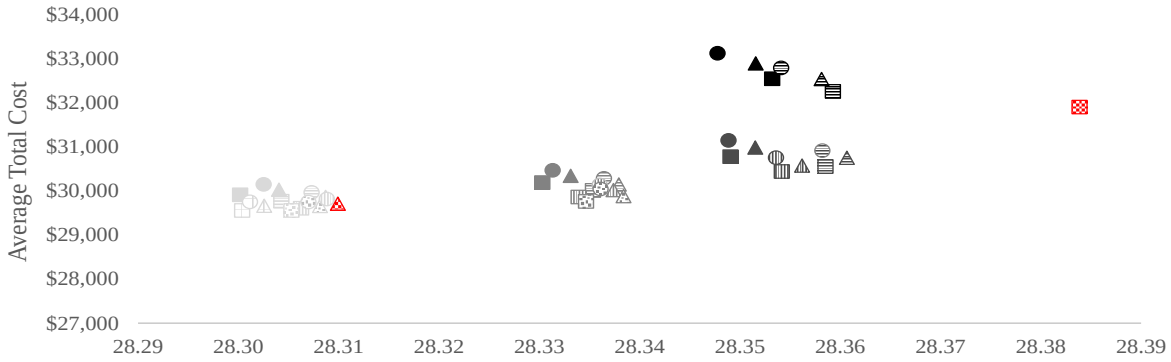
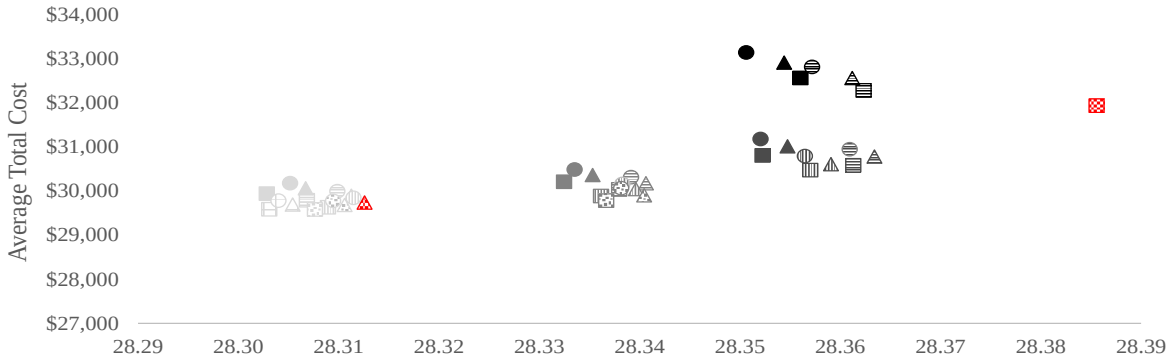


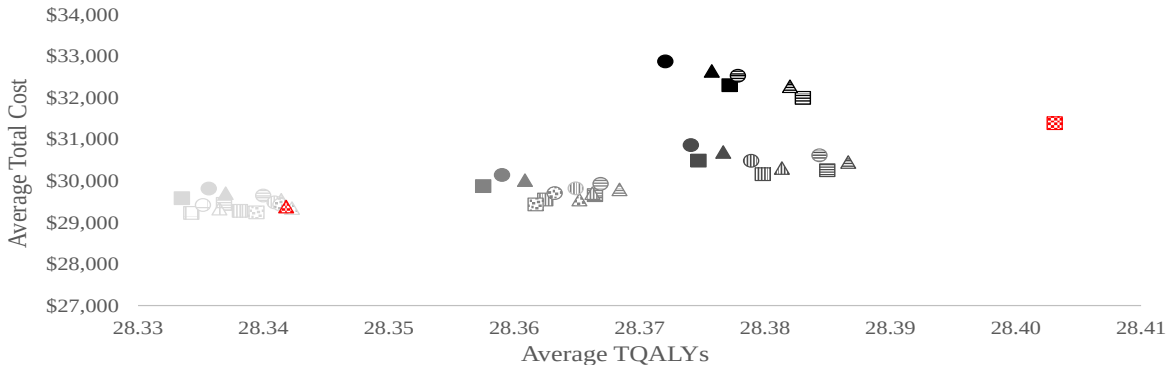
Figure 3.15: *Static screening policies* with varying colonoscopy screening sensitivities for adenomatous polyps and cancer for high-risk female patients given stopping ages of 75, 80, 85



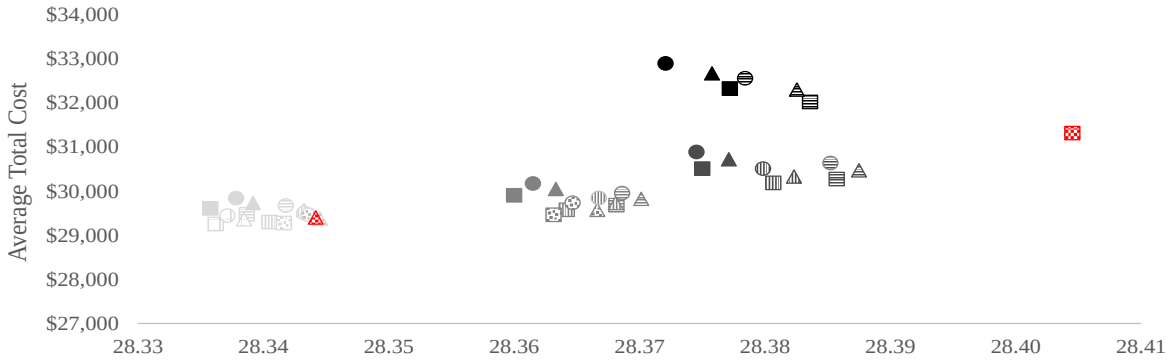
(a) Sensitivity of 80% for polyps and 85% for cancer



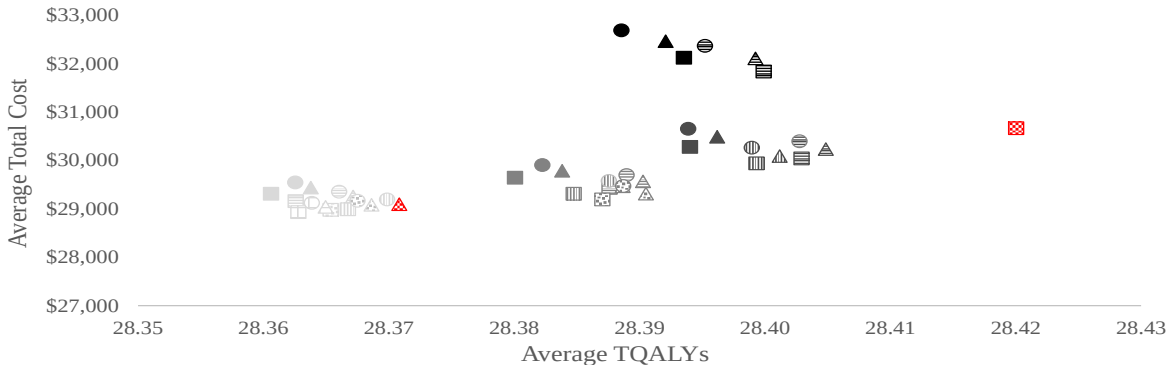
(b) Sensitivity of 80% for polyps and 90% for cancer



(c) Sensitivity of 85% for polyps and 90% for cancer (*The base-case*)



(d) Sensitivity of 85% for polyps and 95% for cancer



(e) Sensitivity of 90% for polyps and 95% for cancer

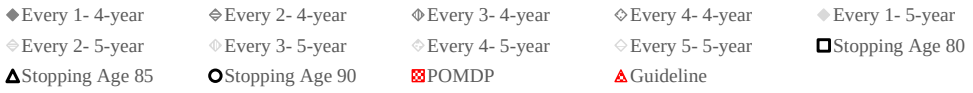
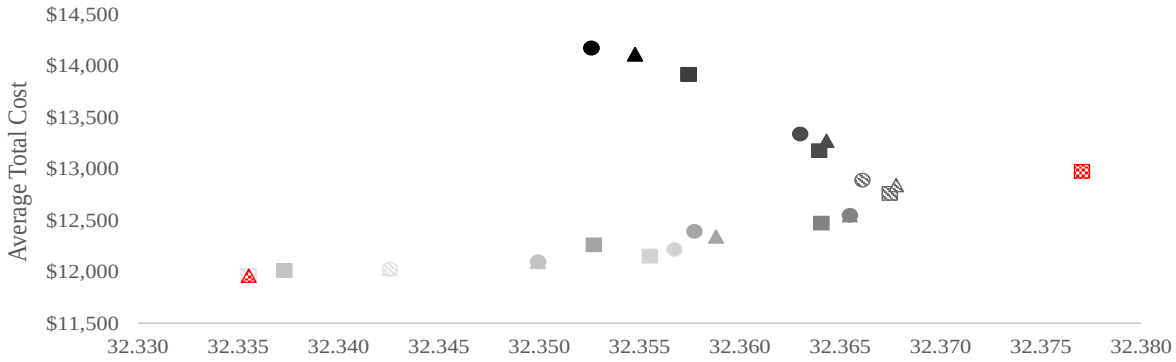
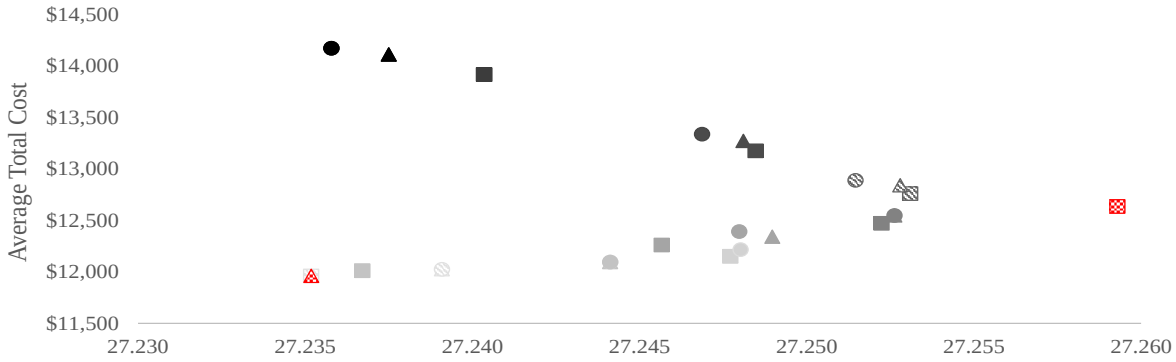


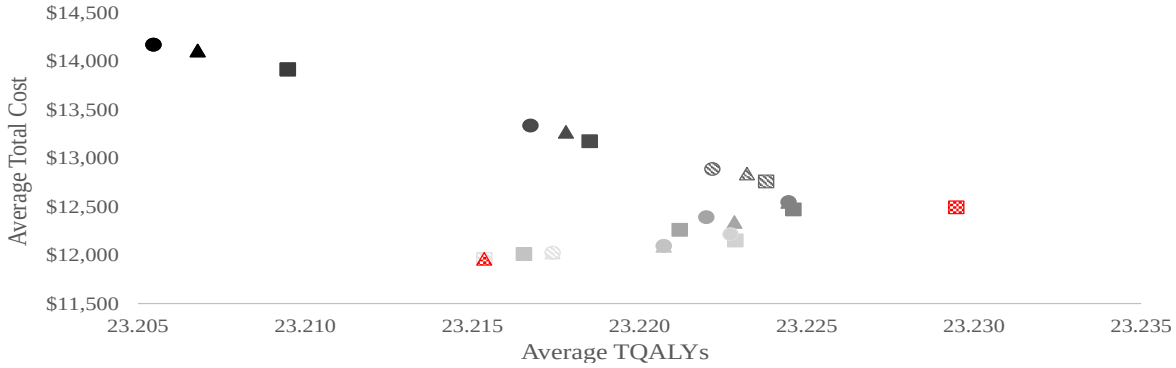
Figure 3.16: *Static screening policies* with varying colonoscopy screening sensitivities for adenomatous polyps and cancer for high-risk male patients given stopping ages of 75, 80, 85



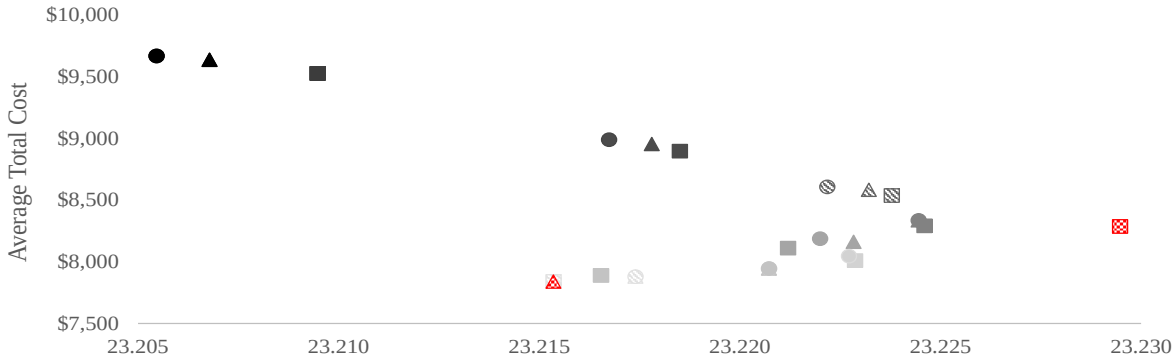
(a) Discount Rates of 0% for TQALYs and 3% for costs



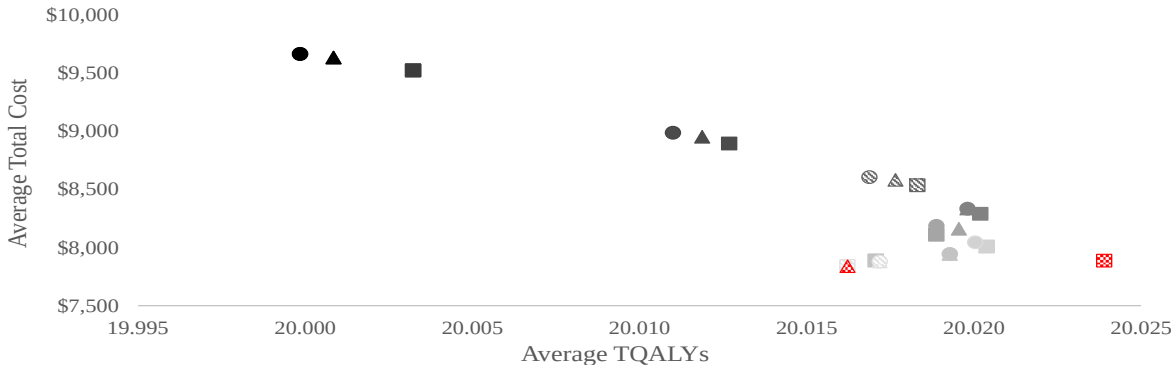
(b) Discount Rates of 1% for TQALYs and 3% for costs



(c) Discount Rates of 2% for TQALYs and 3% for costs



(d) Discount Rates of 2% for TQALYs and 5% for costs



(e) Discount Rates of 3% for TQALYs and 5% for costs

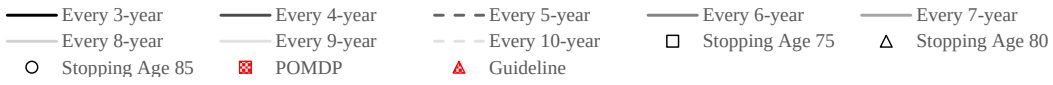
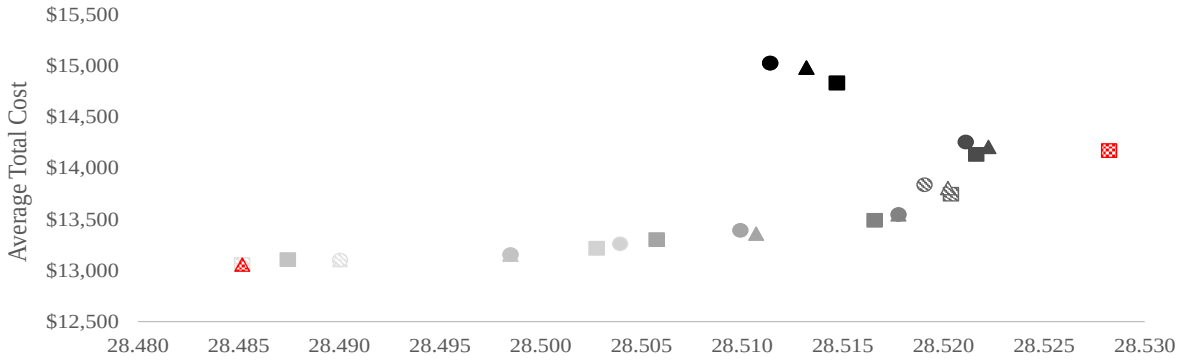
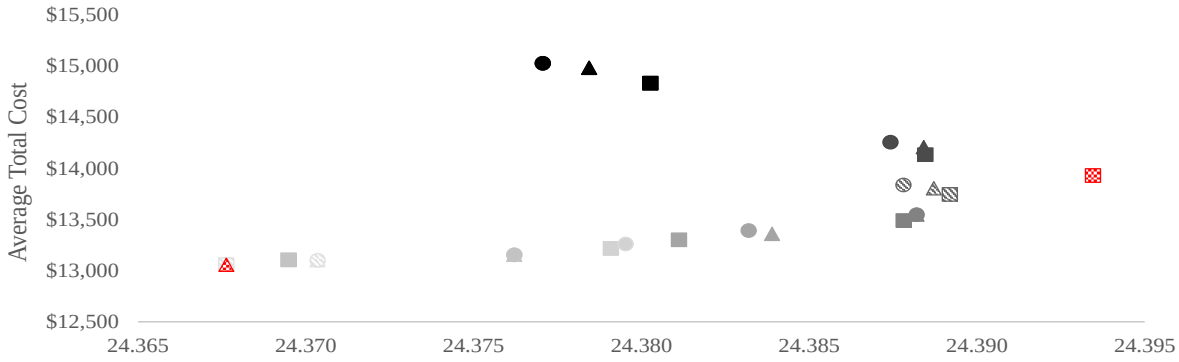


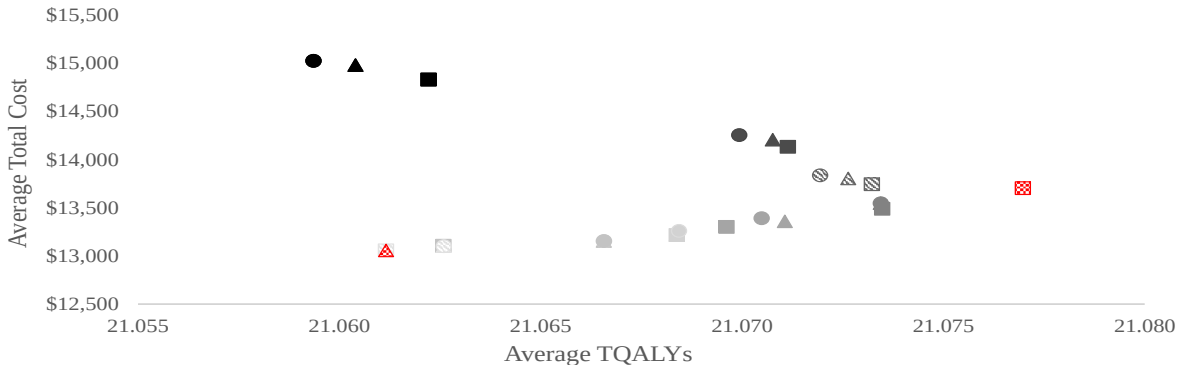
Figure 3.17: *Static screening policies* with varying discount rates for TQALYs and costs for low-risk female patients given stopping ages of 75, 80, 85



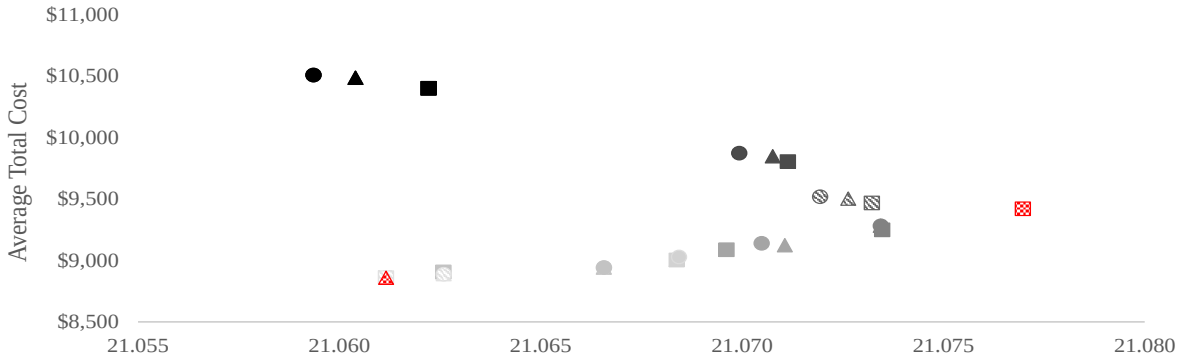
(a) Discount Rates of 0% for TQALYs and 3% for costs



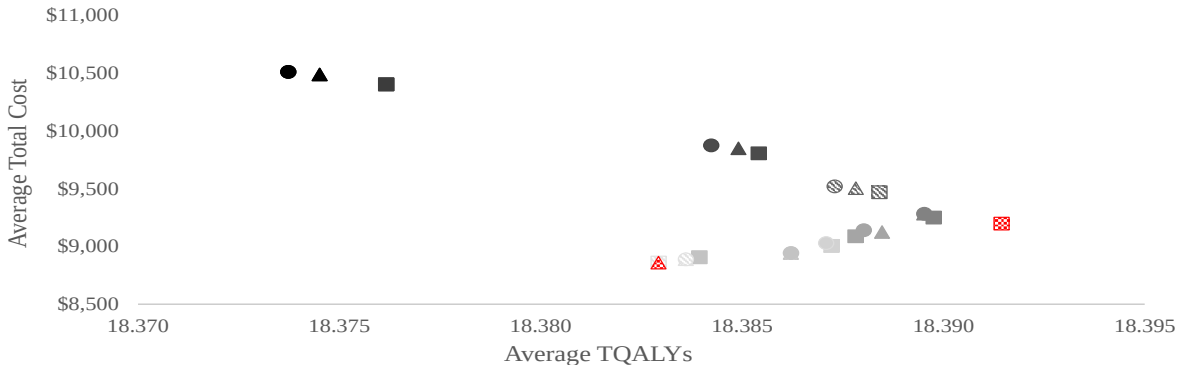
(b) Discount Rates of 1% for TQALYs and 3% for costs



(c) Discount Rates of 2% for TQALYs and 3% for costs



(d) Discount Rates of 2% for TQALYs and 5% for costs



(e) Discount Rates of 3% for TQALYs and 5% for costs

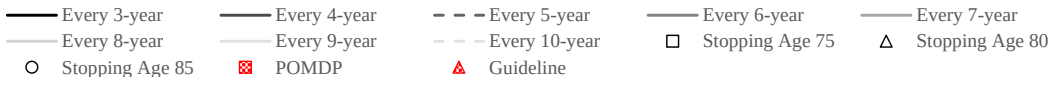
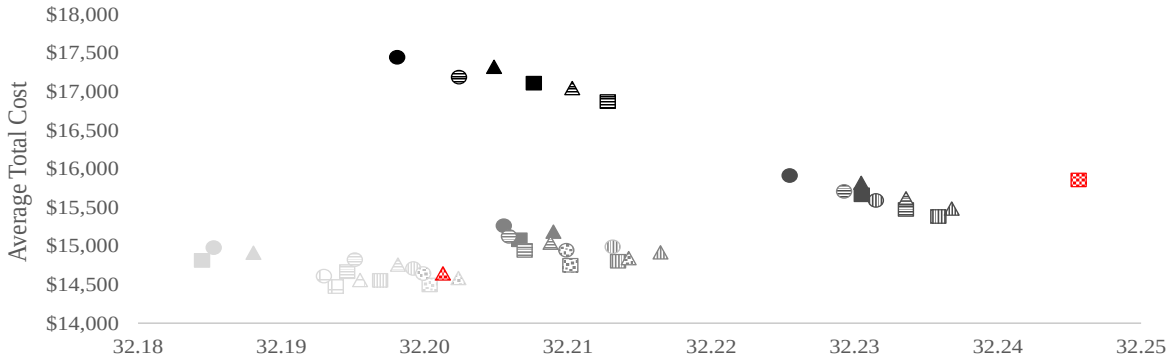
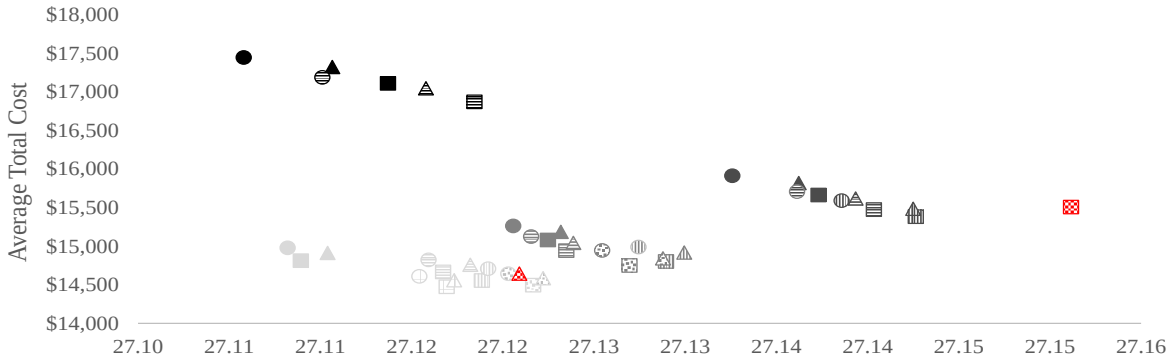


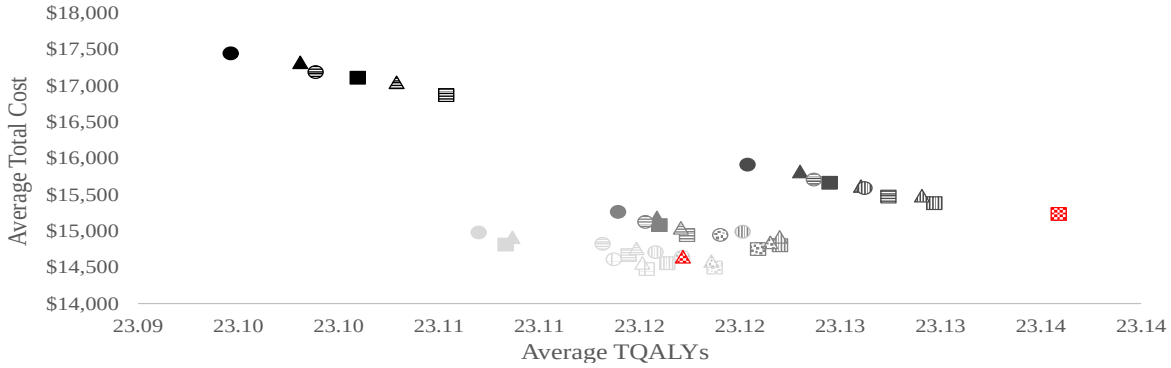
Figure 3.18: *Static screening policies* with varying discount rates for TQALYs and costs for low-risk male patients given stopping ages of 75, 80, 85



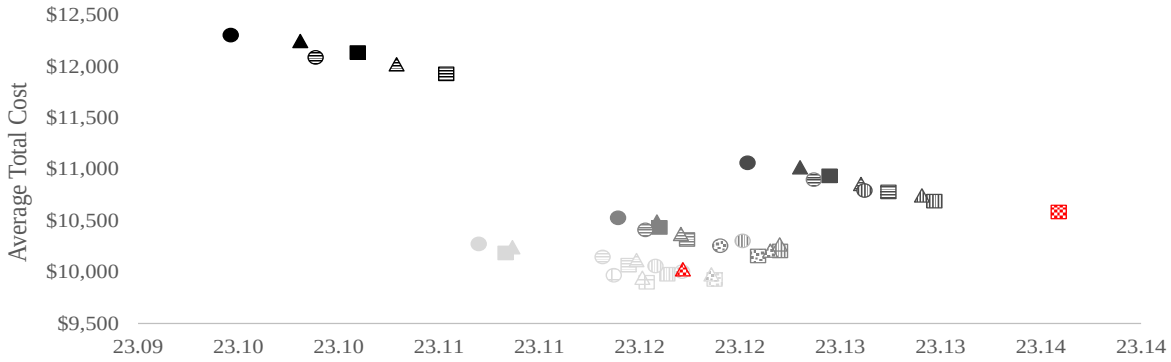
(a) Discount Rates of 0% for TQALYs and 3% for costs



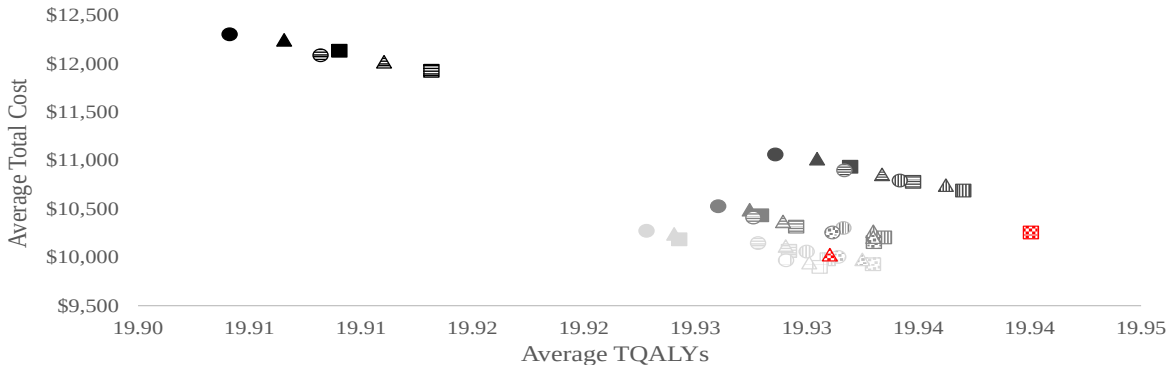
(b) Discount Rates of 1% for TQALYs and 3% for costs



(c) Discount Rates of 2% for TQALYs and 3% for costs



(d) Discount Rates of 2% for TQALYs and 5% for costs



(e) Discount Rates of 3% for TQALYs and 5% for costs

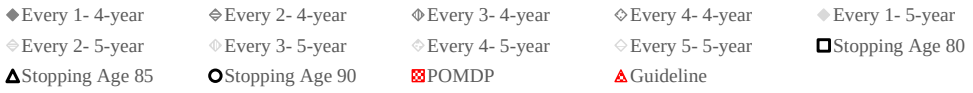
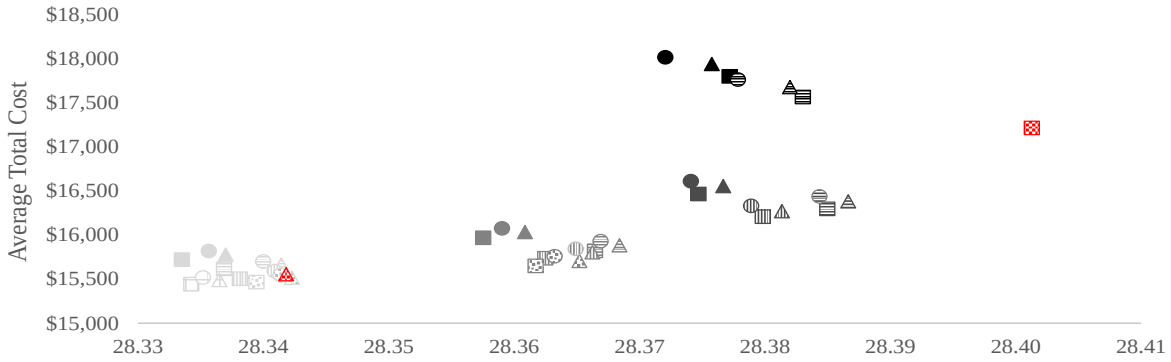
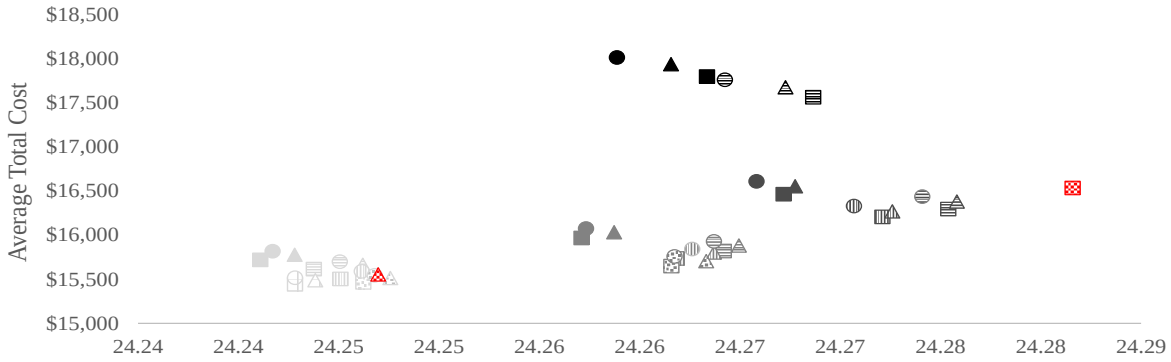


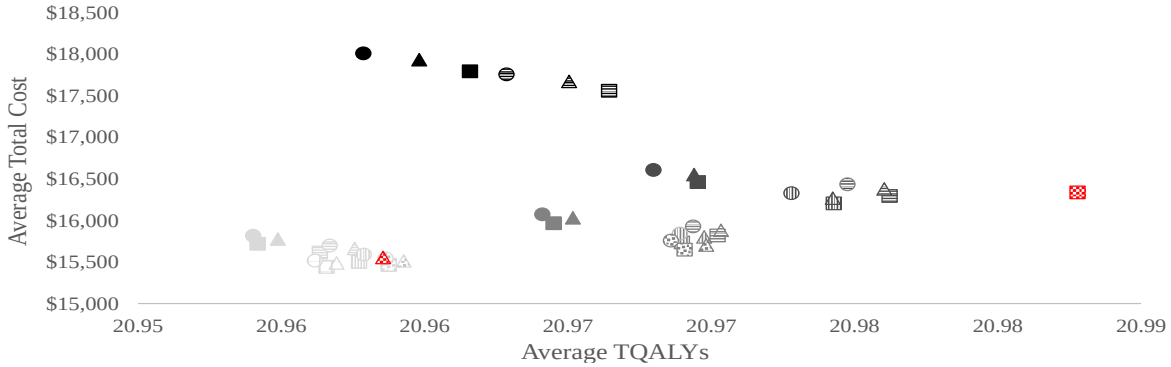
Figure 3.19: *Static screening policies* with varying discount rates for TQALYs and costs for high-risk female patients given stopping ages of 80, 85, 90



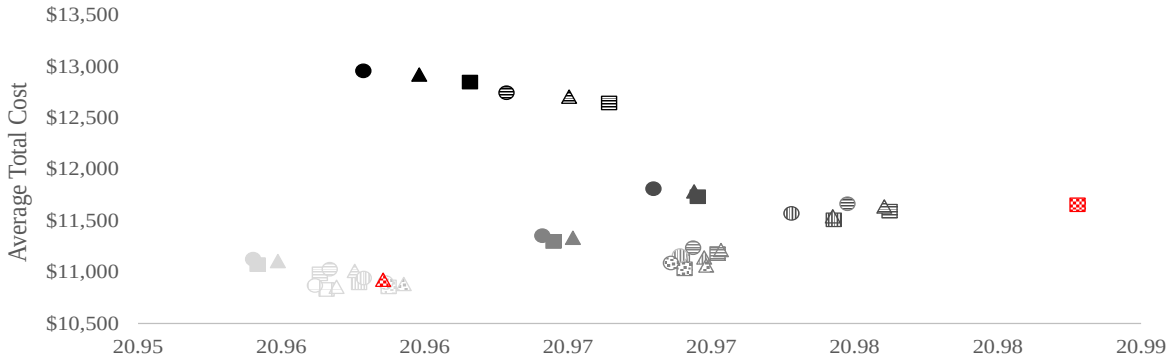
(a) Discount Rates of 0% for TQALYs and 3% for costs



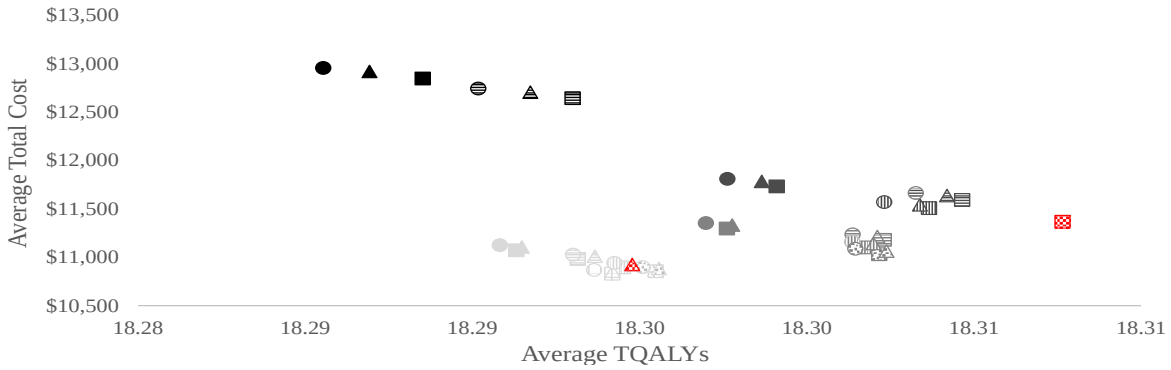
(b) Discount Rates of 1% for TQALYs and 3% for costs



(c) Discount Rates of 2% for TQALYs and 3% for costs



(d) Discount Rates of 2% for TQALYs and 5% for costs



(e) Discount Rates of 3% for TQALYs and 5% for costs

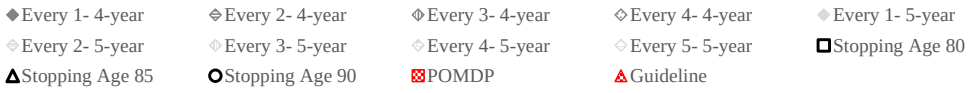


Figure 3.20: *Static screening policies* with varying discount rates for TQALYs and costs for high-risk male patients given stopping ages of 80, 85, 90

3.4.4 Age-dependent Periodic Policies

In this section, we analyze *dynamic-periodic colonoscopy screening policies* which recommend patients undergoing periodic colonoscopy screening at age-group- and gender-specific frequencies for low-risk and high-risk patients. We establish the alternative dynamic-periodic policies by dividing the age range between age-50 and the stopping age into two and three sub-intervals; i.e., 2 and 3 age-groups for 1-switch and 2-switch policies, respectively.

For dynamic-periodic policy type, we analyze all possible ways to divide the time horizon into sub-intervals of length 5-year or multiples thereof because (i) dividing decision horizon in the order of 5-year intervals is a common practice in the CRC prevention and surveillance literature (Zauber et al., 2009; Lieberman et al., 2000), (ii) division of the intervals into shorter intervals seems clinically impractical. The complete list of age intervals is given in Table 3.9. For each division option in Table 3.9, we consider all possible combinations of dynamic-periodic policies in the form of the current guidelines. That is, 1) we consider all possible ways of selecting a single screening frequency for each sub-interval from the set of $\{3, 4, \ldots, 10\}$ for low-risk patients; 2) we consider all possible ways of selecting two screening frequencies, i.e., f_1 after a polypectomy and f_2 after a negative colonoscopy; $f_1 \leq f_2$ and $f_1, f_2 \in \{1, 2, \ldots, 6\}$ for high-risk patients.

Table 3.9: Age-dependent periodic policy list

2-age-group (1-switch)	50-60	50-65	50-70	50-75	50-80		
3-age-group (2-switch)	50-60-70	50-60-75	50-60-80	50-65-70	50-65-75	50-65-80	50-70-80

Sections 3.4.4.1 and 3.4.4.2 present our findings about the best-performing age-dependent periodic policies for high- and low-risk patients, respectively. In order to help readers follow the figures presented in these sections easily, we provide a reference list here explaining the key elements of the figure formatting and notation.

- **Marker type** defines the age group scenario:

  - Square: 2-age group
  - Diamond: 3-age group
  - Circle: Guideline (dashed filled) and Optimal POMDP policy (gingham filled)

- **Marker color** defines at what age to switch the screening frequency after age-50 (i.e., the first switch age).
  - the color gets lighter when the switch age increases, e.g., black at age-60, dark gray at age-65, etc.

- **Marker fill type** defines at what age to switch the screening frequency after the first switch age (i.e., the second switch age).

- **Labeled policies, i.e.,** π_i, $i = \{1, 2, 3, 4\}$ represent the Pareto-frontier policies that are cost-effective compared to a reference point which, in our analysis, is the previous cost-effective Pareto-frontier solution starting from the guideline policy. The cost-effectiveness is determined using the incremental cost-effectiveness ratio (ICER) given in Equation 3.11.

- **The best-performing policy, i.e.,** π_4, is the Pareto-frontier policy that performs closest to the optimal POMDP policy in terms of TQALYs regardless of its ICER value.

- **Screening frequencies associated with Policy** π_i are given in parenthesis and ordered with respect to age intervals with commas in between. For high-risk patients, hyper is used between the screening frequencies after a polypectomy and after a negative result. E.g., $\pi_i(3, 4, 5)$ denotes the screening policy that suggests every 3-, 4-, and 5-year screening for the low-risk patients at the ages in age intervals 1, 2, and 3, respectively; whereas π_i(3-6,4-5) denotes the screening policy that suggests every 3-, and 4-year screening after polypectomy and 6- and 5-year screening after a negative result for the high-risk patients at the ages in age intervals 1, and 2, respectively.

3.4.4.1 The Best Performing Age-dependent Periodic Policies for High-Risk Patients

High-risk patients are those who have previously developed at least one adenomatous polyp (Winawer et al., 2003). The screening frequency recommended by the guidelines for high-risk patients is every 3-year after a polypectomy and every 5-year thereafter until another positive finding (Siegel et al., 2017). Let colonoscopy every f_1^i-year be the screening frequency after a polypectomy for high-risk patients in age-group i, and every f_2^i-year be the frequency followed after a negative test result from the previous colonoscopy. The structure of the guidelines motivates us to consider dynamic-periodic policies with decreasing screening intervals after negative colonoscopy results, i.e., $f_1^i \leq f_2^i$. Moreover, it is intuitive to consider decreasing screening frequencies as patients age especially for high-risk patients, because the expected benefit from preventing CRC through colonoscopy screening decreases in age due to reduced expected remaining lifetime. Thus, we assume $f_j^i / leq f^(i+1)_j$.

Although these assumptions allow us to reduce the number of different screening frequency combinations simulated, we still have 1176 different combinations generated from the frequency sets of $i_j \in \{1, 2, \cdots, 6\}$, $\forall i \in \{1, 2, 3\}$ and $\forall j \in \{1, 2\}$ under two gender, 12 different age intervals (given in Table 3.9) and 3 different stopping age $(80, 85, 90)$ options for the base case parameters. For each dynamic-periodic policy combination and patient setting, the simulation is run for $100,000$ patients. Because we have conducted a comprehensive and computationally expensive numerical analysis, in this section, we only present the performances of the Pareto-efficient policies in the spectrum of total expected cost and expected TQALYs.

Figures 3.21 and 3.22 present the performances of the 1-switch dynamic-periodic policies on the Pareto-efficiency spectrum which perform better than the AGA guideline in terms of TQALYs for high-risk female and male patients, respectively. Figures 3.21 and 3.22 show that the number of Pareto-efficient solutions are quite few compared to the number of all policies evaluated. That is, there are 25 (30), 36 (22), and 26 (21) Pareto-efficient solutions for female (male) patients given stopping ages of $80, 85$, and 90, respectively. This observation is important since we also want to simplify the policy evaluation process for

the clinicians.

As Figures 3.21 and 3.22 show, the 1-switch periodic screening policies provide greater improvement in TQALYs both for female and male patients over the guideline performances. Note that π_1 policies perform at a lower cost than the guidelines while providing significant improvements in terms of TQALYs for both genders and all stopping ages. Policy π_1 for each age-group and gender is important because it requires less aggressive screening than the guideline, e.g., 6.34, 6.6, and 6.95 colonoscopies on average for the stopping ages 80, 85, and 90 for female patients compared to 7 colonoscopies on average by the guideline, they still increase the TQALYs. Specifically, considering only 1-switch policies, Figures 3.21 and 3.22 suggest that clinicians can achieve TQALYs improvements without any additional costs by adopting every 4-5 year screening in the first age interval for all stopping ages and 5-5 year screening for the stopping ages 80, 85 and 6-6 year screening for the stopping age 90 in the second age interval as opposed to 3-5 year screening suggested by guidelines. Although there may be other policies with better TQALYs compared to the guidelines with no additional cost, they may not achieve this under all gender and stopping age combinations.

The above finding indicates that the performance of a screening policy depends not only how many colonoscopy screening is performed but also when those colonoscopy screenings are performed. That is, if clinicians follow an age-dependent screening frequency after the first polypectomy while using the same frequency with the guideline after a negative test result, we can achieve a significant improvement in terms of both TQALYs and average cost. Moreover, scheduling fewer colonoscopy screening can also improve the utilization of existing resources which may indirectly further improve the performances of the screening policies.

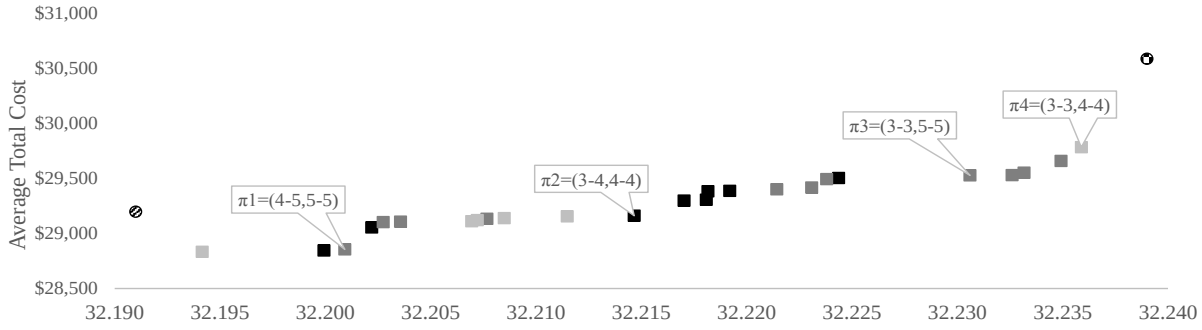
Similar observations are also made for male patients in Figure 3.22. Note that the policies, which improve the guideline with less or no cost for male patients, perform very closely to each other, thus, we select the policies with the highest TQALYs among them as π_1 policies. Although the TQALY improvements achieved by π_1 policies for male patients are more significant than those for female patients, the number of colonoscopies scheduled

for male patients is still not greater than the guideline.

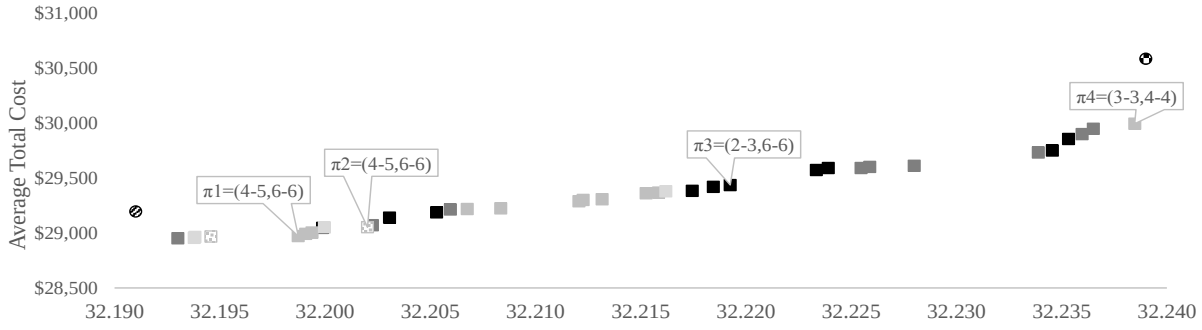
Figures 3.21 and 3.22 also imply that the stopping age for colonoscopy screening significantly affects the policy structure in terms of age-grouping. This is intuitive because when the stopping age increases there are more years that we can schedule colonoscopy screenings. The possibility of scheduling additional screening decreases the importance of variability in the policy structure regarding to the age threshold of screening frequency switches. More specifically, 32% (23%), 44% (43%), and 24% (34%) of the Pareto-efficient policies switch to a different screening frequency at age 60, 65, and 70 when the stopping age is 80 whereas 27% (23%), 25% (36%), 30% (36%), 10% (5%), and 5% (0%) of the Pareto-efficient policies switch to a different screening frequency at age 60, 65, 70, 75 and 80 when the stopping age is 85.

The 2-age group Pareto-efficiency analysis provides a strong evidence of improvements which is more pronounced for high-risk females than male patients. That is, the Pareto-efficient solutions are scattered smoothly in Figure 3.21 than the ones in Figure 3.22. Moreover, the gap between the optimal POMDP and the Pareto-efficient 2-age group policy is larger for male patients. This is mainly because the survival rate is greater for high-risk females, therefore, the benefits of colonoscopy screening is more pronounced for female patients.

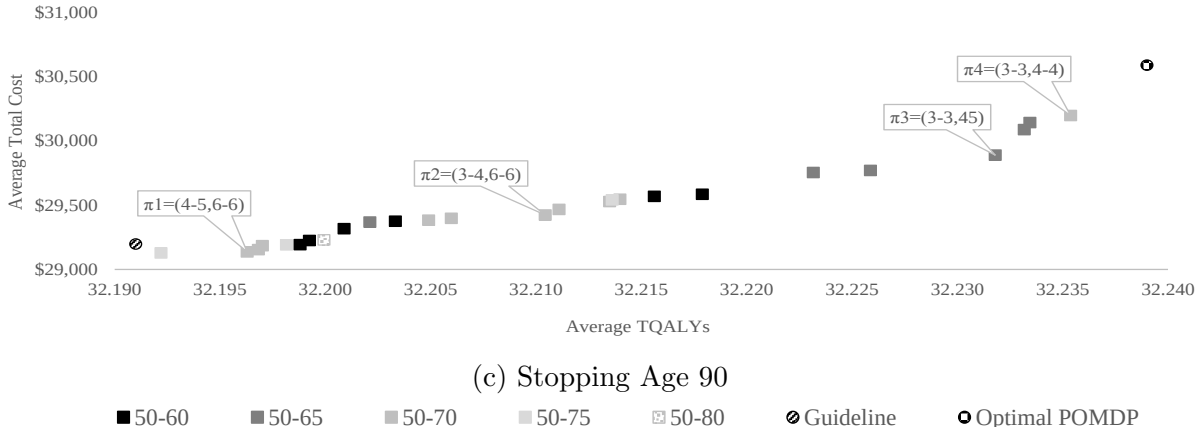
Another important observation from Figures 3.21 and 3.22 is about to what extend 1-switch policies may close the optimality gap between the simple-to-implement and optimal POMDP policies. Policies denoted as π_4 are the most cost-effective policies in closing the optimality gap. They suggest every 3-3 (2-2) years screening in the first age interval and 4-4 (2-3) years screening in the second age interval with a switch age of 70 (65), 60 (65), and 70 (65) for female patients (male) given the stopping ages of 80, 85, and 90, respectively. They provide significant improvements over the guideline and perform very closely to the optimal POMDP policy. When we compare their complexity and the performances with the guideline, we observe that the combined effect of the screening age and the frequency on the performance is more important than following different screening frequencies after polypectomy and the negative test result, i.e, f^i_j, $i, j \in \{1, 2\}$ values are equal or at most



(a) Stopping Age 80

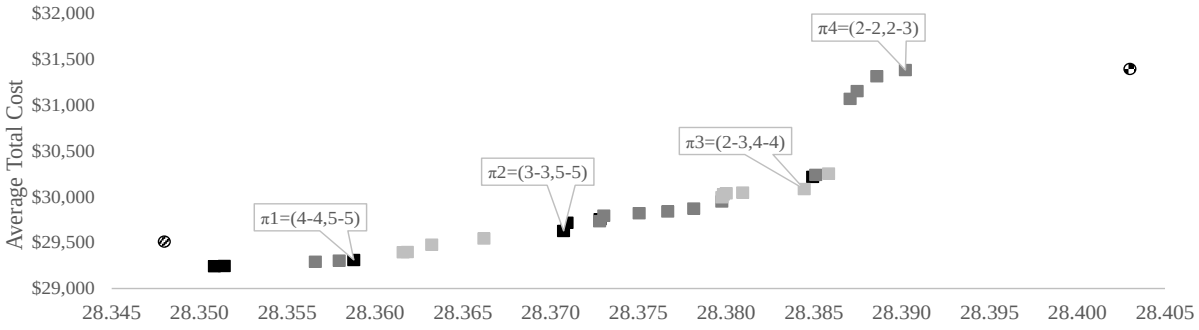


(b) Stopping Age 85

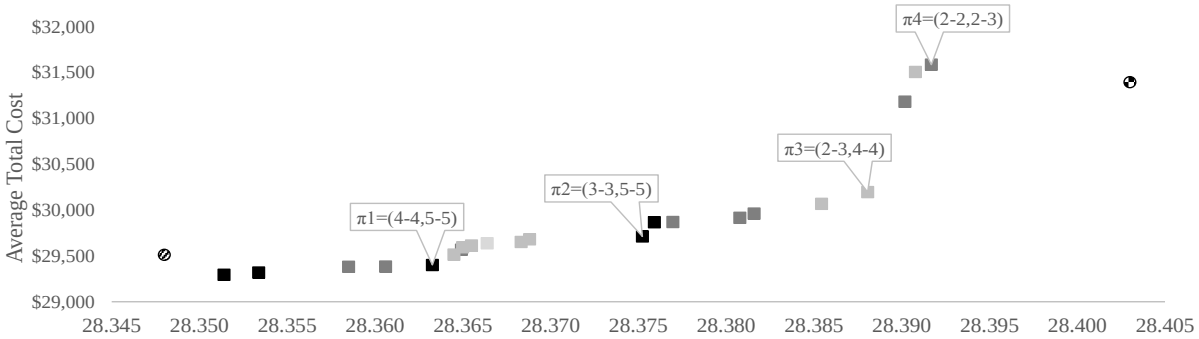


(c) Stopping Age 90

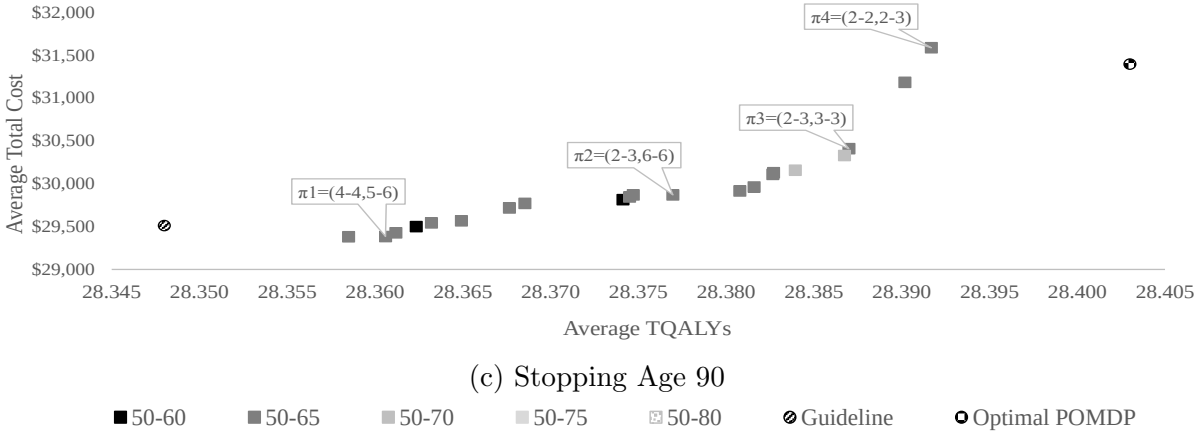
Figure 3.21: Pareto-frontier *age-dependent periodic screening policies* under 2-age group scenarios for high-risk female patient given stopping ages of 80, 85, and 90



(a) Stopping Age 80



(b) Stopping Age 85



(c) Stopping Age 90

Figure 3.22: Pareto-frontier *age-dependent periodic screening policies* under 2-age group scenarios for high-risk male patient given stopping ages of 80, 85, 90

one year apart for π_4 policies for high-risk patients.

Our analysis of 1-switch periodic-dynamic policies shows that there is still room for improvement in terms of TQALYs both for female and male patients. Thus, we next divide the age interval after age-50 into three sub-intervals. Figures 3.23 and 3.24 present our analysis on 3-age-group screening policies respectively for female and male patients. We observe that π_1 policies of 3-age group outperform the performances of both the guideline and the 2-age group π_1 policies with no or very small additional cost for female patients except for the stopping age 80. When the stopping age is 80, π_1 policy of 2-age group, (4-5, 5-5) with the switch age of 65, provides greater improvement than that of 3-age, (4-5, 5-5, 6-6) with the switch ages of 65, 75, at a smaller cost. This finding implies that following a simpler policy can sometimes provide better performance than a complex one depending on the stopping age.

For the stopping ages 85 and 90, the π_1 policies for 2- and 3-age-groups require approximately the same number of colonoscopy screenings; however, later group can achieve better TQALYs by changing the screening frequency from 4-5 years to 5-5 years between the ages 65 and 80 and following the same frequency of 6-6 years after age-80. That is, for 2-age group π_1 policies are (4-5, 5-5) with the switch age of 70 whereas they are (4-5, 5-5, 6-6) with the switch ages of 65 and 80 for 3-age group. This observation supports our claim that considering the combined effect of age and screening frequency in policy development, we can achieve better performances without increasing the number of colonoscopies performed.

Our observations about 3-age group π_1 policies for male patients are similar to those for female patients with slight differences. The 3-age group policy for male patients suggests (4-4, 5-5, 5-6) with the switch ages (60, 75) and (60, 80) given the stopping ages 80 and 85, 90, respectively. The π_1 policy with the stopping age of 85 provide the greatest improvement over the guideline among all stopping ages followed by the stopping ages 90, and 80. For male patients, we realize the larger effect of the stopping age on screening policy performances because the survival rates are lower for male patients.

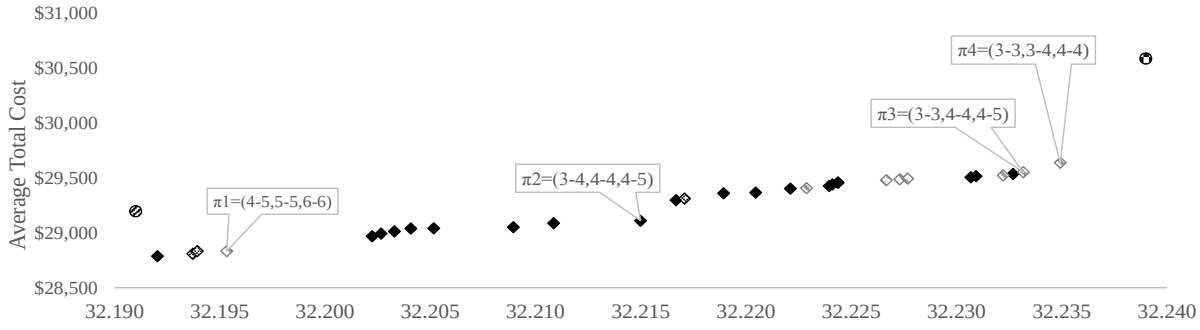
Similar to the 2-age group policies, the improvements are more significant for high-risk

female patients than male patients. Although the Pareto-efficiency curve is steeper for male patients than for female patients, the gap between the optimal POMDP and the best performing Pareto-efficient 3-age group policy is larger for male patients. We again see the effect of survival probability on the performances; however, dividing the age interval into three rather than two sub-intervals still increases the TQALYs achieved by the periodic policies for both females and males significantly.

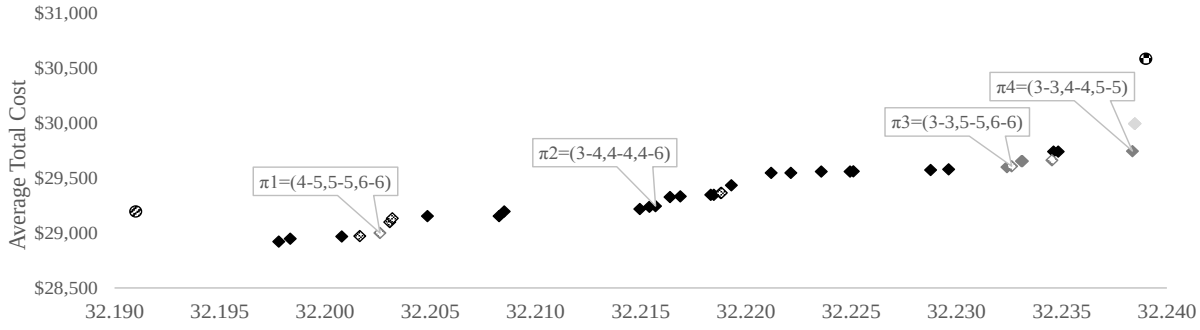
π_4 policies suggest every 3-3 years screening in the first age interval, 4-4 years screening in the second age interval and 5-5 years screening in the third age interval with the first switch age of 65 for all stopping ages, and the second switch ages of 75 and 80 for the stopping ages 80, and 85, 90, respectively, for the female patients. On the other hand, π_4 policies suggest male patients undergo colonoscopy screening with one year shorter screening intervals, i.e, every 2-2 ,3-3, and 4-4 years screening, respectively, in the first, second, and third age intervals with the switch ages of 65 and 75 for all stopping ages.

Figures 3.23 and 3.24 show that there are age-dependent periodic policies that perform very closely and yet less costly than the optimal POMDP policy in some cases, e.g., Figure 3.23b $\pi_4 = (3-3, 4-4, 5-5)$ achieve 99.99% TQALYs that the optimal POMDP achieves with 3% less expected cost for female patients given the stopping age of 85. Although the gap seems larger for the male patients the percentage TQALYs achieved is still greater than 99% of the optimal POMDP.

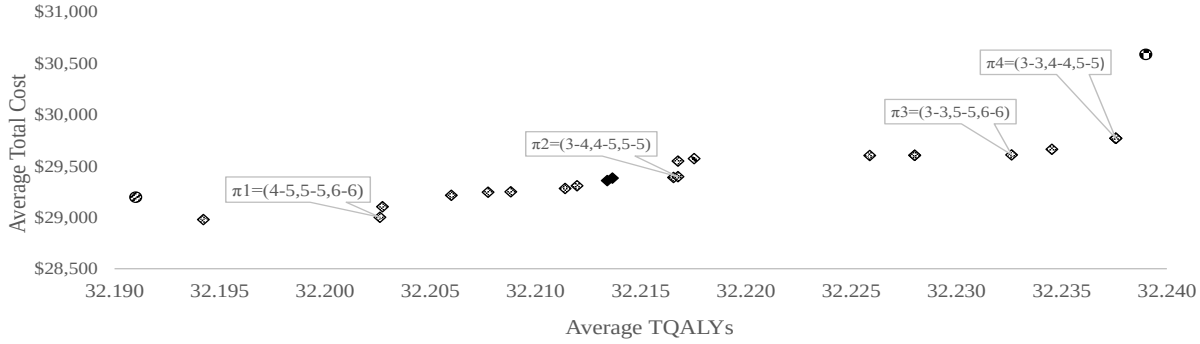
We present the Pareto-efficient dynamic-periodic policies considering both one- and two-switch policies under all possible considered stopping age and switch age combination in Figure 3.25 for high-risk patients. In this figure, there are a greater number of 3-age-group Pareto-efficient policies than that of 2-age-group for both genders on the final Pareto curve. This is mainly because by dividing the age interval into three, we provide a higher variety of age-dependent policies that can schedule the colonoscopy screenings over the intervals more flexibly. Figures 3.25a and 3.25b show that the greater proportion of Pareto-efficient policies stop screening at age-80 and age-85 respectively for female and male patients. There is no Pareto-efficient policy that terminates colonoscopy screenings at age-90 for any of the genders. This is intuitive when we consider the life expectancy



(a) Stopping Age 80



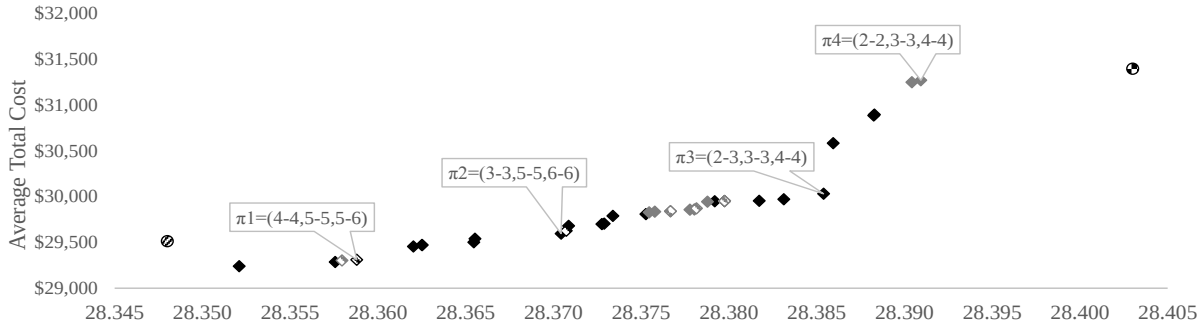
(b) Stopping Age 85



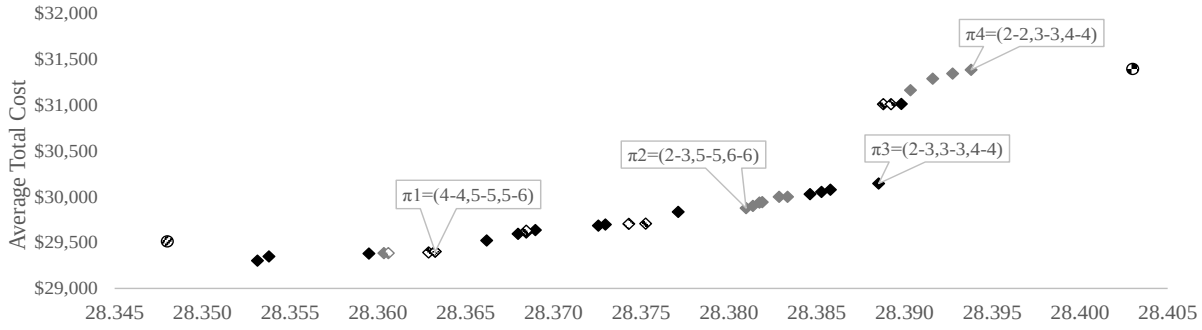
(c) Stopping Age 90

◆50-60-70 ◇50-60-75 ◈50-60-80 ◆50-65-75 ◈50-65-80 ◈50-70-80 ⊘Guideline ◔Optimal POMDP

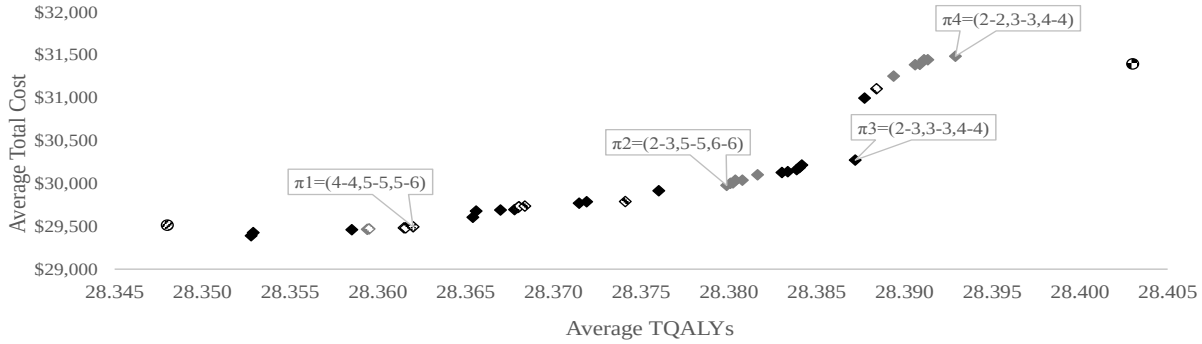
Figure 3.23: Pareto-frontier *age-dependent periodic screening policies* under 3-age group scenarios for high-risk female patient given stopping ages of 80, 85, 90



(a) Stopping Age 80



(b) Stopping Age 85



(c) Stopping Age 90

◆50-60-70 ◇50-60-75 ◈50-60-80 ◆50-65-75 ◈50-65-80 ◆50-70-80 ⊘Guideline ◔Optimal POMDP

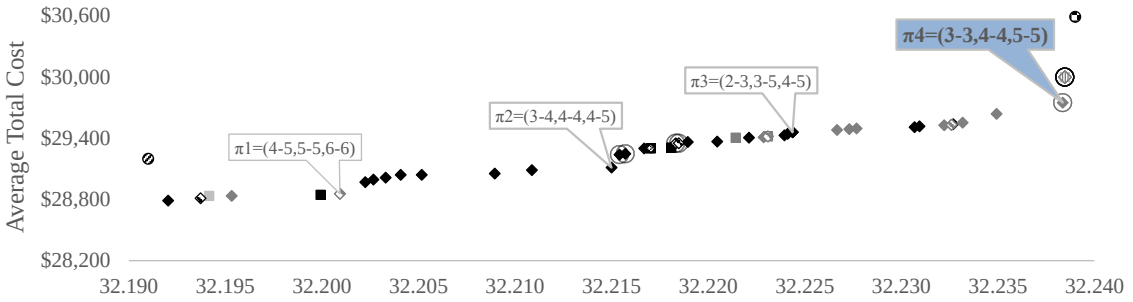
Figure 3.24: Pareto-frontier *age-dependent periodic screening policies* under 3-age group scenarios for high-risk male patient given stopping ages of 80, 85, 90

after ages 80 and 85, and the disutility of colonoscopy screening. For both genders, the most observed switch ages are 60 and 70 respectively for the first and second switch ages. This is also reasonable when one think about the probabilities of developing polyp and CRC. Note that the policy structures are quite similar for all π_i, $i = 1, \cdots, 4$ in Figures 3.25a and 3.25b. That is, the screening frequency for high-risk patients is usually around 3-5 years. Although these intervals are consistent with those suggested by the current guidelines, the policies proposed in our analysis smartly using these intervals in a partially dynamic manner to further improve health outcomes compared to those in the literature.

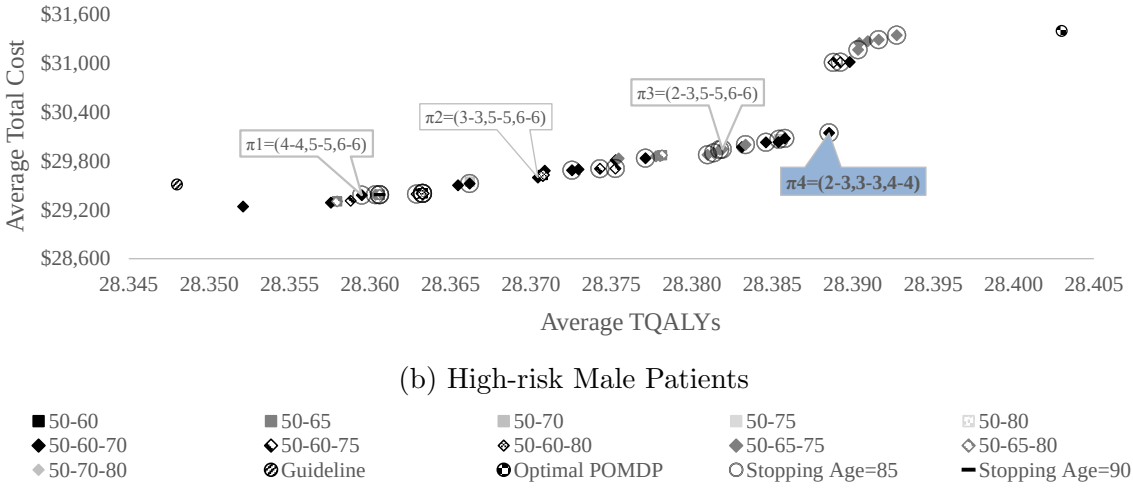
We give the details of promising Pareto-efficient policies and the periodic policies suggesting 1 or 2 more colonoscopy screenings than the guidelines in Table 3.10. We choose the promising Pareto-efficient policies based on their ICER values ($< \$50,000$ is accaptable) among the Pareto-efficient policies that suggest at most 2 more colonoscopy on average than the guidelines. When the alternatives satisfying the aforementioned condition perform closely to each other, we choose the policy with 2-age group (if any) to reduce the complexity in practice.

For female and male patients, the promising Pareto-efficient policies outperform the guidelines in terms of TQALYs yet result in reasonable or no additional expected costs. When we compare the periodic policies and the optimal POMDP, we see that the promising periodic-static policies can attains most of the TQALYs improvements achieved by the optimal policies with less expected costs. If we also consider the additional cost of adopting the optimal policy in practice, i.e., the cost of complications and mistakes due to applying a completely dynamic policy based on every-changing risk-thresholds, we can conclude that the proposed promising policies can be preferred over the optimal POMDP policies in terms of their costs. However, the optimal POMDP policies' benefits should be reviewed carefully. They provide improvements in terms of not only TQALYs but also CRC risk and mortality. When we consider these metrics for the promising policies, we observe that they perform a little shy of the optimal POMDP. Therefore, the optimal POMDP policies are still the benchmark that we should aim at achieving their performances from all other perspectives including CRC risk and mortality. Therefore, we conclude that promising dynamic-periodic policies with 2-switch points may be sufficient to close the optimality

gap in TQALYs, we still need more flexible simpler-to-implement policy structures to close the gap with optimal POMDP policy in terms of other secondary metrics such as CRC risk and CRC mortality.



(a) High-risk Female Patients



(b) High-risk Male Patients

Figure 3.25: Pareto-frontier *age-dependent periodic screening policies* for high-risk patients given stopping ages of 80, 85, 90

Table 3.10: Comparison of the US guidelines, POMDP policies derived in Erenay et al. (2014), and Pareto-efficient Dynamic Periodic Policies for high-risk patients

		Guideline	π_1	π_2	π_3	π_4	+1 Col	+2 Col	POMDP*
Males									
	Total QALYs	28.348	28.360	28.371	28.381	28.390	28.375		28.403
	No. of colonoscopies	6.43	6.56	7.02	7.87	8.47	7.43		10.58
	Lifetime CRC risk	3.93%	3.76%	3.81%	3.27%	2.99%	3.45%	π_4	2.59%
	CRC mortality	1.17%	1.15%	1.31%	1.05%	0.90%	1.09%		0.75%
	Total Cost ($)	29,513	29,380	29,596	29,877	30,145	29,707		31,397
	Screening Interval	4.50	4.36	3.71	3.48	3.29	3.73		2.54
Females									
	Total QALYs	32.191	32.192	32.215	32.224	32.238	32.233		32.239
	No. of colonoscopies	7.00	6.05	7.13	7.65	8.58	8		10.18
	Lifetime CRC risk	3.56%	4.17%	3.62%	3.62%	3.01%	3.46%	NA	2.66%
	CRC mortality	1.16%	1.47%	1.25%	1.29%	0.96%	1.23%		0.76%
	Total Cost ($)	29,198	28,788	29,111	29,428	29,748	29,552		30,589
	Screening Interval	4.50	4.77	4.02	3.63	3.52	3.44		2.97

* Source: Erenay et al. (2014)

NA refers to the case that no Pareto-efficient solution satisfies the definition of the policy.

3.4.4.2 The Best Performing Age-dependent Periodic Policies for Low-Risk Patients

In this section, we analyze the age-dependent periodic policies for low-risk patients. Recall that a *low-risk patient* is an asymptomatic average-risk patient with no personal or family history of CRC (Winawer et al., 2003). Guidelines recommend that low-risk patients undergo colonoscopy screening in every 10 years starting from age-50 (USPSTF, 2008). Erenay et al. (2014) show that the guideline's performance for low-risk patients can be improved by following a personalized colonoscopy screening policy, which schedules colonoscopies with dynamically varying intervals based on age and other factors such as gender and previous screening findings. Therefore, like we do for high-risk patients in Section 3.4.4.1, we perform a similar Pareto-efficiency analysis on age-dependent periodic policies for low-risk patients.

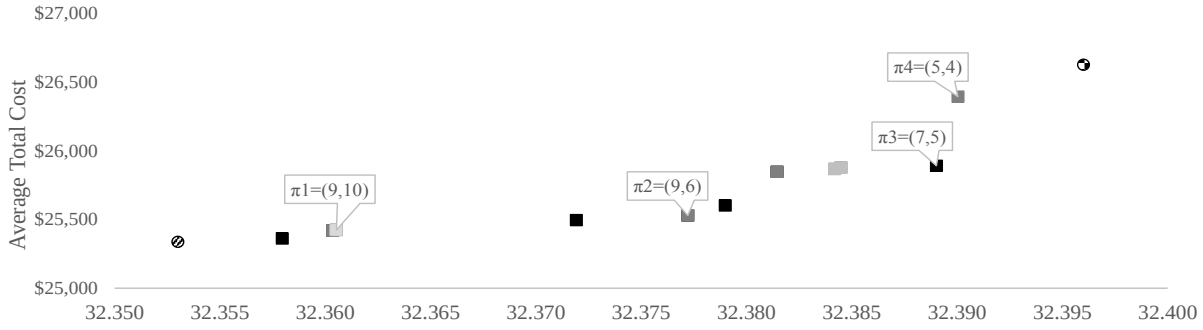
Let the screening frequencies that define age-dependent periodic policies for low-risk patients be denoted by f_i for the age intervals $i = 1, 2, 3$. We assume that f_i, $\forall i$ can take any value from the set that $\{3, 4, \cdots, 10\}$. These frequencies are determined because there is no need to follow (i) a more aggressive screening policy for low-risk patients than that is suggested by the guidelines for high-risk patients, or (ii) a less aggressive screening policy than the policy suggested by the guidelines for low-risk patients (USPSTF, 2008). We also assume that the screening frequencies are not necessarily in an increasing order based on their corresponding screening intervals as for high-risk patients, e.g., both $(3, 5, 7)$ and $(7, 3, 5)$ can be followed for low-risk patients. We make this assumption because we want to observe the true relationship between the age, or equivalently risk of polyp and CRC development, and screening frequency for low-risk patients. Although this increases the number of screening frequency combinations that we need to evaluate for low-risk patients, the number of combinations for low-risk patients is still less than that for high-risk patients. This is because, in the case of high-risk patients, the screening policies for each age-group specify two screening intervals (after a polypectomy v.s. after a negative colonoscopy), while the policies for low-risk patients specify a single screening interval for each age-group. Note that, for low-risk patients, the screening frequency also changes after detection of a polyp or a cancer and the patient moves to the high-risk or post-cancer patient group.

As noted, a low-risk patient can develop a polyp and become a high-risk patient. In this case, the screening policy followed in the high-risk level for this formerly low-risk patient is also important on the aforementioned policies' performances in expected TQALYs, costs, lifetime CRC risk, etc. In our simulation analysis, we assume when a low-risk patient turns into a high-risk patient after a polyp detection/removal, the promising age-dependent screening policy, that is shown perform closest to the optimal one in Section 3.4.4.1, is followed afterwards. That is, the screening policy for high-risk level is (3-3, 4-4, 5-5) with the switch ages of 65 and 75 if the formerly low-risk patient is female, and (2-3, 3-3, 4-4) with the switch ages of 60 and 70 if the formerly low-risk patient is male (refer to the π_4 policies in Figure 3.25).

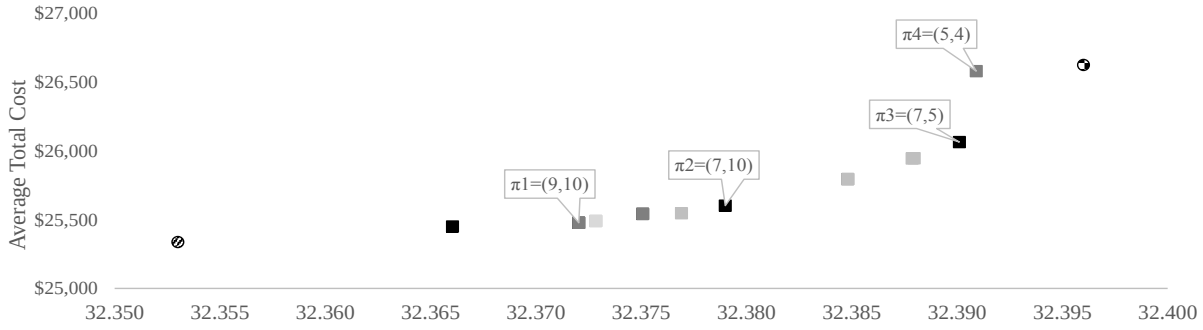
The performances of the Pareto-efficient 2-age-group (one-switch) dynamic-periodic

policies performing better than the guidelines in terms of TQALYs for low-risk female and male patients are shown in Figures 3.26 and 3.27, respectively. The improvements over the guideline performances achieved by the 2-age-group dynamic-periodic screening policies are significant in both figures. π_1 policies have similar structures for each gender with all stopping ages. That is, they suggest every 9-year screening before ages 70, 60, and 75, respectively, with the stopping ages 75, 80, and 85, and 10-year screening afterwards for female patients whereas it is every 10-year until age 60 and 7-year afterwards for male patients regardless of the stopping age. As these policies show, the screening frequency is higher for male patients than female patients because the probability of developing CRC is smaller for female patients in both low- and high-risk levels than for male patients (Howlader et al., 2012). Although the screening frequency is higher for female patients, the average number of colonoscopy screenings performed is very similar for both genders with all stopping ages (i.e., it is around 4.7 colonoscopies). This is also expected since the life expectancy for female patients is higher than that for male patients which results in a longer time period over which colonoscopy screenings are scheduled. The difference in life expectancy for males and females also affects the change in the switch ages with the stopping ages. The switch age is 60 for male patients regardless of the stopping age because the average life expectancy is not significantly different beyond ages 75, 80, and 85. However, it changes with the stopping age for female patients as 70, 60, and 75, respectively, with respect to the stopping ages of 75, 80, and 85, which follows the trend in the ratio between the life expectancy and the cumulative probability of developing a polyp/CRC. That is, the increase in the cumulative probability of developing a polyp or CRC is more effective in screening decisions when the life expectancy is higher.

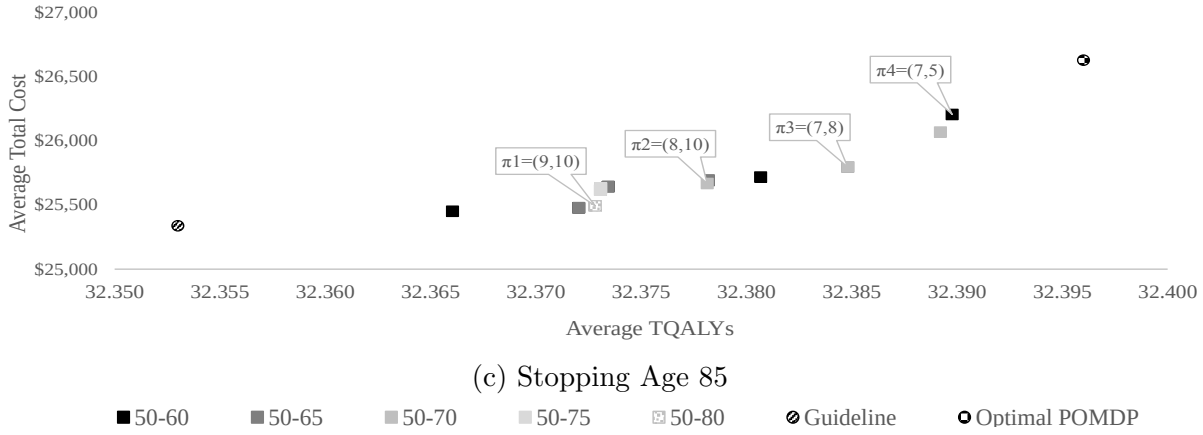
Remember that the π_1 policies for high-risk patients improve the guidelines with no cost for both genders and almost all stopping ages; however, there is no π_1 policy for low-risk patients that improve the guideline with no or negligibly small cost. We believe that this results from the trade-off between the benefit and cost of scheduling additional colonoscopy screenings. The dynamic-periodic policies for both genders schedule approximately one more colonoscopy screenings than the guideline; however, the benefit of this one additional screening is not pronounced enough to be observed in the expected costs. This observation



(a) Stopping Age 75

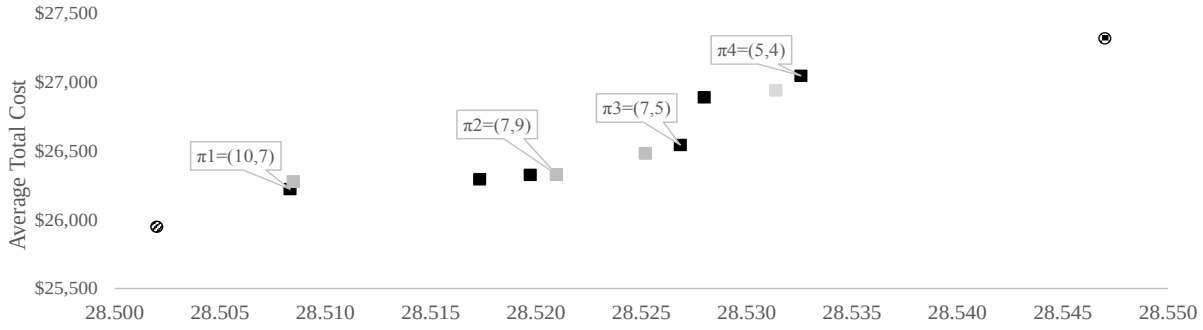


(b) Stopping Age 80

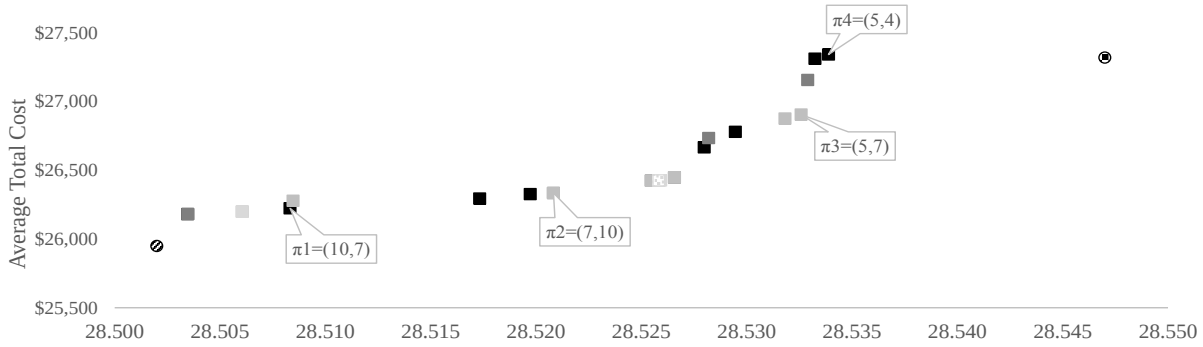


(c) Stopping Age 85

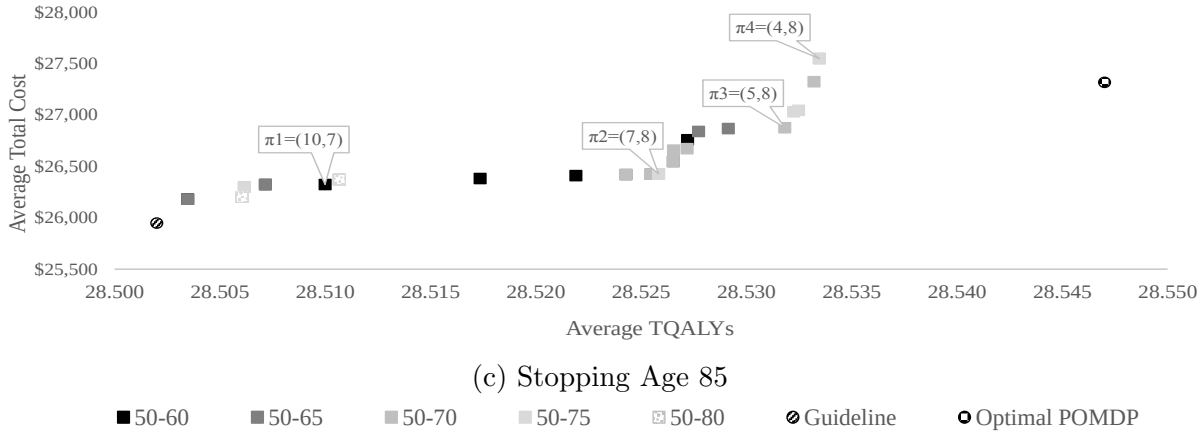
Figure 3.26: Pareto-frontier *age-dependent dynamic periodic screening policies* under 2-age group scenarios for low-risk female patient given stopping ages of 75, 80, 85



(a) Stopping Age 75



(b) Stopping Age 80



(c) Stopping Age 85

Figure 3.27: Pareto-frontier *age-dependent dynamic periodic screening policies* under 2-age group scenarios for low-risk male patient given stopping ages of 75, 80, 85

is also in line with the findings of Erenay et al. (2014) that the optimal POMDP policies provide greater improvements for high-risk patients than low-risk patients in expense of lower costs. This is mainly because of the fact that the high-risk patients are more prone to develop a polyp or cancer later on their life than the low-risk patients, thus, the benefit of colonoscopy screening is more pronounced for them. Additionally, the screening frequencies in the first and second age intervals are not quite different (i.e., 9 vs. 10 for females and 10 vs. 7 for males). This can be also another reason why we do not see the effect of considering different age groups enough on the performance improvements.

π_4 policies suggest every 5-year screening in the first age interval and 4-year screening in the second age interval with a switch age of 65 for female patients and 60 for male patients given the stopping ages of 75 and 80, respectively. When the stopping age is 85, the π_4 policies are screen every 7- and 5-year, every 4- and 8-year with the switch ages of 60 and 75, respectively for female and male patients. Although π_4 policies achieve significant improvements over the guideline, their performances are not close enough to the performances of the optimal POMDP policies. For female patients they perform closer to the optimal POMDP policies than the policies for male patient do; however, the gap is still significant. Therefore, we proceed with analyzing 3-age group dynamic-periodic policies to fill this gap for low-risk patients.

Figures 3.28 and 3.29 present the performances of the 3-age group dynamic-periodic screening policies, respectively, for low-risk female and male patients. The performances of the selected 3-age group Pareto-efficient policies are very similar to those of 2-age group policies. The π_1 policies suggest every 10-year screening between the ages 50-60 for both female and male patients and 9-year and 7-year screenings between the ages 60-70, respectively, for female and male patients regardless of the stopping age. The screening frequencies from the second switch age until the stopping age are 3, 7, and 8, respectively, with the stopping ages 75, 80, and 85 for female patients, and 9 for male patients with all stopping ages. Although the screening policy is the same with all stopping ages for male patients, the performance is better for stopping age 85 since more colonoscopy screenings can be performed over a longer time horizon (i.e., 4.68 for the stopping ages 75 and 80 whereas it is 4.91 for 85). Here recall that the screening policy remains the same with

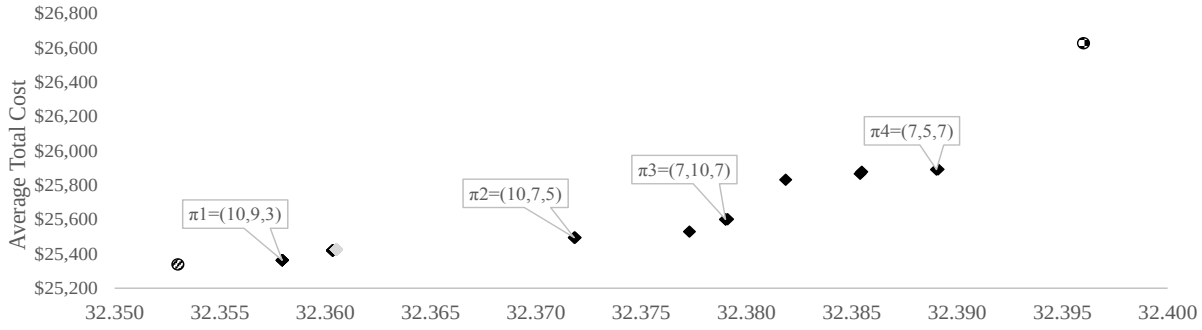
all stopping ages for male patients due to the low life expectancy for male patients as explained before. π_1 policies suggest increasing the screening frequency in the last age interval until the stopping age for the female patients and keeping it at every 9-year for the male patients. However, they actually suggest not performing any screening after age-78 by forcing the policy to postpone the screenings in the last interval after the stopping age. This is an interesting result because although our analysis enable policies to consider 3 different age interval and schedule aggressive colonoscopy screenings in these intervals, the Pareto-efficiency analysis shows if we want a policy that operates at a similar cost as the guidelines, then, considering only 2 age groups can be enough.

π_4 policies recommend patients undergo colonoscopy screening every 7- (5-), 5- (4-), and 7-year (3-year), respectively, in the first, second and third age intervals for female (male) patients when the screenings are terminated at age-75. Whereas the dynamic-periodic screening frequencies are 5, 4, and 5 for both genders with the stopping age of 80. Although this remains the same for male patients when the stopping age increases to 85, the screening frequency increases to 9 years in the last age interval for the female patients. This may seem conflicting with the fact that the decision on the screening frequency is mainly based on the risk of developing CRC (i.e., the screenings are more frequent in the age interval if the risk of developing CRC is higher). However, this change from the screening policy 5-4-5 to 5-4-9 is to prevent any additional disutility by keeping the number of colonoscopy screening scheduled the same, e.g., it is around 6.842 for both policies with the stopping ages 80 and 85. The switch ages are 60 and 70 for males at all stopping ages. However, it is 65 and 75 at stopping ages 75 and 80, and 60 and 75 at stopping age 85 for female patients. Figures 3.28 and 3.29 show that π_4 policies significantly outperform the guideline policies and provide most of the TQALYs that the optimal POMDP policies attain at a lower cost.

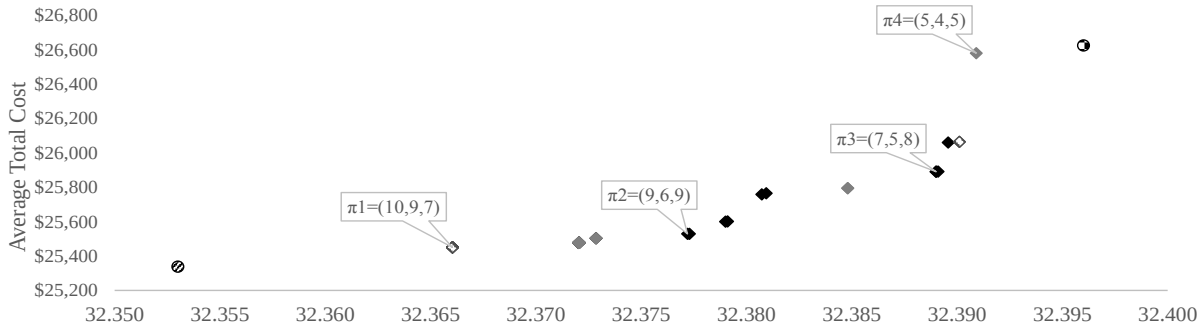
The Pareto-efficient dynamic-periodic policies for low-risk patients provide promising performances and have very simple screening patterns compared to the optimal POMDP policies. Moreover, the performance difference in the TQALYs achieved by the dynamic-periodic policies for low-risk patients are not significant between female and male patients compared to the difference between high-risk female and male patients. These all together

imply that the dynamic-periodic policies have important potential to be followed in practice by the clinicians which is one major objective of this study.

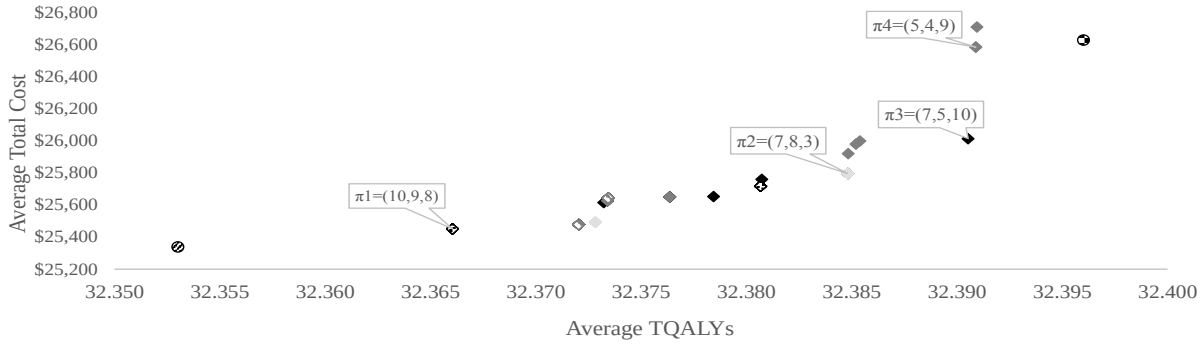
To sum up our findings for low-risk patients, we provide Figure 3.30 that presents Pareto-efficient age-dependent Dynamic-periodic policies with the stopping ages of 75, 80, and 85 for low-risk patients. The π_4 policies in Figures 3.30a and 3.30b provide noticeably high performances in terms of TQALYs compared to the guideline policies for low-risk female and male patients, respectively. These policies suggest more frequent screening between ages 60 and 70 (i.e., (5,4,6) vs. (7,5,10)) than the frequencies at other ages. The screening frequencies are higher for male patients than female patients because of their higher risk of developing CRC. On the other hand, the stopping age to terminate colonoscopy screenings is higher for female patients because of their higher life expectancy (i.e., 90 vs. 85). Although the TQALYs of π_4 policy is closer to that of the optimal POMDP for female patients than male patients, the π_4 policy for male patients performs very close to the optimal POMDP in other performance measures, e.g., life time CRC risks are 2.02% with the π_4 and 2.01% with the POMDP for male patients where these values are, respectively, 2.17% and 1.95% for female patients. This is because the π_4 policy suggests more frequent screening for male patients than females compared to the guidelines (note that they both are less frequent than the optimal POMDP); however, the improvements in TQALYs is not that close to the optimal solution for male patients as opposed to the improvements in other performance metrics such as CRC risk. This is again because the fact that the benefit of more frequent screening in terms of TQALYs is not pronounced enough due to the lower life expectancy of male patient, i.e., the TQALYs improvement via reducing CRC risk to below 2.02% with periodic screening does not worth the additional colonoscopy disutility, which should can be further reduced by more dynamic colonoscopy scheduling. The majority of Pareto-efficient policies suggest dividing the age interval after age-50 into three age groups with the switch ages of 60 and 70 in Figures 3.30a and 3.30b. For female patients, there are more policies with the stopping age 85 in the Pareto-frontier; whereas, the stopping age 90 appears more frequently among the Pareto-efficient policies for males compared to females. The number of 3-age group policies are greater than that of 2-age group for both genders.



(a) Stopping Age 75

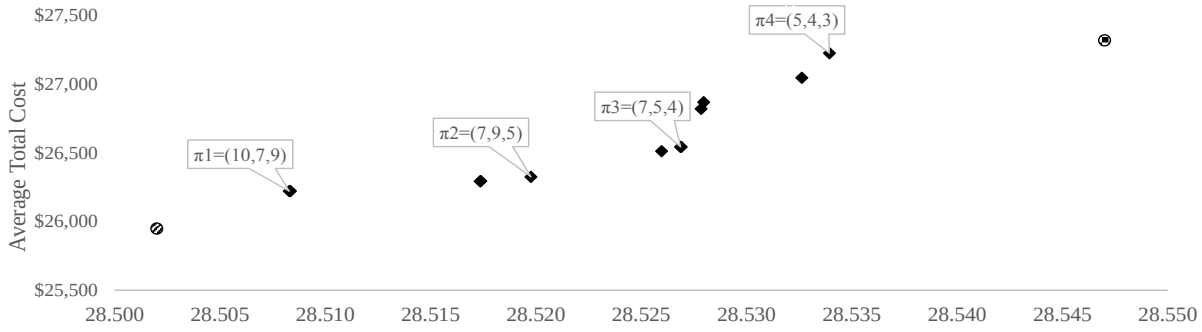


(b) Stopping Age 80

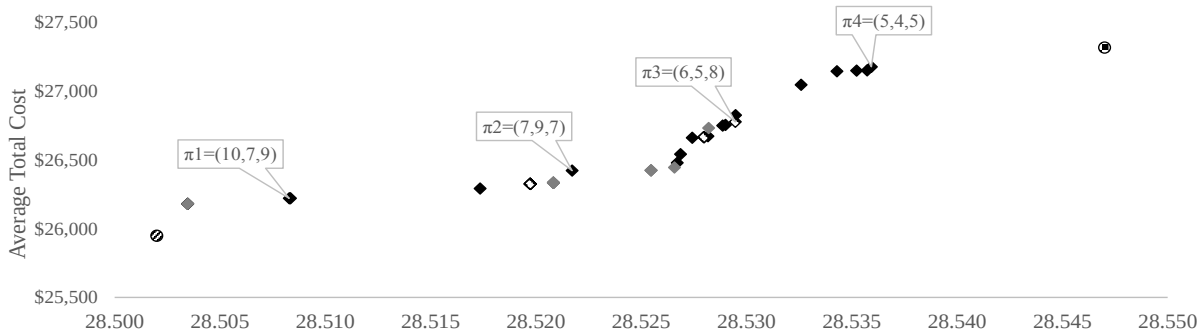


(c) Stopping Age 85

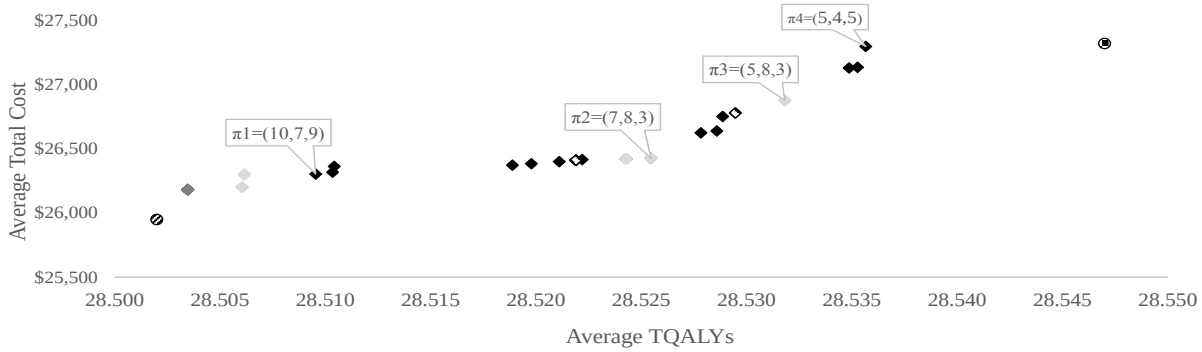
Figure 3.28: Pareto-frontier *age-dependent dynamic periodic screening policies* under 3-age group scenarios for low-risk female patient given stopping ages of 75, 80, 85



(a) Stopping Age 75



(b) Stopping Age 80



(c) Stopping Age 85

Figure 3.29: Pareto-frontier *age-dependent dynamic periodic screening policies* under 3-age group scenarios for low-risk male patient given stopping ages of 75, 80, 85

When determining the best performing age-dependent Dynamic-periodic policy for low-risk patients, we consider π_4 policies in Figures 3.30a and 3.30b and the Pareto-efficient policies suggesting one and two more colonoscopy screenings than the guidelines in Table 3.12. As for high-risk patients, we limit the number of additional colonoscopy screenings performed by the dynamic-periodic policies compared to the guidelines with at most 2 additional screenings. The best performing Pareto-efficient policy is defined as the policy that provides the highest TQALYs while satisfying the condition on the number of colonoscopy screenings. Note that π_4 policies provide the highest TQALYs for both genders; however, the π_4 policy for male patients schedule 2.68 more colonoscopy screenings than the guidelines on average. That's why we suggest clinicians to follow the π_4 policy for low-risk female patients and the Pareto-efficient policy that suggest two additional colonoscopy (*+2 Col*, $(7, 5, 8)$) for low-risk male patients.

Based on our analysis in Sections 3.4.4.1 and 3.4.4.2, we determine the best performing age-dependent periodic policies for low- and high-risk patients as given in Table 3.11. The policies are (7,5,10) and (5,4,6) for low-risk females and males with the stopping ages of 85 and 80 and the age interval of 50-60-70, respectively (see Figure 3.30). We follow the screening policies (3-3,4-4,5-5) and (2-3,3-3,4-4) for high-risk females and males with the stopping age of 85 and the age intervals of 50-65-75 and 50-60-70, respectively, as in Figure 3.25.

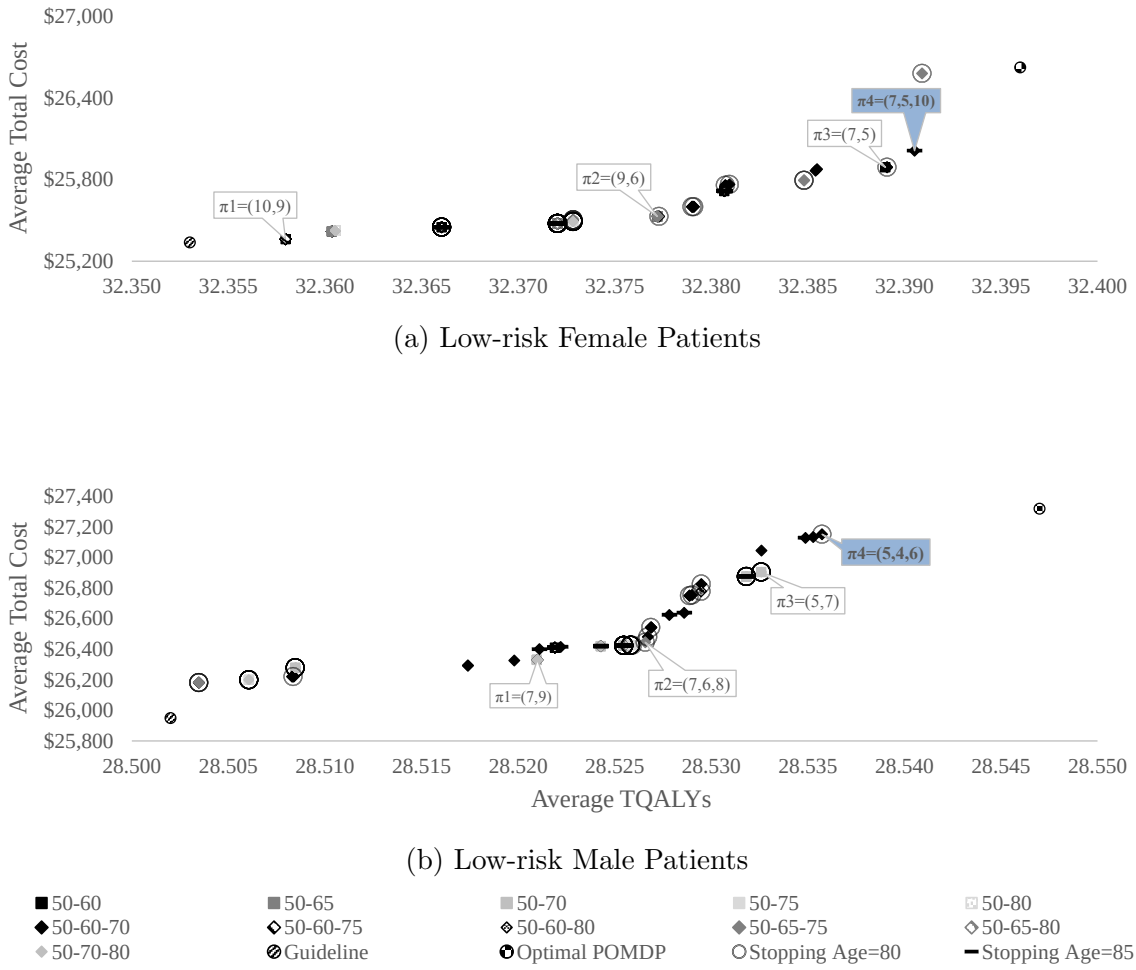
Table 3.11: The best performing age-dependent policies for low- and high-risk patients

	Policy, Stopping Age, Age intervals	
	Females	Males
Low-risk patients	(7,5,10), 85, 50-60-70	(5,4,6), 80, 50-60-70
High-risk patients	(3-3,4-4,5-5), 85, 50-65-75	(2-3,3-3,4-4), 85, 50-60-70

Table 3.12: Comparison of the US guidelines, POMDP policies derived in Erenay et al. (2014), and Pareto-efficient Dynamic-Periodic Policies for low-risk patients

	Guideline	π_1	π_2	π_3	π_4	+1 Col	+2 Col	POMDP*
Males								
Low-risk patients								
Total QALYs	28.502	28.521	28.527	28.533	28.536	28.509	28.529	28.547
No. of colonoscopies	3.72	4.90	5.31	6.18	6.68	4.74	5.69	6.96
Lifetime CRC risk	3.41%	2.74%	2.41%	2.19%	2.02%	2.76%	2.25%	2.01%
CRC mortality	1.24%	0.97%	0.78%	0.73%	0.69%	0.95%	0.73%	0.67%
Total Cost ($)	25,949	26,330	26,446	26,905	27,153	26,278	26,638	27,319
Screening Interval	8.45	6.06	5.93	4.71	4.33	6.75	5.6	4.12
High-risk patients								
Females								
Low-risk patients								
Total QALYs	32.353	32.358	32.377	32.389	32.391	32.381		32.396
No. of colonoscopies	4.00	4.14	4.76	5.57	5.87	5.22		7.05
Lifetime CRC risk	3.16%	3.12%	2.66%	2.38%	2.17%	2.44%	π_4	1.95%
CRC mortality	1.24%	1.23%	0.99%	0.88%	0.75%	0.87%		0.65%
Total Cost ($)	25,338	25,363	25,529	25,891	26,013	25,717		26,627
Screening Interval	8.37	8.11	7.13	5.65	5.85	6.84		4.52
High-risk patients								

* Source: Erenay et al. (2014)



(a) Low-risk Female Patients

(b) Low-risk Male Patients

Figure 3.30: Pareto-frontier *age-dependent dynamic periodic screening policies* for low-risk patients given stopping ages of 75, 80, 85

3.4.5 Robustness of The Periodic Policies

We conduct a robustness analysis with varying (i) costs, (ii) sensitivity of colonoscopy screening, and (iii) discount rates to see how the selected best performing periodic policies in Table 3.11 perform under different scenario settings.

Table 3.13 presents the average costs and the ICER values of the best performing periodic policies under low-, base-, high-cost scenario settings, respectively denoted by cost multipliers of 1, 2, and 3. In our ICER calculation, we assume the guideline policy as a base, thus, the cost-effectiveness discussed in this section is in comparison with the guideline policies. The best performing policies are cost-effective under all scenario settings and for both female and male patients with one exception. The best performing policy does not remain cost-effective when we assume high-cost for low-risk male patients. This observation is very important because it shows that the best performing periodic policy may result in a performance which is not as good as its performance with the base case parameters. Therefore, the periodic policies should be cautiously selected considering the system characteristic when adopting them in practice.

Table 3.13: Robustness analysis with cost parameters for low- and high-risk patients

	Cost Multiplier	Avg. Treatment Cost	Avg. Screening Cost	Avg. Total Cost	ICER
Female	1.00	$9,964	$1,297	$10,744	$5,577
	2.00	$20,880	$5,100	$26,013	$20,186
	3.00	$32,747	$11,410	$45,113	$43,376
Male	1.00	$10,122	$1,459	$11,474	$6,820
	2.00	$21,055	$5,740	$27,153	$24,786
	3.00	$32,801	$12,845	$46,779	$53,696

(a) Low-risk Patients

	Cost Multiplier	Avg. Treatment Cost	Avg. Screening Cost	Avg. Total Cost	ICER
Female	1.00	$10,224	$1,938	$11,682	$4,137
	2.00	$21,968	$7,613	$29,748	$15,498
	3.00	$35,232	$17,023	$53,511	$33,923
Male	1.00	$10,384	$1,904	$12,221	$4,322
	2.00	$22,159	$7,481	$30,145	$15,877
	3.00	$35,325	$16,731	$53,518	$34,559

(b) High-risk Patients

Our findings from the robustness analysis with sensitivity of colonoscopy screening are

presented in Table 3.14. When the colonoscopy screening increases either for polyps or cancer, the best performing policies perform greater number of screenings for both risk levels and genders. The increase in the number of colonoscopies performed is greater (i) for low-risk patients than high-risk patients and (ii) for females than males. (i) is because the best performing policies do not perform close to the optimal for low-risk patients, thus, there is still room for improvement by utilizing the colonoscopy screenings further. That is, when the screening sensitivity is high, the effect of disutility that prevents having more aggressive periodic policies on the Pareto-efficiency set for low-risk patients decreases. On the other hand, (ii) is due to females being more prompt to developing CRC.

Table 3.14: Robustness analysis with sensitivity of colonoscopy for low- and high-risk patients

	Sensitivity of Colonoscopy	Avg. TQALYs	Avg. Number of Screening	CRC Risk	CRC Mortality	Avg. Lifetime	Avg. Screening Interval	Avg. Treatment Cost	Avg. Screening Cost	Avg. Total Cost
Female	80,85	32.3769	5.8442	0.0233	0.0084	32.4129	5.8698	$20,936	$5,070	$26,138
	80,90	32.3793	5.8451	0.0233	0.0082	32.4153	5.8691	$20,978	$5,071	$26,158
	85,90	32.3905	5.8677	0.0217	0.0075	32.4268	5.8522	$20,880	$5,100	$26,013
	85,95	32.3913	5.8686	0.0217	0.0073	32.4275	5.8513	$20,914	$5,101	$26,021
	90,95	32.4022	5.8890	0.0202	0.0068	32.4384	5.8365	$20,817	$5,127	$25,877
Male	80,85	28.5203	6.6619	0.0220	0.0077	28.6280	4.3328	$21,115	$5,716	$27,296
	80,90	28.5222	6.6626	0.0220	0.0075	28.6298	4.3324	$21,153	$5,717	$27,310
	85,90	28.5357	6.6810	0.0202	0.0069	28.6435	4.3250	$21,055	$5,740	$27,153
	85,95	28.5363	6.6816	0.0202	0.0068	28.6441	4.3246	$21,094	$5,741	$27,171
	90,95	28.5498	6.6985	0.0185	0.0062	28.6574	4.3182	$20,971	$5,762	$27,006

(a) Low-risk Patients

	Sensitivity of Colonoscopy	Avg. TQALYs	Avg. Number of Screening	CRC Risk	CRC Mortality	Avg. Lifetime	Avg. Screening Interval	Avg. Treatment Cost	Avg. Screening Cost	Avg. Total Cost
Female	80,85	32.2160	8.5737	0.0325	0.0104	32.3272	3.5218	$22,103	$7,595	$30,005
	80,90	32.2194	8.5751	0.0325	0.0102	32.3306	3.5215	$22,166	$7,596	$30,036
	85,90	32.2384	8.5784	0.0301	0.0096	32.3490	3.5234	$21,968	$7,613	$29,748
	85,95	32.2398	8.5794	0.0301	0.0095	32.3504	3.5230	$22,024	$7,614	$29,764
	90,95	32.2539	8.5812	0.0280	0.0087	32.3641	3.5249	$21,872	$7,628	$29,547
Male	80,85	28.3642	8.4597	0.0327	0.0102	28.5307	3.2867	$22,270	$7,464	$30,401
	80,90	28.3675	8.4612	0.0327	0.0099	28.5339	3.2862	$22,353	$7,465	$30,437
	85,90	28.3885	8.4646	0.0299	0.0090	28.5546	3.2878	$22,159	$7,481	$30,145
	85,95	28.3892	8.4654	0.0299	0.0088	28.5552	3.2874	$22,210	$7,482	$30,163
	90,95	28.4076	8.4688	0.0274	0.0081	28.5734	3.2887	$22,029	$7,497	$29,913

(b) High-risk Patients

The robustness analysis on the discount rates can also be used to analyze how ICER

changes; however, since the change is proportional the best performing policies remain cost-effective compared to the guidelines. However, changes in discount rates may result in a change in the structure of the best performing policies similar to the change that is observed in our sensitivity analysis for static periodic policies (see Figures 3.17-3.20).

Table 3.15: Robustness analysis with discount rates on TQALYs and costs for low- and high-risk patients

	Discount Rate	Avg. TQALYs	Avg. Lifetime	Avg. Treatment Cost	Avg. Screening Cost	Avg. Total Cost
Female	1,1	32.3905	32.4268	$20,880	$5,100	$26,013
	1,0.97	32.3905	32.4268	$8,877	$3,334	$12,440
	0.99,0.97	27.2721	27.3274	$8,877	$3,334	$12,440
	0.98,0.97	23.2398	23.3026	$8,877	$3,334	$12,440
	0.98,0.95	23.2398	23.3026	$5,375	$2,636	$8,218
	0.97,0.95	20.0321	20.0961	$5,375	$2,636	$8,218
Male	1,1	28.5357	28.6435	$21,055	$5,740	$27,153
	1,0.97	28.5357	28.6435	$9,828	$3,917	$13,999
	0.99,0.97	24.4004	24.5023	$9,828	$3,917	$13,999
	0.98,0.97	21.0814	21.1758	$9,828	$3,917	$13,999
	0.98,0.95	21.0814	21.1758	$6,305	$3,155	$9,651
	0.97,0.95	18.3945	18.4810	$6,305	$3,155	$9,651

(a) Low-risk Patients

	Discount Rate	Avg. TQALYs	Avg. Lifetime	Avg. Treatment Cost	Avg. Screening Cost	Avg. Total Cost
Female	1,1	32.2384	32.3490	$21,968	$7,613	$29,748
	1,0.97	32.2384	32.3490	$9,694	$5,267	$15,231
	0.99,0.97	27.1448	27.2676	$9,694	$5,267	$15,231
	0.98,0.97	23.1316	23.2561	$9,694	$5,267	$15,231
	0.98,0.95	23.1316	23.2561	$6,075	$4,285	$10,584
	0.97,0.95	19.9387	20.0594	$6,075	$4,285	$10,584
Male	1,1	28.3885	28.5546	$22,159	$7,481	$30,145
	1,0.97	28.3885	28.5546	$10,641	$5,251	$16,230
	0.99,0.97	24.2779	24.4327	$10,641	$5,251	$16,230
	0.98,0.97	20.9780	21.1207	$10,641	$5,251	$16,230
	0.98,0.95	20.9780	21.1207	$6,991	$4,303	$11,545
	0.97,0.95	18.3060	18.4369	$6,991	$4,303	$11,545

(b) High-risk Patients

3.5 Bi-criteria Constrained MDP Model

As we discuss in Section 3.1, colonoscopy screening is an effective tool to reduce colorectal cancer incidence and death rates remarkably due to its specific nature of removing the lesions simultaneously with their detection. Recent statistics show that early detection of the cancer is especially important for colorectal cancer considering that 5-year survival rate is 91% if the cancer is detected in the early stage, however it decreases to 11% for the patients with the cancer detected in the late stage. With this motivation oftentimes clinicians may perform unnecessarily frequent screenings which is in a conflict with the guidelines. Although colonoscopy, the most-accurate screening method, reduces the colorectal cancer occurrence risk thus the mortality rates significantly, there are costs and disutility associated with it. The number of colonoscopy screenings that a patient undergoes has a direct effect on the costs and disutility, yet, on the risk of colorectal cancer and mortality. Conventionally, the cost of colonoscopy and TQALY's are the main considerations when determining a screening schedule, especially from the perspective of third-party healthcare payers. Too frequent screenings were discouraged in US before the health care reform legislation, however now, all health insurance plans are required to provide colorectal cancer screening tests (at no cost except when a polyp is removed) for adults starting from age 50 and continuing until age 75. On the other hand, the disutility of undergoing colonoscopy screenings remains same. Therefore, determining screening policies is still an open question from the multi-objective research point of view. In this section, we aim to answer the question that "What is the optimal screening policy that maximizes TQALY's while requiring less frequent colonoscopy screening and resulting in less costs or?"

We propose a conceptual idea of a bi-criteria constraint MDP (BCC-MDP) model that can be utilized to achieve the objective of maximizing TQALYs with less costs, or with a constraint on the maximum number of colonoscopy screenings performed. This model is actually very similar to the model developed in Erenay et al. (2014) in terms of how the CRC progression is modeled. It is very different from Erenay et al. (2014)'s model at the same time because of the limit on the number of colonoscopy screenings performed. That is, Erenay et al. (2014) assume that performing a colonoscopy screening is an option

at the beginning of each year, thus, the decisions are evaluated using the POMDP model each year and ours also evaluates the decisions at the beginning of each year. However, in order to eliminate the partially observable states by recording the past history, ours only keeps the history during the years from the last positive screening while knowing that there are a few of them because of the constraint on the maximum number of colonoscopy screenings. Note that without this constraint, it is not computationally efficient, even possible for larger problem instances, to enumerate all possible combinations of the ages when a positive screening result was observed last. Therefore, limiting the number of colonoscopies together with the specific nature of colonoscopy that allows removing the lesions simulatanously provides a unique opportunity to reduce a POMDP model to an MDP which also allows us to find the optimal solution as opposed to finding an approximate solution as discussed in Section 3.2.2.

When modeling BCC-MDP, we utilize the definitions of Erenay et al. (2014) for the immediate rewards and belief states since they only depend on how the natural progression of CRC is assumed. In addition to the health states, we define a state set that includes the past history keeping track of the test results. Let H_t denote this set at time t, then, $H_t = \{H_t^1, H_t^2, \ldots, H_t^{NC}\}$ where NC denotes the maximum number of colonoscopy screenings allowed. We derive the transition probabilities following the definition of this set and thus reduce the POMDP to MPD.

3.6 Conclusion

Colonoscopy screening prevents and early detects CRC, one of the most common and deadliest cancers in the world. Considering that the risk of developing CRC significantly increases after age 50, and North American population is aging, the colonoscopy screening and follow-up policies employed by the gastroenteritis play a vital role in the well-being of this aging population. Existing clinical guidelines recommend colonoscopy screening protocols that are shown to be cost-effective in CRC prevention and early detection; however, almost half of the practitioners do not follow these guidelines, indicating the controversy

around the best CRC screening practices. Several studies analyze alternative CRC screening protocols using simulation and mathematical models. Especially, dynamic alternative protocols can significantly increase health outcomes improvements due to CRC screening and follow-up. However, under dynamic policies, colonoscopy screening and surveillance intervals significantly vary in factors such as age, gender, and personal history, which are harder to implement for clinicians. In this chapter, we propose a set of simpler-to-implement colorectal cancer (CRC) screening policies and present a detailed analysis showing how these policies perform under different scenario settings.

We employ a patient-level discrete-event simulation model, built and validated using real data, to mimic CRC progression in asymptomatic low-risk and high-risk individuals, and estimate the expected life-years, age-based risk of having CRC, CRC mortality, costs associated with CRC screening, and the number of required colonoscopies for each screening protocol. We evaluate the performances of all relevant simpler-to-implement colonoscopy protocols, including the periodic screening protocols currently used by practitioners, all feasible periodic policies with n period switch times (for n=1,2). Our analysis also identifies under what parameter setting which alternative simpler policies are sufficient to provide close-to-optimal performance. We use the performances of the optimal POMDP policies derived in Erenay et al. (2014) for benchmarking. The optimal POMDP policies take age, gender, and risk-level into account and optimize the screening decisions that maximize the TQALYs. These policies improve the guidelines significantly in terms of all performance measures including risk of CRC and mortality, and extend the benefits of colonoscopy screenings for all patients. However, their structure is noticeably complex since the POMDP model captures many aspects of the problem and make decisions accordingly. Therefore, we determine the simpler-to-implement policies that perform closely to the corresponding optimal POMDP policies while following a very simple policy structure at most with two periodic screening frequencies.

The simpler-to-implement policies that are promoted in this study improve the guideline significantly for both genders and risk-levels. The improvement is more prominent for female patients than male patients and for high-risk patients than low-risk patients especially when their performances are compared to the corresponding optimal POMDP

policies for the majority of the cases. The policies using a single screening frequency that is independent from patients' age, static-periodic policies, provide very similar TQALYs to that of the optimal POMDP for high-risk female patients at a lower costs; however, the gap between the best performing static-policy and the optimal POMDP is significant for others. This finding shows that considering the patients' age when making screening decisions is necessary if we aim at providing close-to-optimal performances. The performances of the policies dividing the age interval into two sub-intervals, 2-age group dynamic-periodic policies, also perform very closely to the optimal POMDP with reasonable increase in cost. The gap is negligible for high-risk patients, especially for high-risk females, whereas the findings indicate there is still room for improvement for the male patients. The policies with 3-age group, the ones following 3 different screening frequencies in 3 sub-age-intervals, fill this gap considerably with reasonably small costs; however, the improvement is still not satisfying for male patients. This finding is consistent with the fact that CRC progression is more aggressive for male patients whereas the benefits of colonoscopy screening can only be achieved by following very personalized screening decisions as the optimal POMDP model does due to lower life expactancy of males. In other words, it is really important for males to catch any CRC related lesions on time while minimizing the disutility of colonoscopy so the relatively shorter life period can be utilized as much as possible. This, however, may seem in conflict, which is indeed true. To handle this kind of conflicting objectives, we propose using a bi-criteria constraint MDP formulation that can maximize the benefits of colonoscopy screening while restricting the number of colonoscopies performed, for example. Therefore, we strongly suggest policy makers to utilize the modeling framework proposed in this study when informing the guidelines.

References

H. Abouee-Mehrizi, O. Baron, O. Berman, and V. Sarhangian. Threshold-based allocation policies for inventory management of red blood cells. *Working paper*, 2017.

ACS. Cancer facts and figures 2018, american cancer society. 2018.

D.A. Ahlquist, D.J. Sargent, C.L. Loprinzi, T.R. Levin, D.K. Rex, D.J. Ahnen, K. Knigge, M.P. Lance, L.J. Burgart, S.R. Hamilton, et al. Stool DNA and occult blood testing for screen detection of colorectal neoplasia. *Annals of Internal Medicine*, 149(7):441–450, 2008.

M.K. Akmatov, M. Kretzschmar, A. Krämer, and R.T. Mikolajczyk. Timeliness of vaccination and its effects on fraction of vaccinated population. *Vaccine*, 26(31):3805–3811, 2008.

Oguzhan Alagoz. Optimizing cancer screening using partially observable markov decision processes. *Tutorials in Operations Research*, 2011.

Oguzhan Alagoz, Turgay Ayer, and Fatih Safa Erenay. Operations research models for cancer screening. *Wiley encyclopedia of operations research and management science*, 2011.

Oguzhan Alagoz, Jagpreet Chhatwal, and Elizabeth S Burnside. Optimal policies for reducing unnecessary follow-up mammography exams in breast cancer diagnosis. *Decision Analysis*, 10(3):200–224, 2013.

American Cancer Society. Colorectal cancer facts and figures 2018. *American Cancer Society*, 2018.

Elizabeth Arias. United states life tables, 2004. *National vital statistics reports*, 56(9): 1–40, 2007.

T.M. Assi, S.T. Brown, A. Djibo, B.A. Norman, J. Rajgopal, J.S. Welling, S.I. Chen, R.R. Bailey, S. Kone, H. Kenea, et al. Impact of changing the measles vaccine vial size on Niger's vaccine supply chain: A computational model. *BMC Public Health*, 11(1):425, 2011.

Turgay Ayer, Pinar Keskinocak, and Julie Swann. Research in public health for efficient, effective, and equitable outcomes. In *Bridging Data and Decisions*, pages 216–239. INFORMS, 2014.

Mehmet US Ayvaci, Oguzhan Alagoz, and Elizabeth S Burnside. The effect of budgetary restrictions on breast cancer diagnostic decisions. *Manufacturing & Service Operations Management*, 14(4):600–617, 2012.

A.T. Bach and J.A. Goad. The role of community pharmacy-based vaccinations in the USA: Current practice and future directions. *Integrated Pharmacy Research and Practice*, 4: 67–77, 2015.

M. Bakker, J. Riezebos, and R.H. Teunter. Review of inventory systems with deterioration since 2001. *European Journal of Operational Research*, 221(2):275–284, 2012.

O. Baron, O. Berman, and D. Perry. Continuous review inventory models for perishable items ordered in batches. *Mathematical Methods of Operations Research*, 72(2):217–247, 2010.

Richard E Bellman and Stuart E Dreyfus. *Applied dynamic programming*. Princeton university press, 2015.

Matthew W Bending, Paul Trueman, Karin V Lowson, Hazel Pilgrim, Paul Tappenden, Jim Chilcott, and Janine Tappenden. Estimating the direct costs of bowel cancer services

provided by the national health service in england. *International journal of technology assessment in health care*, 26(4):362–369, 2010.

Dimitri P Bertsekas, Dimitri P Bertsekas, Dimitri P Bertsekas, and Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.

S. Black, J. Eskola, C.A. Siegrist, N. Halsey, N. MacDonald, B. Law, E. Miller, N. Andrews, J. Stowe, D. Salmon, et al. Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic h1n1 influenza vaccines. *The Lancet*, 374(9707):2115–2122, 2010.

D.E. Bloom, D. Canning, and M. Weston. The value of vaccination. *World Economics-Henley on Thames*, 6(3):15, 2005.

Duncan P Boldy and P Ciaran O'Kane. Health operational research—a selective overview. *European Journal of Operational Research*, 10(1):1–9, 1982.

Vikram Boolchand, Gregory Olds, Joseph Singh, Pankaj Singh, Amitabh Chak, and Gregory S Cooper. Colorectal screening after polypectomy: a national survey study of primary care physicians. *Annals of internal medicine*, 145(9):654–659, 2006.

SC Brailsford, PR Harper, B Patel, and M Pitt. An analysis of the academic literature on simulation and modelling in health care. *Journal of simulation*, 3(3):130–140, 2009.

M.L. Brandeau. Allocating resources to control infectious diseases. In *Operations Research and Health Care*, pages 443–464. Springer, 2005.

P.A. Briss, L.E. Rodewald, A.R. Hinman, A.M. Shefer, R.A. Strikas, R.R. Bernier, V.G. Carande-Kulis, H.R. Yusuf, S.M. Ndiaye, S.M. Williams, et al. Reviews of evidence regarding interventions to improve vaccination coverage in children, adolescents, and adults. *American Journal of Preventive Medicine*, 18(1):97–140, 2000.

Edmund K Burke, Patrick De Causmaecker, Greet Vanden Berghe, and Hendrik Van Landeghem. The state of the art of nurse rostering. *Journal of scheduling*, 7(6):441–499, 2004.

Elizabeth S Burnside, Jagpreet Chhatwal, and Oguzhan Alagoz. What is the optimal threshold at which to recommend breast biopsy? *PloS one*, 7(11):e48820, 2012.

A. Burton, R. Monasch, B. Lautenbach, M. Gacic-Dobo, M. Neill, R. Karimov, L. Wolfson, G. Jones, and M. Birmingham. Who and unicef estimates of national infant immunization coverage: Methods and processes. *Bulletin of the World Health Organization*, 87 (7):535–541, 2009.

A. Burton, R. Kowalski, M. Gacic-Dobo, R. Karimov, and D. Brown. A formal representation of the who and unicef estimates of national immunization coverage: A computational logic approach. 2012.

Michael D Cabana, Cynthia S Rand, Neil R Powe, Albert W Wu, Modena H Wilson, Paul-Andre C Abboud, and Haya R Rubin. Why don't physicians follow clinical practice guidelines?: A framework for improvement. *Jama*, 282(15):1458–1465, 1999.

Leslie Anne Campbell, John T Blake, George Kephart, Eva Grunfeld, and Donald MacIntosh. Understanding the effects of competition for constrained colonoscopy services with the introduction of population-level colorectal cancer screening: A discrete event simulation model. *Medical Decision Making*, 37(2):253–263, 2017.

Scott B Cantor, Stephen J Spann, Robert J Volk, Melchor P Cardenas, and Michael M Warren. Prostate cancer screening: a decision analysis. *The Journal of family practice*, 41(1):33–41, 1995.

Tugba Cayirli and Emre Veral. Outpatient scheduling in health care: A review of literature*. *Production and Operations Management*, 12(4):519, 2003.

CCS. Canadian cancer statistics 2018, canadian cancer society's advisory committee on cancer statistics. 2018.

CDC. National vital statistics reports, national center for health statistics at cdc. 2018.

Mucahit Cevik, Turgay Ayer, Oguzhan Alagoz, and Brian L Sprague. Analysis of mammography screening policies under resource constraints. *Production and Operations Management*, 27(5):949–972, 2018.

Brenda Cheang, Haibing Li, Andrew Lim, and Brian Rodrigues. Nurse rostering problems—-a bibliographic survey. *European Journal of Operational Research*, 151(3): 447–460, 2003.

Jagpreet Chhatwal, Oguzhan Alagoz, and Elizabeth S Burnside. Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Operations research*, 58(6):1577–1591, 2010.

Stephen E Chick, Hamed Mamani, and David Simchi-Levi. Supply chain coordination and influenza vaccination. *Operations Research*, 56(6):1493–1506, 2008.

Palanivel Chinnakali, Vaman Kulkarni, S Kalaiselvi, and Baridalyne Nongkynrih. Vaccine wastage assessment in a primary care setting in urban india. *Journal of Pediatric Sciences*, 4(1):1–6, 2012.

S.H. Cho. The optimal composition of influenza vaccines subject to random production yields. *Manufacturing & Service Operations Management*, 12(2):256–277, 2010.

CISNET. Colorectal Cancer Model Profiles. http://cisnet.cancer.gov/colorectal/profiles.html, April 23 2012.

J. Clemens, J. Holmgren, S.H.E. Kaufmann, and A. Mantovani. Ten years of the global alliance for vaccines and immunization: Challenges and progress. *Nature Immunology*, 11(12):1069–1072, 2010.

P. Colrain. Study of EVM assessments from 57 GAVI-eligible countries, 2012-2013. Technical report, WHO UNICEF Reports, 2013.

José Cruzado, Fernando I Sánchez, José María Abellán, Francisco Pérez-Riquelme, and Fernando Carballo. Economic evaluation of colorectal cancer (crc) screening. *Best practice & research Clinical gastroenterology*, 27(6):867–880, 2013.

T. Dai, S.H. Cho, and F. Zhang. Contracting for on-time delivery in the US influenza vaccine supply chain. *Manufacturing & Service Operations Management*, 2016.

Ö.O. Dalgıç, O.Y. Özaltın, W.A. Ciccotelli, and F.S. Erenay. Deriving effective vaccine allocation strategies for pandemic influenza: Comparison of an agent-based simulation and a compartmental model. *To appear in PloS one*, 2017.

Mark S Daskin and Latoya K Dean. Location of health care facilities. In *Operations research and health care*, pages 43–76. Springer, 2005.

M. C. Degli-Atti, M. C. Rota, A. Bella, S. Salmaso, ICONA Study Group, et al. Do changes in policy affect vaccine coverage levels? Results of a national study to evaluate childhood vaccination coverage and reasons for missed vaccination in Italy. *Vaccine*, 22 (31):4351–4357, 2004.

B.T. Denton, O. Alagoz, A. Holder, and E.K. Lee. Medical decision making: Open research challenges. *IIE Transactions on Healthcare Systems Engineering*, 1(3):161–167, 2011.

S. Dentzer. Vaccines, children, and fulfilling human potential. *Health Affairs*, 30(6):1006–1006, 2011.

A. Dhamodharan and R.A. Proano. Determining the optimal vaccine vial size in developing countries: A monte carlo simulation approach. *Health Care Management Science*, 15(3): 188–196, 2012.

A. Dhamodharan, R.A. Proano, and S. Kumar. A stochastic approach to determine the optimal vaccine vial size. In *IIE Annual Conference Proceedings*, page 1. Institute of Industrial Engineers-Publisher, 2011.

David Dobrzykowski, Vafa Saboori Deilami, Paul Hong, and Seung-Chul Kim. A structured analysis of operations and supply chain management research in healthcare (1982–2011). *International Journal of Production Economics*, 147:514–530, 2014.

P.K. Drain, C.M. Nelson, and J.S. Lloyd. Single-dose versus multi-dose vaccine vials for immunization programmes in developing countries. *Bulletin of the World Health Organization*, 81(10):726–731, 2003.

Fatih Safa Erenay. *Optimal screening policies for colorectal cancer prevention and surveillance*. PhD thesis, University of Wisconsin–Madison, 2010.

Fatih Safa Erenay, Oguzhan Alagoz, Ritesh Banerjee, and Robert R Cima. Estimating the unknown parameters of the natural history of metachronous colorectal cancer using discrete-event simulation. *Medical Decision Making*, 31(4):611–624, 2011.

Fatih Safa Erenay, Oguzhan Alagoz, and Adnan Said. Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management*, 16(3):381–400, 2014.

Fatih Safa Erenay, Oguzhan Alagoz, Ritesh Banerjee, Adnan Said, and Robert R Cima. Cost-effectiveness of alternative colonoscopy surveillance strategies to mitigate metachronous colorectal cancer incidence. *Cancer*, 122(16):2560–2570, 2016.

Andreas T Ernst, Houyuan Jiang, Mohan Krishnamoorthy, Bowie Owens, and David Sier. An annotated bibliography of personnel scheduling and rostering. *Annals of Operations Research*, 127(1-4):21–144, 2004.

Anthony S Fauci. Seasonal and pandemic influenza preparedness: science and countermeasures. *The Journal of infectious diseases*, 194(Supplement_2):S73–S76, 2006.

A. Fleischhacker, A. Ninh, and Y. Zhao. Positioning inventory in clinical trial supply chains. *Production and Operations Management*, 24(6):991–1011, 2015.

Samuel Fomundam and Jeffrey W Herrmann. A survey of queuing theory applications in healthcare. 2007.

A.L. Frazier, G.A. Colditz, C.S. Fuchs, and K.M. Kuntz. Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA*, 284(15):1954–1961, 2000.

R. Ganeshan and T.P. Harrison. An introduction to supply chain management. Technical report, 1995.

GAVI. Value of vaccination, cost-effective. http://www.gavi.org/about/value/cost-effective, 2012. [Online; accessed 2015-07-10].

GAVI. Supply and procurement–the detailed product profiles (dpps). http://www.gavi.org/library/gavi-documents/supply-procurement/, 2015a. [Online; accessed 2016-09-21].

GAVI. Pentavalent vaccine support. http://www.gavi.org/support/nvs/pentavalent/, 2015b. [Online; accessed 2016-06-10].

GAVI. Detailed product profiles. http://www.gavi.org/about/market-shaping/detailed-product-profiles/, 2017. [Online; accessed 2016-10-10].

J.A. Goad, M.S. Taitel, L.E. Fensterheim, and A.E. Cannon. Vaccinations administered during off-clinic hours at a national community pharmacy: Implications for increasing patient access and convenience. *The Annals of Family Medicine*, 11(5):429–436, 2013.

S Lucas Goede, Aafke HC van Roon, Jacqueline CIY Reijerink, Anneke J van Vuuren, Iris Lansdorp-Vogelaar, J Dik F Habbema, Ernst J Kuipers, Monique E van Leerdam, and Marjolein van Ballegooijen. Cost-effectiveness of one versus two sample faecal immunochemical testing for colorectal cancer screening. *Gut*, pages gutjnl–2011, 2012.

James S Goodwin, Amanpal Singh, Nischita Reddy, Taylor S Riall, and Yong-Fang Kuo. Overuse of screening colonoscopy in the medicare population. *Archives of internal medicine*, 171(15):1335–1343, 2011.

S.K. Goyal and B.C. Giri. Recent trends in modeling of deteriorating inventory. *European Journal of Operational Research*, 134(1):1–16, 2001.

Linda V Green, Joao Soares, James F Giglio, and Robert A Green. Using queueing theory to increase the effectiveness of emergency department provider staffing. 2006.

S. Guichard, K. Hymbaugh, B. Burkholder, S. Diorditsa, C. Navarro, S. Ahmed, and M.M. Rahman. Vaccine wastage in Bangladesh. *Vaccine*, 28(3):858–863, 2010.

Murat M Günal and Michael Pidd. Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4(1):42–51, 2010.

Evrim Didem Güneş and E Lerzan Örmeci. Or applications in disease screening. In *Operations Research Applications in Health Care Management*, pages 297–325. Springer, 2018.

Evrim Didem Güneş, Egemen Lerzan Örmeci, and Derya Kunduzcu. Preventing and diagnosing colorectal cancer with a limited colonoscopy resource. *Production and Operations Management*, 24(1):1–20, 2015.

Diwakar Gupta and Brian Denton. Appointment scheduling in health care: Challenges and opportunities. *IIE transactions*, 40(9):800–819, 2008.

Ü. Gürler and B.Y. Özkaya. Analysis of the (s, S) policy for perishables with a random shelf life. *IIE Transactions*, 40(8):759–781, 2008.

Gerald Haidinger, Thomas Waldhoer, and Christian Vutuc. Self-reported colonoscopy screening in austria. *European Journal of Cancer Prevention*, 17(4):354–357, 2008.

J.B. Harris, M. Gacic-Dobo, R. Eggers, D.W. Brown, and S.V. Sodha. Global routine vaccination coverage, 2013. *The Morbidity and Mortality Weekly Report (MMWR)*, 63 (46):1055–8, 2014.

Ulrike Haug, Amy B Knudsen, and Karen M Kuntz. How should individuals with a false-positive fecal occult blood test for colorectal cancer be managed? a decision analysis. *International journal of cancer*, 131(9):2094–2102, 2012.

Lu He, Yinglei Li, and Sung Hoon Chung. Markov chain based modeling and analysis of colonoscopy screening processes. In *IIE Annual Conference. Proceedings*, pages 740–745. Institute of Industrial and Systems Engineers (IISE), 2017.

Alexis Heaton, Kirstin Krudwig, Tina Lorenson, Craig Burgess, Andrew Cunningham, and Robert Steinglass. Doses per vaccine vial container: An understated and underestimated driver of performance that needs more evidence. *Vaccine*, 35(17):2272–2278, 2017.

Kurt Heidenberger. Strategic investment in preventive health care: Quantitative modelling for programme selection and resource allocation. *Operations-Research-Spektrum*, 18(1): 1–14, 1996.

N Howlader, AM Noone, M Krapcho, N Neyman, R Aminou, SF Altekruse, CL Kosary, J Ruhl, Z Tatalovich, H Cho, et al. Seer cancer statistics review, 1975-2009 (vintage 2009 populations), national cancer institute. bethesda, md. *MD, USA*, 2012.

Health Protection and Promotion Act, Population and Public Health Division, Ministry of Health and Long-Term Care, Ontario, Canada: HPPA. Vaccine storage and handling protocol, 2016. http://www.health.gov.on.ca/en/pro/programs/publichealth/oph_standards/docs/vaccine_storage_handling.pdf, 2016. [Online; accessed 2017-01-10].

Peter JH Hulshof, Nikky Kortbeek, Richard J Boucherie, Erwin W Hans, and Piet JM Bakker. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. *Health systems*, 1(2):129–175, 2012.

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

S.H. Jacobson, E.C. Sewell, and R.A. Proano. An analysis of the pediatric vaccine supply shortage problem. *Health Care Management Science*, 9(4):371–389, 2006.

Sheldon H Jacobson, Edward C Sewell, Robert Deuson, and Bruce G Weniger. An integer programming model for vaccine procurement and delivery for childhood immunization: a pilot study. *Health Care Management Science*, 2(1):1–9, 1999.

Sheldon H Jacobson, Edward C Sewell, Daniel A Allwine, Enrique A Medina, and Bruce G Weniger. Designing pediatric vaccine formularies and pricing pediatric combination vaccines using operations research models and algorithms. *Expert review of vaccines*, 2 (1):15–19, 2003.

D.T. Jamison, J.G. Breman, A.R. Measham, G. Alleyne, M. Claeson, D.B. Evans, P. Jha, A. Mills, and P. Musgrove. *Disease control priorities in developing countries*. World Bank Publications, 2006.

JB Jun, Sheldon H Jacobson, and JR Swisher. Application of discrete-event simulation in health care clinics: A survey. *Journal of the operational research society*, 50(2):109–123, 1999.

Cengiz Kahraman and Y Ilker Topcu. *Operations Research Applications in Health Care Management*. Springer, 2018.

L. Kamara, J.B. Milstien, M. Patyna, P. Lydon, A. Levin, and L. Brenzel. Strategies for financial sustainability of immunization programs: A review of the strategies from 50 national immunization program financial sustainability plans. *Vaccine*, 26(51):6717–6726, 2008.

I.Z. Karaesmen, A. Scheller-Wolf, and B. Deniz. Managing perishable and aging inventories: Review and future research directions. In *Planning production and inventories in the extended enterprise*, pages 393–436. Springer, 2011.

Korina Katsaliaki and Navonil Mustafee. Applications of simulation within the healthcare context. *Journal of the Operational Research Society*, 62(8):1431–1451, 2011.

Carrie N Klabunde, Paul S Frame, Ann Meadow, Elizabeth Jones, Marion Nadel, and Sally W Vernon. A national survey of primary care physicians' colorectal cancer screening recommendations and practices. *Preventive medicine*, 36(3):352–362, 2003.

Carrie N Klabunde, Sally W Vernon, Marion R Nadel, Nancy Breen, Laura C Seeff, and Martin L Brown. Barriers to colorectal cancer screening: a comparison of reports from primary care physicians and average-risk adults. *Medical care*, pages 939–944, 2005.

Carrie N Klabunde, David Lanier, Marion R Nadel, Caroline McLeod, Gigi Yuan, and Sally W Vernon. Colorectal cancer screening by primary care physicians: recommendations and practices, 2006–2007. *American journal of preventive medicine*, 37(1):8–16, 2009.

A. Klenke, L. Mattner, et al. Stochastic ordering of classical discrete distributions. *Advances in Applied probability*, 42(2):392–410, 2010.

A.B. Knudsen, P.M. McMahon, and G.S. Gazelle. Use of modeling to evaluate the cost-effectiveness of cancer screening programs. *Journal of Clinical Oncology*, 25(2):203–208, 2007.

A.B. Knudsen, I. Lansdorp-Vogelaar, C.M. Rutter, J.E. Savarino, M. Van Ballegooijen, K.M. Kuntz, and A.G. Zauber. Cost-effectiveness of computed tomographic colonography screening for colorectal cancer in the Medicare population. *Journal of the National Cancer Institute*, 102(16):1238–1252, 2010.

Amy B Knudsen, Ann G Zauber, Carolyn M Rutter, Steffie K Naber, V Paul Doria-Rose, Chester Pabiniak, Colden Johanson, Sara E Fischer, Iris Lansdorp-Vogelaar, and Karen M Kuntz. Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the us preventive services task force. *Jama*, 315(23):2595–2609, 2016.

L.J. Kornish and R.L. Keeney. Repeated commit-or-defer decisions with a deadline: The influenza vaccine composition. *Operations Research*, 56(3):527–541, 2008.

R. Kotagiri, R.A. Proaño, and M.E. Kuhl. Vaccine vendor managed inventory: The effect of using birth forecast. In *IIE Annual Conference Proceedings*, page 1. Institute of Industrial Engineers-Publisher, 2011.

Alex H Krist, Resa M Jones, Steven H Woolf, Sarah E Woessner, Daniel Merenstein, J William Kerns, Walter Foliaco, and Paul Jackson. Timing of repeat colonoscopy: disparity between guidelines and endoscopists' recommendation. *American journal of preventive medicine*, 33(6):471–478, 2007.

Ernst J Kuipers, Thomas Rösch, and Michael Bretthauer. Colorectal cancer screening—optimizing current strategies and new directions. *Nature reviews Clinical oncology*, 10(3):130, 2013.

Karen M Kuntz, Iris Lansdorp-Vogelaar, Carolyn M Rutter, Amy B Knudsen, Marjolein Van Ballegooijen, James E Savarino, Eric J Feuer, and Ann G Zauber. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. *Medical Decision Making*, 31(4):530–539, 2011.

C Lakshmi and Sivakumar Appa Iyer. Application of queueing theory in health care: A literature review. *Operations Research for Health Care*, 2(1):25–39, 2013.

I. Lansdorp-Vogelaar, M. van Ballegooijen, A.G. Zauber, R. Boer, J. Wilschut, and J.D.F. Habbema. At what costs will screening with CT colonography be competitive? A cost-effectiveness approach. *International Journal of Cancer*, 124(5):1161–1168, 2009a.

I. Lansdorp-Vogelaar, M. van Ballegooijen, A.G. Zauber, R. Boer, J. Wilschut, S.J. Winawer, and J.D.F. Habbema. Individualizing colonoscopy screening by sex and race. *Gastrointestinal Endoscopy*, 70(1):96–108, 2009b.

Iris Lansdorp-Vogelaar, Marjolein Van Ballegooijen, Ann G Zauber, J Dik F Habbema, and Ernst J Kuipers. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. *JNCI: Journal of the National Cancer Institute*, 101(20):1412–1422, 2009c.

B.Y. Lee, B.A. Norman, T.M. Assi, S.I. Chen, R.R. Bailey, J. Rajgopal, S.T. Brown, A.E. Wiringa, and D.S. Burke. Single versus multi-dose vaccine vials: An economic computational model. *Vaccine*, 28(32):5292–5300, 2010.

B.Y. Lee, T.M. Assi, K. Rookkapan, D.L. Connor, J. Rajgopal, V. Sornsrivichai, S.T. Brown, J.S. Welling, B.A. Norman, S.I. Chen, et al. Replacing the measles ten-dose vaccine presentation with the single-dose presentation in Thailand. *Vaccine*, 29(21): 3811–3817, 2011.

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

David A Lieberman, David G Weiss, John H Bond, Dennis J Ahnen, Harinder Garewal, William V Harford, Dawn Provenzale, Steve Sontag, Tom Schnell, Theodore E Durbin, et al. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. *New England Journal of Medicine*, 343(3):162–168, 2000.

G.J. Lim and E.K. Lee. *Optimization in medicine and biology*, volume 3. CRC Press, 2008.

F. Loeve, R. Boer, GJ Van Oortmarssen, M. Van Ballegooijen, and JDF Habbema. The miscan-colon simulation model for the evaluation of colorectal cancer screening. *Computers and Biomedical Research*, 32(1):13–33, 1999.

F. Loeve, M.L. Brown, R. Boer, M. van Ballegooijen, and G.J. van Oortmarssen. Endoscopic colorectal cancer screening: A cost-saving analysis. *Journal of the National Cancer Institute*, 92(7):557–563, 2000.

Franka Loeve, Rob Boer, Ann G Zauber, Marjolein van Ballegooijen, Gerrit J van Oortmarssen, Sidney J Winawer, and J Dik F Habbema. National polyp study data: evidence for regression of adenomas. *International journal of cancer*, 111(4):633–639, 2004.

Millie D Long and Bruce E Sands. When do you start and when do you stop screening for colon cancer in inflammatory bowel disease? *Clinical Gastroenterology and Hepatology*, 16(5):621–623, 2018.

Walter E Longo, Katherine S Virgo, Frank E Johnson, Charles A Oprian, Anthony M Vernava, Terence P Wade, Maureen A Phelan, William G Henderson, Jennifer Daley, and Shukri F Khuri. Risk factors for morbidity and mortality after colectomy for colon cancer. *Diseases of the colon & rectum*, 43(1):83–91, 2000.

Shail Maheshwari, Tushar Patel, and Parita Patel. Screening for colorectal cancer in elderly persons: who should we screen and when can we stop? *Journal of aging and health*, 20 (1):126–139, 2008.

Hamed Mamani, Stephen E Chick, and David Simchi-Levi. A game-theoretic model of international influenza vaccination coordination. *Management Science*, 59(7):1650–1670, 2013.

Jack S Mandel, Timothy R Church, John H Bond, Fred Ederer, Mindy S Geisser, Steven J Mongin, Dale C Snover, and Leonard M Schuman. The effect of fecal occult-blood screening on the incidence of colorectal cancer. *New England Journal of Medicine*, 343 (22):1603–1607, 2000.

David McCoy, Gayatri Kembhavi, Jinesh Patel, and Akish Luintel. The bill & melinda gates foundation's grant-making programme for global health. *The Lancet*, 373(9675): 1645–1653, 2009.

Geoffrey McNicoll. World population ageing 1950-2050. *Population and Development Review*, 28(4):814–816, 2002.

J. Medlock and A.P. Galvani. Optimizing influenza vaccine distribution. *Science*, 325 (5948):1705–1708, 2009.

Reinier GS Meester, Chyke A Doubeni, Ann G Zauber, S Luuk Goede, Theodore R Levin, Douglas A Corley, Ahmedin Jemal, and Iris Lansdorp-Vogelaar. Public health impact of achieving 80% colorectal cancer screening rates in the united states by 2018. *Cancer*, 121(13):2281–2285, 2015.

Reinier GS Meester, Elisabeth FP Peterse, Amy B Knudsen, Anne C de Weerdt, Jennifer C Chen, Anna P Lietz, Andrea Dwyer, Dennis J Ahnen, Rebecca L Siegel, Robert A Smith, et al. Optimizing colorectal cancer screening by race and sex: Microsimulation analysis ii to inform the american cancer society colorectal cancer screening guideline. *Cancer*, 2018.

M.H. Mofrad, L.M. Maillart, B.A. Norman, and J. Rajgopal. Dynamically optimizing the administration of vaccines from multi-dose vials. *IIE Transactions*, 46(7):623–635, 2014.

M.H. Mofrad, G.G.P. Garcia, L.M. Maillart, B.A. Norman, and J. Rajgopal. Customizing immunization clinic operations to minimize open vial waste. *Socio-Economic Planning Sciences*, 54:1–17, 2016.

E.R. Moxon, P. Das, B. Greenwood, D.L. Heymann, R. Horton, O.S. Levine, S. Plotkin, and G. Nossal. A call to action for the new decade of vaccines. *The Lancet*, 378(9788): 298–302, 2011.

Pauline A Mysliwiec, Martin L Brown, Carrie N Klabunde, and David F Ransohoff. Are physicians doing too much colonoscopy? a national survey of colorectal surveillance after polypectomy. *Annals of Internal Medicine*, 141(4):264–271, 2004.

S. Nahmias. Perishable inventory theory: A review. *Operations Research*, 30(4):680–708, 1982.

NCI. National Cancer Institute. http://www.cancer.gov/types/colorectal/screening-fact-sheet, 2014. [Online; accessed 13-Oct-2015].

Aileen R Neilson and David K Whynes. Cost-effectiveness of screening for colorectal cancer: a simulation model. *Mathematical Medicine and Biology: A Journal of the IMA*, 12(3-4): 355–367, 1995.

R.M. Ness, A.M. Holmes, R. Klein, and R. Dittus. Cost-utility of one-time colonoscopic screening for colorectal cancer at various ages. *The American Journal of Gastroenterology*, 95(7):1800–1811, 2000.

Lawrence Nicholson, Asoo J Vakharia, and S Selcuk Erenguc. Outsourcing inventory management decisions in healthcare: Models and application. *European Journal of Operational Research*, 154(1):271–290, 2004.

Jesse N Nodora, William D Martz, Erin L Ashbeck, Elizabeth T Jacobs, Patricia A Thompson, and María Elena Martínez. Primary care physician compliance with colorectal cancer screening guidelines. *Cancer Causes & Control*, 22(9):1277–1287, 2011.

Ronald Anthony Nosek and James P Wilson. Queuing theory and customer satisfaction: a review of terminology, trends, and applications to pharmacy practice. *Hospital pharmacy*, 36(3):275–279, 2001.

Kevin C Oeffinger, Elizabeth TH Fontham, Ruth Etzioni, Abbe Herzig, James S Michaelson, Ya-Chen Tina Shih, Louise C Walter, Timothy R Church, Christopher R Flowers, Samuel J LaMonte, et al. Breast cancer screening for women at average risk: 2015 guideline update from the american cancer society. *Jama*, 314(15):1599–1614, 2015.

Björn Ohlsson and Birger Pålsson. Follow-up after colorectal cancer surgery. *Acta oncologica*, 42(8):816–826, 2003.

World Health Organization et al. Who vaccine-preventable diseases: monitoring system: 2008 global summary. 2008.

World Health Organization et al. Global vaccine action plan 2011-2020. *Global Vaccine Action Plan 2011-2020.*, 2013.

World Health Organization et al. Measles vaccines: Who position paper, april 2017–recommendations. *Vaccine*, 2017.

E Lerzan Örmeci, Evrim Didem Güneş, and Derya Kunduzcu. A modeling framework for control of preventive services. *Manufacturing & Service Operations Management*, 18(2): 227–244, 2015.

O.Y. Ozaltin, O.A. Prokopyev, A.J. Schaefer, and M.S. Roberts. Optimizing the societal benefits of the annual influenza vaccine: A stochastic programming approach. *Operations Research*, 59(5):1131–1143, 2011.

S. Ozawa, M.L. Stack, D.M. Bishai, A. Mirelman, I.K. Friberg, L. Niessen, D.G. Walker, and O.S. Levine. During the "decade of vaccines",the lives of 6.4 million children valued at $231 billion could be saved. *Health Affairs*, 30(6):1010–1020, 2011.

J. Papastergiou, C. Folkins, W. Li, and J. Zervas. Community pharmacist–administered influenza immunization improves patient access to vaccination. *Canadian Pharmacists Journal/Revue des Pharmaciens du Canada*, 147(6):359–365, 2014.

D. Parmar, E.M. Baruwa, P. Zuber, and S. Kone. Impact of wastage on single and multi-dose vaccine vials: Implications for introducing pneumococcal vaccines in developing countries. *Human Vaccines*, 6(3):270–278, 2010.

PATH. Point-of-service data to drive vaccine supply chains. http://www.path.org/publications/detail.php?i=2054, 2012. [Online; accessed 2016-06-10].

Elisabeth FP Peterse, Reinier GS Meester, Rebecca L Siegel, Jennifer C Chen, Andrea Dwyer, Dennis J Ahnen, Robert A Smith, Ann G Zauber, and Iris Lansdorp-Vogelaar. The impact of the rising colorectal cancer incidence in young adults on the optimal age to start screening: microsimulation analysis i to inform the american cancer society colorectal cancer screening guideline. *Cancer*, 2018.

William P Pierskalla and David J Brailer. Applications of operations research in health care delivery. *Handbooks in operations research and management science*, 6:469–505, 1994.

W.P. Pierskalla. Supply chain management of blood banks. In *Operations Research and Health Care*, pages 103–145. Springer, 2005.

H. Pilgrim, P. Tappenden, J. Chilcott, M. Bending, P. Trueman, A. Shorthouse, and J. Tappenden. The costs and benefits of bowel cancer service developments using discrete event simulation. *Journal of the Operational Research Society*, 60(10):1305–1314, 2008.

Pharmacists and Immunization Working Group: PIWG. Pharmacists and publicly funded vaccines in B.C.: General information. http://www2.gov.bc.ca/assets/gov/health/health-drug-coverage/pharmacare/vaccine-guide.pdf, 2015. [Online; accessed 2017-01-10].

S.A. Plotkin. Vaccines: The fourth century. *Clinical and Vaccine Immunology*, 16(12): 1709–1719, 2009.

J Preater. Queues in health. *Health Care Management Science*, 5(4):283–283, 2002.

R.A. Proano, S.H. Jacobson, and W. Zhang. Making combination vaccines more accessible to low-income countries: The antigen bundle pricing problem. *Omega*, 40(1):53–64, 2012.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

Abdur Rais and Ana Viana. Operations research in healthcare: a survey. *International transactions in operational research*, 18(1):1–31, 2011.

J. Rajgopal, D.L. Connor, T.M. Assi, B.A. Norman, S.I. Chen, R.R. Bailey, A.R. Long, A.R. Wateska, K.M. Bacon, S.T. Brown, et al. The optimal number of routine vaccines to order at health clinics in low or middle income countries. *Vaccine*, 29(33):5512–5518, 2011.

A. Ramirez-Nafarrate, J.D. Lyon, J.W. Fowler, and O.M. Araz. Point-of-dispensing location and capacity optimization via a decision support system. *Production and Operations Management*, 24(8):1311–1328, 2015.

S.D. Ramsey, J. Wilschut, R. Boer, and M. van Ballegooijen. A decision-analytic evaluation of the cost-effectiveness of family history-based colorectal cancer screening programs. *The American Journal of Gastroenterology*, 105(8):1861–1869, 2010.

Smithers Rapra. Global market for pre-filled syringes forecast to rise to $6.6 billion by 2020. http://www.smithersrapra.com/news/2015/february/global-market-for-pre-filled-syringes-forecast-to, 2015. [Online; accessed 2016-10-15].

N. Ravichandran. Stochastic analysis of a continuous review perishable inventory system with positive lead time and poisson demand. *European Journal of Operational Research*, 84(2):444–457, 1995.

John F Repede and John J Bernardo. Developing and validating a decision support system for locating emergency medical vehicles in louisville, kentucky. *European journal of operational research*, 75(3):567–581, 1994.

M.J. Robbins and S.H. Jacobson. Pediatric vaccine procurement policy: The monopsonist's problem. *Omega*, 39(6):589–597, 2011.

M.J. Robbins and S.H. Jacobson. Analytics for vaccine economics and pricing: Insights and observations. *Expert Review of Vaccines*, 14(4):605–616, 2015.

S. Roberts, L. Wang, R. Klein, R. Ness, and R. Dittus. Development of a simulation model of colorectal cancer. *ACM Transactions on Modeling and Computer Simulation*, 18(1): 1–30, 2007.

S.D. Roberts. Tutorial on the simulation of healthcare systems. In *Simulation Conference (WSC), Proceedings of the 2011 Winter*, pages 1403–1414. IEEE, 2011.

C.M. Rutter and J.E. Savarino. An evidence-based microsimulation model for colorectal cancer: validation and application. *Cancer Epidemiology Biomarkers & Prevention*, 19 (8):1992–2002, 2010.

C.M. Rutter, D.L. Miglioretti, and J.E. Savarino. Evaluating risk factor assumptions: a simulation-based approach. *BMC Medical Informatics and Decision Making*, 11(1): 55–64, 2011.

V. Sarhangian, H. Abouee-Mehrizi, O. Baron, O. Berman, N.M. Heddle, and R. Barty. Reducing the age of transfused red blood cells in hospitals: Ordering and allocation policies. *Vox Sanguinis*, 2016.

Christina E Saville, Honora K Smith, and Katarzyna Bijak. Operational research techniques applied throughout cancer care services: a review. *Health Systems*, pages 1–22, 2018.

Andrew J Schaefer, Matthew D Bailey, Steven M Shechter, and Mark S Roberts. Modeling medical treatment using markov decision processes. In *Operations research and health care*, pages 593–612. Springer, 2005.

Robert E Schoen, Paul F Pinsky, Joel L Weissfeld, Lance A Yokochi, Timothy Church, Adeyinka O Laiyemo, Robert Bresalier, Gerald L Andriole, Saundra S Buys, E David Crawford, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *New England Journal of Medicine*, 366(25):2345–2357, 2012.

Eline H Schreuders, Arlinda Ruco, Linda Rabeneck, Robert E Schoen, Joseph JY Sung, Graeme P Young, and Ernst J Kuipers. Colorectal cancer screening: a global overview of existing programmes. *Gut*, 64(10):1637–1649, 2015.

S. Setia, H. Mainzer, M.L. Washington, G. Coil, R. Snyder, and B.G. Weniger. Frequency and causes of vaccine wastage. *Vaccine*, 20(7):1148–1156, 2002.

Aasma Shaukat, Steven J Mongin, Mindy S Geisser, Frank A Lederle, John H Bond, Jack S Mandel, and Timothy R Church. Long-term mortality after screening for colorectal cancer. *New England Journal of Medicine*, 369(12):1106–1114, 2013.

A. Shefer, P. Briss, L. Rodewald, R. Bernier, R. Strikas, H. Yusuf, S. Ndiaye, S. Williams, M. Pappaioanou, and A.R. Hinman. Improving immunization coverage rates: An evidence-based review of the literature. *Epidemiologic Reviews*, 21(1):96–142, 1999.

J. Shi, O. Alagoz, F.S. Erenay, and Q. Su. A survey of optimization models on cancer chemotherapy treatment planning. *Annals of Operations Research*, 221(1):331–356, 2014.

Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2016. *CA: a cancer journal for clinicians*, 66(1):7–30, 2016.

Rebecca L Siegel, Kimberly D Miller, Stacey A Fedewa, Dennis J Ahnen, Reinier GS Meester, Afsaneh Barzi, and Ahmedin Jemal. Colorectal cancer statistics, 2017. *CA: a cancer journal for clinicians*, 67(3):177–193, 2017.

Frank A Sonnenberg and J Robert Beck. Markov models in medical decision making: a practical guide. *Medical decision making*, 13(4):322–338, 1993.

M.L. Stack, S. Ozawa, D.M. Bishai, A. Mirelman, Y. Tam, L. Niessen, D.G. Walker, and O.S. Levine. Estimated economic benefits during the "decade of vaccines" include treatment savings, gains in labor productivity. *Health Affairs*, 30(6):1021–1028, 2011.

Tyler Stevens and Carol A Burke. Colonoscopy screening in the elderly: when to stop? *The American journal of gastroenterology*, 98(8):1881, 2003.

CE Stevenson. Statistical models for cancer screening. *Statistical methods in medical research*, 4(1):18–32, 1995.

S. Subramanian, G. Bobashev, and R.J. Morris. When budgets are tight, there are better options than colonoscopies for colorectal cancer screening. *Health Affairs*, 29(9):1734–1740, 2010.

Sujha Subramanian, Georgiy Bobashev, and Robert J Morris. Modeling the cost-effectiveness of colorectal cancer screening: policy guidance based on patient preferences and compliance. *Cancer Epidemiology and Prevention Biomarkers*, 18(7):1971–1978, 2009.

Sapna Syngal, Jane C Weeks, Deborah Schrag, Judy E Garber, and Karen M Kuntz. Benefits of colonoscopic surveillance and prophylactic colectomy in patients with hereditary nonpolyposis colorectal cancer mutations. *Annals of internal medicine*, 129(10):787–796, 1998.

T.D. Szucs. Health economic research on vaccinations and immunisation practices: An introductory primer. *Vaccine*, 23(17):2095–2103, 2005.

A. Tafazzoli, S. Roberts, R. Klein, R. Ness, and R. Dittus. Probabilistic cost-effectiveness comparison of screening strategies for colorectal cancer. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 19(2):6–29, 2009.

E. Tekin, Ü. Gürler, and E. Berk. Age-based vs. stock level control policies for a perishable inventory system. *European Journal of Operational Research*, 134(2):309–329, 2001.

S. D. Torun and N. Bakırcı. Vaccination coverage and reasons for non-vaccination in a district of Istanbul. *BMC Public Health*, 6(1):1, 2006.

S. Tunc, O. Alagoz, and E. Burnside. Opportunities for operations research in medical decision making. *IEEE Intelligent Systems*, 29(3):59, 2014.

UNICEF. Cold chain weight and volume calculator. https://www.unicef.org/supply/index_51098.html, 2015. [Online; accessed 2016-10-10].

UNICEF and WHO. Immunization summary: A statistical reference containing data through 2013. http://www.who.int/immunization/monitoring_surveillance/routine/coverage/en/index4.html, 2014. [Online; accessed 2015-07-10].

USCB. US Census Bureau population estimates: National characteristics. http://www.census.gov/popest/data/intercensal /national/nat2010.html, April 5 2012.

USPSTF. Screening for colorectal cancer: Us preventive services task force recommendation statement, us preventive services task force and others. *Annals of internal medicine*, 149(9):627, 2008.

M. van Ballegooijen, C.M. Rutter, A.B. Knudsen, A.G. Zauber, J.E. Savarino, I. Lansdorp-Vogelaar, R. Boer, E.J. Feuer, J.D.F. Habbema, and K.M. Kuntz. Clarifying differences in natural history between models of screening the case of colorectal cancer. *Medical Decision Making*, 31(4):540–549, 2011.

Frank Van Hees, Sameer D Saini, Iris Lansdorp-Vogelaar, Sandeep Vijan, Reinier GS Meester, Harry J de Koning, Ann G Zauber, and Marjolein van Ballegooijen. Personalizing colonoscopy screening for elderly individuals based on screening history, cancer risk, and comorbidity status could increase cost effectiveness. *Gastroenterology*, 149(6): 1425–1437, 2015.

David J Vanness, Amy B Knudsen, Iris Lansdorp-Vogelaar, Carolyn M Rutter, Ilana F Gareen, Benjamin A Herman, Karen M Kuntz, Ann G Zauber, Marjolein Van Ballegooijen, Eric J Feuer, et al. Comparative economic evaluation of data from the acrin national ct colonography trial with three cancer intervention and surveillance modeling network microsimulations. *Radiology*, 261(2):487–498, 2011.

Sandeep Vijan, Erica W Hwang, Timothy P Hofer, and Rodney A Hayward. Which colon cancer screening test? a comparison of costs, effectiveness, and compliance. *The American journal of medicine*, 111(8):593–601, 2001.

Sandeep Vijan, Inku Hwang, John Inadomi, Roy KH Wong, J Richard Choi, John Napierkowski, Jonathan M Koff, and Perry J Pickhardt. The cost-effectiveness of ct colonography in screening for colorectal neoplasia. *The American journal of gastroenterology*, 102(2):380, 2007.

I. Vogelaar, M. Van Ballegooijen, D. Schrag, R. Boer, S.J. Winawer, J.D.F. Habbema, and A.G. Zauber. How much can current interventions reduce colorectal cancer mortality in the US? *Cancer*, 107(7):1624–1633, 2006.

D.G. Walker, R. Hutubessy, and P. Beutels. Who guide for standardisation of economic evaluations of immunization programmes. *Vaccine*, 28(11):2356–2359, 2010.

B.G. Weniger, R.T. Chen, S.H. Jacobson, E.C. Sewell, R. Deuson, J.R. Livengood, and W.A. Orenstein. Addressing the challenges to immunization practice with an economic algorithm for vaccine selection. *Vaccine*, 16(19):1885–1897, 1998.

WHO. Guidelines for estimating costs of introducing new vaccines into the national immunization system. http://apps.who.int/iris/bitstream/10665/67342/1/WHO_V-B_02.11_eng.pdf, 2002.

WHO. Monitoring vaccine wastage at country level: Guidelines for programme managers. http://apps.who.int/iris/bitstream/10665/68463/1/WHO_VB_03.18.Rev.1_eng.pdf, 2005. [Online; accessed 2015-07-10].

WHO. World health organization, fact sheets, the top 10 causes of death. http://www.who.int/mediacentre/factsheets/fs310/en/index2.html, 2012. [Online; accessed 2015-07-07].

WHO. Operational guidelines. http://www.searo.who.int/india/topics/routine_immunization/Operational_Guidelines_for_introduction_Hib_as_Pentavalent_vaccine_2013.pdf, 2013.

WHO. WHO policy statement: Multi-dose vial policy (MDVP), handling of multi-dose vaccine vials after opening. http://apps.who.int/iris/bitstream/10665/135972/1/WHO_IVB_14.07_eng.pdf, 2014a. [Online; accessed 2015-07-10].

WHO. Immunization supply chain and logistics. http://www.who.int/immunization/programmes_systems/supply_chain/en, 2014b. [Online; accessed 2016-06-10].

WHO. Who vaccine preventable disease monitoring system, 2016 global summary. Technical report, WHO, 2016a.

WHO. WHO prequalified vaccines. https://extranet.who.int/gavi/PQ_Web/, 2016b. [Online; accessed 2016-09-12].

WHO. World health organization, who recommendations for routine immunization - summary tables. http://www.who.int/immunization/policy/immunization_tables/en/, 2017. [Online; accessed 2017-07-18].

WHO, GAVI, UNICEF, and Bill & Melinda Gates Foundations. Global vaccine action plan 2011-2020. http://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/, 2015. [Online; accessed 2015-07-10].

UNICEF WHO. Global immunization data 2014. Technical report, WHO UNICEF Reports, 2015.

J.A. Wilschut, J.D.F. Habbema, M.E. van Leerdam, L. Hol, I. Lansdorp-Vogelaar, E.J. Kuipers, and M. van Ballegooijen. Fecal occult blood testing when colonoscopy capacity is limited. *Journal of the National Cancer Institute*, 103(23):1741–1751, 2011a.

Janneke A Wilschut, Lieke Hol, Evelien Dekker, Jan B Jansen, Monique E van Leerdam, Iris Lansdorp-Vogelaar, Ernst J Kuipers, J Dik F Habbema, and Marjolein van Ballegooijen. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology*, 141(5):1648–1655, 2011b.

Janneke A Wilschut, Ewout W Steyerberg, Monique E van Leerdam, Iris Lansdorp-Vogelaar, J Dik F Habbema, and Marjolein van Ballegooijen. How much colonoscopy screening should be recommended to individuals with various degrees of family history of colorectal cancer? *Cancer*, 117(18):4166–4174, 2011c.

Joanne AP Wilson. Colon cancer screening in the elderly: when do we stop? *Transactions of the American Clinical and Climatological Association*, 121:94, 2010.

Sidney Winawer, Robert Fletcher, Douglas Rex, John Bond, Randall Burt, Joseph Ferrucci, Theodore Ganiats, Theodore Levin, Steven Woolf, David Johnson, et al. Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. *Gastroenterology*, 124(2):544–560, 2003.

Sidney J Winawer. Colorectal cancer screening. *Best Practice & Research Clinical Gastroenterology*, 21(6):1031–1048, 2007.

Grace Hui-Min Wu, Yi-Ming Wang, Amy Ming-Fang Yen, Jau-Min Wong, Hsin-Chih Lai, Jane Warwick, and Tony Hsiu-Hsi Chen. Cost-effectiveness analysis of colorectal cancer screening with stool dna testing in intermediate-incidence countries. *BMC cancer*, 6(1): 1, 2006.

K Robin Yabroff, Angela B Mariotto, Eric Feuer, and Martin L Brown. Projections of the costs associated with colorectal cancer care in the united states, 2000–2020. *Health economics*, 17(8):947–959, 2008.

K Robin Yabroff, Carrie N Klabunde, Gigi Yuan, Timothy S McNeel, Martin L Brown, Dana Casciotti, Dennis W Buckman, and Stephen Taplin. Are physicians' recommendations for colorectal cancer screening guideline-consistent? *Journal of general internal medicine*, 26(2):177–184, 2011a.

K Robin Yabroff, Jennifer Lund, Deanna Kepka, and Angela Mariotto. Economic burden of cancer in the united states: estimates, projections, and future research. *Cancer Epidemiology Biomarkers & Prevention*, 20(10):2006–2014, 2011b.

R. Yaesoubi and T. Cohen. Dynamic health policies for controlling the spread of emerging infections: Influenza as an example. 2011a.

R. Yaesoubi and T. Cohen. Generalized markov models of infectious disease spread: A novel framework for developing dynamic health policies. *European Journal of Operational Research*, 215(3):679–687, 2011b.

R. Yaesoubi and S.D. Roberts. How much is a health insurer willing to pay for Colorectal Cancer screening tests? In *Simulation Conference, 2008. WSC 2008. Winter*, pages 1624–1631. IEEE, 2008.

R. Yaesoubi and S.D. Roberts. A game-theoretic framework for estimating a health purchaser's willingness-to-pay for health and for expansion. *Health Care Management Science*, 13(4):358–377, 2010.

W. Yang, M. Parisi, B.J. Lahue, M.J. Uddin, and D. Bishai. The budget impact of controlling wastage with smaller vials: A data driven model of session sizes in Bangladesh, India (Uttar Pradesh), Mozambique, and Uganda. *Vaccine*, 32(49):6643–6648, 2014.

H. Yarmand, J.S. Ivy, B. Denton, and A.L. Lloyd. Optimal two-phase vaccine allocation to geographically different regions under uncertainty. *European Journal of Operational Research*, 233(1):208–219, 2014.

HR Yun, LJ Lee, JH Park, YK Cho, YB Cho, WY Lee, HC Kim, HK Chun, and SH Yun. Local recurrence after curative resection in patients with colon and rectal cancers. *International journal of colorectal disease*, 23(11):1081–1087, 2008.

Michel Zaffran, Jos Vandelaer, Debra Kristensen, Bjørn Melgaard, Prashant Yadav, KO Antwi-Agyei, and Heidi Lasher. The imperative for stronger vaccine supply and logistics systems. *Vaccine*, 31:B73–B80, 2013.

Jane Zapka, Carrie N Klabunde, Stephen Taplin, Gigi Yuan, David Ransohoff, and Sarah Kobrin. Screening colonoscopy in the us: attitudes and practices of primary care physicians. *Journal of general internal medicine*, 27(9):1150–1158, 2012.

Ann G Zauber, Iris Lansdorp-Vogelaar, Amy B Knudsen, Janneke Wilschut, Marjolein van Ballegooijen, and Karen M Kuntz. Evaluating test strategies for colorectal cancer screening: a decision analysis for the us preventive services task force. *Annals of internal medicine*, 149(9):659–669, 2008.

Ann G Zauber, Iris Lansdorp-Vogelaar, Amy B Knudsen, Janneke Wilschut, Marjolein van Ballegooijen, and Karen M Kuntz. Evaluating test strategies for colorectal cancer screening—age to begin, age to stop, and timing of screening intervals. 2009.

Ann G Zauber, Sidney J Winawer, Michael J O'Brien, Iris Lansdorp-Vogelaar, Marjolein van Ballegooijen, Benjamin F Hankey, Weiji Shi, John H Bond, Melvin Schapiro, Joel F Panish, et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *New England Journal of Medicine*, 366(8):687–696, 2012.

Jingyu Zhang, Jennifer E Mason, Brian T Denton, and William P Pierskalla. Applications of operations research to the prevention, detection, and treatment of disease. *Encyclopedia of Operations Research and Management, Wiley*, 2011.

APPENDICES

Appendix A

Proofs of Structural Properties

Recall that $V(t, \mathbf{q}, h) = \max\{g_1(t, \mathbf{q}, h), \cdots, g_n(t, \mathbf{q}, h), I_{h_c}\nu(t-1, \mathbf{q})\}$. The property given in Equation A.1 guarantees that proving the monotonicity of $g_i(t, \mathbf{q}, h), \forall\, i \in \mathbb{I}$ is sufficient to prove that $V(t, \mathbf{q}, h)$ is monotone.

$$\max(a', b') \leq \max(a, b') \leq \max(a, b) \text{ if } a \geq a',\ b \geq b' \tag{A.1}$$

A.1 Proof of Proposition 1

As $\nu(t, \mathbf{q})$ is independent of h, showing that $g_i(t, \mathbf{q}, h)$ is non-increasing in h for all $h \in \mathbb{H}$ is sufficient to prove that the optimal value function $V(t, \mathbf{q}, h)$ is non-increasing in h.

Proof by induction: We start the proof with a preliminary property. Suppose that a type i vial is opened at timeslot h due to the arrival of a new vaccine demand. Then either there will be less than or equal to z_i vaccine demands at and after timeslot h, or the opened vial will be depleted before the end of the day and a new vaccine demand arrival

will occur. Therefore, the following relation holds for any given $i \in \mathbb{I}$ and $h \in \mathbb{H}$:

$$\sum_{d=1}^{z_i} p_{D_h}(d) + \sum_{y=h+z_i}^{\eta} p_{Y_h^i}(y) = 1 \quad \Rightarrow \sum_{y=h+z_i}^{\eta} p_{Y_h^i}(y) = 1 - \sum_{d=1}^{z_i} p_{D_h}(d) = \sum_{d=z_i+1}^{\eta-h+1} p_{D_h}(d). \tag{A.2}$$

Step 1: First, we show that $g_i(t, \mathbf{q}, h)$ is non-increasing in h for any $t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$, $i \in \mathbb{I}$ where $\sum_{i=1}^{n} q_i \leq 1$. Note that if $q_i = 0$, $g_i(t, \mathbf{q}, h) = 0$ for all $h \in \mathbb{H}$, thus, the monotonicity holds for all $t \in \mathbb{T}$. If $q_i = 1$, then;

$$\begin{aligned}
g_i(t, \mathbf{q}, h+1) &= \sum_{y=h+z_i+1}^{\eta} (z_i + V(t, \mathbf{q} - \mathbf{e_i}, y)) p_{Y_{h+1}^i}(y) + \sum_{d=1}^{z_i} (d + \nu(t-1, \mathbf{q} - \mathbf{e_i})) p_{D_{h+1}}(d) \\
&= \sum_{y=h+z_i+1}^{\eta} z_i p_{Y_{h+1}^i}(y) + \sum_{d=1}^{z_i} d p_{D_{h+1}}(d) \\
&= \sum_{y=z_i+1}^{\eta-h} z_i p_{D_{h+1}}(y) + \sum_{d=1}^{z_i} d p_{D_{h+1}}(d) \qquad \text{(A.3)} \\
&\leq \sum_{y=z_i+1}^{\eta-h+1} z_i p_{D_h}(y) + \sum_{d=1}^{z_i} d p_{D_h}(d) \qquad \text{(A.4)} \\
&= g_i(t, \mathbf{q}, h)
\end{aligned}$$

Equation A.3 follows the relation shown in Equation A.2. Note that the right-hand sides of Equation A.3 and Inequality A.4 are equal to the expected value of two truncated binomial distributions. Inequality A.4 holds due to a property of truncated binomial distribution: let X_1 and X_2 be two independent truncated binomial random variables with the same truncation level and parameters (n_1, p) and (n_2, p), respectively, then $E(X_1) \leq E(X_2)$ when $n_1 \leq n_2$ (Klenke et al., 2010).

Step n: Suppose $V(t, \mathbf{q}, h)$ is non-increasing in h for all $t \in \mathbb{T}$ and $\mathbf{q} \in \mathbb{Q}$ such that $\sum_{i=1}^{n} q_i \leq n-1$ (*the induction assumption*). We now show that $g_i(t, \hat{\mathbf{q}}, h)$ is non-increasing

in h for any $t \in \mathbb{T}$, $\hat{\mathbf{q}} \in \mathbb{Q}$, $i \in \mathbb{I}$ where $\sum_{i=1}^{n} \hat{q}_i = n$.

The following preliminaries are required to prove monotonicity in Step n.

$$\begin{aligned} P_{Y_{h+1}^i}(\eta+1) &= 1 - \sum_{y=h+1+z_i}^{\eta} P_{Y_{h+1}^i}(y) \\ &= 1 - \sum_{y=h+1+z_i}^{\eta} P_{Y_h^i}(y-1) + (P_{Y_h^i}(\eta) - \end{aligned} \tag{A.5}$$

$$= P_{Y_h^i}(\eta+1) + P_{Y_h^i}(\eta), \quad \text{for all } h \in \mathbb{H}, \tag{A.6}$$

$$\nu(t-1,\mathbf{q}) \leq V(t,\mathbf{q},\eta), \quad \text{for all } t \in \mathbb{T},\ \mathbf{q} \in \mathbb{Q} \text{ s.t. } \sum_{i=1}^{n} q_i \leq n-1 \tag{A.7}$$

Equation A.5 is due to the fact that $P_{Y_{h+1}^i}(y+1) = P_{Y_h^i}(y)$. Moreover, $\nu(t-1,\mathbf{q}) \leq V(t,\mathbf{q},\eta)$ because we assume that the last patient may arrive at the beginning of timeslot η, thus, this inequality holds. Note that if $\hat{q}_i = 0$, $g_i(t,\hat{\mathbf{q}},h) = 0$ for all $h \in \mathbb{H}$, thus, the monotonicity holds. If $\hat{q}_i > 0$, then;

$$
\begin{aligned}
g_i(t,\hat{\mathbf{q}},h+1) &= \sum_{y=h+z_i+1}^{\eta}(z_i+V(t,\hat{\mathbf{q}}-\mathbf{e_i},y))p_{Y^i_{h+1}}(y)+\sum_{d=1}^{z_i}(d+\nu(t-1,\hat{\mathbf{q}}-\mathbf{e_i}))p_{D_{h+1}}(d)\\
&= \sum_{y=h+z_i+1}^{\eta} z_i p_{Y^i_{h+1}}(y)+\sum_{d=1}^{z_i} d p_{D_{h+1}}(d)+\sum_{y=h+z_i+1}^{\eta} V(t,\hat{\mathbf{q}}-\mathbf{e_i},y)p_{Y^i_{h+1}}(y)\\
&\quad+\sum_{d=1}^{z_i}\nu(t-1,\hat{\mathbf{q}}-\mathbf{e_i})p_{D_{h+1}}(d)\\
&= \sum_{y=z_i+1}^{\eta-h} z_i p_{D_{h+1}}(y)+\sum_{d=1}^{z_i} d p_{D_{h+1}}(d)+\sum_{y=h+z_i}^{\eta-1} V(t,\hat{\mathbf{q}}-\mathbf{e_i},y+1)p_{Y^i_h}(y)\\
&\quad+\nu(t-1,\hat{\mathbf{q}}-\mathbf{e_i})p_{Y^i_{h+1}}(\eta+1) \qquad \text{(A.8)}\\
&\leq \sum_{y=z_i+1}^{\eta-h+1} z_i p_{D_h}(y)+\sum_{d=1}^{z_i} d p_{D_h}(d)+\sum_{y=h+z_i}^{\eta-1} V(t,\hat{\mathbf{q}}-\mathbf{e_i},y)p_{Y^i_h}(y)\\
&\quad+\nu(t-1,\hat{\mathbf{q}}-\mathbf{e_i})[p_{Y^i_h}(\eta)+p_{Y^i_h}(\eta+1)] \qquad \text{(A.9)}\\
&\leq \sum_{y=z_i+1}^{\eta-h+1} z_i p_{D_h}(y)+\sum_{d=1}^{z_i} d p_{D_h}(d)+\sum_{y=h+z_i}^{\eta} V(t,\hat{\mathbf{q}}-\mathbf{e_i},y)p_{Y^i_h}(y)\\
&\quad+\nu(t-1,\hat{\mathbf{q}}-\mathbf{e_i})p_{Y^i_h}(\eta+1) \qquad \text{(A.10)}\\
&= \sum_{y=h+z_i}^{\eta}(z_i+V(t,\hat{\mathbf{q}}-\mathbf{e_i},y))p_{Y^i_h}(y)+\sum_{d=1}^{z_i}(d+\nu(t-1,\hat{\mathbf{q}}-\mathbf{e_i}))p_{D_h}(d)\\
&= g_i(t,\hat{\mathbf{q}},h)
\end{aligned}
$$

As in Step 1, Equation A.8 follows the relation shown in Equation A.2 and $P_{Y^i_{h+1}}(y+1) = P_{Y^i_h}(y)$. The transition between Equations A.8 and A.9 is obtained from the property derived in Equation A.6 and the property of truncated binomial distribution used in Inequality A.4. Note that $V(t,\hat{\mathbf{q}}-\mathbf{e_i},y+1)$ in Equation A.8 is replaced with $V(t,\hat{\mathbf{q}}-\mathbf{e_i},y)$ in Equation A.9 following the induction assumption that $V(t,\hat{\mathbf{q}}-\mathbf{e_i},h+1) \leq V(t,\hat{\mathbf{q}}-\mathbf{e_i},h)$, $\forall \mathbf{q} \in \mathbb{Q}$ s.t. $\mathbf{q} = \hat{\mathbf{q}} - \mathbf{e_i}$, $\forall i \in \mathbb{I}$, and thus, $\sum_{i=1}^{n} q_i = n-1$. Moreover, the transition to Equation A.10 follows Equation A.7. Hence, Proposition 1 holds for all $t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$. ■

A.2 Proof of Theorem 1

Proposition 1 shows that if $1 \leq h_1 \leq h_2 \leq \eta$ and $h_c = 0$, then $g_i(t, \mathbf{q}, h_1) \geq g_i(t, \mathbf{q}, h_2)$ for all $i \in \mathbb{I}$, $t \in \mathbb{T}$, $\mathbf{q} \in \mathbb{Q}$. Following this, once $\nu(t-1, \mathbf{q}) \geq g_i(t, \mathbf{q}, h)$ at any $h \in \mathbb{H}$, then; $\nu(t-1, \mathbf{q}) \geq g_i(t, \mathbf{q}, \hat{h})$ for all $i \in \mathbb{I}$, $t \in \mathbb{H}$, $\mathbf{q} \in \mathbb{Q}$, $\hat{h} \in \mathbb{H}$ such that $\hat{h} \geq h$. Thus, for any $h \geq h^*_N(t, \mathbf{q}) = \min_h \left\{ \nu(t-1, \mathbf{q}) \geq \max_{i \in \mathbb{I}} \{g_i(t, \mathbf{q}, \mathbf{h})\} \right\}$, $A^*_{(t, \mathbf{q}, h)} = N$ must hold because $\nu(t-1, \mathbf{q})$ is guaranteed to be greater than or equal to $g_i(t, \mathbf{q}, h)$ for those h values by Proposition 1. ■

A.3 Proof of Proposition 2

A.3.1 Continuous Service Scenario

Step 1: We first show that $V(t, \mathbf{q}, h)$ is non-decreasing in t for all $h \in \mathbb{H}$ and $\mathbf{q} \in \mathbb{Q}$ such that $\sum_{i=1}^n q_i = 1$. Recall that $V(t, \mathbf{0}, h) = 0$ for all $t \in \mathbb{T}$ and $h \in \mathbb{H}$ regardless of the service scenario and $g_i(t, \mathbf{q}, h) = 0$ for all $t \in \mathbb{T}$, $h \in \mathbb{H}$, $i \in \mathbb{I}$, and $\mathbf{q} \in \mathbb{Q}$ s.t. $q_i = 0$ and $\sum_{i=1}^n q_i = 1$. When $h_c = 0$ and $\sum_{i=1}^n q_i = 1$, $V(t, \mathbf{q}, h)$ is independent from t for all $t \in \mathbb{T}$, $h \in \mathbb{H}$ as shown in Equations A.11 and A.12.

$$V(t, \mathbf{q}, h) = \max\{g_1(t, \mathbf{q}, h), \cdots, g_n(t, \mathbf{q}, h)\} = g_i(t, \mathbf{q}, h), \quad \text{if } q_i = 1,\ q_j = 0\ \forall j \neq i,\ i, j \in \mathbb{I} \tag{A.11}$$

$$\begin{aligned} g_i(t, \mathbf{q}, h) &= \sum_{y=h+z_i}^{\eta} (z_i + V(t, \mathbf{0}, y)) p_{Y_h^i}(y) + \sum_{d=1}^{z_i} (d + \nu(t-1, \mathbf{0})) p_{D_h}(d) \\ &= \sum_{y=h+z_i}^{\eta} z_i p_{Y_h^i}(y) + \sum_{d=1}^{z_i} d p_{D_h}(d), \end{aligned} \tag{A.12}$$

Since Equation A.12 is independent of t, Proposition 2 holds when $h_c = 0$ and $\sum_{i=1}^n q_i = 1$.

Step n: Suppose $V(t, \mathbf{q}, h)$ is non-decreasing in t for all $h \in \mathbb{H}$ and $\mathbf{q} \in \mathbb{Q}$ such that

$\sum_{i=1}^{n} q_i \leq n-1$ (*the induction assumption*). We now show that $V(t+1, \mathbf{q}, h) \geq V(t, \mathbf{q}, h)$ for all $t \in \mathbb{T}$, $h \in \mathbb{H}$ and $\mathbf{q} \in \mathbb{Q}$ such that $\sum_{i=1}^{n} q_i = n$.

$$\begin{aligned} g_i(t+1, \mathbf{q}, h) &= \sum_{y=h+z_i}^{\eta} (z_i + V(t+1, \mathbf{q} - \mathbf{e_i}, y)) p_{Y_h^i}(y) + \sum_{d=1}^{z_i} (d + \nu(t, \mathbf{q} - \mathbf{e_i})) p_{D_h}(d), \\ &\geq \sum_{y=h+z_i}^{\eta} (z_i + V(t, \mathbf{q} - \mathbf{e_i}, y)) p_{Y_h^i}(y) + \sum_{d=1}^{z_i} (d + \nu(t-1, \mathbf{q} - \mathbf{e_i})) p_{D_h}(d), \\ &\geq g_i(t, \mathbf{q}, h) \end{aligned} \tag{A.13}$$

Considering the induction assumption, the transition to Equation A.13 holds as $V(t+1, \mathbf{q} - \mathbf{e_i}, h) \geq V(t, \mathbf{q} - \mathbf{e_i}, h)$ for all $h \in \mathbb{H}$, $\mathbf{q} \in \mathbb{Q}$, $i \in \mathbb{I}$.

Hence, $V(t+1, \mathbf{q}, h) = \max\{g_1(t+1, \mathbf{q}, h), \cdots, g_n(t+1, \mathbf{q}, h)\} \geq \max\{g_1(t, \mathbf{q}, h), \cdots, g_n(t, \mathbf{q}, h)\} = V(t, \mathbf{q}, h)$ for all $t \in \mathbb{T}$, $h \in \mathbb{H}$ and $\mathbf{q} \in \mathbb{Q}$ such that $\sum_{i=1}^{n} q_i = n$. ■

A.3.2 Terminated Service Scenario

Let $\mathcal{P}_{\bar{t}}^*$ be the optimal policy that specifies the optimal actions from $\bar{t}$ to the end of the current replenishment cycle when $h_c = 0$, i.e., $\mathcal{P}_{\bar{t}}^* = \{A_s^* \in \mathbb{A}_s : s = (t, \mathbf{q}, h) \in \mathbb{S}, \bar{t} \geq t \geq 1\}$. Because termination of vaccination service is allowed, we can use the optimal policy for $t = \bar{t}$ (i.e., $\mathcal{P}_{\bar{t}}^*$) to generate a feasible policy for any t' such that $T \geq t' \geq \bar{t}$. We refer to this feasible policy as $\mathcal{P}_{t'}^f$ and define it for all $s = (t, \mathbf{q}, h) \in \mathbb{S}$ as follows:

$$\mathcal{P}_{t'}^f(s) = \begin{cases} \mathcal{P}_{\bar{t}}^*(s) = A_s^* & \text{if } 1 \leq t \leq \bar{t} \\ N, & \text{if } \bar{t} < t < t' \end{cases} \tag{A.14}$$

Let $V^{\mathcal{P}_{t'}^f}(t', \mathbf{q}, h)$ refer to the expected total number of vaccine doses administered under policy $\mathcal{P}_{t'}^f$. Since the performance of the $\mathcal{P}_{t'}^f$ in expected demand covered matches to that

of $\mathcal{P}_{\bar{t}}^{*}$, the monotonicity of the optimal value function in t holds. That is:

$$V(\bar{t}, \mathbf{q}, h) = V^{\mathcal{P}_{t'}^{f}}(t', \mathbf{q}, h) \leq V(t', \mathbf{q}, h) \tag{A.15}$$

The last inequality above holds because the value function of any feasible policy is smaller than that of the optimal policy in a maximization problem. ■

A.4 Proof of Proposition 3

Because the proofs of Propositions 3.a and 3.b are very similar, we provide the proof of Proposition 3.a. and omit the other one for brevity.

A.4.1 Continuous Service Scenario

This proof adopts *the proof by induction* method for which we consider that there are $T \times \eta$ decision epochs (steps) in a replenishment cycle. Therefore, Step n refers to the number of decision epochs starting from state $s = (t, \mathbf{q}, h)$ until the next replenishment cycle, i.e., $n = (t \times \eta) + (\eta - h)$.

Step 1: We start with showing that $V(1, \mathbf{q}, h) \geq V(1, \mathbf{q} - \mathbf{r}, h)$ for all $\mathbf{q} \in \mathbb{Q}$, $\mathbf{r} = (r_1 \cdots r_n)$ such that $\sum_{i=1}^{n} r_i \geq 0$ and $r_i \leq q_i \ \forall i \in \mathbb{I}$, $h \in \mathbb{H}$. Note that in the last day of replenishment cycle, the expected number of vaccine doses administered starting from timeslot h to the beginning of next replenishment cycle is the same regardless of what size of a vial is opened at timeslot h for all $h \in \mathbb{H}$. Since no terminal reward assigned to the ending inventory in a replenishment cycle, any size of a vial can be opened when demand is observed. Thus, $V(1, \mathbf{q}, h)$ is equal to the expectation of a truncated binomial distribution.

Let $L^{\mathbf{q}}$ be the total number of doses on-hand at the beginning of timeslot h when the

available inventory is $\mathbf{q}$, i.e., $L^{\mathbf{q}} = \sum_{i=1}^{n} z_i q_i$. Then;

$$
\begin{aligned}
V(1,\mathbf{q},h) =& \sum_{d=1}^{L^{\mathbf{q}}} d p_{D_h}(d) + \sum_{d=L^{\mathbf{q}}+1}^{\eta-h} L^{\mathbf{q}} p_{D_h}(d) \\
=& \sum_{d=1}^{L^{\mathbf{q}-\mathbf{r}}} d p_{D_h}(d) + \sum_{d=L^{\mathbf{q}-\mathbf{r}}+1}^{L^{\mathbf{q}}} d p_{D_h}(d) + \sum_{d=L^{\mathbf{q}}+1}^{\eta-h} L^{\mathbf{q}} p_{D_h}(d) && \text{(A.16)} \\
\geq& \sum_{d=1}^{L^{\mathbf{q}-\mathbf{r}}} d p_{D_h}(d) + \sum_{d=L^{\mathbf{q}-\mathbf{r}}+1}^{L^{\mathbf{q}}} L^{\mathbf{q}-\mathbf{r}} p_{D_h}(d) + \sum_{d=L^{\mathbf{q}}+1}^{\eta-h} L^{\mathbf{q}-\mathbf{r}} p_{D_h}(d) && \text{(A.17)} \\
=& \sum_{d=1}^{L^{\mathbf{q}-\mathbf{r}}} d p_{D_h}(d) + \sum_{d=L^{\mathbf{q}-\mathbf{r}}+1}^{\eta-h} L^{\mathbf{q}-\mathbf{r}} p_{D_h}(d) && \text{(A.18)} \\
=& V(1,\mathbf{q}-\mathbf{r},h)
\end{aligned}
$$

Note that the transition between Equations A.16 and A.17 is due to the fact that $L^{\mathbf{q}} = \sum_{i=1}^{n} z_i q_i \geq \sum_{i=1}^{n} z_i (q_i - r_i) = L^{\mathbf{q}-\mathbf{r}}$. Hence, the proposition holds when $t = 1$.

Step n: Suppose that $V(t,\mathbf{q},h) \geq V(t,\mathbf{q}-\mathbf{r},h)$ for all $t \in \mathbb{T}$, $h \in \mathbb{H}$ such that $(t \times \eta) + (\eta - h) < (\hat{t} \times \eta) + (\eta - \hat{h})$ (*the induction assumption*). We now show that $V(\hat{t},\mathbf{q},\hat{h}) \geq V(\hat{t},\mathbf{q}-\mathbf{r},\hat{h})$ for all $\mathbf{q} \in \mathbb{Q}$, $\mathbf{r} = (r_1 \cdots r_n)$ such that $\sum_{i=1}^{n} r_i \geq 0$ and $r_i \leq q_i$ for all $i \in \mathbb{I}$ in Step n (i.e., $n = (\hat{t} \times \eta) + (\eta - \hat{h})$). Let $V(\hat{t},\mathbf{q},\hat{h}) = g_i(\hat{t},\mathbf{q},\hat{h})$, $i \in \mathbb{I}$.

$$
\begin{aligned}
g_i(\hat{t},\mathbf{q},\hat{h}) =& \sum_{y=h+z_i}^{\eta} (z_i + V(\hat{t},\mathbf{q}-\mathbf{e_i},y)) p_{Y_h^i}(y) + \sum_{d=1}^{z_i} (d + \nu(\hat{t}-1,\mathbf{q}-\mathbf{e_i})) p_{D_h}(d) && \text{(A.19)} \\
\geq& \sum_{y=h+z_i}^{\eta} (z_i + V(\hat{t},\mathbf{q}-\mathbf{r}-\mathbf{e_i},y)) p_{Y_h^i}(y) + \sum_{d=1}^{z_i} (d + \nu(\hat{t}-1,\mathbf{q}-\mathbf{r}-\mathbf{e_i})) p_{D_h}(d), && \text{(A.20)} \\
=& g_i(\hat{t},\mathbf{q}-\mathbf{r},\hat{h})
\end{aligned}
$$

Note that the transition from Equation A.19 to A.20 is due to the induction assumption guaranteeing that (i) $V(\hat{t},\mathbf{q}-\mathbf{e_i},y) \geq V(\hat{t},\mathbf{q}-\mathbf{r}-\mathbf{e_i},y)$ for all $y \geq \hat{h}$ and (ii) $\nu(\hat{t}-1,\mathbf{q}-\mathbf{e_i}) =$

$\sum_{h=1}^{\eta+1} V(\hat{t}-1, \mathbf{q}-\mathbf{e_i}, h)p_X(h) \geq \sum_{h=1}^{\eta+1} V(\hat{t}-1, \mathbf{q}-\mathbf{r}-\mathbf{e_i}, h)p_X(h) = \nu(\hat{t}-1, \mathbf{q}-\mathbf{r}-\mathbf{e_i})$.

Hence, Proposition 3 holds when $h_c = \eta + 1$. ■

A.4.2 Terminated Service Scenario

We follow a similar proof structure as the one in Section A.3.2. Let $\mathcal{P}^*_{\mathbf{q}-\mathbf{r}}$ denote the optimal policy under the terminated service scenario that specifies the optimal actions from state s to the end of the replenishment cycle when the inventory on-hand in state $s = (t, \mathbf{q}-\mathbf{r}, h)$ is $\mathbf{q}-\mathbf{r}$ and $\mathbf{r} = (r_1 \cdots r_n)$ such that $\sum_{i=1}^{n} r_i \geq 0$, $r_i \leq q_i \; \forall \; i \in \mathbb{I}$. Suppose $\mathcal{P}^f_{\mathbf{q}}$ is the feasible policy given in Equation A.21 when the inventory on-hand in state s is $\mathbf{q}$.

$$\mathcal{P}^f_{\mathbf{q}}(s') = \begin{cases} \mathcal{P}^*_{\mathbf{q}-\mathbf{r}}(s') = A^*_{s'} & \text{if } q'_i \leq q_i - r_i,\ t' \leq t,\ (t \times \eta) + (\eta - h) \leq (t' \times \eta) + (\eta - h') \\ N, & \text{otherwise} \end{cases} \quad \forall s' \in \mathbb{S} \tag{A.21}$$

Let $V^{\mathcal{P}^*_{\mathbf{q}-\mathbf{r}}}(t, \mathbf{q}-\mathbf{r}, h)$ refer to the expected total number of vaccine doses administered under policy $\mathcal{P}^*_{\mathbf{q}-\mathbf{r}}$. Since the performance of the $\mathcal{P}^*_{\mathbf{q}-\mathbf{r}}$ in expected demand covered matches to that of $\mathcal{P}^f_{\mathbf{q}}$, the monotonicity of the optimal value function in $\mathbf{q}$ holds. That is:

$$V(t, \mathbf{q}-\mathbf{r}, h) = V^{\mathcal{P}^*_{\mathbf{q}-\mathbf{r}}}(t, \mathbf{q}-\mathbf{r}, h) = V^{\mathcal{P}^f_{\mathbf{q}}}(t, \mathbf{q}, h) \leq V(t, \mathbf{q}, h) \tag{A.22}$$

Above the last inequality holds because the value function of any feasible policy is smaller than that of the optimal policy in a maximization problem. ■

Appendix B

The Specifics of The Pediatric Vaccines

We conduct our numerical analysis using data about four routine pediatric vaccines: measles, DTP-HepB-Hib (Pentavalent), yellow fever (YF), and BCG vaccines. Table B.1 presents some of the specifics of these vaccines (GAVI, 2017). Note that the information in Table B.1 combines the information from both GAVI (2017) and WHO (2017). This combined information is for a general vaccine administration setting and may differ based on country's and/or patient's risk profile, manufacturer of the vaccines, etc. Open shelf-lives are reported as six to eight hours for the vaccines not containing an effective preservative. Opened multi-dose vaccine vials should be discarded at the end of the immunization session which takes six hours on average. If the storage conditions satisfy the requirements listed in WHO (2014a), pentavalent vaccine can be used for the subsequent immunization session up to a maximum of 28 days based on the assumption that stock is replenished once a month. However, the maximum performance of the preservative is not guaranteed. Considering that our study specifically focuses on the vaccine administration settings in developing countries where lack of appropriate storage conditions is presence, we assume that the open shelf-life is 6-8 hours for pentavalent vaccine as well. The column showing the stability of the vaccines refers to the vaccine's durability against heat exposure.

Table B.1: The pediatric vaccine specifics

Vaccine type	Presentation (doses)	WHO recommended schedule†			Shelf-life (months)	Open shelf-life‡	Stability
		Dose	Age (months)				
			Routine	Campaign			
Measles	1^o, 2^o, 5^o, 10, 20^o	2	9-18	9-60	24, 2-8 ^{0}C	6-8 hrs	Med
Pentavalent	1, 2, 5, 10	3	1.5	NA	36, 2-8 ^{0}C	28 days	Med
BCG	10^o, 20^o	1	> 0	NA	24, 2-8 ^{0}C	6-8 hrs	Med, High
YF	2^o, 5, 10, 20	1	9-12	NA	24, 2-8 ^{0}C	6-8 hrs	Med

†: only provides number of doses in primary series and age interval for the vaccine.
‡: See WHO (2014a) for the details of multi-dose vial policy of WHO.
NA stands for *not available*.
> 0 stands for *as soon as after birth*.
^{o}: vial presentations that are not on the GAVI's menu for vaccines. The vaccines on the GAVI's menu are offered by GAVI and listed in the country's application portal as well as in the WHO's pre-qualified vaccine list.

Appendix C

Determining the Total Number of Timeslots in a Clinic Day (η)

We validate the Poisson arrival assumption by setting the number of timeslots in a clinic day to a sufficiently large value such that the probability that more than one arrival occurs in a timeslot is negligible. We estimate the performance measures, $\phi(T, \mathbf{Q})$ and $\omega(T, \mathbf{Q})$ under the optimal policies for various η values, i.e., $\eta \in \{32, 96, 480, 960, 1920\}$ in an $8-$hour clinic day. The probability that an arrival occurs in a timeslot, p, also varies with η for a given daily demand rate as $p = \mu/\eta$. Table C.1 shows that the performance improvement achieved by increasing η is negligible when $\eta \geq 480$. Hence, we can conclude that $\eta = 480$ is sufficiently large for our analyses under the Poisson arrival assumption.

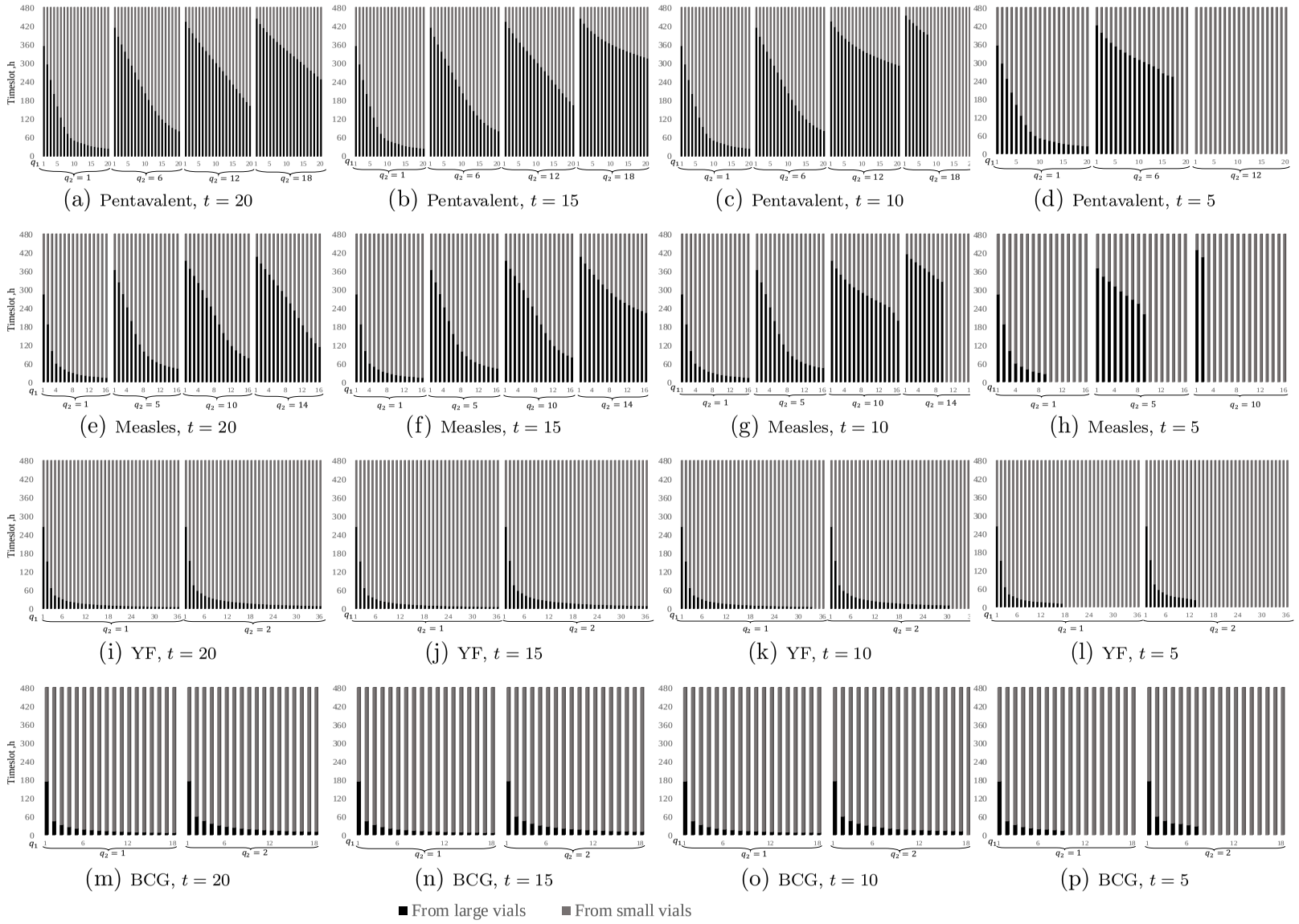
Table C.1: Performances of the optimal policies with various η values and base case T, Q, and μ

				$\eta = 32$	$\eta = 96$	$\eta = 480$	$\eta = 960$	$\eta = 1920$
Pentavalent Vaccine	*Terminated service*	$\alpha_1 = 18\%$	$\phi(T, \mathbf{Q})$	94.21	93.12	92.71	92.66	92.64
			$\omega(T, \mathbf{Q})$	5.72	6.79	7.19	7.24	7.26
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	89.24	88.34	87.99	87.95	87.92
			$\omega(T, \mathbf{Q})$	10.59	11.49	11.83	11.87	11.89
	Continuous service	$\alpha_1 = 18\%$	$\phi(T, \mathbf{Q})$	91.68	90.37	89.91	89.86	89.83
			$\omega(T, \mathbf{Q})$	8.28	9.58	10.04	10.09	10.12
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	70.91	71.59	71.77	71.79	71.80
			$\omega(T, \mathbf{Q})$	29.09	28.41	28.23	28.21	28.20
Measles Vaccine	*Terminated service*	$\alpha_1 = 18\%$	$\phi(T, \mathbf{Q})$	92.02	91.11	90.76	90.71	90.69
			$\omega(T, \mathbf{Q})$	7.91	8.80	9.15	9.19	9.21
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	89.24	88.34	87.99	87.95	87.92
			$\omega(T, \mathbf{Q})$	10.59	11.49	11.83	11.87	11.89
	Continuous service	$\alpha_1 = 18\%$	$\phi(T, \mathbf{Q})$	82.10	81.92	81.79	81.78	81.77
			$\omega(T, \mathbf{Q})$	17.90	18.08	18.21	18.22	18.23
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	70.91	71.59	71.77	71.79	71.80
			$\omega(T, \mathbf{Q})$	29.09	28.41	28.23	28.21	28.20
YF Vaccine	*Terminated service*	$\alpha_1 = 82\%$	$\phi(T, \mathbf{Q})$	88.43	87.97	87.80	87.78	87.77
			$\omega(T, \mathbf{Q})$	11.44	11.85	12.01	12.03	12.03
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	57.33	57.68	57.79	57.80	57.81
			$\omega(T, \mathbf{Q})$	42.67	42.32	42.21	42.20	42.19
	Continuous service	$\alpha_1 = 82\%$	$\phi(T, \mathbf{Q})$	80.06	80.15	80.16	80.16	80.16
			$\omega(T, \mathbf{Q})$	19.94	19.85	19.84	19.84	19.84
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	54.98	54.87	54.79	54.78	54.77
			$\omega(T, \mathbf{Q})$	45.02	45.13	45.21	45.22	45.23
BCG Vaccine	*Terminated service*	$\alpha_1 = 82\%$	$\phi(T, \mathbf{Q})$	87.07	85.90	85.45	85.40	85.37
			$\omega(T, \mathbf{Q})$	12.93	14.10	14.54	14.60	14.63
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	57.33	57.68	57.79	57.80	57.81
			$\omega(T, \mathbf{Q})$	42.67	42.32	42.21	42.20	42.19
	Continuous service	$\alpha_1 = 82\%$	$\phi(T, \mathbf{Q})$	68.74	69.31	69.45	69.47	69.47
			$\omega(T, \mathbf{Q})$	31.26	30.69	30.55	30.53	30.53
		$\alpha_1 = 0\%$	$\phi(T, \mathbf{Q})$	54.98	54.87	54.79	54.78	54.77
			$\omega(T, \mathbf{Q})$	45.02	45.13	45.21	45.22	45.23

Appendix D

Optimal Vaccine Administration Decisions with Continuous Service

Figure D.1 shows the optimal actions at all timeslots on particular days ($t \in \{5, 10, 15, 20\}$) for various inventory levels (q_1, q_2) and vaccine types under the base case continuous service scenario ($T = 20$, $Q = 220$, $\eta = 480$, and $\mu = 11$). The optimal policies generally follow a monotonic pattern in terms of timeslot thresholds for switching to small vials, and administer doses more conservatively when the total number of doses on-hand is not sufficiently large compared to the total expected demand until the next replenishment cycle.



(a) Pentavalent, $t = 20$ (b) Pentavalent, $t = 15$ (c) Pentavalent, $t = 10$ (d) Pentavalent, $t = 5$

(e) Measles, $t = 20$ (f) Measles, $t = 15$ (g) Measles, $t = 10$ (h) Measles, $t = 5$

(i) YF, $t = 20$ (j) YF, $t = 15$ (k) YF, $t = 10$ (l) YF, $t = 5$

(m) BCG, $t = 20$ (n) BCG, $t = 15$ (o) BCG, $t = 10$ (p) BCG, $t = 5$

Figure D.1: The optimal vaccine administration decisions for various (t, q_1, q_2) combinations under base case parameter scenario with continuous service

Appendix E

Base Case Performance of the Optimal Policies

Recall that $\bar{\alpha}_1$ is the minimum α_1 value for which the optimal policy can attain more than 95% of the improvement in $\phi(T, \mathbf{Q})$ achieved by the optimal policy with $\alpha_1 = 100\%$. We present the required $\bar{\alpha}_1$ values and their performances for all vaccine types in Table E.1. Note that $\pi(T, \mathbf{Q})_\alpha$, $\phi(T, \mathbf{Q})_\alpha$, and $\omega(T, \mathbf{Q})_\alpha$ refer to the expected total cost, percentage demand covered, and percentage open-vial wastage associated with the optimal policy with $\bar{\alpha}_1 = \alpha_1$. Table E.1 shows that when the size of large vials decreases, the corresponding $\bar{\alpha}_1$ value also decreases, e.g. $\bar{\alpha_1}$ is 18% for measles vaccines held in 5-dose and 10-dose vials; whereas, $\bar{\alpha_1}$ is 82% for YF vaccines held in 5-dose and 20-dose vials under the base case scenario.

Table E.1: Comparison of the optimal policies for single and multiple vial size cases in terms of relative performance differences in $\pi(T, \mathbf{Q})$, $\phi(T, \mathbf{Q})$, and $\omega(T, \mathbf{Q})$

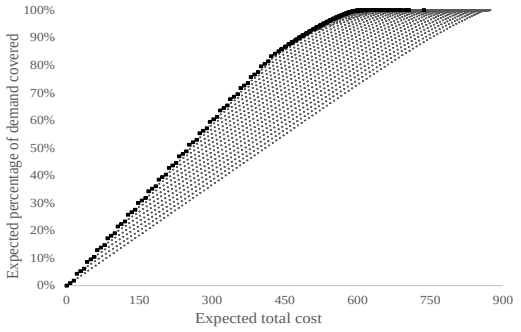
			Relative difference between $\alpha_1 = 0\%$ and $\alpha_1 = \bar{\alpha}_1$			Relative difference between $\alpha_1 = 0\%$ and $\alpha_1 = 100\%$		
		$\bar{\alpha}_1$	$\pi(T, \mathbf{Q})^1$	$\phi(T, \mathbf{Q})^1$	$\omega(T, \mathbf{Q})^1$	$\pi(T, \mathbf{Q})^2$	$\phi(T, \mathbf{Q})^2$	$\omega(T, \mathbf{Q})^2$
Pentavalent	$h_c = \eta + 1$	18%	14%	18%	−18%	75%	23%	−24%
	$h_c = 0$		14%	5%	−5%	75%	8%	−9%
Measles	$h_c = \eta + 1$	18%	23%	10%	−10%	127%	13%	−13%
	$h_c = 0$		23%	3%	−3%	128%	4%	−4%
YF	$h_c = \eta + 1$	82%	19%	26%	−26%	23%	30%	−30%
	$h_c = 0$		18%	30%	−30%	22%	34%	−34%
BCG	$h_c = \eta + 1$	82%	52%	15%	−15%	64%	17%	−17%
	$h_c = 0$		52%	30%	−30%	63%	30%	−30%

$\pi(T, \mathbf{Q})^1 = \frac{\pi(T, \mathbf{Q})_{\bar{\alpha}_1} - \pi(T, \mathbf{Q})_{0\%}}{\pi(T, \mathbf{Q})_{0\%}} \times 100\%$, $\pi(T, \mathbf{Q})^2 = \frac{\pi(T, \mathbf{Q})_{100\%} - \pi(T, \mathbf{Q})_{0\%}}{\pi(T, \mathbf{Q})_{0\%}} \times 100\%$, $\phi(T, \mathbf{Q})^1 = \phi(T, \mathbf{Q})_{\bar{\alpha}_1} - \phi(T, \mathbf{Q})_{0\%}$
$\phi(T, \mathbf{Q})^2 = \phi(T, \mathbf{Q})_{100\%} - \phi(T, \mathbf{Q})_{0\%}$, $\omega(T, \mathbf{Q})^1 = \omega(T, \mathbf{Q})_{\bar{\alpha}_1} - \omega(T, \mathbf{Q})_{0\%}$, $\omega(T, \mathbf{Q})^2 = \omega(T, \mathbf{Q})_{100\%} - \omega(T, \mathbf{Q})_{0\%}$.

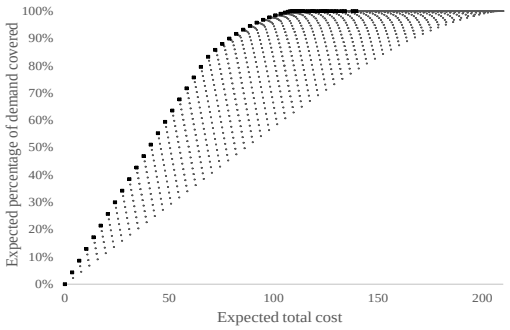
Appendix F

The Performances of the Optimal Vaccine Administration Policies for Possible (α_1, Q) Pairs

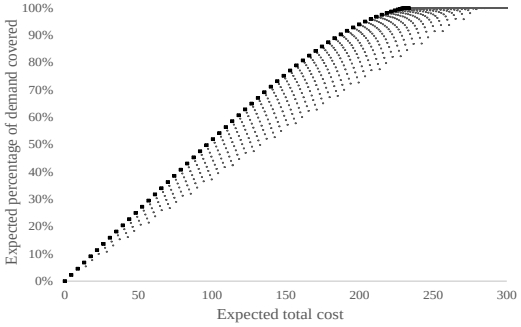
Figures F.1 and F.2 show the performances of the optimal policies for possible (α_1, Q) pairs where $1 \leq Q \leq 200 \times z_1 + 100 \times z_2$ under the terminated ($h_c = 0$) and and continuous ($h_c = \eta{+}1$) service scenarios, respectively. The Pareto-efficient (α_1, Q) pairs are highlighted with bold symbols. The number of Pareto-efficient pairs is quite limited compared to the number of feasible (α_1, Q) pairs. Therefore, policy makers may better view their alternatives for vaccine stock management by focusing on the reasonable sections of these Pareto-frontiers. The numbers of feasible (α_1, Q) combinations with continuous service are similar to those with terminated service. However, the pairs scatter along a larger area under the continuous service scenario.



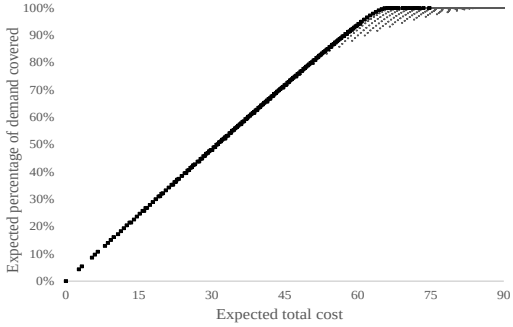
(a) Pentavalent - 2-dose and 10-dose vials



(b) Measles - 5-dose and 10-dose vials

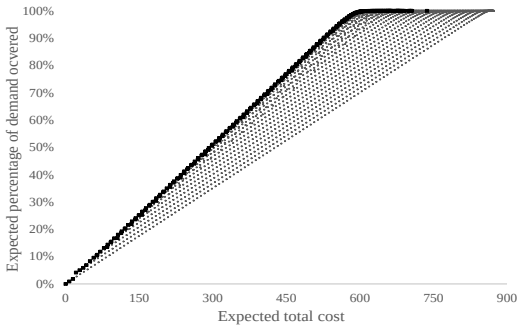


(c) YF - 5-dose and 20-dose vials

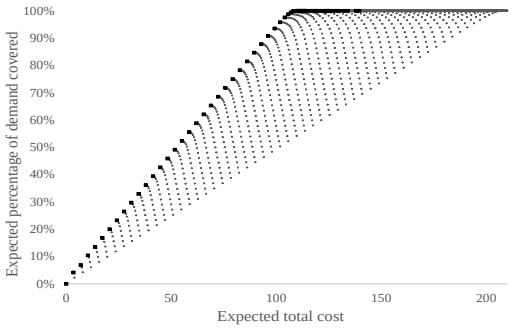


(d) BCG - 10-dose and 20-dose vials

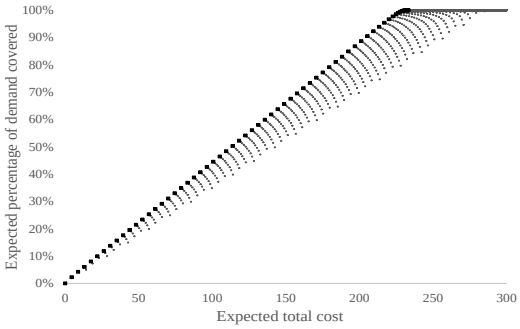
Figure F.1: The performances of the optimal vaccine administration policies for possible (α_1, Q) pairs in terms of expected percentage demand covered and total cost given the base case values of T, μ, and η with terminated service



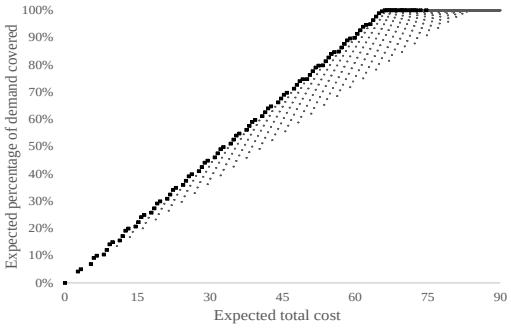
(a) Pentavalent - 2-dose and 10-dose vials



(b) Measles - 5-dose and 10-dose vials



(c) YF - 5-dose and 20-dose vials



(d) BCG - 10-dose and 20-dose vials

Figure F.2: The performances of the optimal vaccine administration policies for possible (α_1, Q) pairs in terms of expected percentage demand covered and total cost given the base case values of T, μ, and η with continuous service

Appendix G

Heuristic Policy Development Procedure

In order to achieve the coverage rates that the proposed policies provide, cooperation of healthcare practitioners is very important. However, it may not be practical for healthcare practitioners to keep track of the optimal actions which depend on the number of days remaining until the next replenishment, the number of vials on-hand, and the timeslot of the day. To reduce the complexities regarding the practical implications of the proposed policies, we derive simple vaccine administration policies using two different vial sizes with a continuous service scenario, i.e., $\alpha_1 > 0\%, h_c = \eta + 1$. The simple policies suggest using a single switching timeslot threshold for any $\mathbf{q} = (q_1, q_2)$ during each sub-horizon of equal length that together constitute a replenishment cycle. We determine the number of sub-horizons such that the length of a sub-horizon has a practical representation, e.g., a 20-day replenishment cycle which denotes a monthly vaccine administration service can be divided into 4 sub-horizons so that each sub-horizon corresponds to a week in practice. Simple vaccine administration policies assume continuous service. We first determine a list of candidate switching timeslots for each sub-horizon from the optimal timeslots of the first day of the sub-horizon. The motivation for considering only the first day of a sub-horizon to approximate a single switching timeslot for that sub-horizon is based on

Proposition 2 and the numerical observation implying that the optimal policies administer doses conservatively. Since the policies use large vials for a shorter duration of time on each day when there are fewer days left until the next replenishment the doses wasted during each sub-horizon is at its minimum level which may help administering more doses during the overall horizon. Although focusing only on each sub-horizon's first day helps reducing the number of switching timeslots to consider, there can still be $q_1 q_2$ different switching timeslots for each sub-horizon. To limit this number, we select the candidates among the optimal switching timeslots of each sub-horizon's first day for the inventory levels that are greater than a most likely minimum inventory level for the sub-horizon, i.e., $q_1 z_1 + q_2 z_2 > Q_i$ for sub-horizon i. We calculate Q_i using the equation $Q^i = Q - \sum_{j=1}^{i-1}(1 + \omega^{CP}(T, \mathbf{Q})\mu T^j)$ where μT^j is the expected demand satisfied during the sub-horizon j of T_j days and $\omega^{CP}(T, \mathbf{Q})\mu T^j$ is an approximate upper bound for the expected number of doses wasted during sub-horizon j. We use $\omega^{CP}(T, \mathbf{Q})$ which denotes the wastage rate of the current practice using only large vials with a continuous service ($\alpha_1 = 0\%, h_c = \eta + 1$) to calculate an approximate upper bound for the wastage because the highest wastage rate is observed from the current practice in our numerical analyses. We further eliminate some of these timeslots satisfying $q_1 z_1 + q_2 z_2 > Q_i$,when determining the simple policy. This elimination is performed based on Proposition 3 and the numerical observation stating that the switching timeslots increase as q_2 increases. Therefore, we only consider the median of the optimal switching timeslots of q_1 values for each candidate q_2 value (e.g., for $t = 20$, $q_2 = 10$, we consider the median of $h_1^*(20, 10, q_1), q_1 = 1, \cdots, Q_1$). By this, we can reduce the number of candidates for each week to the number of q_2 values that satisfy $q_1 z_1 + q_2 z_2 > Q_i$. Among all possible combinations of the candidate switching timeslots for each sub-horizon, we only consider the combinations that satisfy the numerical observation that the switching timeslots increase as t decreases (i.e., switching timeslot of sub-horizon $i \leq$ that of sub-horizon *i+1*). Lastly, we calculate the expected demand covered during a replenishment cycle for the simple policies considered and select the policy with the highest coverage as the best simple policy.