8-2017

Personalized Decision Modeling for Intervention and Prevention of Cancers

Fan Wang
University of Arkansas, Fayetteville

Personalized Decision Modeling for Intervention and Prevention of Cancers

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Engineering

by

Fan Wang
Tianjin University
Bachelor of Science in Industrial Engineering, 2007
University of Arkansas
Master of Science in Industrial Engineering, 2012

August 2017
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Dr. Shengfan Zhang
Dissertation Director

Dr. Justin Chimka
Committee Member

Dr. Murat Kurt
Committee Member

Dr. Edward Pohl
Committee Member

Abstract

Personalized medicine has been utilized in all stages of cancer care in recent years, including the prevention, diagnosis, treatment and follow-up. Since prevention and early intervention are particularly crucial in reducing cancer mortalities, personalizing the corresponding strategies and decisions so as to provide the most appropriate or optimal medical services for different patients can greatly improve the current cancer control practices. This dissertation research performs an in-depth exploration of personalized decision modeling of cancer intervention and prevention problems. We investigate the patient-specific screening and vaccination strategies for breast cancer and the cancers related to human papillomavirus (HPV), representatively. Three popular healthcare analytics techniques, Markov models, regression-based predictive models, and discrete-event simulation, are developed in the context of personalized cancer medicine. We discuss multiple possibilities of incorporating patient-specific risk into personalized cancer prevention strategies and showcase three practical examples. The first study builds a Markov decision process model to optimize biopsy referral decisions for women who receives abnormal breast cancer screening results. The second study directly optimizes the annual breast cancer screening using a regression-based adaptive decision model. The study also proposes a novel model selection method for logistic regression with a large number of candidate variables. The third study addresses the personalized HPV vaccination strategies and develops a hybrid model combining discrete-event simulation with regression-based risk estimation. Our findings suggest that personalized screening and vaccination benefit patients by maximizing life expectancies and minimizing the possibilities of dying from cancer. Preventive screening and vaccination programs for other cancers or diseases, which have clearly identified risk factors and measurable risk, may all benefit from patient-specific policies.

Acknowledgments

I would like to thank my advisor, Dr. Shengfan Zhang, for her continued mentorship, support, guidance and patience during my five years in the Industrial Engineering doctoral program. She was always here when I needed help for anything. I am especially grateful for his patience and understanding.

I also would like to thank other dissertation committee members, Drs. Edward Pohl, Justin Chimka, and Murat Kurt, for all their guidance, help and advice on my study and dissertation.

A special thank goes out to the Industrial Engineering staff, Karen Standley, Tamara Ellenbecker, and Carrie Pennington for their tremendous help during my master and doctoral studies.

Special thanks are extended to my fellow graduate students for bringing ease and happiness to me from 2010 to 2017.

Finally, I would like to express my sincere gratitude to the University of Arkansas for letting me be a student here.

Dedication

To my wife and parents.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Published Paper

Chapter 3:

Wang, F., Zhang, S., & Henderson, L. M. (2017). Adaptive decision-making of breast cancer mammography screening: A heuristic-based regression model. Omega. Advance online publication. doi.org/10.1016/j.omega.2017.05.001

1 Introduction

Over the last few years, personalized cancer medicine has emerged as a popular topic in the healthcare community. The term personalized cancer medicine generally refers to the medical practices for cancers that categorize patients into different groups and tailor medical decisions, services, and products to individual characteristics of patients according to their risk of cancers or predicted clinical responses. Although "personalized medicine" is usually described as providing "the right patient with the right drug at the right dose at the right time", it is involved in all stages of care, including prevention, diagnosis, treatment and follow-up (US Food and Drug Administration 2013). Personalized medicine provides different patients with customized and appropriate prevention, screening, and treatment strategies, which are more effective and cause fewer side effects compared with the standard options.

When it comes to the cancer medicine, personalizing prevention and early intervention is extraordinarily more critical for cancer patients than for non-life-threating diseases patients. On one hand, cancer prevention directly reduces or eliminates the chances of developing cancers and cancer-related mortalities. On the other hand, cancer survival heavily depends upon early diagnosis. According to the American Cancer Society (ACS), during 2017, there will be 1,688,780 newly diagnosed cancer cases and 600,920 cancer deaths in the U.S. (ACS 2017). However, a substantial proportion of cancer cases could be effectively prevented. For example, certain cancers caused by infectious agents, could be prevented through vaccinations, such as human papillomavirus (HPV) vaccines for cervical cancer prevention and hepatitis B virus (HBV) vaccines for liver cancer prevention. In addition, routine screening is known to help reduce mortalities for breast, colon, rectum, cervix, prostate and lung cancers by early detections (ACS 2017). Therefore, personalizing cancer prevention and invention strategies and providing

the most appropriate or optimal medical services for different patients can greatly improve the current cancer control practices.

This dissertation concentrates on the decision making processes of the two types of intervention and prevention approaches, screenings and vaccinations, and aims to optimally determine the personalized strategies for patients. For cancer screening, we focus on the screening mammography of breast cancer, which is the most commonly diagnosed cancer in women worldwide, and model both pre- and post-screening decisions at the individual level. For vaccination, HPV-related cancers are analyzed with a special focus on cervical cancer, which is the third most common female cancer worldwide (Wardle et al. 2015). We model the personalized HPV vaccination from the perspective of HPV-related cancer prevention. Three mainstream methodologies in healthcare analytics are adopted in this dissertation, including Markov models, regression-based predictive model, and discrete-event simulation. We show how to apply these popular approaches in the context of personalized cancer prevention and propose some novel hybrid model frameworks.

In Chapter 2, we explore the personalization of post-screening biopsy referral decision. A Markov decision process (MDP) model is developed to provide the optimal follow-up referral decision when a patient receives abnormal screening result. The optimal policy is generated by maximizing life expectancy generated based on every woman's specific breast risk characteristics. The post-screening referral decisions can be converted to pre-screening decisions under some assumptions. The model not only considers post-screening biopsy decisions but also offer optimal surveillance decisions for treated breast cancer patients.

Chapter 3 presents a screening decision model that provides an adaptive screening strategy while considering the tradeoff between advantage and disadvantages of screening. We present a

two-stage decision framework: (1) age-specific breast cancer risk estimation, and (2) annual mammography screening decision-making based on the estimated risk. Whether a woman should receive a mammogram is determined based on her breast cancer risk at her current age. Our "on-line" pre-screening decision is adaptive to a woman's latest health status, which causes less bias in reflecting the individual risk of every woman. Our optimal decisions outperform the existing mammography screening guidelines in terms of the average loss of life expectancy.

Chapter 4 shifts the focus to the personalization of HPV vaccination policy. This study models the impact of HPV vaccination at different ages on every individual woman and track women's courses of life to estimate the vaccination's clinical outcomes in terms of prevented HPV-related cancers. With the purpose of providing patient-specific HPV vaccination strategies, especially personalized catch-up vaccination policies, we develop a discrete-event simulation model to evaluate multiple clinical consequences after a woman gets vaccinated based on a number of personal attributes. As our simulation model works at the individual level, we build an HPV risk estimation model reflecting every woman's HPV risk, which dynamically changes over time, to support the lifetime simulation model. We use such a hybrid model to estimate the clinical consequences and proves that the optimal HPV vaccination catch-up policy should be personalized for different women.

References

American Cancer Society. (2017) Cancer Facts & Figures 2017. Atlanta: American Cancer Society.

US Food and Drug Administration. (2013). Paving the way for personalized medicine: FDA's role in a new era of medical product development. Silver Spring, MD: US Food and Drug Administration.

Wardle, J., Robb, K., Vernon, S., & Waller, J. (2015). Screening for prevention and early diagnosis of cancer. American Psychologist, 70(2), 119.

2 Personalized Biopsy Referral Decision Modeling for Breast Cancer Screening and Surveillance Mammography in the Presence of Cancer Regression

Abstract

Mammography is currently the recommended method for population-based breast cancer screening. It is also used as a surveillance approach to monitor for recurrent breast cancers for patients already treated for early-stage breast cancer. However, all abnormal mammograms should be followed up with additional tests, e.g., biopsy exams, to confirm the presence of cancer due to the low specificity associated with mammography. Although biopsy is a relatively accurate approach with negligibly low false-positive rate, it is painful and often results in negative side-effects. This study aims to make personalized decision about whether abnormal screening and surveillance mammogram findings should be referred for follow-up biopsy immediately according to women's personal risk characteristics. We develop a discrete-time finite-horizon Markov decision process model to optimize biopsy referral policies, which maximizes women's total life expectancy. Our study provides personalized optimal follow-up biopsy referral policies for both screening mammography and surveillance mammography based on a woman's risk factors. We present the optimal biopsy referral policies for a typical high-risk and a typical low-risk woman under different scenarios of clinical history. These optimal screening policies vary with the personal risk factors significantly. Our results also demonstrate that the optimal biopsy policies for surveillance mammography fluctuate dramatically over years. In addition, different types of treatments for in situ breast cancer result in distinct optimal post-treatment biopsy referral policies. This study suggests that existing screening and surveillance mammography guidelines can be improved. The optimal schedule of mammography follow-ups should be determined on a patient-specific basis.

2.1 Introduction

As the most common non-cutaneous form of cancer and the second leading cause of cancer mortality among U.S. women (DeSantis et al. 2011), breast cancer has raised serious concerns from the public health community. According to the American Cancer Society (ACS), in 2017 approximately 316,120 women will be diagnosed with breast cancer and about 40,610 women will die from this cause. Although the advances in treatments have reduced breast cancer mortality rates, whether the cancer is curable still relies heavily on the early detection (Berry et al.). The ACS reported that the 5-year relative survival rate is 99% for early stage (i.e. localized breast cancer) and 23% for late stage (i.e. distant breast cancer) (Smith et al, 2011).

Mammography is the de facto standard for breast cancer screening practice. The ACS, the American College of Radiology (ACR) and the American Medical Association (AMA) recommend that U.S. women should receive annual mammograms beginning at age 40 or 45. There are a large number of clinical trials and population-based evaluations suggesting that mammography can significantly reduce breast cancer mortality. Specifically, Tabar et al. (2003) reported that screening mammography decreased the breast cancer mortality in women aged 40–69 by 44%.

Mammography is not merely an effective screening procedure to early detect new cases of breast cancers, but also intensely used as a surveillance approach to monitor for recurrent breast cancer for patients already treated for early-stage breast cancer. Numerous studies showed that breast cancer survivors, especially those who received breast-conserving surgeries (i.e. lumpectomy or lumpectomy followed by radiotherapy), remain at high risk of second breast cancers for a long time, including recurrences of the treated tumors and new primary breast cancers (Clarke et al. 2005, Solin et al. 2005). Shaitelman et al. (2005) conducted a long-term follow-up study on women with a history of ductal carcinoma in situ (DCIS) breast cancer and

found that these patients still have a risk of tumor recurrence for up to 20 years after treatments. The American Society of Clinical Oncology (ASCO) recommends women treated with breast-conserving surgery (BCS) start annual surveillance mammograms one year after the initial diagnosis and no earlier than 6-months after the radiation therapy is completed (Khatcheressian et al. 2006). In recent studies, annual surveillance mammograms after treatments of breast cancers were proving to significantly reduce breast cancer mortality (Lash et al. 2007).

However, mammography is not perfect. Whether a mammogram result is positive (i.e. abnormal) is subjectively decided by a radiologist based on his/her professional skills and experiences. Thus possible false interpretations are usually inevitable. Previous studies showed that the risk of false positive (i.e. radiologists labeled the mammograms as positive but the cancer is actually absent) is significant. The positive predictive value (PPV) of mammography lies between 1.4% and 9.1%, which means the likelihood that the breast cancer really exists is lower than 10% when a woman gets a positive mammogram result (Breast Cancer Surveillance Consortium 2009).

To confirm the existence of cancer cells, all abnormal findings will be followed up with additional testing, such as a biopsy exam, to confirm whether the cancer is present. Biopsy is a medical test that conducts a pathological examination on the breast tissue sample obtained from the suspicious breast area through a needle or a surgical procedure. It is the current "gold standard" of evaluating suspicious screening results (Bruening et al. 2009). Although biopsy is an accurate test with negligibly low false-positive rate (Parker et al. 1994), it is often painful and may result in side-effects since it is an invasive procedure placing patients at the risk of morbidities and even mortalities (Bruening et al. 2009). Due to the low PPV of mammography, it is estimated that 70% to 90% of breast biopsies are performed unnecessarily for benign diseases

(Zhi et al. 2007). These unnecessary follow-up exams not only resulted in large medical expenses, but also caused the patients to suffer from pain, anxiety and multiple side-effects.

In addition to the inherent limitations of mammography technique, the occurrences of spontaneous regression of breast cancers may also stultify some of the follow-up biopsy exams. Zhang and Ivy (2012) summarized medical exploration of breast cancer spontaneous regression in the literature. These medical cases showed that some cancers may disappear in the absence of treatment. One observational study found that 22% of breast cancer tumors are likely to resolve themselves without any treatment (Zahl et al. 2008). There are limited analytical studies addressing this phenomenon though. Fryback et al. (2006) incorporated regression in their simulation model. They assigned approximately 40% of the breast tumors to be limited malignant potential tumors, which progressed to a maximum of 1cm in diameter and then regressed after 2 years if undetected. Zhang and Ivy (2012) studied the impact of regression on lifetime breast cancer mortality. They suggested the regression rate around 20% may make a difference in mortality under different breast cancer treatment decisions. These studies implied that even if the screening mammography were perfect in detecting breast cancer, a further biopsy test may still not be necessary in the case of breast tumor regression.

Moreover, a number of prior studies showed that different women may face different risks of breast cancer. Gail et al. (1989,1999) estimated women's risk of breast cancer based on a series of personal risk attributes, including age, race, age at menarche, age at birth of first child, number of first-degree family history of breast cancer, and number of previous breast biopsy exams. Barlow et al. (2006) established a breast cancer risk prediction model for women receiving screening mammography using factors including age, race, body mass index (BMI), prior breast procedure, menopause status, breast density, family history of breast cancer, and a

prior false-positive mammogram. For early-stage breast cancer (i.e. in situ cancer) patients who undergo BCS, numerous studies reported that different types of treatments (i.e. lumpectomy followed with radiotherapy and lumpectomy alone) result in significantly different post-treatment recurrence risks (Boyages et al. 1999, Viani et al. 2007). Particularly, prior studies found that different treatments result in opposite risks of recurrence on the ipsilateral and the contralateral breasts: the addition of radiotherapy to lumpectomy reduced the risk of ipsilateral breast cancer recurrence (IBTR) while increasing the risk of contralateral breast cancer recurrence (CBTR) (Viani et al. 2007). Nevertheless, several studies argued that the risk of recurrence after treatment varies over time. (Shaitelman et a. 2012, Habel et al. 1997). Therefore, although multiple healthcare organizations provided standard guidelines for both breast cancer screening and surveillance practices, it is important to make personalized decision about whether an abnormal mammogram finding should be referred for biopsy according to the woman's specific risk characteristic, which motivates the goal of this paper: determining whether a patient should undergo a biopsy when she receives a positive mammogram result.

There are several studies seeking to optimize the trade-off between the negative effects of breast cancer screening and patients' long-term benefits of early diagnosis. Maillart et al. (2008) employed a partially observable Markov decision process (POMDP) model to evaluate different screening mammography policies by measuring life-time breast cancer mortality risk and expected mammogram counts. Ayer et al. (2012) formulated a POMDP model to propose an optimal personalized screening mammography schedule based on their personal risk characteristics associated with breast cancer, including age, race, age at menarche, age at first live birth, and prior screening history. Chhatwal et al. (2010) investigated how to make biopsy referral decision to maximize patients total expected quality-adjusted life years (QALY). They

used a finite-horizon discrete-time Markov decision process (MDP) model to offer optimal biopsy referral policies for patients with different breast cancer risk scores (i.e. a woman's current probability of cancer). In addition, Ayvaci et al. (2012) applied an MDP model to optimizing biopsy referral decisions for different breast cancer risk scores under budgetary restrictions.

Our study is different from previous studies and contributes to the literature in the following ways: 1) our model considers the complex pathology of breast cancer by incorporating the potential spontaneous regression of breast cancer; 2) our multidimensional model directly provides personalized biopsy referral policy for every individual patient with different combination of risk factors, instead of estimated risk scores; and 3) our model addresses dynamic personalized biopsy referral policy of surveillance mammograms after treatments for early-stage breast cancers (i.e. in situ cancer). We use an MDP model with history-dependent states to provide optimal policies for early-stage breast cancer patients at different years after different treatments. To the best of our knowledge, this is the first analytical study that optimizes surveillance mammography policies after breast cancer treatments in terms of total expected quality-adjusted life years. We also combine screening mammograms for new cancer cases with surveillance mammograms for breast cancer follow-ups into an integrated model, which is expected to provide a comprehensive solution of the breast cancer prevention and control problem.

The remainder of this paper is organized as follows. Section 2.2 describes the formulation of the MDP model. We also present several structural properties associated with the model. In Section 2.3, we apply our model in two example cases to demonstrate how our method provides personalized optimal biopsy referral policies. Finally, we discuss our results and the limitations

of this paper in Section 2.4. The main medical terms used throughout the whole paper are summarized in Table 1.

Table 1 The definitions of the main medical terms

Medical Terms	Acronym	Definition
Breast-conserving surgery	BCS	An operation to remove the breast cancer but not the breast itself.
Contralateral breast tumor recurrence	CBTR	The cancer recurrence occurs in the opposite breast.
In situ cancer		Non-invasive breast cancer. The abnormal cells have not spread outside the duct to other tissues in the breast.
Invasive cancer		Cancer that has spread from where it started in the breast into surrounding, healthy tissue.
Ipsilateral breast tumor recurrence	IBTR	The cancer recurrence occurs in the treated breast.
Lumpectomy		A surgery to remove a tumor in a breast and a small amount of normal tissue around it
Mastectomy		A surgery to remove part of or entire breast.
Positive predictive value	PPV	The probability that a person with a positive test result has, or will get, the disease.
Radiotherapy		The use of high-energy radiation from x-rays, gamma rays, neutrons, protons, and other sources to kill cancer cells and shrink tumors.
Salvage therapy		Salvage therapy is any therapy that is done in an attempt to cure cancer following the failure of an initial treatment.

Source: Dictionary of cancer terms (National cancer institute 2015)

2.2 Material and Methods

2.2.1 Decision Process

In this study, we assume that women strictly undergo yearly screening mammography from age 40 as recommended by the ACR and AMA guidelines. A typical screening mammography process in one year consists of two main stages (Figure 1). The first stage is to conduct an annual mammogram exam. If the exam result is negative, the woman is considered to be healthy and recommended to do nothing and just wait till the next scheduled mammogram. If the woman receives a positive mammogram, in the next stage, a decision has to be made about whether this

patient should be referred to a biopsy exam. Due to biopsy's extremely low false-positive rate (Parker et al. 1994), we assume biopsy is perfect. Thus, patients with positive biopsy findings will be scheduled for treatment immediately. Negative findings indicate absence of breast cancer, so that the patients should continue receiving annual mammograms.

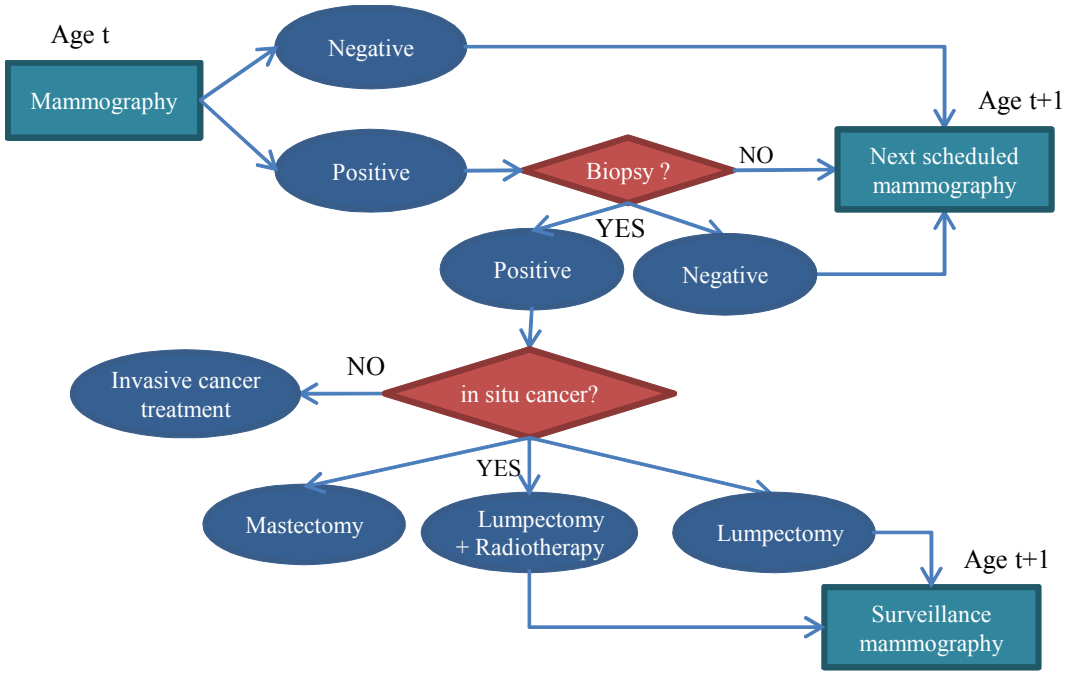


Figure 1 Annual screening process of breast cancer

Figure 1 also demonstrates how we model the process of surveillance mammography for in situ cancer patients after being treated. As Figure 1 shows, we differentiate two types of local treatments: mastectomy and breast-conserving surgery (i.e. lumpectomy alone and lumpectomy with radiotherapy). For in situ patients who have received mastectomy treatments, surveillances after mastectomy vary from person to person. For example, patients treated with prophylactic bilateral mastectomy are no longer recommended to undergo routine follow-ups. Therefore, our decision process only considers surveillance mammography after lumpectomy with or without radiotherapy. For the patients who have received breast-conserving surgery, according to the ASCO guideline on breast cancer follow-up and management (Khatcheressian et al. 2006), the

first post-treatment mammogram should be conducted one year after the initial mammogram. The ASCO also recommends these patients undergo yearly mammography thereafter. Hence we assume that all the patients treated with breast-conserving surgery will carry on with annual mammography after completion of treatments.

Our study focuses on the biopsy referral decision when a woman receives a positive mammogram. In our decision model, whether a woman should be referred to biopsy is determined based on six breast cancer risk factors, including the woman's age, race, age at birth of first child, first-degree family history of breast cancer, menopause status and clinical history of in situ breast cancer (Table 2). The first five factors are common risk factors and their risk associated with breast cancer is estimated by Barlow et al. (2006). We exclude the dynamic breast cancer risk attributes (e.g. BMI) that may change over time, since the levels of these dynamic factors in the future are unknown. The six risk factors produce 23,040 different combinations in total.

Table 2 Risk factors of breast cancer considered in the model

Risk factors	**Level**
Age	40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74,75-79, 80-100;
Race	White, Black, Asian, Hispanic, Other;
Age at birth of first child	<30, >=30 or Nulliparous;
First-degree family history of breast cancer	Yes, No;
Menopause status	Premenopausal, Postmenopausal;
Clinical history of in situ breast cancer	
Had the woman been diagnosed with in situ breast cancer before?	Yes, No;
If yes, what was the treatment type?	Mastectomy, Lumpectomy, Lumpectomy plus radiotherapy;
How many years have passed since the treatment?	1, 2, 3…20, 20+;

2.2.2 MDP Model

We develop a discrete time finite-horizon Markov decision process (MDP) model to optimize biopsy referral policies, which maximize women's total life expectancy. For every woman with specific background, we solve the MDP model optimally using patient-specific transition probabilities that are estimated based on the combination of her risk factors. We formulate the MDP model with the following components.

Decision epochs: $t = 40, 41, \ldots 100$, which represent a woman's age at the current time period. Since both screening and surveillance mammograms are performed annually according to the ASCO, a biopsy decision will also be made on a yearly basis after a mammogram result starting at age 40. According to the U.S. life tables (Arias 2012), we employ age 100 as the terminating point.

States: $s_t \in S = \{\mathrm{N}, \mathrm{P}, \mathrm{N}_\mathrm{L}^n, \mathrm{P}_\mathrm{L}^n, \mathrm{N}_\mathrm{R}^n, \mathrm{P}_\mathrm{R}^n, \mathrm{M}, \mathrm{I}, \mathrm{D}\}, n = 1, 2, \ldots, 20+$. We use 89 states to describe a woman's state after receiving a mammogram at time *t.* N and P represent that a woman without clinical history of breast cancer receives a negative mammogram and a positive mammogram, respectively. Two history-dependent states, $\mathrm{N_L}^n$ and $\mathrm{P_L}^n$, represent that the in situ patients treated with lumpectomy alone (L) receive negative and positive surveillance mammograms respectively. The superscript *n* denotes the number of years after completion of the lumpectomy treatment. Since women with a history of in situ cancer have risk of recurrence up to 20 years after treatment (Shaitelman et al. 2012), *n* can take values from 1 to 20. If no recurrence occurs up to 20 years after the treatment, we use a single value "20+" to represent all the succeeding years. Thus, there are 21 states denoted by $\mathrm{N_L}^n$ or $\mathrm{P_L}^n$. Similarly, $\mathrm{N_R}^n$ and $\mathrm{P_R}^n$ represent in situ patients treated with lumpectomy plus radiotherapy (R) receive negative and positive surveillance mammograms, respectively. Likewise, superscript *n* can take values from 1 to 20+.

In our model, M (mastectomy), I (invasive cancer) and D (death) are treated as absorbing states, which means once a woman enters one of these states, she will quit the decision process. Due to the high pathological complexity of treatments for invasive cancer, we do not consider the surveillance and follow-ups for invasive cancer patients either. Thus we use absorbing state I to represent that a woman is diagnosed with invasive breast cancer. The absorbing state D includes deaths from breast cancer and other causes. Considering 21 possible values of n in N_L^n, P_L^n, N_R^n and P_R^n, our model actually includes 89 states.

Actions: $a_t \in \{\mathrm{W}, \mathrm{B}\}$. There are two options being considered: waiting (i.e. doing nothing and waiting until the next annual mammogram) or doing a biopsy exam immediately. It is noted that waiting is always the optimal action when the state shows a negative screening result.

Transition probability: $p_t(s_{t+1}|s_t, a_t)$. Transition probability represents the possibility of transition from the state s_t to the state s_{t+1} at decision epoch t, given that action a_t is taken.

Intermediate rewards: $r_t(s_t, a_t)$. The intermediate reward is life years assigned to the woman at time t given that she is in state s_t and action a_t is taken. Thus the intermediate reward equals the one year minus the disutility of the action a_t at current age. We assume disutility of action W is zero and only address the disutility caused by a biopsy exam. Let d_t denote the disutility of biopsy at age t, and then the immediate reward is defined as

$$r_t(s_t, a_t) = \begin{cases} 1 & \text{if } a_t = \mathrm{W}, \\ 1 - d_t & \text{if } a_t = \mathrm{B}, \end{cases} \quad \text{for } s_t \neq \mathrm{M, I, D.}$$

Lump-sum rewards: $R_t(s_t)$. If a woman enters one of the three absorbing states, she will be assigned with one-time lump-sum reward, which equals the total life expectancy calculated based on her current age and state. In addition, we implement the "half-cycle correction" by assuming that, on average, a transition from any transient state to an absorbing state occurs

halfway through each decision epoch, instead of at the beginning or the end of an age (Sonnenberg et al. 1993). Let $e_t(s_t)$ denote the life expectancy for a woman at age t with state s_t. The lump-sum reward can be calculated by

$$R_t(s_t) = \begin{cases} e_t(s_t), & \text{if } s_t = \text{M, I,} \\ 0.5, & \text{if } s_t = \text{D.} \end{cases}$$

Value function: $V_t(s_t)$. Value function calculates the maximum total life expectancy given that the woman is at age *t* and her current state is s_t. For the states associated with negative mammograms, value functions are deterministic since action W is always selected. For the absorbing states M, I and D, the value functions are equal to their corresponding lump-sum rewards at current age *t.* Thus, the general value function is given by

$$V_t(s_t) = \begin{cases} \max\left\{\left[r_t(s_t,\text{W}) + \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}\,|\,s_t,\text{W})\right], \left[r_t(s_t,\text{B}) + \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}\,|\,s_t,\text{B})\right]\right\} & \text{if } s_t = \text{P}, \text{P}_\text{L}^n, \text{P}_\text{R}^n, \\ r_t(s_t,\text{W}) + \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1}\,|\,s_t,\text{W}) & \text{if } s_t = \text{N}, \text{N}_\text{L}^n, \text{N}_\text{R}^n \\ R_t(s_t) & \text{if } s_t = \text{M, I, D} \\ & \text{for } t = 40, 41, \cdots, 99. \end{cases} \quad (1)$$

For $s_t = \text{P}, \text{P}_\text{L}^n, \text{P}_\text{R}^n$, the optimal action $a_t^*(s_t)$ for age *t* at state s_t is determined by

$$a_t^*(s_t) = \underset{a_t \in [\text{W, B}]}{\arg\max}\, V_t(s_t), \quad \text{for t} = 40, 41, \cdots, 99. \quad (2)$$

Since t = 100 is the end of the decision horizon, we assume there will be no future reward accumulating for woman at age 100, i.e.

$$\text{V}_{100}(\text{S}_{100}) = 0.$$

2.2.3 State Transition

State transitions are action dependent, and we introduce the one-step transition diagram for the biopsy action in Figure 2.

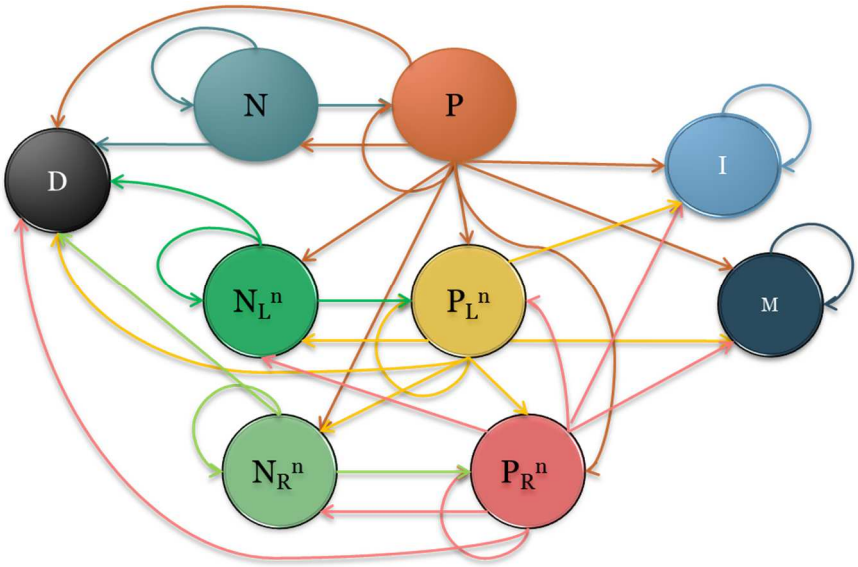


Figure 2 The transition diagram for action B "do biopsy"

The state transitions from P to N_L^n, P_L^n, N_R^n, and P_R^n imply diagnoses of in situ breast cancer and treatments. For instance, the transition from P to N_L^1 involves five stages: (a) the woman with a positive mammogram chooses to do biopsy; (b) the positive mammogram is proving to be a true-positive finding by the biopsy exam; (c) this true-positive finding is identified as in situ cancer; (d) this patient is then treated with lumpectomy; and (e) in the first surveillance mammography after completion of the treatment, the patient receives a negative mammogram.

Similarly, for states P_L^n and P_R^n, if biopsy exams confirm their findings are true-positive, their transitions also involve implicit treatment stages. As an example, Figure 3 illustrates the one-step transition processes starting from P_L^n with a tree structure. On one hand, if the finding at the P_L^n state is confirmed by the biopsy exam to be a false positive mammogram, then the woman is considered cancer-free at the beginning of the current decision epoch. She may stay healthy or get a cancer recurrence during the current year. However, regardless of her true health status, the result of the next annual mammogram can be either negative N_L^{n+} or positive P_L^{n+1}, (no treatment at P_L^n). On the other hand, if the finding at P_L^n turns out to be a true positive, then the woman has a cancer recurrence. For an invasive recurrence, the patient enters the absorbing state I. For an in

situ recurrence, we subsequently distinguish the types of treatment for the recurrence, which is referred to as salvage therapy. The main treatment approaches of salvage therapy for in situ cancers are the same as those for new breast cancer (i.e. lumpectomy alone, lumpectomy with radiotherapy, and mastectomy) (Solin et al. 2001). We assume all these in situ recurrences can be treated successfully (Fong et al. 2011). After completion of the salvage therapy, the first surveillance mammogram can be negative or positive, which are denoted as N_L^1 or P_L^1 for lumpectomy, and N_R^1or P_R^1 for lumpectomy with radiotherapy, respectively.

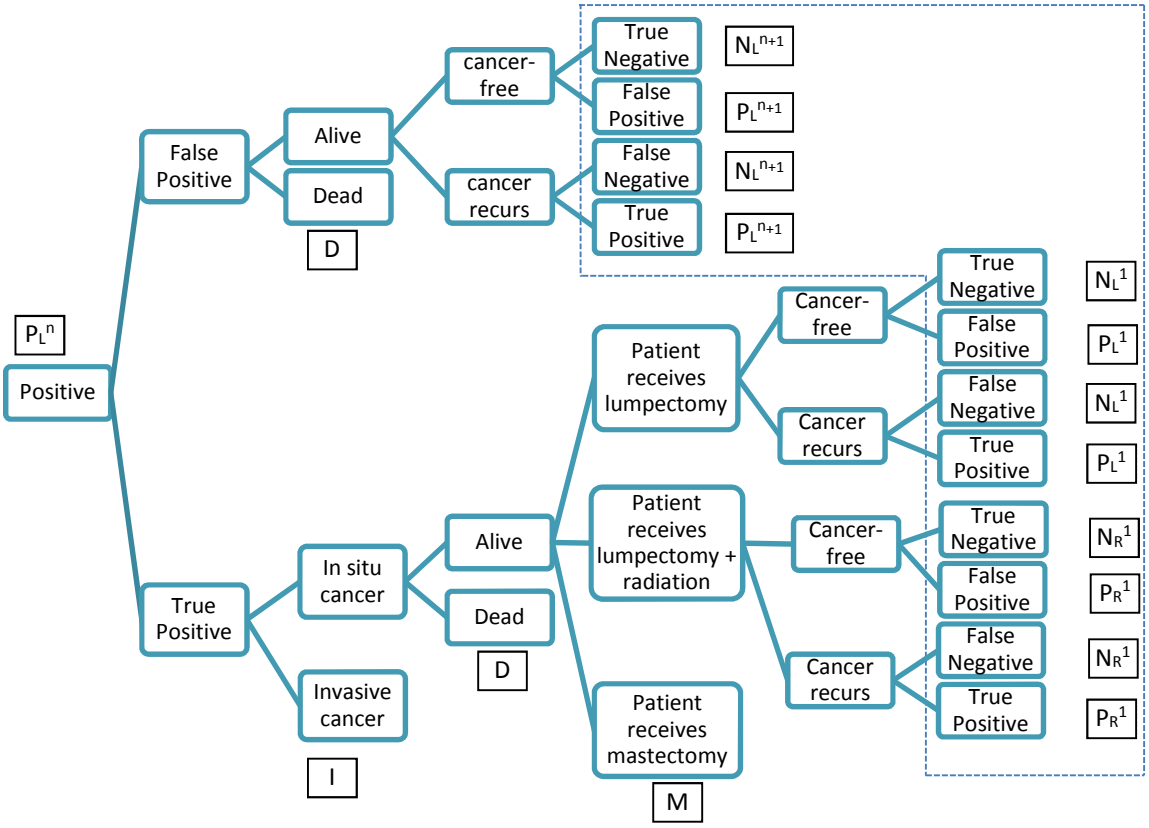


Figure 3 One-step state transitions starting from P_L^n

With the tree structure, all the relevant transition probabilities are easy to calculate. For instance, to calculate the transition probability from P_L^n to I under action B (i.e. $p_t(\mathrm{I}|\mathrm{P}_\mathrm{L}^n, \mathrm{B})$), we only need to multiply all the relevant probabilities from node P_L^n to node I, namely, multiplying true positive rate by the proportion of invasive cancers in all breast cancers. Details of the

calculation of all the transition probabilities are presented in the Appendix A.

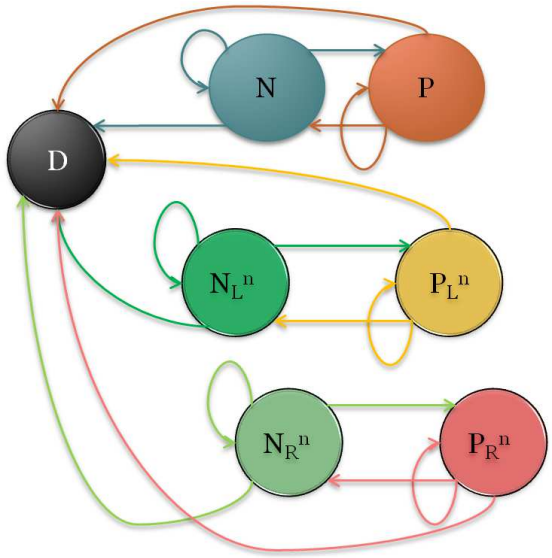


Figure 4 The transition diagram for action W "wait"

Figure 4 is the transition diagram for the women who choose the action "wait". We assume that women will persistently wait until the next mammograms. As no biopsy will be performed on the women, the presence of cancer will be unknown. Given that spontaneous regressions may occur on some of breast cancer tumors if undetected (Fryback et al. 2006), our model considers the cancer regression in the transitions after the action "wait" has been taken or after a false-negative mammogram. While spontaneous regressions of invasive cancers were rarely found in the medical literature, reported cases of spontaneous regression of in situ cancers are not infrequent (Joensuu and Lundin 2004). Hence we assume spontaneous regressions can only occur at in situ stage. Figure 5 shows an example of how cancer regressions get involved in the transitions starting from P under the action "wait". As the figure suggests, the regression may occur when P is in fact a true positive result and corresponds to an in situ breast tumor. In addition to P, the possibility of spontaneous regression is implicated in all transient states (i.e. N, P, N_L^n, P_L^n, N_R^n and P_R^n) in Figure 4.

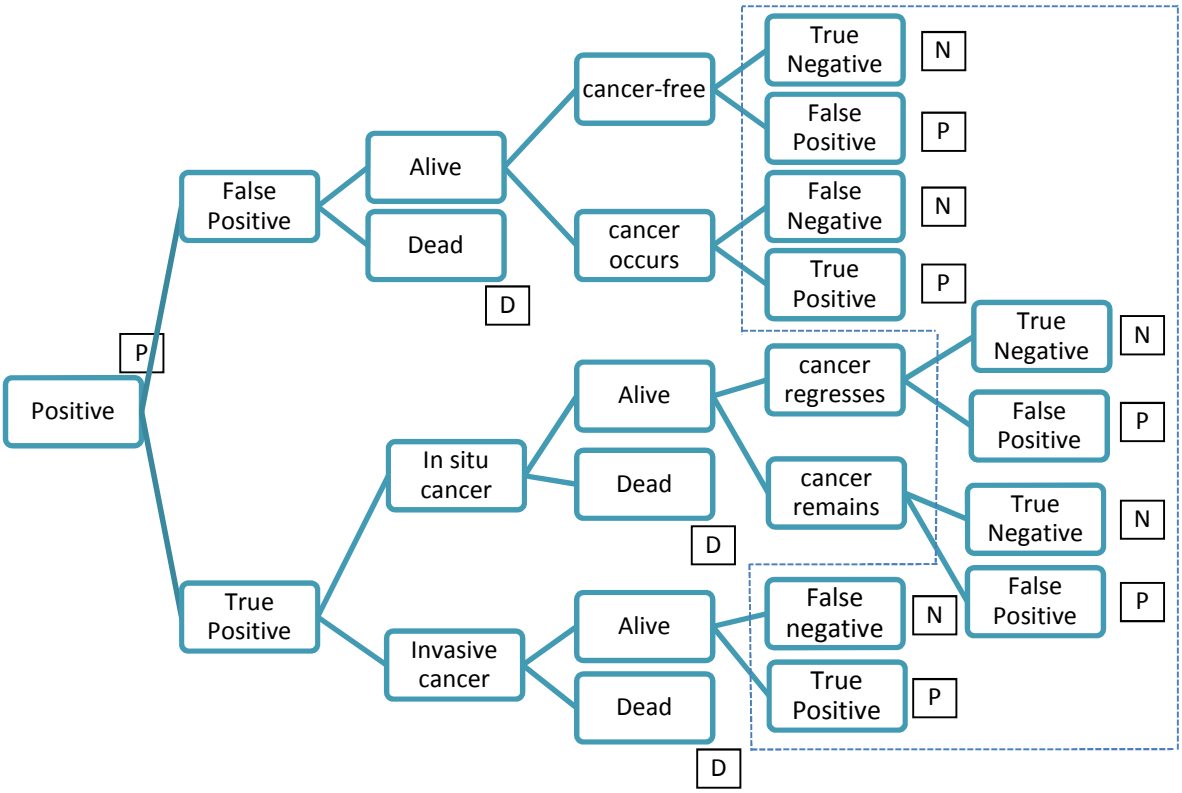


Figure 5 The transitions starting from P in one decision epoch

There are two different types of deaths, death from breast cancer and death from other causes in the state transitions. We assume that only patients with invasive cancer may die from breast cancer in one-step transition. In situ cancer, as an early stage of breast cancer, is assumed not to result in death within one year. In an earlier study using the Carolina Mammography Registry data (Zhang 2011), the mortality rate is found to be close to zero. As a result, we separate deaths from in situ cancer and invasive cancer in Figure 5.

2.2.4 Structural Properties

In this section, we present several structural properties of the MDP model to investigate how the optimal solution is affected by model inputs. Based on the medical facts, we first introduce several assumptions used to prove structural properties.

ASSUMPTION 1. *Given two patients A and B at the same age t, if any one of the inequalities in the following holds, then all the other inequalities hold.*

(b) $p_t^A(P_L^1|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(P_L^1|P_{L\square R}^n,\mathrm{B}), p_t^A(N_L^1|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(N_L^1|P_{L\square R}^n,\mathrm{B}), p_t^A(P_R^1|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(P_R^1|P_{L\square R}^n,\mathrm{B}),$
$p_t^A(N_R^1|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(N_R^1|P_{L\square R}^n,\mathrm{B}), p_t^A(\mathrm{M}|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(\mathrm{M}|P_{L\square R}^n,\mathrm{B}), p_t^A(\mathrm{I}|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(\mathrm{I}|P_{L\square R}^n,\mathrm{B}),$
$p_t^A(\mathrm{D}|P_{L\square R}^n,\mathrm{B}) \geq p_t^B(\mathrm{D}|P_{L\square R}^n,\mathrm{B}), p_t^A(P_{L\square R}^{n+1}|P_{L\square R}^n,\mathrm{W}) \leq p_t^B(P_{L\square R}^{n+1}|P_{L\square R}^n,\mathrm{W}), p_t^A(N_{L\square R}^{n+1}|P_{L\square R}^n,\mathrm{W}) \leq p_t^B(N_{L\square R}^{n+1}|P_{L\square R}^n,\mathrm{W}),$
and $p_t^A(\mathrm{D}|P_{L\square R}^n,\mathrm{W}) \geq p_t^B(\mathrm{D}|P_{L\square R}^n,\mathrm{W})$; where $P_{L\square R}^n = P_L^n$ or P_R^n.

For every inequality, the left-hand side corresponds to patient A's transition probability from a pre-clinical stage to breast cancer diagnosis, and the right-hand side corresponds to that of patient B. Condition (a) presents that all the transition probabilities from a screening positive mammogram to other possible post-treatment states including death are higher for patient A than B. Condition (b) includes all the transition probabilities from a surveillance positive mammogram to other possible post-treatment states. If patient A and patient B both receive breast-conserving surgeries (i.e. lumpectomy or lumpectomy with radiotherapy), then post-treatment recurrence risk of patient A is not lower than that of patient B.

Assumption 1 reflects that patient A always has a faster deterioration rate than patient B in terms of both incidence of new primary breast cancer and cancer recurrence. It can be viewed as: if patient A's transition probability from any one of the pre-clinical stages (i.e. screening mammogram or surveillance mammogram) to breast cancer diagnosis or death is higher than or equal to that of patient B, then all her other transition probabilities from pre-clinical stages to breast cancer diagnosis is not lower than those of patient B. This assumption also implies that if patient A's risk of developing new breast cancer is higher than or equal to that of patient B, then once they are diagnosed with in situ cancer, the risk of cancer recurrence of patient A is not lower than that of patient B. It is consistent with prior medical literature (Buist et al. 2010).

ASSUMPTION 2. *Let*

$$V_t(\mathrm{T,B}) = V\left(P_L^1\right)\frac{P_t(P_L^1\,|\,\mathrm{P,B})}{P_t\left(P_L^1|\mathrm{P,B}\right)+P_t\left(N_L^1|\mathrm{P,B}\right)+P_t\left(P_R^1|\mathrm{P,B}\right)+P_t\left(N_R^1|\mathrm{P,B}\right)+P_t\left(\mathrm{M}|\mathrm{P,B}\right)+P_t(\mathrm{I}\,|\,\mathrm{P,B})}$$
$$+V\left(N_L^1\right)\frac{P_t(N_L^1\,|\,\mathrm{P,B})}{P_t\left(P_L^1|\mathrm{P,B}\right)+P_t\left(N_L^1|\mathrm{P,B}\right)+P_t\left(P_R^1|\mathrm{P,B}\right)+P_t\left(N_R^1|\mathrm{P,B}\right)+P_t\left(\mathrm{M}|\mathrm{P,B}\right)+P_t(\mathrm{I}\,|\,\mathrm{P,B})}$$
$$+V\left(P_R^1\right)\frac{P_t(P_R^1\,|\,\mathrm{P,B})}{P_t\left(P_L^1|\mathrm{P,B}\right)+P_t\left(N_L^1|\mathrm{P,B}\right)+P_t\left(P_R^1|\mathrm{P,B}\right)+P_t\left(N_R^1|\mathrm{P,B}\right)+P_t\left(\mathrm{M}|\mathrm{P,B}\right)+P_t(\mathrm{I}\,|\,\mathrm{P,B})}$$
$$+V\left(N_R^1\right)\frac{P_t(N_R^1\,|\,\mathrm{P,B})}{P_t\left(P_L^1|\mathrm{P,B}\right)+P_t\left(N_L^1|\mathrm{P,B}\right)+P_t\left(P_R^1|\mathrm{P,B}\right)+P_t\left(N_R^1|\mathrm{P,B}\right)+P_t\left(\mathrm{M}|\mathrm{P,B}\right)+P_t(\mathrm{I}\,|\,\mathrm{P,B})}$$
$$+V\left(\mathrm{M}\right)\frac{P_t(M\,|\,\mathrm{P,B})}{P_t\left(P_L^1|\mathrm{P,B}\right)+P_t\left(N_L^1|\mathrm{P,B}\right)+P_t\left(P_R^1|\mathrm{P,B}\right)+P_t\left(N_R^1|\mathrm{P,B}\right)+P_t\left(\mathrm{M}|\mathrm{P,B}\right)+P_t(\mathrm{I}\,|\,\mathrm{P,B})}$$
$$+V\left(\mathrm{I}\right)\frac{P_t(\mathrm{I}\,|\,\mathrm{P,B})}{P_t\left(P_L^1|\mathrm{P,B}\right)+P_t\left(N_L^1|\mathrm{P,B}\right)+P_t\left(P_R^1|\mathrm{P,B}\right)+P_t\left(N_R^1|\mathrm{P,B}\right)+P_t\left(\mathrm{M}|\mathrm{P,B}\right)+P_t(\mathrm{I}\,|\,\mathrm{P,B})},$$

and

$$V_t\left(\mathrm{NP,B}\right) = V\left(\mathrm{N}\right)\frac{P_t(\mathrm{N}|\mathrm{P,B})}{P_t\left(\mathrm{N}|\mathrm{P,B}\right)+P_t(\mathrm{P}|\mathrm{P,B})}+V\left(\mathrm{P}\right)\frac{P_t(\mathrm{P}|\mathrm{P,B})}{P_t\left(\mathrm{N}|\mathrm{P,B}\right)+P_t(\mathrm{P}|\mathrm{P,B})},$$

then $V_t\left(\mathrm{NP,B}\right) \geq V_t\left(\mathrm{T,B}\right) \geq V_t\left(\mathrm{D}\right)$.

In this assumption, we denote the post- treatment total life expectancy of a woman with biopsy-verified true positive mammogram by $V_t\left(\mathrm{T,B}\right)$, where T and B represents treatment and biopy, respectively. $V_t\left(\mathrm{NP,B}\right)$ corresponds to the total life expectancy of a women with biopsy-verified false positive mammogram. NP means the women will continue undergoing mammography, which will turn out to be either negative or positive. The assumption indicates that the post-biopsy life expectancy after a negative (i.e. benign) biopsy result is always higher than or equal to that after a positive (i.e. malignant) biopsy result. The same assumption has been made in some other breast cancer screening studies (Chhatwal et al 2010).

Theorem 1. *For any given patient, there exists a threshold of biopsy disutility* d_t^*, $\forall t$, *such that*

$$a_t^*\left(s_t\right) = \begin{cases} W & \text{if } d_t > d_t^*, \\ B & \text{if } d_t \leq d_t^*. \end{cases} \tag{3}$$

According to Equation (1) in Section 2.2,

$$d_t^* = \sum_{s_t+1 \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B}) - \sum_{s_t+1 \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}).$$

Thus d_t^* is the gain of life expectancy brought by doing biopsy. Theorem 1 implies that only when the disutility of biopsy does not exceed the gain of life expectancy brought by a biopsy exam, will doing biopsy be an optimal choice. Therefore, d_t^* provides a measurement to assess how likely a patient makes a biopsy decision. Obviously, a higher d_t^* indicates that the patient is more apt for biopsy relative to another patient with a lower d_t^*.

Proposition 1. *Let $V_t^A(s_t)$ and $V_t^B(s_t)$ be the value functions of patient A and patient B, respectively. Under assumption 1, $V_t^A(s_t) \leq V_t^B(s_t),\ \forall t$ and $\forall s_t$.*

Proposition 1 confirm the intuition that the higher risk of breast cancer is, the lower life expectancy of a patient will be.

Proposition 2. *Let $R_t^1(I)$ and $R_t^2(I)$ be the two possible lump-sum rewards at state I for a patient, respectively. If $R_t^1(I) \geq R_t^2(I)$, then $V_t^1(s_t) \geq V_t^2(s_t)$, $\forall t$ and $\forall s_t$.*

Corollary 1. Let $R_t^A(M)$ and $R_t^B(M)$ be the lump-sum *rewards of state M for patient A patient B,* respectively. If $R_t^A(M) \geq R_t^B(M)$, then $V_t^A(s_t) \geq V_t^B(s_t)$, $\forall t$ and $\forall s_t$.

Proposition 2 and corollary 1 mean that $V_t(s_t)$ is nondecreasing in $R_t(I)$ as well as $R_t(M)$. The clinical significance of these two conclusions is that the better prognosis for invasive cancer patients or in situ cancer patients who received mastectomy, the higher life expectancy will be for all women undergoing annual mammography.

The proofs of all the conclusions above are presented in Appendix B.

2.3 Results

In this section, we first describe how we use the data from various sources to estimate the model

parameters in Section 2.3.1. Then Section 2.3.2 presents the computational results by implementing the MDP model. Finally, we perform sensitivity analyses on several model parameters in Section 2.3.3.

2.3.1 Parameter Estimation

Since there is no single dataset that reports all the clinical data regarding epidemiology, diagnosis and treatment of breast cancer, we resort to multiple sources to estimate the model parameters (Table 3). In addition, the value of regression rate is assumed to be 20% in our base case analysis (Zhang and Ivy 2012). In the remainder of this section, we explain how we obtain the parameter estimation for calculating transition probabilities, intermediate reward and lump-sum reward.

Table 3 The data source for parameter estimation

Parameter	Data source	Parameter type
Incidence	Breast Cancer Surveillance Consortium (BCSC) dataset	Case-specific
Prevalence	BCSC dataset	Age-specific
Sensitivity, specificity and PPV of mammography	BCSC 2009	Age-specific
Mortality for the general population	Arias 2012	Age-specific
One-year mortality of invasive cancers	Zhang 2011	Age-specific
Percentages of treatment types for in situ cancer	Buist et al. 2010	Age-specific
Breast cancer recurrent rate	Wapnir et al. 2011, Shaitelman et al. 2012	Year-specific
Life expectancy for invasive cancer patients	Laboratory for Quantitative Medicine	Age-specific

The incidences of in situ breast cancers and invasive cancers are estimated using a raw dataset from the BCSC, which consists of 1,007,660 women's screening mammograms from 1996 through 2002. This dataset reports one observation randomly selected from every woman's screening history. Each observation reports a woman's age (by 5-year age group), menopausal status, race, age at first birth, family history of breast cancer and diagnosis information within 1 year of the screening mammogram. For women with a specific combination of risk levels, we treat them as one risk group, and use the diagnosis information to derive the incidence rate of

breast cancer, as well as the proportions of in situ cancers and invasive cancers for this group. Since only the women aged 35–84 are included in this dataset, we use the incidence of woman aged 80–84 as the incidence for all the women older than 80. We also use this assumption to process all other data that do not cover all age groups of 40–100 to fit our decision horizon.

Recurrence rate of breast cancer after treatment is different from the incidence of new breast cancer. In addition, the magnitude of recurrence risk fluctuates over the years after completion of treatment. Due to scarcity of data, we assume the recurrence rates to be the same for all women. We adopt the yearly incidence rate of recurrence for each of the 20 years after completion of lumpectomy with radiotherapy reported by Shaitelman et al. (2012). The yearly recurrence rates associated with lumpectomy alone are estimated based on Viani et al.'s meta-analysis (2007), which investigated the effect of the addition of radiation therapy to lumpectomy for in situ cancer on IBTR (ipsilateral breast tumor recurrence) and CBTR (contralateral breast tumor recurrence). Let $P_L^n(\text{IBTR})$, $P_L^n(\text{CBTR})$, $P_R^n(\text{IBTR})$ and $P_R^n(\text{CBTR})$ denote the probabilities of IBTR and CBTR after lumpectomy and the probabilities of IBTR and CBTR after lumpectomy with radiation in the *n*th year after treatment, respectively. We have

$$P_R^n(\text{IBTR})/P_L^n(\text{IBTR}) = 0.6,$$
$$P_R^n(\text{CBTR})/P_L^n(\text{CBTR}) = 1.53.$$

Since no study discusses age-specific annual death rates of untreated invasive breast cancer, we estimate these rates for invasive cancer patients based on the one-year treated mortality using the Carolina Mammography Registry data (Zhang 2011). To adjust for the untreated breast cancer, we use Verkooijen et. al's study (Verkooijen et al. 2005) on the effect of refusal of treatment on mortality, i.e., $P_t(\text{DUC}) = 3P_t(\text{DC})$, where $P_t(\text{DUC})$ and $P_t(\text{DC})$ are the annual mortality rates for untreated invasive cancer patients and treated invasive cancer patients at age *t*,

respectively.

Table 4 The distribution of treatment types by age

Treatment	First-time cancer and CBTR					IBTR
	40-49	50-59	60-69	70-79	≥80	
Lumpectomy alone	25.8%	27%	31%	37.6%	53.2%	0
Lumpectomy with radiotherapy	37.2%	37%	35%	28.4%	16.8%	0
Mastectomy	37%	36%	34%	34%	30%	100%

Source: Baxter et al. 2004
IBTR = ipsilateral breast tumor recurrence, CBTR = contralateral breast tumor recurrence

As explained in Section 2.2.3, the state transitions involve implicit treatment stage. Here we differentiate the probabilities of choosing different types of treatments for new cancers from those of choosing different salvage treatments for cancer recurrence. According to Zujewski et al. (2011), the use of breast-conserving surgery (BCS), including lumpectomy with or without radiotherapy, is correlated with age. Older patients are more likely to receive BCS than younger patients, while the use of radiotherapy following a BCS decreased with age. We adopt the treatment distribution of new in situ breast cancers reported by Baxter et al. (2004). With respect to cancer recurrences, mastectomy is considered to be the standard salvage therapy when IBTR occurs for patients who had BCS (Jatoi and Kaufmann 2010). An earlier study also reported that the vast majority of the patients who experienced in situ IBTR were treated with mastectomy (Solin et al. 2001). Although several cases of further BCS in patients previously receiving lumpectomy without radiotherapy have been reported, Harris and Solin found that a second BCS and mastectomy treatment result in nearly equivalent post-treatment survivals for in situ recurrence (Harris and Solin 2000). Therefore, we assume all the patients previously treated with lumpectomy with or without radiation will choose mastectomy as the salvage therapy when in situ IBTR occurs. However, for CBTR, to the best of our knowledge, there is no study available discussing the distribution of treatment types. Since CBTR is a tumor occurring in the

contralateral breast, which is independent of the primary breast tumor, we assume it has the same treatment distribution as the new cancer. Table 4 summarizes all the probability distributions of treatment types in different cases.

2.3.1.1 Parameters for State Transitions

In this section, we introduce parameter estimation for intermediate and lump-sum rewards. Biopsy is an invasive procedure that may place the patients at risk of morbidities and even mortalities. Thus, we assume the disutility of doing biopsy to be 2 weeks (Velanovich et al. 1995)**.** As a woman gets older, the risk and side-effects caused by biopsy also become more harmful. Hence, we assume the disutility associated with biopsy is inversely proportional to the age-specific EQ-5D scores, which reflects varying negative impacts of biopsy on women's health at different ages. EQ-5D is a utility-based measure of health status and widely used in clinical and economic evaluation of health care (EuroQol Group 1990). Here we employ the age-specific mean EQ-5D score reported by Saarni et al. (2006). All the disutility estimates are presented in Table 5.

Table 5 Disutility of Biopsy

Age group	40-44	45-54	55-64	65-74	75-84	>85
EQ-5D values	0.911	0.870	0.816	0.776	0.633	0.446
Disutility of biopsy (week)	2	2.09	2.23	2.35	2.88	4.09

The lump-sum rewards for absorbing states $e_t(\mathrm{M})$ and $e_t(\mathrm{I})$, are determined based on a patient's life expectancy. Hillner et al. (1996) showed that in situ cancers treated with mastectomy reduces women's life expectancy by 1.8% over a 20-year period. Based on this finding, we assume that the discounting factor of mastectomy that shortens women's life expectancy is 1.8%. Therefore, $e_t(\mathrm{M})$ is equal to the discounted population-based life expectancy of American women aged *t*, which is available in the United States life tables (Arias

2012). In addition, we use the breast cancer outcome calculator (http://www.lifemath.net/cancer/breastcancer/outcome/index.php) developed by the Laboratory for Quantitative Medicine to estimate the life expectancies (i.e. $e_t(\mathrm{I})$) for treated invasive cancer patients. We use ages and the mean size of screening detected invasive tumors (BCSC 2009) as the inputs for the calculation.

2.3.2 Experiment Results

Table 6 The personal status of the risk factors for example cases

Risk factors		Patient A	Patient B
Age		40	40
Race		white	Asian
Age at birth of first child		>30	<30
First-degree family history of breast cancer		Yes	No
Menopause status		Premenopausal	Premenopausal
Clinical history of in situ breast cancer			
	Scenario 1	None	None
	Scenario 2	in situ cancer is detected at age 45	in situ cancer is detected at age 45
	Scenario 3	in situ cancer is detected at age 70	in situ cancer is detected at age 70

In this section, we present two example cases to show the optimal biopsy referral policies for different women. The personal breast cancer risk attributes of the two examples are listed in Table 6. In these two case studies, we assume both patient A and patient B will turn into postmenopausal at age 51 based on the mean age of natural menopause reported by Kato et al.(1998). Patient A and patient B are representative of high-risk and low-risk women, respectively. Based on the BCSC dataset, the breast cancer incidence of the low-risk women represented by patient A is estimated to be 0.0049, while that of high-risk women represented by patient B is 0.0108. Therefore, patient A is expected to undergo more intensive screenings and follow-up biopsy exams for detecting breast cancer than patient B.

We study three scenarios to show the optimal policies for women with different clinical histories. Table 7 presents the optimal biopsy referral policies for the two women without clinical history on a 60-year horizon (i.e. from age 40 through age 99). In order to reach the highest life expectancy, these two women should follow the optimal policies until they reach age 100 or they are diagnosed with breast cancer. While patient A is recommended to receive 46 biopsies among 60 positive mammograms, patient B is only recommended to undergo 42 biopsies. This outcome is consistent with the assumption that patient A bears higher breast cancer risk than patient B does.

Table 7 The optimal biopsy referral policies for two patients in scenario 1 from age 40 to 90

Age	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
Patient A	W	W	W	W	W	B	B	B	B	B	W	W	B	B	B	B	B	B	B	B	B	B	B	B	B	W	W	W	W	W
Patient B	W	W	W	W	W	W	W	W	W	W	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	W	W	W	W	W
Age	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	91	92	93	94	95	96	97	98	99	
Patient A	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	W	W	
Patient B	B	B	B	W	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	W	W	

B= do biopsy, W=wait.

In this model, the optimal decisions at different ages depend on multiple variables, including the incidence rate, ratio of invasive cancer to all breast cancers, disutility of biopsy, death rate from invasive cancer, and distribution of treatment types. On the one hand, the monotonicity of these variables are not consistent. On the other hand, even some of these variables are not monotonic and fluctuate over age. As a result, our optimal policies demonstrate non-monotone patterns. For instance, patient A is recommended to do biopsy from age 45 but skip biopsy at ages 50 and 51. According to the data, the incidence rates of breast cancer for patient A from age 40 to 44 are lower than those after age 44, which explains why patient A should start follow-up biopsy from age 45. However, the incidence rates of breast cancer at ages 50 and 51 are not significantly different from those between ages 45 and 60. This optimal action “wait” at ages 50

and 51 is probably because of a relatively low proportion of invasive cancers among new cancer cases at these two ages for patient A, since a lower possibility of developing malignant breast cancer means less necessity for biopsy exam. In a similar manner, the optimal policy flips between “W” and “B” with age to reach an optimal tradeoff among different non-monotone determinants.

Our results partly explain the controversy on screening guidelines in the public health community. While the ACS recommends women to start receiving routine mammography from age 45, the U.S. Preventive Services Task Force (USPTF) and the American College of Preventive Medicine (ACMP) suggest the starting age of mammography should be 50 (Oeffinger et al. 2015, Siu 2016). However, our results show that the first years of starting biopsies are age 50 and age 45 for the patients with low risk and high risk, respectively. Since both women are not recommended to do biopsy before these two ages, no matter if the mammogram is positive or negative, their mammography decisions should also start from age 50 and age 45 respectively. This result also aligns with some individual studies, which argue that women should begin mammography screenings between age 45 and 50 (Antman and Shea 1999).

The patient B’s optimal referral policy recommends women suspend biopsies at age 50 and 51 after five consecutive biopsy recommendations between age 45 and age 49, which implicitly reflects the USPSTF’s concern that the potential benefit of frequent breast cancer screening before age 50 cannot offset the cumulative harms caused by potential false-positive results and potential over-diagnosis of mammograms as well as follow-up biopsies.

It is worth noting that both patients are recommended to skip mammograms between ages 65 and 69, consistent with some clinical studies indicating that recall rates were significantly lower for women aged 65–69 compared with women aged 50–64 (Moss et al. 2011). The experimental

results also echo some other researchers' efforts on the improvement of screening mammography. For instance, we recommend patient A skip biopsies as well as mammograms in 4 out of 9 years from age 66 to age 74. In addition, patient B, who bears relatively lower risk of breast cancer, is recommended to skip 5 out of 9 biopsies as well as mammograms during the same interval. To some extent, it coincides with Braithwaite et al.'s study on women undergoing annual mammograms between ages 66 and 89, which suggests that women aged 66 to 74 years should reduce the frequency of routine screenings to biennial (Braithwaite et al. 2013). Therefore, from the perspective of breast cancer prevention, age 66 to age 74 can be viewed as a "low-risk interval" that does not need intensive screenings.

Table 8 The 20-year optimal post-treatment biopsy referrals for patients in scenarios 2 and 3

state	P_L^1	P_L^2	P_L^3	P_L^4	P_L^5	P_L^6	P_L^7	P_L^8	P_L^9	P_L^{10}	P_L^{11}	P_L^{12}	P_L^{13}	P_L^{14}	P_L^{15}	P_L^{16}	P_L^{17}	P_L^{18}	P_L^{19}	P_L^{20}
Age	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
Patient A	B	B	B	B	W	W	W	W	W	B	B	B	B	B	W	W	W	W	W	W
Patient B	B	B	B	B	W	W	W	W	W	B	B	B	B	B	W	W	W	W	W	W
state	P_R^1	P_R^2	P_R^3	P_R^4	P_R^5	P_R^6	P_R^7	P_R^8	P_R^9	P_R^{10}	P_R^{11}	P_R^{12}	P_R^{13}	P_R^{14}	P_R^{15}	P_R^{16}	P_R^{17}	P_R^{18}	P_R^{19}	P_R^{20}
Age	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
Patient A	B	B	B	B	W	W	W	W	W	W	W	W	W	W	B	B	B	B	B	B
Patient B	B	B	B	B	W	W	W	W	W	W	W	W	W	W	B	B	B	B	B	B

state	P_L^1	P_L^2	P_L^3	P_L^4	P_L^5	P_L^6	P_L^7	P_L^8	P_L^9	P_L^{10}	P_L^{11}	P_L^{12}	P_L^{13}	P_L^{14}	P_L^{15}	P_L^{16}	P_L^{17}	P_L^{18}	P_L^{19}	P_L^{20}
Age	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Patient A	B	B	B	B	B	W	W	W	W	W	B	W	W	W	B	B	B	B	B	W
Patient B	B	B	B	B	B	W	W	W	W	W	B	W	W	W	B	B	B	B	B	W
state	P_R^1	P_R^2	P_R^3	P_R^4	P_R^5	P_R^6	P_R^7	P_R^8	P_R^9	P_R^{10}	P_R^{11}	P_R^{12}	P_R^{13}	P_R^{14}	P_R^{15}	P_R^{16}	P_R^{17}	P_R^{18}	P_R^{19}	P_R^{20}
Age	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Patient A	B	B	B	B	W	W	W	W	W	W	W	W	W	W	B	B	B	W	W	W
Patient B	B	B	B	B	W	W	W	W	W	W	W	W	W	W	B	B	B	W	W	W

B= do biopsy, W=wait.

Next, we investigate the cases of the patients with clinical history of in situ cancer. Since we use the same yearly recurrence rates for all patients, we expect there is little difference between different patient's optimal post-treatment biopsy referral policies. Therefore, only age, number of years after treatment, and treatment option are determinants of the optimal biopsy policies for these cases. In scenarios 2 and 3, two patients are diagnosed with in situ breast cancer at age 45 and age 70, respectively. Table 8 presents the optimal biopsy referral policies for the two scenarios under different treatments. As expected, the two patients have identical optimal biopsy referral policies if they are in the same post-treatment year and received the same treatment. Moreover, neither lumpectomy alone nor lumpectomy with radiotherapy indicates a significant predominance in the total number of optimal biopsy decisions.

2.3.3 Sensitivity Analysis

We conduct two sensitivity analyses by varying two parameters: regression rate and disutility of biopsy. Since regression rate reflects "self-healing" property of breast tumors, the larger the regression rate is, the less likely that the patient should receive a follow-up biopsy. Figure 6 shows the decreasing trend of total number of optimal biopsy decisions with the increase of regression rate. The inflection points of regression rate for patient B is approximately 35%, while varying regression rate for patient A does not impact her optimal biopsy referral policy.

These results suggest that the spontaneous regression of breast tumors does not necessarily influence our biopsy referral policy. As compared to patient B, patient A's risk of developing breast cancer is too high to be counteracted by potential spontaneous regression of in situ tumor. It is worth mentioning that the decrease in the total number of optimal biopsy decisions predominantly occurs in young women in our finding. The proportion of in situ cancers among all breast cancer cases is very low among very elderly (Barlow et al. 2006). Particularly, in our

dataset, the percentage of in situ cancers is close to zero for the patients older than 80. Therefore, the spontaneous regression of in situ tumors has little effect on the biopsy decisions for older women.

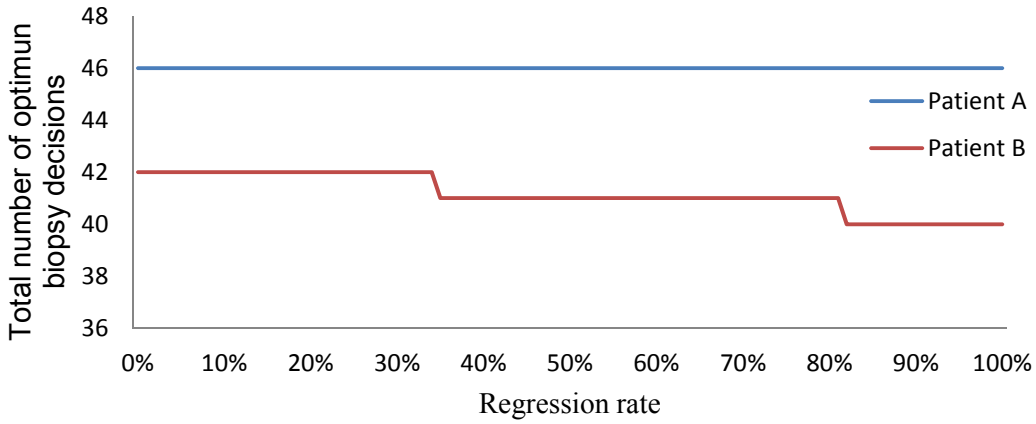


Figure 6 The trend of total number of optimum biopsies from age 40 through 100 when varying the regression rate

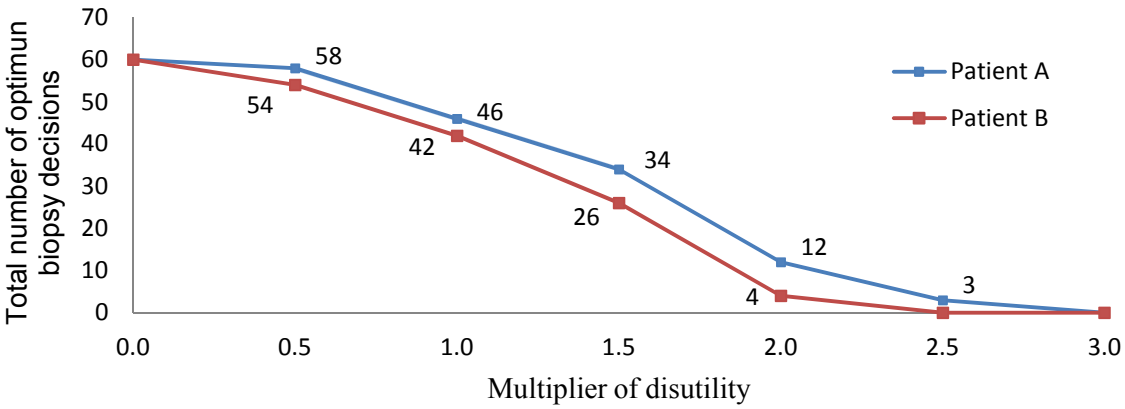


Figure 7 The trend of total number of optimum biopsies from age 40 through 100 when varying the biopsy disutility

We then vary the magnitude of biopsy disutility by multiplying the base value (i.e. 2 weeks at age 40) using different coefficients (Fig 7). When the multiplier is set to zero, which means biopsy has no negative impact on patient's health so that doing biopsy would not reduce intermediate life years at the current decision epoch, the optimal action is always undertaking biopsy. As the disutility of biopsy increases, the patient is increasingly likely opt to "wait" when

receiving a positive mammogram, since the reduction of mortality by choosing biopsy is no longer sufficient to offset inherent harm and risk posed on health brought by the biopsy exam. When the disutility of biopsy increases threefold (6 weeks at age 40), both patient A and patient B should never do biopsy in their lifetime.

2.4 Conclusion and Limitations

2.4.1 Discussion

This study provides personalized optimal follow-up biopsy referral policies for both screening mammography and surveillance mammography. We aim at analyzing the common over-diagnosis problem in the current annual mammography practice of breast cancer. This problem is formulated as an MDP model with a value function of maximizing total life expectancy. Although we focus on decisions of biopsy referring after positive mammograms, our policy is in fact equivalent to a mammography guideline. If a patient is recommended not to do biopsy in the current year, the corresponding mammogram is actually also unnecessary, since no matter whether the outcome is positive or negative, the woman will wait until the next mammogram. Previous medical literature suggests that the dose required for a mammogram is so small that the associated risk of radiation exposure is negligible (Berry et al. 2005). Therefore, it is reasonable to assume that mammography is a harmless exam without significant disutility. Removing the mammograms associated with optimal decision of "wait" from our model has no effect on model output. As a result, our study suggests that existing mammography guidelines proposed by the healthcare organizations, which recommend periodic screening mammography for all women, can be improved. The schedule of mammography should be determined on a patient-specific basis.

We also discuss the influence of potential spontaneous regression of breast tumor on the

biopsy referral decision. Despite lack of sufficient direct evidence, there have been concerns that spontaneous regression of tumor may lead to a percentage of over-diagnosis in the breast cancer screening practice. Our sensitivity analysis finds that regression has no effect on high risk women, but a certain level of regression rate may reduce the need for biopsy exams in low-risk women.

In addition, our model addresses the surveillance mammography after completion of treatment for in situ cancers. Our case studies demonstrate that surveillance mammography schedules can also be personalized. Although we have only three-dimensional determinants (i.e. age, number of years after treatment, and treatment type) for biopsy referral policy of following surveillance mammography, our results still reveal the differences between the optimal policies of people with different combinations of determinants. Previous studies showed that the recurrence risk after breast-conserving surgery varies nonlinearly over time and risk factors for recurrence need to be viewed differently based on the time frame of follow-ups (Shaitelman et a. 2012, Braithwaite at al. 2013). Our experimental results demonstrate that the optimal biopsy policies fluctuate dramatically over years, which are consistent with the non-monotone pattern of post-treatment recurrence risk over a long-term time horizon.

Moreover, although radiotherapy after lumpectomy is proving to significantly reduce the risk of local recurrence after treatment of in situ cancer, our study does not show that patients with lumpectomy plus radiotherapy will perform less biopsies than those with lumpectomy alone, which is a somewhat counter-intuitive result since the good performance of postoperative radiotherapy in reducing local recurrence has been emphasized for years.

2.4.2 Limitations and Future work

There are several limitations in our study. First, we use a general set of yearly recurrence rates

for all patients. In fact, previous studies found several major risk factors associated with breast cancer are also related to recurrence rate after initial diagnosis of in situ cancer (Buist et al.2010). Future work using patient-specific yearly recurrence rates may provide more realistic personalized biopsy referral polices for surveillance mammography. Second, our MDP model only considers two options, skipping or doing biopsy. As a matter of fact, there exist more approaches to early detect breast cancer, such as breast self-examination. Therefore, our biopsy referral policy is very conservative, since introducing self-examination would potentially reduce mortality brought by skipping mammograms and follow-up biopsies. Including breast self-examination as an option in the model will represent the real situations better. Third, in order to fit our MDP model, the risk factors we used to categorize patients do not include any dynamic breast cancer risk factors (e.g. BMI and breast density), which may change over time. For instance, since we do not know a woman's BMI at her next age, it is very difficult to determine her breast cancer risk in the following year as well as all future years when we consider BMI as a determinant of breast cancer risk. As a result, we are unable to calculate the associated transition probabilities. There are a lot of dynamic breast cancer risk factors identified in our raw dataset, including BMI, prior breast procedure, breast density, result of last mammogram and hormone therapy use. We plan to incorporate the effects of these dynamic factors in our future work. Lastly, the experimental results of the optimal biopsy referral policies are completely data-driven, which are inevitably impacted by the statistical noise and numerical errors in the input data. As a result, the optimal policy may not be a generally optimal. The flips in the optimal policies are therefore insufficiently justified. Our future research will employ some data-smoothing techniques to mitigate these undesirable impacts from the input data.

Acknowledgement

Data collection and sharing was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040, HHSN261201100031C). A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/

References

American Cancer Society. (2017) Cancer Facts & Figures 2017.Atlanta: American Cancer Society;

Antman, K., & Shea, S. (1999). Screening mammography under age 50. *JAMA: the journal of the American Medical Association*, *281*(16), 1470-1472.

Arias, E. (2012). United States life tables, 2008. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 54(14), 1-40.

Ayvaci, M. U., Alagoz, O., & Burnside, E. S. (2012). The effect of budgetary restrictions on breast cancer diagnostic decisions. Manufacturing & Service Operations Management, 14(4), 600-617.

Ayer, T., Alagoz, O., & Stout, N. K. (2012). OR Forum—A POMDP Approach to Personalize Mammography Screening Decisions. Operations Research, 60(5), 1019-1034.

Barlow, W. E., White, E., Ballard-Barbash, R., Vacek, P. M., Titus-Ernstoff, L., Carney, P. A., ... & Kerlikowske, K. (2006). Prospective breast cancer risk prediction model for women undergoing screening mammography. Journal of the National Cancer Institute, 98(17), 1204-1214.

Baxter, N. N., Virnig, B. A., Durham, S. B., & Tuttle, T. M. (2004). Trends in the treatment of ductal carcinoma in situ of the breast. *Journal of the National Cancer Institute*, *96*(6), 443-448.

Berry, D. A., Cronin, K. A., Plevritis, S. K., Fryback, D. G., Clarke, L., Zelen, M., ... & Feuer, E. J. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine*, *353*(17), 1784-1792.

Boyages, M. B. B. S., Delaney, M. B. B. S., & Taylor, M. B. B. S. (1999). Predictors of local recurrence after treatment of ductal carcinoma in situ. Cancer, 85(3), 616-628.

Braithwaite, D., Zhu, W., Hubbard, R. A., O'Meara, E. S., Miglioretti, D. L., Geller, B., ... & Kerlikowske, K. (2013). Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates?. Journal of the National Cancer Institute, 105(5), 334-341.

Breast Cancer Surveillance Consortium (2009) Performance Measures for 1,960,150 Screening Mammography Examinations from 2002 to 2006 by Age. Retrieved July 25, 2015 from http://breastscreening.cancer.gov/data/performance/screening/2009/perf_age.html

Breast Cancer Surveillance Consortium (2009) Cancers for 2,264,089 Screening Mammography Examinations from 2002 – 2006. Retrieved July 25, 2015 from http://breastscreening.cancer.gov/data/benchmarks/screening/2009/table4.html

Bruening, W., Schoelles, K., Treadwell, J., Launders, J., Fontanarosa, J., & Tipton, K. (2009). Comparative effectiveness of core-needle and open surgical biopsy for the diagnosis of breast lesions.

Buist, D. S., Abraham, L. A., Barlow, W. E., Krishnaraj, A., Holdridge, R. C., Sickles, E. A., ... & Geller, B. M. (2010). Diagnosis of second breast cancer events after initial diagnosis of early stage breast cancer. Breast cancer research and treatment, 124(3), 863-873.

Chhatwal, J., Alagoz, O., & Burnside, E. S. (2010). Optimal breast biopsy decision-making based on mammographic features and demographic factors. Operations research, 58(6), 1577-1591.

Clarke, M., Collins, R., Darby, S., Davies, C., Elphinstone, P., Evans, E., ... & Wang, Y. (2005). Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, *366*(9503), 2087.

DeSantis, C., Siegel, R., Bandi, P., & Jemal, A. (2011). Breast cancer statistics, 2011. *CA: A Cancer Journal for Clinicians*, *61*(6), 408-418.

EuroQol Group. (1990). EuroQol-a new facility for the measurement of health-related quality of life. Health Policy, 16(3), 199-208.

Fong, J., Kurniawan, E. D., Rose, A. K., Mou, A., Collins, J. P., Miller, J. A., & Mann, G. B. (2011). Outcomes of screening-detected ductal carcinoma in situ treated with wide excision alone. Annals of surgical oncology, 18(13), 3778-3784.

Fryback, D. G., Stout, N. K., Rosenberg, M. A., Trentham-Dietz, A., Kuruchittham, V., & Remington, P. L. (2006). The Wisconsin breast cancer epidemiology simulation model. JNCI Monographs, 2006(36), 37-47.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mulvihill, J. J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, *81*(24), 1879-1886.

Gail, M. H., Costantino, J. P., Bryant, J., Croyle, R., Freedman, L., Helzlsouer, K., & Vogel, V. (1999). Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *Journal of the National Cancer Institute*, *91*(21), 1829-1846.

Habel, L. A., Moe, R. E., Daling, J. R., Holte, S., Rossing, M. A., & Weiss, N. S. (1997). Risk of contralateral breast cancer among women with carcinoma in situ of the breast. Annals of surgery, 225(1), 69.

Harris, E. E., & Solin, L. J. (2000). The diagnosis and treatment of ductal carcinoma in situ of the breast. The breast journal, 6(2), 78-95.

Hillner, B. E., Desch, C. E., Carlson, R. W., Smith, T. J., Esserman, L., & Bear, H. D. (1996). Trade-offs between survival and breast preservation for three initial treatments of ductal carcinoma-in-situ of the breast. Journal of clinical oncology, 14(1), 70-77.

Jatoi, I., & Kaufmann, M. (2010). Management of breast diseases. Springer.

Joensuu, H., & Lundin, J. (2004). Spontaneous Regression of Cancerous Tumors Detected by Mammography Screening—Reply. JAMA: The Journal of the American Medical Association, 292(21), 2580-2580.

Kato, I., Toniolo, P., Akhmedkhanov, A., Koenig, K. L., Shore, R., & Zeleniuch-Jacquotte, A. (1998). Prospective study of factors influencing the onset of natural menopause. Journal of clinical epidemiology, 51(12), 1271-1276.

Khatcheressian, J. L., Wolff, A. C., Smith, T. J., Grunfeld, E., Muss, H. B., Vogel, V. G., ... & Davidson, N. E. (2006). American Society of Clinical Oncology 2006 update of the breast cancer follow-up and management guidelines in the adjuvant setting. Journal of Clinical Oncology, 24(31), 5091-5097.

Lash, T. L., Fox, M. P., Buist, D. S., Wei, F., Field, T. S., Frost, F. J., ... & Silliman, R. A. (2007). Mammography surveillance and mortality in older breast cancer survivors. Journal of clinical oncology, 25(21), 3001-3006.

Maillart, L. M., Ivy, J. S., Ransom, S., & Diehl, K. (2008). Assessing dynamic breast cancer screening policies. Operations Research, 56(6), 1411-1427.

Moss, S. M., Brown, J., Garvican, L., Coleman, D. A., Johns, L. E., Blanks, R. G., ... & Gamble, P. (2001). Routine breast screening for women aged 65–69: results from evaluation of the demonstration sites. British journal of cancer, 85(9), 1289.

National cancer institute. Dictionary of cancer terms. Retrieved July 25, 2015 from http://www.cancer.gov/dictionary

Oeffinger, K. C., Fontham, E. T., Etzioni, R., Herzig, A., Michaelson, J. S., Shih, Y. C. T., ... & Wolf, A. M. (2015). Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. Jama, 314(15), 1599-1614.

Parker, S. H., Burbank, F., Jackman, R. J., Aucreman, C. J., Cardenosa, G., Cink, T. M., ... & Garver, P. R. (1994). Percutaneous large-core breast biopsy: a multi-institutional study. Radiology, 193(2), 359-364.

Saarni, S. I., Härkänen, T., Sintonen, H., Suvisaari, J., Koskinen, S., Aromaa, A., & Lönnqvist, J. (2006). The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. Quality of Life Research, 15(8), 1403-1414.

Schootman, M., Jeffe, D. B., Reschke, A. H., & Aft, R. L. (2003). Disparities related to socioeconomic status and access to medical care remain in the United States among women who never had a mammogram. Cancer Causes & Control, 14(5), 419-425.

Shaitelman, S. F., Wilkinson, J. B., Kestin, L. L., Ye, H., Goldstein, N. S., Martinez, A. A., & Vicini, F. A. (2012). Long-Term Outcome in Patients With Ductal Carcinoma In Situ Treated With Breast-Conserving Therapy: Implications for Optimal Follow-up Strategies. International Journal of Radiation Oncology* Biology* Physics, 83(3), e305-e312.

Siu, A. L. (2016). Screening for breast cancer: US Preventive Services Task Force recommendation

Smith, R. A., Cokkinides, V., Brooks, D., Saslow, D., Shah, M., & Brawley, O. W. (2011). Cancer screening in the United States, 2011

Solin, L. J., Fourquet, A., Vicini, F. A., Haffty, B., Taylor, M., McCormick, B., ... & Schultz, D. J. (2001). Salvage treatment for local recurrence after breast‐conserving surgery and radiation as initial treatment for mammographically detected ductal carcinoma in situ of the breast. Cancer, 91(6), 1090-1097.

Solin, L. J., Fourquet, A., Vicini, F. A., Taylor, M., Olivotto, I. A., Haffty, B., ... & Hwang, W. T. (2005). Long‐term outcome after breast‐conservation treatment with radiation for mammographically detected ductal carcinoma in situ of the breast. Cancer, 103(6), 1137-1146.

Sonnenberg, F. A., & Beck, J. R. (1993). Markov models in medical decision making a practical guide. Medical decision making, 13(4), 322-338.

Tabar, L., Yen, M. F., Vitak, B., Chen, H. H., Smith, R. A., & Duffy, S. W. (2003). Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet*, *361*(9367), 1405.

Velanovich, V. (1995). Immediate biopsy versus observation for abnormal findings on mammograms: an analysis of potential outcomes and costs. *The American journal of surgery*, *170*(4), 327-332.

Verkooijen, H. M., Fioretta, G. M., Rapiti, E., Bonnefoi, H., Vlastos, G., Kurtz, J., ... & Bouchardy, C. (2005). Patients' refusal of surgery strongly impairs breast cancer survival. Annals of surgery, 242(2), 276.

Viani, G. A., Stefano, E. J., Afonso, S. L., De Fendi, L. I., Soares, F. V., Leon, P. G., & Guimarães, F. S. (2007). Breast-conserving surgery with or without radiotherapy in women with ductal carcinoma in situ: a meta-analysis of randomized trials. Radiat oncol, 2(1), 28.

Wapnir, I. L., Dignam, J. J., Fisher, B., Mamounas, E. P., Anderson, S. J., Julian, T. B., ... & Wolmark, N. (2011). Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. Journal of the National Cancer Institute, 103(6), 478-488.

Zahl, P. H., Mæhlen, J., & Welch, H. G. (2008). The natural history of invasive breast cancers detected by screening mammography. Archives of internal medicine, 168(21), 2311.

Zhang, S., & Ivy, J., (2012) Analytic Modeling of Breast Cancer Spontaneous Regression. Proceedings of the 2012 Industrial and Systems Engineering Research Conference, Orlando, FL, 2012.

Zhang S. (2011) Modeling the Complexity of Breast Cancer under Conditions of Uncertainty. PhD dissertation, North Carolina State University.

Zhi, H., Ou, B., Luo, B. M., Feng, X., Wen, Y. L., & Yang, H. Y. (2007). Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions. Journal of Ultrasound in Medicine, 26(6), 807-815.

Zujewski, J. A., Harlan, L. C., Morrell, D. M., & Stevens, J. L. (2011). Ductal carcinoma in situ: trends in treatment over time in the US. Breast cancer research and treatment, 127(1), 251-257.

Appendix

A. Estimation for Transition Probabilities

Parameters

$P(DO)$ = annual death probability (excluding death of breast cancer)

$P(DUC)$ = annual death probability of untreated invasive breast cancer patient

$P(DC)$ = annual death probability of treated invasive cancer patient

$P(NC-C)$ = annual incidence rate of new breast cancer

$P(NC-NC) = 1- P(NC-C)$

$P(ins|C)$ = proportion of in situ cancers among all breast cancers

$P(inv|C)$ = proportion of invasive cancers among all breast cancers

$P(NC|Negative)$ = percentage of negative mammograms in which no cancer exists.

$P(C|Negative)$ = percentage of negative mammograms in which cancers exist.

$P(NC|Positive)$ = percentage of positive mammograms in which no cancer exists.

$P(C|Positive)$ = percentage of positive mammograms in which cancers exist.

$P(Negative|NC)$ = percentage of non-cancers that are interpreted as negative mammograms

$P(Positive|NC)$ = percentage of non-cancers that are interpreted as negative mammograms

$P(Negative|C)$ = percentage of cancers that are interpreted as negative mammograms

$P(Positive|C)$ = percentage of cancers that are interpreted as negative mammograms

$P(R)$ = annual regression rate of in situ cancer

$P(L|ins)$ = probability of choosing lumpectomy treatment for in situ patients

$P(R|ins)$ = probability of choosing lumpectomy with radiotherapy treatment for in situ patients

$P(M|ins)$ = probability of choosing mastectomy treatment for in situ patients

$P_L^n(IBTR)$ = recurrence rate of IBTR in the *n*th year after completion of lumpectomy

$P_L^n(CBTR)$ = recurrence rate of CBTR in the *n*th year after completion of lumpectomy

$P_R^n(IBTR)$ = recurrence rate of IBTR in the *n*th year after completion of lumpectomy with radiotherapy

$P_R^n(CBTR)$ = recurrence rate of CBTR in the *n*th year after completion of lumpectomy with radiotherapy

$P(ins, IBTR|C_L^n)$ = proportion of in situ IBTRs in all the recurrences in the *n*th year after completion of lumpectomy

$P(inv, IBTR|C_L^n)$ = proportion of invasive IBTRs in all the recurrences in the *n*th year after completion of lumpectomy

$P(ins, CBTR|C_L^n)$ = proportion of in situ IBTRs in all the recurrences in the *n*th year after completion of lumpectomy

$P(inv, CBTR|C_L^n)$ = proportion of invasive CBTRs in all the recurrences in the *n*th year after completion of lumpectomy

$\mathrm{P}(\mathrm{ins}, \mathrm{IBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})$ = proportion of in situ IBTRs in all the recurrences in the *n*th year after completion of lumpectomy with radiotherapy

$\mathrm{P}(\mathrm{inv}, \mathrm{IBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})$ = proportion of invasive IBTRs in all the recurrences in the *n*th year after completion of lumpectomy with radiotherapy

$\mathrm{P}(\mathrm{ins}, \mathrm{CBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})$ = proportion of in situ IBTRs in all the recurrences in the *n*th year after completion of lumpectomy with radiotherapy

$\mathrm{P}(\mathrm{inv}, \mathrm{CBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})$ = proportion of invasive IBTRs in all the recurrences in the *n*th year after completion of lumpectomy with radiotherapy

Probabilities

Starting from N

- $\mathrm{p}(\mathrm{N}|\mathrm{N}, \mathrm{W}) = \mathrm{P}(\mathrm{NC}|\mathrm{Negative})\big(1 - \mathrm{P}(\mathrm{DO})\big)\big(\mathrm{P}(\mathrm{NC} - \mathrm{NC})\mathrm{P}(\mathrm{Negative}|\mathrm{NC}) + \mathrm{P}(\mathrm{NC} - \mathrm{C})\mathrm{P}(\mathrm{Negative}|\mathrm{C})\big) + \mathrm{P}(\mathrm{C}|\mathrm{Negative})\big((\mathrm{P}(\mathrm{ins}|\mathrm{C})\big(1 - \mathrm{P}(\mathrm{DO})\big)\big(\mathrm{P}(\mathrm{R})\mathrm{P}(\mathrm{Negative}\,|\,\mathrm{NC}) + (1 - \mathrm{P}(\mathrm{R}))\mathrm{P}(\mathrm{Negative}\,|\,\mathrm{NC})\big) + \mathrm{P}(\mathrm{inv}|\mathrm{C})(1 - \mathrm{P}(\mathrm{DO})(1 - \mathrm{P}(\mathrm{DUC}))\mathrm{P}(\mathrm{Negative}|\mathrm{C})\big)$
- $\mathrm{p}(\mathrm{D}|\mathrm{N}, \mathrm{W}) = \mathrm{P}(\mathrm{NC}|\mathrm{Negative})\mathrm{P}(\mathrm{DO}) + \mathrm{P}(\mathrm{C}|\mathrm{Negative})\Big(\mathrm{P}(\mathrm{ins}|\mathrm{C})\mathrm{P}(\mathrm{DO}) + \mathrm{P}(\mathrm{inv}|\mathrm{C})\Big(1 - \big(1 - \mathrm{P}(\mathrm{DO})\big)\big(1 - \mathrm{P}(\mathrm{DUC})\big)\Big)\Big)$
- $\mathrm{p}(\mathrm{P}|\mathrm{N}, \mathrm{W}) = 1 - \mathrm{p}(\mathrm{N}|\mathrm{N}, \mathrm{W}) - \mathrm{p}(\mathrm{D}|\mathrm{N}, \mathrm{W})$

Starting from P

Action: B

- $p(P|P,B) = P(NC|Positive)(1 - P(DO))(P(NC - NC)P(Positive|NC) + P(NC - C)P(Positive|C))$
- $p(N|P,B) = P(NC|Positive)(1 - P(DO)(P(NC - NC)P(Negative|NC) + P(NC - C)P(Negative|C))$
- $p(P_L^1|P,B) = P(C|Positive)P(ins|C)P(L|ins)(1 - P(DO))\left(\left(1 - \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)\right)P(Positive|C) + \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)P(Positive|NC)\right)$
- $p(N_L^1|P,B) = P(C|Positive)P(ins|C)P(L|ins)(1 - P(DO))\left(\left(1 - \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)\right)P(Negative|C) + \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)P(Negative|NC)\right)$
- $p(P_R^1|P,B) = P(C|Positive)P(ins|C)P(R|ins)(1 - P(DO))\left(\left(1 - \left(1 - P_R^1(IBTR)\right)\left(1 - P_R^1(CBTR)\right)\right)P(Positive|C) + \left(1 - P_R^1(IBTR)\right)\left(1 - P_R^1(CBTR)\right)P(Positive|NC)\right)$
- $p(N_R^1|P,B) = P(C|Positive)P(ins|C)P(R|ins)(1 - P(DO))\left(\left(1 - \left(1 - P_R^1(IBTR)\right)\left(1 - P_R^1(CBTR)\right)\right)P(Positive|C) + \left(1 - P_R^1(IBTR)\right)\left(1 - P_R^1(CBTR)\right)P(Positive|NC)\right)$
- $p(M|P,B) = P(C|Positive)P(ins|C)P(M|ins)$
- $p(I|P,B) = P(C|Positive)P(inv|C)$
- $p(D|P,B) = 1 - p(P|P,B) - p(N|P,B) - p(P_L^1|P,B) - p(N_L^1|P,B) - p(P_R^1|P,B) - p(N_R^1|P,B) - p(M|P,B) - p(I|P,B)$

Action: W

- $p(N|P,W) = P(NC|Positive)\big(1 - P(DO)\big)\big(P(NC - NC)P(Negative|NC) + P(NC - C)P(Negative|C)\big) + P(C|Positive)\big((P(ins|C)\big(1 - P(DO)\big)\big(P(R)P(Negative\,|\,NC) + (1 - P(R))P(Negative\,|\,NC)\big) + P(inv|C)(1 - P(DO)(1 - P(DUC))P(Negative|C)\big)$
- $p(D|P,W) = P(NC|Positive)P(DO) + P(C|Positive)\Big(P(ins|C)P(DO) + P(inv|C)\Big(1 - \big(1 - P(DO)\big)\big(1 - P(DUC)\big)\Big)\Big)$
- $p(P|P,W) = 1 - p(N|P,W) - p(D|P,W)$

Starting from N_L^n

- $p(N_L^{n+1}|N_L^n,W) = P(NC|Negative)\big(1 - P(DO)\big)\Big(\Big(1 - \big(1 - P_L^n(IBTR)\big)\big(1 - P_L^n(CBTR)\big)\Big)P(Negative|C) + \big(1 - P_L^n(IBTR)\big)\big(1 - P_L^n(CBTR)\big)P(Negative|NC)\Big) + P(C|Negative)\Big(\big(P(ins,IBTR|C_L^n) + P(ins,CBTR|C_L^n)\big)\big(1 - P(DO)\big)\Big(P(R)P(Negative|NC) + \big(1 - P(R)\big)P(Negative|C)\Big) + \big(P(inv,IBTR|C_L^n) + P(inv,CBTR|C_L^n)\big)\big(1 - P(DO)\big)\big(1 - P(DUC)\big)P(Negative|C)\Big)$
- $p(D|N_L^n,W) = P(NC|Negative)P(DO) + P(C|Negative)\Big(\big(P(ins,IBTR|C_L^n) + P(ins,CBTR|C_L^n)\big)P(DO) + \big(P(inv,CBTR|C_L^n) + P(inv,CBTR|C_L^n)\big)\Big(1 - \big(1 - P(DO)\big)\big(1 - P(DUC)\big)\Big)\Big)$
- $p(P_L^{n+1}|N_L^n,W) = 1 - p(N_L^{n+1}|N_L^n,W) - p(D|N_L^n,W)$

Starting from $\mathbf{P_L^n}$

Action: B

- $p(P_L^{n+1}|P_L^n, B) = P(NC|Positive)(1 - P(DO))\left(\left(1 - \left(1 - P_L^n(IBTR)\right)\left(1 - P_L^n(CBTR)\right)\right)P(Positive|C) + \left(1 - P_L^n(IBTR)\right)\left(1 - P_L^n(CBTR)\right)P(Positive|NC)\right)$
- $p(N_L^{n+1}|P_L^n, B) = P(NC|Positive)(1 - P(DO))\left(\left(1 - \left(1 - P_L^n(IBTR)\right)\left(1 - P_L^n(CBTR)\right)\right)P(Negative|C) + \left(1 - P_L^n(IBTR)\right)\left(1 - P_L^n(CBTR)\right)P(Negative|NC)\right)$
- $p(P_L^1|P_L^n, B) = P(C|Positive)(1 - P(DO))(P(ins, CBTR|C_L^n)P(L|ins))\left(\left(1 - \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)\right)P(Positive|C) + \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)P(Positive|NC)\right)$
- $p(N_L^1|P_L^n, B) = P(C|Positive)(1 - P(DO))(P(ins, CBTR|C_L^n)P(L|ins))\left(\left(1 - \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)\right)P(Negative|C) + \left(1 - P_L^1(IBTR)\right)\left(1 - P_L^1(CBTR)\right)P(Negative|NC)\right)$
- $p(P_R^1|P_L^n, B) = P(C|Positive)(1 - P(DO))\,(P(ins, CBTR|C_L^n)P(R|ins))\left(\left(1 - \left(1 - P_R^1(IBTR)\right)\left(1 - P_R^1(CBTR)\right)\right)P(Positive|C) + \left(1 - P_R^1(IBTR)\right)\left(1 - P_R^1(CBTR)\right)P(Positive|NC)\right)$

- $p(N_R^1|P_L^n, B) = P(C|Positive)\big(1 - P(DO)\big)\big(P(ins, CBTR|C_L^n)P(R|ins)\big)\Big(\Big(1 - \big(1 - P_R^1(IBTR)\big)\big(1 - P_R^1(CBTR)\big)\Big)P(Negative|C) + \big(1 - P_R^1(IBTR)\big)\big(1 - P_R^1(CBTR)\big)P(Negative|NC)\Big)$
- $p(I|P_L^n, B) = P(C|Positive)\big(P(inv, IBTR|C_L^n) + P(inv, CBTR|C_L^n)\big)$
- $p(M|P_L^n, B) = P(C|Positive)\big(P(ins, IBTR|C_L^n) + P(ins, CBTR|C_L^n)P(M|ins)\big)$
- $p(D|P_L^n, B) = 1 - p(P_L^{n+1}|P_L^n, B) - p(N_L^{n+1}|P_L^n, B) - p(P_L^1|P_L^n, B) - p(N_L^1|P_L^n, B) - p(P_R^1|P_L^n, B) - p(N_R^1|P_L^n, B) - p(I|P_L^n, B) - p(M|P_L^n, B)$

Action: W

- $p(N_L^{n+1}|P_L^n, W) = P(NC|Positive)\big(1 - P(DO)\big)\Big(\Big(1 - \big(1 - P_L^n(IBTR)\big)\big(1 - P_L^n(CBTR)\big)\Big)P(Negative|C) + \big(1 - P_L^n(IBTR)\big)\big(1 - P_L^n(CBTR)\big)P(Negative|NC)\Big) + P(C|Positive)\Big(\big(P(ins, IBTR|C_L^n) + P(ins, CBTR|C_L^n)\big)\big(1 - P(DO)\big)\Big(P(R)P(Negative|NC) + \big(1 - P(R)\big)P(Negative|C)\Big) + \big(P(inv, IBTR|C_L^n) + P(inv, CBTR|C_L^n)\big)\big(1 - P(DO)\big)\big(1 - P(DUC)\big)P(Negative|C)\Big)$
- $p(D|P_L^n, W) = P(NC|Positive)P(DO) + P(C|Positive)\Big(\big(P(ins, IBTR|C_L^n) + P(ins, CBTR|C_L^n)\big)P(DO) + \big(P(inv, CBTR|C_L^n) + P(inv, CBTR|C_L^n)\big)\Big(1 - \big(1 - P(DO)\big)\big(1 - P(DUC)\big)\Big)\Big)$
- $p(P_L^{n+1}|P_L^n, W) = 1 - p(N_L^{n+1}|P_L^n, W) - p(D|P_L^n, W)$

Starting from N_R^n

- $p(N_R^{n+1}|N_R^n, W) = P(NC|Negative)(1 - P(DO))\left(\left(1 - \left(1 - P_R^n(IBTR)\right)\left(1 - P_R^n(CBTR)\right)\right)P(Negative|C) + \left(1 - P_R^n(IBTR)\right)\left(1 - P_R^n(CBTR)\right)P(Negative|NC)\right) + P(C|Negative)\left(\left(P(ins, IBTR|C_R^n) + P(ins, CBTR|C_L^n)\right)(1 - P(DO))\left(P(R)P(Negative|NC) + (1 - P(R))P(Negative|C)\right) + \left(P(inv, IBTR|C_R^n) + P(inv, CBTR|C_R^n)\right)(1 - P(DO))(1 - P(DUC))P(Negative|C)\right)$
- $p(D|N_R^n, W) = P(NC|Negative)P(DO) + P(C|Negative)\left(\left(P(ins, IBTR|C_R^n) + P(ins, CBTR|C_R^n)\right)P(DO) + \left(P(inv, CBTR|C_R^n) + P(inv, CBTR|C_R^n)\right)\left(1 - \left(1 - P(DO)\right)(1 - P(DUC))\right)\right)$
- $p(P_R^{n+1}|N_R^n, W) = 1 - p(N_R^{n+1}|N_R^n, W) - p(D|N_R^n, W)$

Starting from P_R^n

Action: B

- $p(P_R^{n+1}|P_R^n, B) = P(NC|Positive)(1 - P(DO))\left(\left(1 - \left(1 - P_R^n(IBTR)\right)\left(1 - P_R^n(CBTR)\right)\right)P(Positive|C) + \left(1 - P_R^n(IBTR)\right)\left(1 - P_R^n(CBTR)\right)P(Positive|NC)\right)$
- $p(N_R^{n+1}|P_R^n, B) = P(NC|Positive)(1 - P(DO))\left(\left(1 - \left(1 - P_R^n(IBTR)\right)\left(1 - P_R^n(CBTR)\right)\right)P(Negative|C) + \left(1 - P_R^n(IBTR)\right)\left(1 - P_R^n(CBTR)\right)P(Negative|NC)\right)$

- $p(P_L^1|P_R^n, B) = P(C|Positive)\big(1 - P(DO)\big)\big(P(ins, CBTR|C_R^n)P(L|ins)\big)\Big(\Big(1 - \big(1 - P_L^1(IBTR)\big)\big(1 - P_L^1(CBTR)\big)\Big)P(Positive|C) + \big(1 - P_L^1(IBTR)\big)\big(1 - P_L^1(CBTR)\big)P(Positive|NC)\Big)$
- $p(N_L^1|P_R^n, B) = P(C|Positive)\big(1 - P(DO)\big)\big(P(ins, CBTR|C_R^n)P(L|ins)\big)\Big(\Big(1 - \big(1 - P_L^1(IBTR)\big)\big(1 - P_L^1(CBTR)\big)\Big)P(Negative|C) + \big(1 - P_L^1(IBTR)\big)\big(1 - P_L^1(CBTR)\big)P(Negative|NC)\Big)$
- $p(P_R^1|P_R^n, B) = P(C|Positive)\big(1 - P(DO)\big)\ (P(ins, CBTR|C_R^n)P(R|ins))\Big(\Big(1 - \big(1 - P_R^1(IBTR)\big)\big(1 - P_R^1(CBTR)\big)\Big)P(Positive|C) + \big(1 - P_R^1(IBTR)\big)\big(1 - P_R^1(CBTR)\big)P(Positive|NC)\Big)$
- $p(N_R^1|P_R^n, B) = P(C|Positive)\big(1 - P(DO)\big)\big(P(ins, CBTR|C_L^n)P(R|ins)\big)\Big(\Big(1 - \big(1 - P_R^1(IBTR)\big)\big(1 - P_R^1(CBTR)\big)\Big)P(Negative|C) + \big(1 - P_R^1(IBTR)\big)\big(1 - P_R^1(CBTR)\big)P(Negative|NC)\Big)$
- $p(I|P_R^n, B) = P(C|Positive)\big(P(inv, IBTR|C_R^n) + P(inv, CBTR|C_R^n)\big)$
- $p(M|P_R^n, B) = P(C|Positive)\big(P(ins, IBTR|C_R^n) + P(ins, CBTR|C_R^n)P(M|ins)\big)$
- $p(D|P_R^n, B) = 1 - p(P_R^{n+1}|P_R^n, B) - p(N_R^{n+1}|P_R^n, B) - p(P_L^1|P_R^n, B) - p(N_L^1|P_R^n, B) - p(P_R^1|P_R^n, B) - p(N_R^1|P_R^n, B) - p(I|P_R^n, B) - p(M|P_R^n, B)$

Action: W

- $\mathrm{p}(\mathrm{N}_{\mathrm{R}}^{n+1}|\mathrm{P}_{\mathrm{R}}^{n},\mathrm{W}) = \mathrm{P}(\mathrm{NC}|\mathrm{Positive})\big(1-\mathrm{P}(\mathrm{DO})\big)\Big(\Big(1-\big(1-\mathrm{P}_{\mathrm{R}}^{n}(\mathrm{IBTR})\big)\big(1-\mathrm{P}_{\mathrm{R}}^{n}(\mathrm{CBTR})\big)\Big)\mathrm{P}(\mathrm{Negative}|\mathrm{C})+\big(1-\mathrm{P}_{\mathrm{R}}^{n}(\mathrm{IBTR})\big)\big(1-\mathrm{P}_{\mathrm{R}}^{n}(\mathrm{CBTR})\big)\mathrm{P}(\mathrm{Negative}|\mathrm{NC})\Big)+\mathrm{P}(\mathrm{C}|\mathrm{Positive})\Big(\big(\mathrm{P}(\mathrm{ins},\mathrm{IBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})+\mathrm{P}(\mathrm{ins},\mathrm{CBTR}|\mathrm{C}_{\mathrm{L}}^{\mathrm{n}})\big)\big(1-\mathrm{P}(\mathrm{DO})\big)\Big(\mathrm{P}(\mathrm{R})\mathrm{P}(\mathrm{Negative}|\mathrm{NC})+\big(1-\mathrm{P}(\mathrm{R})\big)\mathrm{P}(\mathrm{Negative}|\mathrm{C})\Big)+\big(\mathrm{P}(\mathrm{inv},\mathrm{IBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})+\mathrm{P}(\mathrm{inv},\mathrm{CBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})\big)\big(1-\mathrm{P}(\mathrm{DO})\big)\big(1-\mathrm{P}(\mathrm{DUC})\big)\mathrm{P}(\mathrm{Negative}|\mathrm{C})\Big)$
- $\mathrm{p}(\mathrm{D}|\mathrm{P}_{\mathrm{R}}^{n},\mathrm{W}) = \mathrm{P}(\mathrm{NC}|\mathrm{Positive})\mathrm{P}(\mathrm{DO})+\mathrm{P}(\mathrm{C}|\mathrm{Positive})\Big(\big(\mathrm{P}(\mathrm{ins},\mathrm{IBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})+\mathrm{P}(\mathrm{ins},\mathrm{CBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})\big)\mathrm{P}(\mathrm{DO})+\big(\mathrm{P}(\mathrm{inv},\mathrm{CBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})+\mathrm{P}(\mathrm{inv},\mathrm{CBTR}|\mathrm{C}_{\mathrm{R}}^{\mathrm{n}})\big)\Big(1-\big(1-\mathrm{P}(\mathrm{DO})\big)\big(1-\mathrm{P}(\mathrm{DUC})\big)\Big)\Big)$
- $\mathrm{p}(\mathrm{P}_{\mathrm{R}}^{n+1}|\mathrm{P}_{\mathrm{R}}^{n},\mathrm{W}) = 1-\ \mathrm{p}(\mathrm{N}_{\mathrm{R}}^{\mathrm{n}+1}|\mathrm{P}_{\mathrm{R}}^{\mathrm{n}},\mathrm{W})-\mathrm{p}(\mathrm{D}|\mathrm{P}_{\mathrm{R}}^{\mathrm{n}},\mathrm{W})$

B. Proofs of Structural Properties

Proof of Theorem 1.

If $d_t > d_t^*$,

$d_t > \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1})p_t(s_{t+1}|s_t,\mathrm{B}) - \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1})p_t(s_{t+1}|s_t,\mathrm{W})$.

$\therefore \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1})p_t(s_{t+1}|s_t,\mathrm{W}) > \sum_{s_{t+1}\in S} v_{t+1}(s_{t+1})p_t(s_{t+1}|s_t,\mathrm{B}) - d_t$;

$\therefore\ r_t(s_t, \mathrm{W}) + \sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}) > \sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B}) - d_t +$

$r_t(s_t, \mathrm{W})$;

$\therefore\ r_t(s_t, \mathrm{W}) + \sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}) > \ r_t(s_t, \mathrm{B}) +$

$\sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B})$;

$\therefore \underset{a_t \in [W,B]}{\operatorname{argmax}}\ V_t(s_t) = \mathrm{W}.$

if $d_t \leq d_t^*$,

$d_t \leq \sum_{s_{t+1} \in S \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B}) - \sum_{s_{t+1} \in S \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}).$

$\therefore \sum_{s_{t+1} \in S \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}) \leq \sum_{s_{t+1} \in S \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B}) - d_t;$

$\therefore\ r_t(s_t, \mathrm{W}) + \sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}) \leq \sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B}) - d_t +$

$r_t(s_t, \mathrm{W})$;

$\therefore\ r_t(s_t, \mathrm{W}) + \sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{W}) \leq \ r_t(s_t, \mathrm{B}) +$

$\sum_{s_{t+1} \in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|s_t, \mathrm{B})$;

$\therefore \underset{a_t \in [W,B]}{\operatorname{argmax}}\ V_t(s_t) = \mathrm{B}\ .$

Proof of Proposition 1.

We prove proposition 1 by induction based on the case $s_t = P$. The proofs for cases $s_t \neq P$ are similar. For t =100, $V_{100}^A(s_t) \leq V_{100}^B(s_t)$ holds because $V_{100}(s_t) = 0$.

Let

$v_t(\mathrm{P},\mathrm{B}) = r_t(\mathrm{P},\mathrm{B}) + \sum_{s+1\in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|\mathrm{P},\mathrm{B})$,

$v_t(\mathrm{W},\mathrm{B}) = r_t(\mathrm{P},\mathrm{W}) + \sum_{s+1\in S} v_{t+1}(s_{t+1}) p_t(s_{t+1}|\mathrm{P},\mathrm{W})$,

$p_t(\mathrm{NP}|\mathrm{P},\mathrm{B}) = p_t(\mathrm{P}|\mathrm{P},\mathrm{B}) + p_T(\mathrm{P}|\mathrm{P},\mathrm{B})$, and

$$p_t(\mathrm{T}|\mathrm{P},\mathrm{B}) = P_t(P_L^1|P,B) + P_t(N_L^1|P,B) + P_t(P_R^1|P,B) + P_t(N_R^1|P,B) + P_t(M|P,B) + P_t(I|P,B).$$

Given any $\mathrm{t} = \mathrm{T}$ and $V_{T+1}^A(s_t) \le V_{T+1}^B(s_t)$, we need to prove $V_T^A(s_T) \le V_T^B(s_T)$.

Using assumption 2,

$$\begin{aligned}
&\mathrm{v}_\mathrm{T}^\mathrm{A}(\mathrm{P},\mathrm{B}) - \mathrm{v}_\mathrm{T}^\mathrm{B}(\mathrm{P},\mathrm{B}) \\
&= \Big(\mathrm{r}_\mathrm{T}(\mathrm{P},\mathrm{B}) + \mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{NP},\mathrm{B})\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{NP}|\mathrm{P},\mathrm{B}) + \mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{T},\mathrm{B})\mathrm{P}_\mathrm{T}^\mathrm{A}(\mathrm{T}|\mathrm{P},\mathrm{B}) \\
&+ \mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{D})\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{D}|\mathrm{P},\mathrm{B}) \Big) \\
&- \Big(\mathrm{r}_\mathrm{T}(\mathrm{P},\mathrm{B}) + \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{NP},\mathrm{B})\mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{NP}|\mathrm{P},\mathrm{B}) + \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{T},\mathrm{B})\mathrm{P}_\mathrm{T}^\mathrm{B}(\mathrm{T}|\mathrm{P},\mathrm{B}) \\
&+ \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{D})\mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{D}|\mathrm{P},\mathrm{B}) \Big) \\
&= \mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{NP},\mathrm{B})\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{NP}|\mathrm{P},\mathrm{B}) - \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{NP},\mathrm{B})\mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{NP}|\mathrm{P},\mathrm{B}) + \mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{T},\mathrm{B})\mathrm{P}_\mathrm{T}^\mathrm{A}(\mathrm{T}|\mathrm{P},\mathrm{B}) \\
&- \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{T},\mathrm{B})\mathrm{P}_\mathrm{T}^\mathrm{B}(\mathrm{T}|\mathrm{P},\mathrm{B}) + \mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{D})\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{D}|\mathrm{P},\mathrm{B}) - \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{D})\mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{D}|\mathrm{P},\mathrm{B})
\end{aligned}$$

According to $\mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{s}_\mathrm{T}) \le \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{s}_\mathrm{T})$, we have

$\mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{NP},\mathrm{B}) \le \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{NP},\mathrm{B})$, $\mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{T},\mathrm{B}) \le \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{T},\mathrm{B})$, and $\mathrm{V}_\mathrm{T}^\mathrm{A}(\mathrm{D}) = \mathrm{V}_\mathrm{T}^\mathrm{B}(\mathrm{D})$

Using assumption 1, we have

$\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{NP}|\mathrm{P},\mathrm{B}) \le \mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{NP}|\mathrm{P},\mathrm{B})$, $\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{T}|\mathrm{P},\mathrm{B}) \ge \mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{T}|\mathrm{P},\mathrm{B})$, and $\mathrm{p}_\mathrm{T}^\mathrm{A}(\mathrm{D}|\mathrm{P},\mathrm{B}) \ge \mathrm{p}_\mathrm{T}^\mathrm{B}(\mathrm{D}|\mathrm{P},\mathrm{B})$

$$\therefore\ \mathrm{v}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{P},\mathrm{B}) - \mathrm{v}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{P},\mathrm{B})$$

$$\leq \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{B})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP}|\mathrm{P},\mathrm{B}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP}|\mathrm{P},\mathrm{B})\right)$$

$$+ \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{T},\mathrm{B})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{T}|\mathrm{P},\mathrm{B}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{T}|\mathrm{P},\mathrm{B})\right) + \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D}|\mathrm{P},\mathrm{B}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D}|\mathrm{P},\mathrm{B})\right)$$

$$\leq \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{B})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP}|\mathrm{P},\mathrm{B}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP}|\mathrm{P},\mathrm{B}) + \mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{T}|\mathrm{P},\mathrm{B}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{T}|\mathrm{P},\mathrm{B})\right.$$

$$\left. + \mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D}|\mathrm{P},\mathrm{B}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D}|\mathrm{P},\mathrm{B})\right) = \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{B})(1-1) = 0$$

Let

$$\mathrm{V}_t(\mathrm{NP},\mathrm{W}) = \mathrm{V}(\mathrm{N})\frac{\mathrm{P}_t(\mathrm{N}|\mathrm{P},\mathrm{W})}{\mathrm{P}_t(\mathrm{N}|\mathrm{P},\mathrm{W}) + \mathrm{P}_t(\mathrm{P}|\mathrm{P},\mathrm{W})} + \mathrm{V}(\mathrm{P})\frac{\mathrm{P}_t(\mathrm{P}|\mathrm{P},\mathrm{W})}{\mathrm{P}_t(\mathrm{N}|\mathrm{P},\mathrm{W}) + \mathrm{P}_t(\mathrm{P}|\mathrm{P},\mathrm{W})}$$

According to $\mathrm{V}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{s}_{\mathrm{T}}) \leq \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{s}_{\mathrm{T}})$, we have $\mathrm{V}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP},\mathrm{W}) \leq \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{W})$

According to assumption 1, we have

$\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP}|\mathrm{P},\mathrm{W}) \leq \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP}|\mathrm{P},\mathrm{W})$, and $\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D}|\mathrm{P},\mathrm{W}) \geq \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D}|\mathrm{P},\mathrm{W})$

$$\therefore \mathrm{v}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{P},\mathrm{W}) - \mathrm{v}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{P},\mathrm{W})$$

$$= \mathrm{V}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP},\mathrm{W})\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP}|\mathrm{P},\mathrm{W}) - \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{W})\mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP}|\mathrm{P},\mathrm{W}) + \mathrm{V}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D})\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D}|\mathrm{P},\mathrm{W})$$

$$- \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D})\mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D}|\mathrm{P},\mathrm{W})$$

$$\leq \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{W})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP}|\mathrm{P},\mathrm{W}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP}|\mathrm{P},\mathrm{W})\right)$$

$$+ \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D}|\mathrm{P},\mathrm{W}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D}|\mathrm{P},\mathrm{W})\right)$$

$$\leq \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{W})\left(\mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{NP}|\mathrm{P},\mathrm{W}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP}|\mathrm{P},\mathrm{W}) + \mathrm{p}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{D}|\mathrm{P},\mathrm{W}) - \mathrm{p}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{D}|\mathrm{P},\mathrm{W})\right)$$

$$= \mathrm{V}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{NP},\mathrm{W})(1-1) = 0$$

$$\therefore V_T^A(s_t) = \max\left\{\mathrm{v}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{P},\mathrm{B}),\ \mathrm{v}_{\mathrm{T}}^{\mathrm{A}}(\mathrm{P},\mathrm{W})\right\} \leq \max\left\{\mathrm{v}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{P},\mathrm{B}),\ \mathrm{v}_{\mathrm{T}}^{\mathrm{B}}(\mathrm{P},\mathrm{W})\right\} = V_T^B(s_t).$$

Proof of Proposition 2.

We prove proposition 2 based on the case $s_t = P$. The proofs for cases $s_t \neq P$ are similar.

$$\mathrm{v}_t^1(\mathrm{P},\mathrm{B}) - \mathrm{v}_t^2(\mathrm{P},\mathrm{B}) = \left(\mathrm{r_t}(\mathrm{P},\mathrm{B}) + \sum_{s_{t+1}\in \mathrm{S-I}} v_{t+1}(s_{t+1}) p_t(s_{t+1}|\mathrm{P},\mathrm{B}) + R_{t+1}^1(I)\right) -$$

$$\left(\mathrm{r_t}(\mathrm{P},\mathrm{B}) + \sum_{s_{t+1}\in \mathrm{S-I}} v_{t+1}(s_{t+1}) p_t(s_{t+1}|\mathrm{P},\mathrm{B}) + R_{t+1}^2(I)\right) = R_{t+1}^1(I) - R_{t+1}^2(I) \geq 0\ ;$$

$$\mathrm{v}_t^1(\mathrm{P},\mathrm{W}) - \mathrm{v}_t^2(\mathrm{P},\mathrm{W}) = 0;$$

$$\therefore V_t^1(s_t) = \max\{\mathrm{v_t^1}(\mathrm{P},\mathrm{B}),\ \mathrm{v_t^1}(\mathrm{P},\mathrm{W})\} \geq \max\{\mathrm{v_t^2}(\mathrm{P},\mathrm{B}),\ \mathrm{v_t^2}(\mathrm{P},\mathrm{W})\} = V_t^2(s_t).$$

Proof of Corollary 1.

The proof is similar to that of proposition 2.

3 Adaptive Decision-Making of Breast Cancer Mammography Screening: A Heuristic-Based Regression Model

Abstract

The American Cancer Society (ACS) updated their breast cancer screening guidelines in late 2015 and recommends that all women have the choice to start annual mammography screenings beginning at age 40. For women ages 45 to 54, the ACS explicitly recommends annual mammograms. However, due to the potential harms associated with screening mammography, such as overdiagnosis and unnecessary work-ups, the best strategy to design an appropriate breast cancer mammography screening schedule remains controversial. Instead of recommending a one-size-fits-all screening schedule, this study identifies a personalized mammography screening strategy adaptive to each woman's age-specific breast cancer risk. We present a two-stage decision framework: (1) age-specific breast cancer risk estimation, and (2) annual mammography screening decision-making based on estimated risk. The results suggest that the optimal combinations of independent variables used in risk estimation are not the same across age groups. Our optimal decision models outperform the existing mammography screening guidelines in terms of the average loss of life expectancy. While most earlier studies improved the breast cancer screening decisions by offering lifetime screening schedules, our proposed model provides an adaptive screening decision aid by age. Since whether or not a woman should receive a mammogram is determined based on her breast cancer risk at her current age, our "on-line" screening policy adapts to a woman's latest health status, which reflects the current individual risk of each woman more accurately.

3.1. Introduction

Breast cancer is the most common non-skin cancer among U.S. women. According to the American Cancer Society (ACS), an estimated 246,660 women will be diagnosed with breast cancer, and an estimated 40,450 women will die from this disease in 2016 (ACS 2016). Routine screening mammography may reduce mortality from breast cancer by detecting the disease at early stages, before the cancer has spread. Several clinical trials and population-based evaluations suggest that mammography may reduce breast cancer mortality significantly (Tabar et al. 2003, Weedon-Fekjær et al. 2014).

Nevertheless, there are potential harms associated with screening mammography, such as overdiagnosis, exposure to radiation, and work-up of positive findings. A cohort study by Hubbard et al. (2011) reported that after ten years of annual screenings, over half of participating women will receive at least one false-positive result. The high false-positive rate of screening mammography (i.e., the mammogram is interpreted as positive but no cancer is present) often results in unnecessary follow-up imaging and biopsy exams, which rule in or out the presence of breast cancer after a positive test result. As an invasive procedure, a biopsy may place a woman at risk of morbidity and, in rare cases, mortality (Bruening et al. 2009). The proportion of women with abnormal mammography findings that are diagnosed with breast cancer is less than 10% (Rosenberg et al. 2006).

Due to the significant benefits and harms associated with screening mammography, designing the most efficient breast cancer screening guideline that maximizes the benefit and minimizes the harms remains controversial in the public health community (Nelson 2010, Tarnay 2012). The ACS updated their breast cancer screening guidelines in late 2015 and now recommends that women begin annual breast cancer screenings beginning at age 45. The guidelines also

recommend that a woman when reaching age 55 should either switch to biennial screenings or continue annual mammography. In addition, the American College of Radiology (ACR) recommends all women begin annual mammography at the age of 40, while the U.S. Preventive Services Task Force (USPSTF) and the American College of Physicians (ACP) advocate beginning screening mammography at age 50 and doing so on a biennial basis (Nelson 2010). In addition, the age at which to cease mammography screening is also debated. Although the ACS and ACR do not specify the age to stop routine screening mammography, the USPSTF and ACP recommend against routine screening for women 75 years or older. Furthermore, there are ongoing discussions on screening frequency and whether it is necessary to perform annual or biennial screening.

The ongoing debate surrounding screening mammography guidelines motivates researchers to pursue a decision policy that finds the optimal trade-offs between the negative effects of screening and patients' long-term benefits of early diagnosis of breast cancer. Kirch and Klein (1974) designed a mathematical model to determine the frequency of mammography screening that minimizes the detection delay for the general population. Their model assumed perfect mammography screening sensitivity and specificity, but in fact, the actual false positive rate of mammography is high. Some additional studies, such as Ozekici and Pliska (1991) and Zelen (1993), considered false positives and false negatives of screening mammography exams in their mathematical models. However, the parameters used in these models were not age-specific, making their solutions less practical since breast cancer risk increases with age. According to Gail and Rimer (1998), an appropriate screening recommendation should reflect each woman's individual risk. Since each woman has different levels of breast cancer risk based on her personal risk factors, breast cancer screening schedules should not be uniform across women. Several

more recent studies addressed this issue by including age- and patient-specific input parameters and generated some effective optimization models for mammography screening policies. Maillart et al. (2008) employed a partially observable Markov process model considering women's age and menopausal status to evaluate different screening mammography policies. Their model used different stages of breast cancer as core states and generated a set of efficient policies in terms of life-time breast cancer mortality and the expected total number of screening mammograms. Chhatwal et al. (2010) focused on how to make biopsy referral decisions after positive screening mammograms to maximize patients total expected quality-adjusted life years (QALYs). They developed a finite-horizon discrete-time Markov decision process (MDP) model to offer optimal biopsy referral policies for patients with different breast cancer risk scores (i.e., a woman's current probability of cancer based on her risk factors and mammographic features). Ayvaci et al. (2012) applied an MDP model to optimize biopsy referral decisions for different breast cancer risk scores under budgetary restrictions. The model by Ayer et al. (2012) is the first screening decision study that directly personalizes mammography screening. They developed a partially observable Markov decision process (POMDP) model that offers optimal screening mammography schedules based on five personal risk factors: age, race, age at menarche, age at first birth and prior screening history. Moreover, similar modeling approaches have also been applied to the screening decisions of some other cancers, such as prostate cancer. Zhang et al. (2012) developed a POMDP to determine optimal biopsy referral decisions for prostate cancer screening based on prostate-specific antigen tests. Erenay et al.'s POMDP model optimized colonoscopy screening policies for colorectal cancer (2014). Besides age, Erenay et al.'s personalized model incorporated both static (i.e., gender) and dynamic factors (i.e., history of

colorectal cancer and polyp). In particular, Alagoz et al. (2011) provided an overall review regarding the applications of various operations research models in cancer screening.

Most of these previous studies utilized Markov modeling approaches, which are inefficient in solving problems with high computational complexity. Since incorporating additional breast cancer risk factors into the model leads to a higher dimensionality of the decision-making framework, a Markov model will inevitably suffer from the so-called curse of dimensionality (Bellman), which refers to the computational complexity that grows exponentially with the dimensionality. Incorporating too many risk factors could cause a Markov model to be computationally intractable. In addition, since all states and transition probabilities between states in Markov models must be precisely pre-specified, it is difficult for these models to process some dynamic risk factors, such as a woman's body mass index (BMI), which are fluctuating over time and thus unpredictable. Hence, these prior studies mainly focused on optimal static lifetime screening policies such that optimal decisions cannot be updated dynamically or adjusted with unpredictable new information.

Our study aims to circumvent the limitations of traditional Markov modeling approaches on this topic by proposing a two-stage individualized mammography screening decision framework that is adaptive to changes in risk factors. We first perform a heuristic-based regression analysis with model selection to evaluate a woman's probability of breast cancer at her current age based on a range of personal risk factors. Then we determine whether this woman should undergo a screening mammogram based on her estimated breast cancer risk at her current age.

Advances in health informatics and analytics in recent years have improved health prediction and management for chronic diseases (West et al. 2005, Misiunas et al. 2016, Bertsimas et al. 2016). In this study, we not only take advantage of logistic regression to eschew the curse of

dimensionality of Markov models, but also discuss the dimensionality reduction of regression models in the context of medical decision-making. We design a novel model selection method for logistic regression from the perspective of making optimal screening decisions. The optimality of a decision is defined in terms of minimal misclassification cost (i.e., the cost of false positives and false negatives), which is a critical concern in medical practice.

The remainder of the paper proceeds as follows: in Section 3.2, we describe the decision-making framework. We then present the numerical results by implementing the model in Section 3.3. In Section 3.4, we discuss the results and their significance to the decision-making of breast cancer mammography screening as well as other disease prevention and treatment problems.

3.2. Methods

In this study, the decision-making process consists of two sub-models: breast cancer risk estimation and decision-making of mammography screening utilization based on the estimated risk. The risk estimation model is a regression model used to predict a woman's probability of developing breast cancer at her current age based on a number of breast cancer risk factors. The risk estimation model is built based on the predictors from the widely accepted Barlow model (Barlow et al. 2006). We improve their model by conducting a model selection with the aim of increasing life expectancy, which is impacted by the false-positive and false-negative prediction errors. According to the estimated probability of developing breast cancer, the next sub-model determines whether this woman should be referred for a screening mammogram or if she should skip the mammogram in the current year and return for screening the following year. In this sub-model, a pre-specified optimal cut-off point of cancer probabilities, which is expected to minimize the woman's loss of life expectancy, serves as a threshold of recommending a screening mammogram. Therefore, the decision-making framework works in an adaptive manner

such that it allows women to input their current risk factor levels, and then the framework generates corresponding optimal decisions regarding mammography screening.

3.2.1. Breast Cancer Risk Estimation Model

The probabilities of breast cancer for women with various risk factors are estimated using a logistic regression model. We formulate the regression model based on the results from the Barlow breast cancer risk model (Barlow et al.2006). The breast cancer risk factors identified in the Barlow model includes age (by 5-year age groups), race (white, black, Asian, native American, and other), Hispanic ethnicity (year/no), number of first degree relatives with breast cancer (0, 1, and 2 or more), age at first birth (<30, >=30, and nulliparous), surgical menopause (yes/no), use of hormone replacement therapy (yes/no), menopausal status (premenopausal/postmenopausal), body mass index (10-25, 25-30, 30-35, and 35 or more), previous breast procedure (yes/no), and last mammographic outcome (positive/negative). However, a number of prior studies revealed that breast cancer risk factors are not independent. Mayberry and Stoddard-Wright (1992) reported that the effects of some breast risk factors on women of different races are not the same, such as family history and age at menarche. For instance, they found that black women who have first-degree relatives and second-degree relatives with breast cancer have similar relative risks of breast cancer, while for white women having first-degree relatives with breast cancer leads to higher relative risk compared to only having second-degree relatives with breast cancer. Cleary and Maihle (1997) reported an inverse association between BMI and the relative risk (RR) of developing breast cancer in premenopausal women, but a positive association in post-menopausal women. Some studies showed that breast cancer risk patterns vary by age and menopausal status (Clavel-Chapelon and Gerber 2002). These studies suggest the existence of interaction effects between the 11 risk

factors described above. Therefore, we consider all possible two-way interaction terms in the regression analysis to more accurately estimate woman's breast cancer risk. The binary response variable is defined by whether a woman has breast cancer in the current year.

A regression model that includes all explanatory variables with interaction terms may lead to overfitting and thereby have poor predictive power (Wears and Lewis 1999). In addition, although all of the 11 risk factors have been associated with breast cancer risk, not all of them may be instrumental for finding the optimal decision of mammography screening at every age based on the prediction of breast cancer risk. Thus, a model selection procedure aimed to reduce dimensionality of explanatory variables and find the optimal regression model is necessary.

Two components are essential to model selection: a criterion or a benchmark for comparing different models and a search strategy for selecting variables. While a criterion serves as a cost function that measures whether a subset of the variables available produces the best model, a search strategy specifies the course of evaluating each subset of explanatory variables according to the criterion.

3.2.1.1. H Measure

There are many traditional model selection criteria in the literature, such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the R^2, the adjusted R^2, and the Mallow's C_p (Kadane and Lazar 2004). In medical applications, however, the performance of binary classification that determines whether a predicted probability of a specific disease should be assigned to an abnormal group (also known as positive) or a normal group (also known as negative) is of special interest to medical decision makers. The receiver operating characteristic (ROC) analysis, which plots the true positive rate versus the false positive rate at various threshold settings (from 0 to 100%), is widely adopted for model selection in the medical

literature (Rosset 2004). Specifically, the area under the curve (AUC) of a ROC curve provides a direct measure to compare models. Nevertheless, a fundamental weakness of the AUC measure is that it assigns two types of error (i.e., false positive and false negative) with the same weights. Neither the AUC nor other traditional criteria pay attention to the relative severity between the two types of errors in prediction and classification models. Thus, their model selection outcomes may be undesired or even unacceptable, especially for some prediction or classification models used to study life-threatening diseases. For the above reasons, this study applies a recently developed statistic, called the H measure (Hand 2009, 2010), as a criterion for model selection.

The H measure introduces the notion of misclassification costs by quantifying the relative severity between the two types of errors. Let τ denote a specified threshold, then the following logic is used to classify a probability output by the breast cancer prediction model as class 1 (cancerous) or class 0 (non-cancerous):

$$Y(x) = \begin{cases} 1, & \text{if } p(x) > \tau \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

where x is the vector of explanatory variables; $p(x)$ is a predicted probability; and $Y(x)$ is the final prediction that whether a woman has breast cancer or not. We normalize the two types of misclassification costs by scaling their numeric values in the range of [0,1]. Let c in [0,1] denote the relative cost of misclassifying a non-cancerous observation as cancer, and thereby $1 - c$ represents the relative cost of misclassifying a cancerous observation as non-cancerous. Then the misclassification cost function of the regression model with given c and τ is defined as follows:

$$L(c, \tau) = c\pi_0\big(1 - \mathrm{Sp}(\tau)\big) + (1 - c)\pi_1\big(1 - \mathrm{Se}(\tau)\big), \tag{2}$$

where $\mathrm{Sp}(\tau)$ is the specificity (i.e., the probability of class 0 observations that are correctly identified as class 0) when the threshold is set to τ; $\mathrm{Se}(\tau)$ is the sensitivity (i.e., the probability of class 1 observations that are correctly classified as class 1) when the threshold is set to τ; and

$\pi_i, i = 0,1$ is the proportion of class i in all observations. In the current study, $\mathrm{Se}(\tau)$ and $\mathrm{Sp}(\tau)$ can be deemed the true positive rates and the true negative rates, respectively. Then we define the minimum weighted cost (MWC) under the cost c as follows,

$$\mathrm{MWC}(c) = \min_{\tau} L(c, \tau)\,. \tag{3}$$

Thus, the minimum weighted cost is reached when finding the optimal threshold τ_c:

$$\tau_c = \operatorname*{argmin}_{\tau} L(c, \tau)\,. \tag{4}$$

Instead of using a constant cost, the H measure specifies a distribution, $w(c)$, to reflect the uncertainty about the exact value of the cost c. Beta distribution with shape parameters $\alpha = \beta = 2$ is usually adopted to represent the cost distribution (Hand 2010). Therefore, the average MWC (L_W) of the breast cancer prediction model is calculated by

$$\mathrm{L}_W = \int_c L(c, \tau_c)\, w(c) dc\,. \tag{5}$$

It is desirable for an index to take larger values for better performance, and ranges between 0 for the worst classification and 1 for perfect classification. Thus, the H measure is defined as

$$\mathrm{H} = 1 - \frac{\mathrm{L}_W}{\mathrm{L}_{\mathrm{W}}^{\max}}\,, \tag{6}$$

where $\mathrm{L}_{\mathrm{W}}^{\max}$ is a constant equal to the maximum value that L_W can reach, which means a trivial classifier is employed to make random classification (i.e., the false positive rate is always identical to the true positive rate).

We calculate the H measure for each candidate prediction model, which is a regression polynomial composed of a non-null subset of the explanatory variables. Hence, the optimal prediction model is the one with the highest H measure.

3.2.1.2. Misclassification Cost

The calculation of the H measure is based on well-defined misclassification costs. In this study, the cost c (the misclassification cost of a false positive result) is defined as the reduction in life expectancy if a woman without breast cancer is classified into the group with an increased risk of cancer and recommended to undergo a screening mammogram exam. The cost $1 - c$ (the misclassification cost of a false negative result) corresponds to the loss of life expectancy when a woman with breast cancer is classified into the group with a low risk of cancer and recommended to skip a screening mammogram. The Breast Imaging Reporting and Data System (BI-RADS) of the ACR, a lexicon used by radiologists for mammography interpretation, describes the level of cancer suspicion for the exam (Sickles et al. 2013). Radiologists interpret and assign screening mammograms into one of the seven BI-RADS assessment categories (D'Orsi et al. 2013). BI-RADS 0 is considered an incomplete mammogram that requires additional imaging. BI-RADS 1 and 2 are associated with negative or benign findings. BI-RADS 3 is a probably benign finding, which needs short interval follow-up. BI-RADS 4 is suspicious abnormality, and BI-RADS 5 is highly suggestive of malignancy. Patients with either BI-RADS 4 or 5 are recommended to undergo an immediate biopsy. BI-RADS 6 is used for biopsy proven breast cancers. In this study, we define a positive screening mammogram as those with a BI-RADS score of 0, 4 or 5. We assume BI-RADS 0 is always followed up by a diagnostic mammogram. We also assume that if a diagnostic mammogram is suspicious (BI-RADS 4 or 5), then a biopsy is recommended. Thus, mammograms that are finally classified as BI-RADS 4 or 5 result in biopsy referrals. The cost c includes the disutilities (i.e., reduction in life expectancy) of an initial screening mammogram, a possible diagnostic mammogram and a possible follow-up

biopsy. Let C_t^P denote the unscaled value of c at age t, then C_t^P represents the cost of a false positive decision for a woman at age t:

$$C_t^P = d_t^M + \mathrm{P}_t(\mathrm{B}_0|\mathrm{NC}) \cdot d_t^M + \mathrm{P}_t(\mathrm{Positive}|\mathrm{NC}) \cdot d_t^B, \quad (7)$$

where d_t^{M} and d_t^{B} are the disutilities (in years) of a mammogram (both screening and diagnostic) and a biopsy at age t, respectively; $P_t(\mathrm{B}_0|\mathrm{NC})$ is the probability that a healthy woman's screening mammogram is labeled as BI-RADS 0; and $P_t(\mathrm{Positive}|\mathrm{NC})$ is the probability that a mammogram is finally assigned to BI-RADS 4 or 5 for a healthy woman without breast cancer.

The misclassification cost of a false negative result is the loss of life expectancy caused by skipping a screening mammogram, which would have detected a cancer for a woman with breast cancer. Let C_t^N denote the unscaled value of $1 - c$ at age t. The misclassification cost of a false negative result is equal to the difference between the life expectancy of skipping a necessary screening mammogram (L_t^{SM}) and undergoing a necessary screening mammogram (L_t^{UM}) for a woman with breast cancer:

$$C_t^N = L_t^{\mathrm{UM}} - L_t^{\mathrm{SM}}, \quad (8)$$

We consider the cost of skipping one necessary screening mammogram (i.e. a mammogram for a woman with cancer) as the loss of life expectancy due to at least a one-year delay in diagnosis and treatment of breast cancer. For a breast cancer patient, such a loss involves two possibilities: (1) an untreated in situ cancer progresses to an invasive cancer; (2) an untreated invasive cancer patient dies from breast cancer. According to Yen et al. (2003), we can categorize in situ breast cancers into two types: the non-progressive in situ cancer (ins_0), which has no propensity to progress to an invasive stage; and the progressive in situ cancer (ins_1), which has the possibility to advance. We model the progression from in situ cancer to invasive cancer as a Poisson process, which is a common assumption used in previous studies (Yen et al,

2003, Jackson et al, 1981). Therefore, a breast cancer patient who skips a screening mammogram at age t has a life expectancy of:

$$L_t^{\mathrm{SM}} = P_t(\mathrm{ins}|\mathrm{C}) \cdot L_t^{\mathrm{SM}}(\mathrm{ins}) + P_t(\mathrm{inv}|\mathrm{C}) \cdot L_t^{\mathrm{SM}}(\mathrm{inv}), \quad (9)$$

where $P_t(\mathrm{ins}|\mathrm{C})$ and $P_t(\mathrm{inv}|\mathrm{C})$ are proportions of in situ cancers and invasive cancers among all breast cancer cases for women at age t; $L_t^{\mathrm{SM}}(\mathrm{ins})$ and $L_t^{\mathrm{SM}}(\mathrm{inv})$ are life expectancies of in situ and invasive patients when skipping a necessary screening mammogram, respectively. As in situ cancers are generally considered not life-threating, we assume an in situ cancer patient will not die from breast cancer within one year of diagnosis (Ozanne et al. 2011). We calculate $L_t^{\mathrm{SM}}(\mathrm{ins})$ as follows:

$$L_t^{\mathrm{SM}}(\mathrm{ins}) = P_t(\mathrm{ins}_0|\mathrm{ins})[1 + L_{t+1}(\mathrm{ins}_0)] + P_t(\mathrm{ins}_1|\mathrm{ins}) \cdot \{\mathrm{e}^{-\lambda_t} \cdot [1 + L_{t+1}(\mathrm{ins}_1)] + (1 - \mathrm{e}^{-\lambda_t}) \cdot [1 + L_{t+1}(\mathrm{inv})]\}, \quad (10)$$

where $P_t(\mathrm{ins}_0|\mathrm{ins})$ and $P_t(\mathrm{ins}_1|\mathrm{ins})$ are the proportions of ins_0 and ins_1 in all in situ cancer cases, respectively; λ_t is the one-year progression rate from ins_1 to invasive cancer; $L_{t+1}(\mathrm{ins}_0)$, $L_{t+1}(\mathrm{ins}_1)$, and $L_{t+1}(\mathrm{inv})$ are life expectancies of patients with non-progressive in situ, progressive in situ, and invasive cancers at age $t+1$, respectively. Note that we calculate the difference of life expectancy between skipping and undergoing a screening mammogram by subtraction, so the loss of life expectancy due to death from a cause other than breast cancer is a redundant term appearing in both the minuend and the subtrahend. Thus, we omit the mortality from other causes in one year and focus on the cancer-specific survival. Similarly, the life expectancy ($L_t^{\mathrm{SM}}(\mathrm{inv})$) of an invasive cancer patient who skips a mammogram is:

$$L_t^{\mathrm{SM}}(\mathrm{inv}) = P_t(\mathrm{UD}) \cdot 0.5 + [1 - P_t(\mathrm{UD})] \cdot [1 + L_{t+1}(\mathrm{inv})], \quad (11)$$

where $\mathrm{P}_t(\mathrm{UD})$ represents the breast cancer-specific death rate of an untreated invasive cancer patient at age t. We assume the life expectancy of a treated patient who dies at age t to be half year (Chhatwal et al. 2010).

To derive L_t^{UM}, we first consider a breast cancer patient who chooses to undergo a screening mammogram and successfully detects her breast cancer. Let L_t^{CD}denote such a patient's life expectancy:

$$L_t^{\mathrm{CD}} = P_t(\mathrm{ins}|\mathrm{C}) \cdot [1 + L_{t+1}(\mathrm{ins}^T)] + P_t(\mathrm{inv}|\mathrm{C}) \cdot \{[1 - P_t(\mathrm{D})] \cdot [1 + L_{t+1}(\mathrm{inv}^T)] + 0.5 \cdot P_t(\mathrm{D})\}, \quad (12)$$

where $L_{t+1}(\mathrm{ins}^T)$and $L_{t+1}(\mathrm{inv}^T)$ are life expectancies of patients treated for in situ cancer and invasive cancer patients at age $t+1$, respectively; and $P_t(\mathrm{D})$ denotes the one-year death rate of treated invasive cancer at age t.

Taking into account the possible false negative mammogram outcome and the disutility of screening, we obtain the life expectancy of a breast cancer patient who chooses to undergo a screening mammogram at age t:

$$L_t^{\mathrm{UM}} = -d_t^{\mathrm{M}} - \mathrm{P}_t(\mathrm{B}_0|\mathrm{C}) \cdot d_t^M + P_t(\mathrm{Positive}|\mathrm{C}) \cdot \left(L_t^{\mathrm{CD}} - d_t^{\mathrm{B}}\right) + P_t(\mathrm{Negative}|\mathrm{C}) \cdot L_t^{\mathrm{SM}}, \quad (13)$$

where $P_t(\mathrm{B}_0|\mathrm{C})$ is the probability that a breast cancer patient's screening mammogram is labeled as BI-RADS 0; $P_t(\mathrm{Positive}|\mathrm{C})$ is the probability that a breast cancer patient is referred for a biopsy, whose screening mammogram or diagnostic mammogram has a BI-RADS score of 4 or 5; and $P_t(\mathrm{Negative}|\mathrm{C})$ is the probability that a breast cancer patient is not recommended for a biopsy when either the screening mammogram or the diagnostic mammogram (when necessary) is classified as BI-RADS 1, 2 or 3. The life expectancy for a woman with biopsy-confirmed breast cancer is $L_{\mathrm{t}}^{\mathrm{CD}}$ subtracting the disutility of the biopsy, the screening mammogram and the possible diagnostic mammogram. If a woman is not referred for a biopsy after a screening

mammogram and/or a diagnostic mammogram, only the disutility of the possible mammograms needs to be considered in life expectancy calculation.

After normalizing the costs obtained from (7) and (8) to c and $1-c$, we obtain the costs of two types of misclassification errors for calculating the H measure. As mentioned above, the costs of the two types of misclassification are both age-specific. Hence, we stratify the optimal risk estimation model by age.

3.2.1.3. Search Strategy

The H measure is used as a criterion to compare different models and identify the optimal combination of explanatory variables. The model selection can be considered as an integer programming problem. In this problem, the H measure is the objective function to be maximized. The solution space consists of all available explanatory variables and their interaction effects. Every feasible solution is a binary configuration of explanatory variables and interaction effects, where values 0 and 1 correspond to being selected and excluded, respectively. Suppose that there is a dependent variable Y, a group of explanatory variables $\mathrm{X_i}$ and their corresponding coefficients β_i as well as the intercept β_0. Given a sample S composed of a group of observations of Y and $(X_1, X_2, \dots, X_n)$, let $\mathrm{H_S}(\cdot)$ denote the H measure function with a sample S. The binary configuration of every explanatory variable is represented by a binary variable x_i, which indicates whether the corresponding explanatory variable X_i is chosen or not. Likewise, every interaction effect is denoted by X_iX_j, controlled by a binary variable x_{ij}. Then the search strategy to find the optimal logistic regression model maximizing H measure can be formulated as a 0-1 integer programming problem as follows,

$$\text{Maximize } \mathrm{H_S}\big(\mathrm{logit}(\Pr(Y|X_1, X_2, \dots, X_n))\big) \tag{14}$$

$$\text{where logit}(\Pr(Y|X_1, X_2 \ldots, X_n)) = \beta_0 + \beta_i \sum_i X_i x_i + \sum_{i \neq j} \beta_{ij} X_i X_j x_{ij}$$

subject to:

$$x_i = 1 \text{ or } 0,$$
$$x_{ij} = 1 \text{ or } 0\,.$$

Solving this 0-1 nonlinear integer programming problem would find the optimal subset of explanatory variables. Traditionally, some nonlinear integer programming problems can be solved by linearization techniques, e.g. simple approximation using piecewise linear functions (Jünger et al. 2009). Considering the complex structure of $H_S(\cdot)$ and the varying values of the objective function dependent upon the given sample S, it is obvious to see that the objective function in (14) cannot be converted to an integer linear programing problem. Thus, it is unable to solve this 0-1 nonlinear integer programming using existing exact algorithms, such as cutting plane and branch and bound methods. Although an exhaustive search that examines all non-null subsets of the explanatory variables can always find the optimal model, such a full-scale search is often not feasible in practice. In medical studies, regression analyses often involve a large number of observations, which make exhaustive searches computationally expensive. Furthermore, the high dimensionality of the explanatory variable space can also result in intractable computational effort. For example, if we model all the 11 breast cancer risk factors and all possible two-way interaction effects between them as the explanatory variables, there are 66 variables being considered, which constitute $2^{66} - 1$ non-null subsets in total. Calculating the H measure for each model and identifying the optimal model is nearly impractical. Therefore, a suboptimal heuristic search strategy is useful in this case.

In recent years, the most commonly used heuristic search strategy of model selection is the greedy search algorithm, which is also referred to as stepwise search (Miller 2002). Since a greedy search usually finds local optima, there are studies using more advanced heuristic algorithms to avoid the shortcoming of the stepwise search. Yang and Honavar (Yang and

Honavar 1998) used a genetic algorithm to perform the model selection and received near-optimal results. Sivagaminathan and Ramakrishnan (2007) developed a hybrid approach of neural networks and ant colony optimization for model selection. Orkcu (2013) found a combination of the genetic algorithm and the simulated annealing algorithm outperforms the original genetic algorithm in most cases.

In this study, aiming at overcoming the local optima of the stepwise model selection methods, we propose a new model selection method for generalized linear model by combining the tabu search algorithm with the H measure. Tabu search was originally introduced by Glover (1989) as an optimization tool applicable to nonlinear covering problems. In various problem settings, tabu search has produced outcomes equal or superior to the best results previously found by other algorithms (Fouskakis and Draper 2002). Furthermore, tabu search has been shown to be very efficient in terms of calculation time in solving many search and optimization problems, even when compared with some advanced heuristic algorithms, such as genetic algorithm and simulated annealing (Augugliaro et al. 1999). As our model has a large number of candidate variables, the calculation time is of great concern. Considering the efficiency and effectiveness of tabu search, we adopt a modified version of tabu search as the search strategy. The modified algorithm consists of three phases: preliminary, intensification, and diversification (Fouskakis and Draper 2002, Glover 1990). In the first phase, this algorithm works as the original tabu search. The preliminary search continues for a specified number of iterations. When it reaches the upper limit of iterations, it moves on to the intensification phase. The intensification starts with the best solution found thus far and clears the tabu list, and then proceeds as in the preliminary phase. If a better solution is found, intensification is restarted with the better solution and an empty tabu list. Such a process repeats until the upper limit of iterations is reached and

then goes to the diversification stage. Intensification allows the algorithm to focus the search around the current best solution. The last phase, diversification, provides a way to explore spaces that have not been visited in the first two phases. The diversification clears the tabu list and sets the most frequent moves as tabu. Then the algorithm starts with a random solution and proceeds as in the preliminary phase. After going through the three phases, the model with the highest H measure found in the whole process is reported as the optimal breast cancer risk estimation model. A detailed algorithm is described in Figure 1.

Tabu Search Model Selection Algorithm

1. Specify the size of the tabu list ***s*** and the number of iterations ***n*.**
2. Fit a model using all variables and use the one with the highest absolute multiplier as initial solution.
3. Evaluate the solution (by calculating the corresponding H measure).

Preliminary Phase

Repeat

(1) Change the solution by removing or adding a variable (switch the value of variable between 0 and 1).

(2) If the objective value > the aspiration value i.e., the best solution so far, then accept it as the current solution.

Else if the element is in the tabu list, then go to the neighboring variable.

Else if the element is not in the tabu list, then accept it as the current solution.

(3) Update the tabu list and the aspiration value.

(4) Check stopping criterion.

If the iterations reach the upper limit ***n***, record the best solution X so far and go to the intensification phase.

Intensification Phase

Start with the best solution X so far and clear the tabu list.

Repeat

Do the preliminary phase, until a solution better than X is found and restart the intensification phase.

If the number of iteration reaches ***n***, go to the diversification phase.

Diversification Phase

(1) Clear the tabu list and set the ***s*** most frequent moves as tabu.

(2) Randomly generate a solution.

Repeat

Do the preliminary phase, until the number of iteration reaches the upper limit ***n***.

Report the best solution found in the three phases.

Figure 1 The model selection algorithm

It is known that, in order for an interaction effect to be significant, at least one of its constituent main effects should be significant under the so-called effect heredity principle (Wu and Hamada 2011). However, since the purpose of the modeling focuses on the predictive performance of each model under the imbalanced misclassification costs, rather than its conventional statistical interpretability, our model selection procedure still allows an interaction effect to be selected without its constituent main effects, which is consistent with the majority of other model selection methods (Cleves 2008, Shmueli 2010, Lenz et al. 2012).

3.2.2. Decision Rule of Accepting or Rejecting a Screening Mammography Exam

After the optimal risk estimation model is identified, the second stage of the decision-making framework is to determine whether a woman should undergo a screening mammogram in the current year based on her estimated breast cancer risk. We adopt the misclassification cost term criterion (MCT) (Greiner 1996) to calculate the optimal cut-off points of cancer probabilities for recommending a screening mammogram. The MCT is a variant of Equation (4). It also incorporates the costs associated with false-positive and false-negative misclassifications to determine the optimal threshold that results in the least expected cost. Let $\emptyset_t$ denote the optimal cut-off point at age t. Only when a woman has a breast cancer risk higher than the optimal cut-off point, should she be recommended to undergo a screening mammogram at the current age. Based on the MCT, the optimal cut-off point for age *t*, $\emptyset_t$, is determined by

$$\min_{\emptyset_t \in [0,100\%]} \text{MCT} = \frac{C_t^N}{C_t^P} \cdot p_t \cdot [1 - \text{Se}(\emptyset_t)] + (1 - p_t) \cdot [1 - \text{Sp}(\emptyset_t)], \tag{15}$$

where p_t is the prevalence of breast cancer in age group t; $\text{Se}(\emptyset_t)$ is the sensitivity associated with $\emptyset_t$ (i.e., the percentage of true positive decisions in all positive classifications); $\text{Sp}(\emptyset_t)$ is the specificity associated with $\emptyset_t$ (i.e., the percentage of true negative decisions in all negative

decisions). Since all parameters involved above are age-specific, we provide the optimal cut-off point for each age group separately.

3.3. Numerical Study

3.3.1. Data and Parameters Estimation

We use the Breast Cancer Surveillance Consortium (BCSC) dataset (Barlow 2006), which includes screening mammography information from 1,007,600 women from January 2000 to December 2009, to execute the regression analysis. The BCSC is a collaborative network of seven mammography registries for studies aiming to assess the delivery and quality of breast cancer screening and related patient outcomes in the United States (2004). Each woman in the BCSC dataset has one screening mammogram and corresponding one-year outcome. The dataset includes each participants' breast cancer diagnosis information within 1 year of the screening mammogram as well as their personal characteristics (including risk factors) as introduced in Section 3.2.1. All these risk factors, except age, are incorporated as explanatory variables in the regression analysis of the risk model. Age is used to stratify the regression models so as to provide age-specific risk estimation models and cut-off points. We adopt the listwise deletion to deal with missing data. In this numerical study, we consider women aged 40-84 since the BCSC data only includes women aged 84 and younger. Whether the woman is diagnosed with breast cancer within one year of the screening mammogram serves as the response variable. We also use this dataset to derive the age-specific prevalence of breast cancer (P_t), as well as the proportions of in situ cancer ($P_t(\text{ins}|\text{C})$) and invasive cancer cases ($P_t(\text{inv}|\text{C})$).

In order to validate the performance of our decision model, we use the holdout method that randomly splits the dataset into a training set and a test set. The training set contains 70% of the observations, while the remaining 30% are used to build the test set. We use the training set to fit

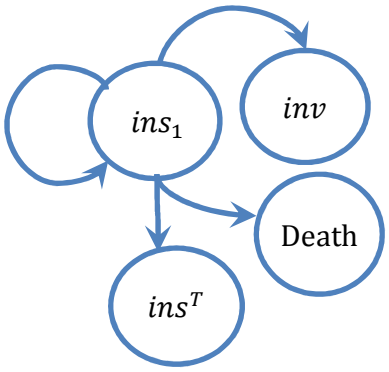
regression models and determine optimal cut-off points. Then the test set is employed to validate our decision model and make comparisons with alternative mammography screening recommendations.

Table 1 Sources of data input for model parameter estimation

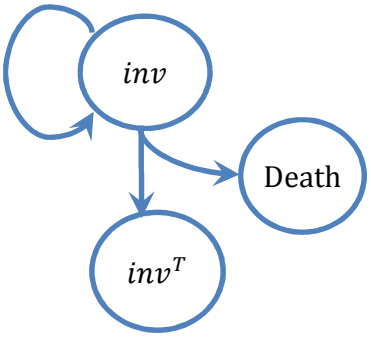
Parameters	Source
λ_t, $\mathrm{P}_t(\mathrm{ins}_0\|\mathrm{ins})$, $\mathrm{P}_t(\mathrm{ins}_1\|\mathrm{ins})$,	Yen et al. (2003)
d_t^{M}, d_t^{B}	Saarni et al. (2006); Ayer et al., (2012)
$P_t(\mathrm{D})$, $P_t(\mathrm{UD})$	Verkooijen et al. (2005); Zhang (2011)
$L_t(\mathrm{inv}^{\mathrm{T}})$,	Beck et al. (1982)
$L_t(\mathrm{ins}^{\mathrm{T}})$, $L_t(\mathrm{ins}_0)$	Arias (2011); Howlader et al. (2012)
$P_t(\mathrm{Positive}\|\mathrm{NC})$, $P_t(\mathrm{Negative}\|\mathrm{C})$, $P_t(\mathrm{Positive}\|\mathrm{C})$	Yankaskas et al. (2005); Weaver et al. (2006)
$\mathrm{P}_t(\mathrm{B}_0\|\mathrm{NC}), P_t(\mathrm{B}_0\|\mathrm{C})_t$	Hubbard et al. (2011)
$L_t(\mathrm{ins}_1)$, $L_t(\mathrm{inv})$,	BCSC (2009); Zhang (2011); Arias (2011); Yen et al. (2003)

Table 1 lists other parameters used in the numerical experiments. In particular, as a woman ages, the risk and side-effects caused by a biopsy become more serious (Burnside et al. 2012). Hence, we assume the disutility of biopsy is inversely proportional to the age-specific EQ-5D scores of the general population (Saarni et al. 2006), which reflects the varying negative impacts of biopsy on women's health at different ages. Since no study provides age-specific annual death rates of untreated invasive breast cancer, we estimate these rates based on the one-year treated death rate of invasive cancer using the data from an earlier study (Zhang 2011). To adjust for the untreated breast cancer, we use Verkooijen et al.'s finding (2005) on the effect of refusal of treatment on mortality, i.e., the death rate of untreated breast cancers is threefold higher than that of treated breast cancers. In addition, we assume that life expectancy of patients treated for in situ cancer ($L_t(\mathrm{ins}^{\mathrm{T}})$) is equal to that of the general population, since the survival rate of in situ cancer is nearly 100% (Recht et al. 1998). Similarly, as a non-progressive in situ cancer will never progress to a life-threating invasive disease even if left untreated (Yen et al. 2003), we also assume the life expectancy of a non-progressive in situ cancer patient ($L_t(\mathrm{ins}_0)$) to be equal to

that of the general population. We calculate the life expectancy of patients treated for invasive cancer ($L_t(\text{inv}^{\text{T}})$) using the DEALE method based on the 10-year survival of invasive breast cancer (Beck et a. 1982, Howlader et al. 2012).



The Markov chain representing progressive in situ cancer



The Markov chain representing progression of invasive cancer

ins_1 = undetected progressive in situ cancer; ins^{T}= in situ cancer patient in treatment;
inv = undetected invasive cancer; inv^{T}= treated invasive cancer;
Death = all-cause death

Figure 2 Markov chain diagrams for estimating $L_t(\text{ins}_1)$ and $L_t(\text{inv})$

In order to estimate $L_t(\text{ins}_1)$ and $L_t(\text{inv})$, we assume that the two types of breast cancer patients will strictly undergo annual screening mammography from t. We use two discrete-time Markov chain models to estimate their expected values. Figure 2 shows the respective transition diagrams of the Markov chain models. The transitions from ins_1 and inv to their corresponding treatment states ($\text{ins}^{\text{T}}$and inv^{T}) are defined as true positive rate of mammography. Thus, a self-loop means the mammogram in the current year fail to detect the in situ cancer or invasive cancer. The transition from ins_1 to "Death" represents the one-year all-cause mortality rate. The transitions between ins_1 and inv represents the one-year progression from in situ cancer to invasive cancer. The treatments and death states are all absorbing states. We assume age 100 is the end of the time horizon, so there are $100 - t$ steps for a process starting with a state at age t.

Thus, the expected lengths of survival counted during the Markov chains looping from age t to age 100 can be used to estimate the life expectancy of $L_t(\text{ins}_1)$ and $L_t(\text{inv})$.

3.3.2. Data Smoothing

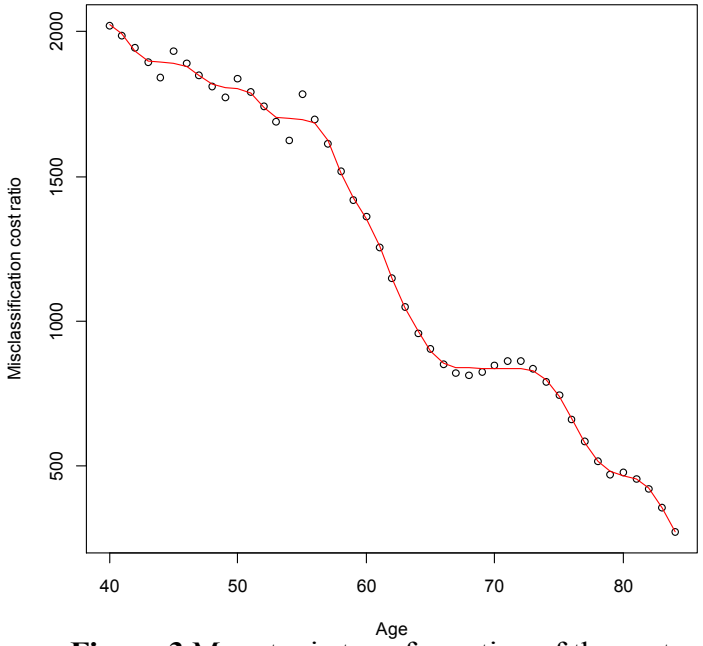


Figure 3 Monotonic transformation of the cost ratios

Braithwaite et al. (Braithwaite et al. 2013) found frequent screenings result in higher risk of false positive mammography results and biopsy referrals without added benefit on older women as compared with younger women. Therefore, it is reasonable to assume that, the ratio between the cost of a false negative result and the cost of a false positive result, which reflects the net benefit of undergoing a mammography screening, will decrease as women age. While the calculated results of the cost ratios by age show a decreasing trend (dots in Figure 3), they are not monotonically decreasing. This may be due to various sources of input data for numerical calculation. In this study, we use a monotone smoothing method proposed by Ramsay (1998) to regularize the raw results of the cost ratios so that they are monotonically decreasing with age.

Suppose x_0 is the lower boundary of age, and the monotonically increasing smoothing function is defined as follows.

$$f(x) = \beta_0 + \beta_1 \int_{x_0}^{x} \mathrm{e}^{\mathrm{w}(u)} \mathrm{d}(u), \tag{16}$$

where β_0 is the value of $f(x_0)$; β_1 is the slope of $f(x)$ at x_0; and $\mathrm{w}(u)$ is a B-spline basis function of order 6 with $\mathrm{w}(x_0)$ equal to zero. β_0 , β_1 and $\mathrm{w}(u)$ are three objects for each smoothed curve estimated from the raw results. The smoothing function iteratively minimizes the least squares fitting criterion (see Ramsay (1998) for details). The monotonically smoothed results of the cost ratios are plotted in Figure 3 with a solid line.

3.3.3. Results

Table 2 Parameters of the tabu search

Size of the tabu list	5	10	20
Iterations in each phase	100	500	1000
Total iterations in three phases	300	1500	3000

We set several scenarios to execute the tabu search algorithm for each age group. Table 2 summarizes the size of the tabu list and the number of iterations in each phase for each scenario. We use three typical levels of total iterations, which represent small, medium and large iteration numbers to test the convergence of the tabu search algorithm. In addition, three different tabu list sizes are employed to ensure a better performance. Therefore, for each age, the model selection is performed 9 times. The best results in 9 runs are selected as the optimal prediction models.

We found the choice of the size of tabu list does not affect the final solutions (results not shown). However, for most age groups, only when the number of iterations in each phase is set to be higher than or equal to 500, does the iteration procedures converge to their final solutions. In contrast, when it is set to 100, the iteration process rarely converges but usually keeps finding better solutions until the end of the intensification phase or the diversification phase. Table 3

shows the combination of explanatory variables of each age's optimal risk prediction model as well as its corresponding value of the H measure.

Although all of the risk factors we included have been associated with increased breast cancer risk in previous studies, the main effects included in the prediction model vary dramatically with age. The main effect selected most often is last mammographic result, which is included in 28 out of 45 age groups. In contrast, the main effect selected least often is current menopausal status, which never appears as a main effect. In addition, as expected, the interaction effects between different risk factors play a vital role in every model. Since we allow our model selection to select an interaction effect without its constituent main effects, some risk factors, which form the principal part of interaction effects of a model, do not appear in the main effects of the model. For example, in the risk model of age 40, although BMI and previous breast procedure are not selected as main effects, they are involved in 8 different interaction effects (i.e., race×bmi, bmi×agefirst, bmi×nrelbc, bmi×brstproc, bmi×lastmamm, race×brstproc, hispanic×brstproc, and brstproc×lastmamm). Such a pattern is very common in other age groups. In particular, although menopausal status never serves as a main effect, it is included in the interaction effects of all age groups between 45 and 54 years. In addition to menopausal status, we also found some risk factors are much more common in the interaction effects than in the main effects, such as the number of first degree relatives with breast cancer. While this factor is the second least selected main effect and only appears in 14 age groups as the main effect (ages 40-44, 47-49, 62, and 80-84), it appears in the interaction effects of 44 out of 45 age groups. These results imply menopausal status and the number of first degree relatives with breast cancer are akin to cofactors that join with another factor to affect a woman's breast cancer risk.

Table 3 Risk estimation models and cut-off probabilities for accepting a mammogram

Age		40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Main effects	menopaus															
	race	√	√	√			√	√	√	√	√	√	√	√	√	√
	hispanic	√	√	√	√	√							√			
	bmi						√	√	√	√	√					
	agefirst	√	√	√	√	√						√	√	√	√	√
	nrelbc	√	√	√	√	√			√	√	√					
	brstproc								√	√	√				√	√
	lastmamm	√	√	√	√	√	√	√	√	√	√	√	√	√		
	surgmeno														√	√
	hrt											√	√	√		
Interaction effects	menopaus×race															
	menopaus×hispanic											√				
	menopaus×bmi								√	√	√				√	√
	menopaus×agefirst						√	√	√	√	√	√	√	√		
	menopaus×nrelbc														√	√
	menopaus×brstproc						√	√	√	√	√	√	√	√		
	menopaus×surgmeno						√	√	√	√	√		√		√	√
	menopaus×lastmamm								√	√	√	√	√	√	√	√
	menopaus×hrt														√	√
	race×hispanic	√	√	√	√	√			√	√	√				√	√
	race×bmi	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	race×agefirst				√	√	√	√	√	√	√	√	√	√	√	√
	race×nrelbc	√	√	√	√	√	√	√	√	√	√				√	√
	race×brstproc	√	√	√	√	√						√	√	√	√	√
	race×lastmamm	√	√	√	√	√	√	√	√	√	√	√		√		
	race×surgmeno						√	√	√	√	√	√		√	√	√
	race×hrt						√	√							√	√
	hispanic×bmi								√	√	√	√	√	√		
	hispanic×agefirst											√	√	√	√	√
	hispanic×nrelbc						√	√	√	√	√	√		√	√	√
	hispanic×brstproc	√	√	√	√	√	√	√	√	√	√	√		√	√	√
	hispanic×lastmamm	√	√	√	√	√										
	hispanic×surgmeno								√	√	√	√	√	√		
	hispanic×hrt											√		√	√	√
	bmi×agefirst	√	√	√	√	√			√	√	√	√	√	√	√	√
	bmi×nrelbc	√	√	√	√	√									√	√
	bmi×brstproc	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	bmi×lastmamm	√	√	√	√	√			√	√	√			√	√	√
	bmi×surgmeno												√	√		
	bmi×hrt						√	√	√	√	√	√	√	√		
	agefirst×nrelbc	√	√	√	√	√									√	√
	agefirst×brstproc															
	agefirst×lastmamm						√	√	√	√	√	√	√	√	√	√
	agefirst×surgmeno						√	√	√	√	√	√	√	√	√	√
	agefirst×hrt						√	√	√	√	√	√	√	√	√	√
	nrelbc×brstproc						√	√								
	nrelbc×lastmamm	√	√	√	√	√	√	√	√	√	√					
	nrelbc×surgmeno											√		√	√	√
	nrelbc×hrt													√	√	√
	brstproc×lastmamm	√	√	√	√	√	√	√	√	√	√	√	√	√		
	brstproc×surgmeno						√	√	√	√	√					
	brstproc×hrt						√	√	√	√	√				√	√
	lastmamm×surgmeno						√	√	√	√	√	√	√	√		
	lastmamm×hrt						√	√	√	√	√	√	√	√	√	√
	surgmeno×hrt						√	√	√	√	√	√	√	√		
H measure		0.146	0.146	0.147	0.145	0.145	0.240	0.239	0.231	0.231	0.231	0.235	0.228	0.273	0.266	0.266
Ratio between C_t^N and C_t^P		2023	1993	1933	1900	1893	1890	1881	1850	1817	1808	1803	1789	1739	1705	1700
Optimal Cut-off point (‰)		1.713	1.713	1.713	1.765	1.765	2.201	2.201	2.315	2.315	2.315	2.727	2.845	3.091	3.127	3.127

(Menopaus =menopausal status, Agefirst=age at first birth, Nrelbc=No. of first degree relatives with breast cancer, Brstproc= Previous breast procedure, Lastmamm= last mammogram, Surgmeno=surgical menopause, Hrt=current hormone therapy)

Table 3 Risk estimation models and cut-off probabilities for accepting a mammogram (Cont.)

Age		55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
Main effects	menopaus															
	race	√	√	√	√	√			√	√	√	√	√	√	√	√
	hispanic				√		√	√								
	bmi	√	√						√	√			√	√	√	√
	agefirst	√	√		√	√	√	√					√	√	√	√
	nrelbc								√							
	brstproc	√	√	√	√					√	√	√	√	√	√	√
	lastmamm	√	√	√	√	√	√	√			√					
	surgmeno								√	√	√					
	hrt				√				√			√	√	√	√	√
Interaction effects	menopaus×race															
	menopaus×hispanic															
	menopaus×bmi															
	menopaus×agefirst															
	menopaus×nrelbc															
	menopaus×brstproc															
	menopaus×surgmeno															
	menopaus×lastmamm															
	menopaus×hrt															
	race×hispanic			√	√	√	√					√	√	√	√	√
	race×bmi						√	√	√	√	√	√	√	√	√	√
	race×agefirst	√	√	√	√	√						√	√	√	√	√
	race×nrelbc	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	race×brstproc							√		√	√	√	√	√	√	√
	race×lastmamm	√	√	√	√	√	√	√	√	√	√					
	race×surgmeno			√		√	√	√	√	√	√	√	√	√	√	√
	race×hrt	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	hispanic×bmi	√	√	√	√											
	hispanic×agefirst												√	√	√	√
	hispanic×nrelbc				√		√									
	hispanic×brstproc								√				√	√	√	√
	hispanic×lastmamm	√	√	√	√	√	√	√	√	√	√					
	hispanic×surgmeno	√	√	√	√	√	√	√	√	√	√					
	hispanic×hrt							√	√	√	√	√				
	bmi×agefirst			√	√	√	√	√	√	√	√	√	√	√	√	√
	bmi×nrelbc			√	√							√	√	√	√	√
	bmi×brstproc	√	√	√	√	√	√	√	√	√		√	√	√	√	√
	bmi×lastmamm							√	√	√	√	√	√	√	√	√
	bmi×surgmeno	√	√	√	√	√	√	√	√	√	√					
	bmi×hrt	√	√	√		√	√	√	√			√	√	√	√	√
	agefirst×nrelbc	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	agefirst×brstproc									√	√	√				
	agefirst×lastmamm	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	agefirst×surgmeno	√	√	√	√	√	√	√	√	√	√					
	agefirst×hrt	√	√	√	√	√	√	√	√	√	√					
	nrelbc×brstproc							√		√	√	√	√	√	√	√
	nrelbc×lastmamm	√	√	√			√	√	√	√	√	√	√	√	√	√
	nrelbc×surgmeno	√	√	√								√	√	√	√	√
	nrelbc×hrt			√	√	√	√	√	√	√						
	brstproc×lastmamm	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	brstproc×surgmeno															
	brstproc×hrt	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	lastmamm×surgmeno	√	√	√		√			√	√	√	√	√	√	√	√
	lastmamm×hrt	√	√	√	√	√	√	√	√	√	√	√				
	surgmeno×hrt															
H measure		0.230	0.230	0.219	0.243	0.228	0.238	0.226	0.235	0.256	0.241	0.237	0.218	0.221	0.221	0.221
Ratio between C_t^N and C_t^P		1698	1687	1623	1509	1426	1355	1260	1149	1047	965	898	854	840	839	838
Optimal Cut-off point (‰)		3.524	3.524	3.461	3.761	3.593	3.725	4.523	4.476	4.687	4.729	5.013	5.132	5.132	5.132	5.132

Table 3 Risk estimation models and cut-off probabilities for accepting a mammogram (Cont.)

Age		70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Main effects	menopaus															
	race															
	hispanic	√				√	√	√	√		√	√	√	√	√	√
	bmi	√	√	√	√						√				√	√
	agefirst		√	√	√	√				√	√					
	nrelbc											√	√	√	√	√
	brstproc							√	√	√		√	√	√		
	lastmamm							√	√			√	√	√	√	√
	surgmeno	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	hrt	√	√	√	√	√	√	√	√	√	√					
Interaction effects	menopaus×race															
	menopaus×hispanic															
	menopaus×bmi															
	menopaus×agefirst															
	menopaus×nrelbc															
	menopaus×brstproc															
	menopaus×surgmeno															
	menopaus×lastmamm															
	menopaus×hrt															
	race×hispanic	√	√	√	√		√	√	√	√	√					
	race×bmi					√	√	√	√	√	√	√	√	√	√	√
	race×agefirst						√	√	√							
	race×nrelbc	√	√	√	√	√				√	√					
	race×brstproc	√	√	√	√	√	√	√	√	√	√					
	race×lastmamm	√	√	√	√	√	√	√	√						√	√
	race×surgmeno	√	√	√	√	√	√	√	√	√	√	√	√	√		
	race×hrt	√	√	√	√	√				√	√					
	hispanic×bmi														√	√
	hispanic×agefirst	√	√	√	√	√	√	√	√	√						
	hispanic×nrelbc									√	√	√	√	√		
	hispanic×brstproc	√	√	√	√	√		√	√	√		√	√	√		
	hispanic×lastmamm										√				√	√
	hispanic×surgmeno						√	√	√		√	√	√	√	√	√
	hispanic×hrt						√	√	√			√	√	√	√	√
	bmi×agefirst	√	√	√	√	√	√			√	√					
	bmi×nrelbc	√	√	√	√	√	√		√	√	√	√	√	√	√	√
	bmi×brstproc								√	√	√	√	√	√	√	√
	bmi×lastmamm	√	√	√	√	√		√							√	√
	bmi×surgmeno	√	√	√	√	√		√		√	√					
	bmi×hrt	√	√	√	√	√				√	√					
	agefirst×nrelbc	√	√	√	√	√		√	√			√	√	√		
	agefirst×brstproc					√	√	√	√	√	√	√	√	√	√	√
	agefirst×lastmamm	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	agefirst×surgmeno						√		√	√	√	√	√	√	√	√
	agefirst×hrt					√	√	√	√						√	√
	nrelbc×brstproc	√	√	√	√							√	√	√	√	√
	nrelbc×lastmamm	√	√	√	√							√	√	√	√	√
	nrelbc×surgmeno	√	√	√	√	√	√	√	√	√						
	nrelbc×hrt	√	√	√	√	√	√			√	√	√	√	√	√	√
	brstproc×lastmamm	√	√	√	√	√	√	√	√	√		√	√	√	√	√
	brstproc×surgmeno															
	brstproc×hrt	√	√	√	√	√		√		√	√	√	√	√	√	√
	lastmamm×surgmeno	√	√	√	√	√				√		√	√	√	√	√
	lastmamm×hrt	√	√	√	√	√		√	√	√		√	√	√	√	√
	surgmeno×hrt														√	√
H measure		0.270	0.270	0.270	0.272	0.311	0.419	0.422	0.425	0.439	0.483	0.550	0.551	0.553	0.546	0.550
Ratio between C_t^N and C_t^P		838	838	838	830	798	740	666	582	515	483	468	454	424	356	272
Optimal Cut-off point (‰)		4.869	4.869	4.869	4.869	4.598	6.253	6.131	6.102	6.195	6.224	6.721	6.721	6.721	6.833	6.833

The interaction effects included in each age group are also remarkably different. In particular, the interaction effect between previous breast procedure and last mammographic result is the most commonly selected predictor, which appears in 42 out of 45 age groups. The second most commonly selected interaction effect is the one between age at first birth and last mammogram (40 age groups), followed by BMI and previous breast procedure (37 age groups). In contrast, the interaction effect between menopausal status and race is excluded from all age groups.

After the monotone smoothing, the ratio between false negative cost and false positive cost strictly decreases with age. The cost ratios for young age groups are as high as 2023 (age 40), but the ratios for older age groups (ages 79-84) fall below 500, implying that the net benefit of undergoing a mammogram will reduce as a woman ages. In addition, there is also a quasi-monotone relationship between age and the optimal cut-off points of predicted probabilities for accepting a mammogram (Figure 4). Since a higher value of the cut-off point implies a lower chance of accepting a mammogram, the figure indicates that as women age, they should adopt a more parsimonious attitude towards annual screening mammography. The curve rises with a significantly higher rate after age 75, which is consistent with the USPSTF's recommendation against routine screening for women 75 years or older.

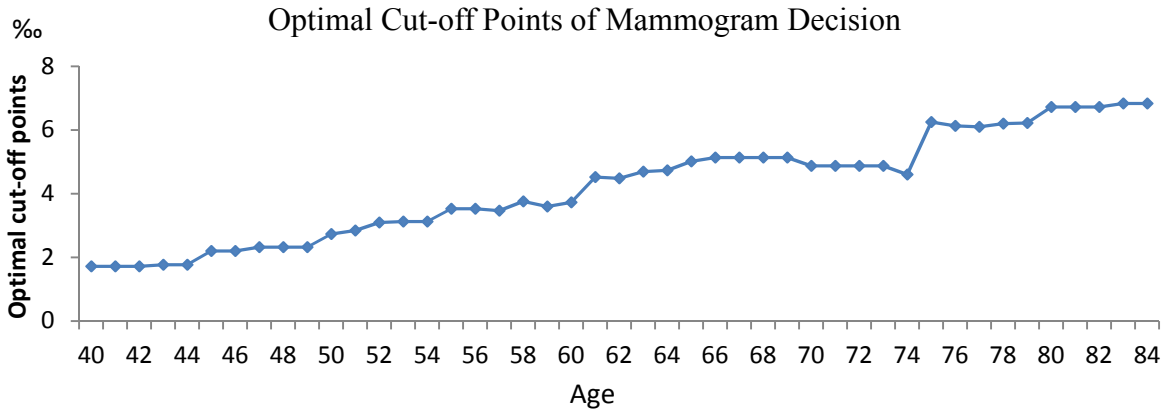


Figure 4 Optimal cut-off points of probabilities for accepting a mammogram by age

3.3.4. Sensitivity Analyses

Due to insufficient data in the literature, two sets of parameters may be subject to variability: disutilities associated with screening (d_t^M and d_t^B) and death rates of untreated invasive cancers ($P_t(UD)$). We perform a sensitivity analyses on these parameters to examine the impact of their variability on the risk model. The sensitivity analyses are conducted on the age groups 40, 60 and 80, which represent the start of screening, mid-point and older age. We adjust these two parameters (increase or decrease by 10%, 30% or 50% each time) to see whether and how the cut-off probability of receiving a mammogram in the optimal risk model may change, with the results summarized in Table 4. "N" suggests that the model stays the same; otherwise, the differences from the original cut-off probabilities are reported. The results suggest that variation in the disutilities associated with screening do not impact the risk models and cut-off points in most cases. However, the risk model is sensitive to the variation in the death rates of untreated invasive cancers for the 60 and 80 years age groups; even just a 10% increase of $P_t(UD)$ results in changed risk models and cut-off points.

Table 4 Sensitivity analyses

Age	Tested Parameter	+50%	+30%	+10%	-10%	-30%	-50%
40	d_t^M, d_t^B	+0.052‰	+0.052‰	N	N	-0.024‰	-0.024‰
	$P_t(UD)$	N	N	N	N	N	N
60	d_t^M, d_t^B	+0.798‰	+0.751‰	N	N	-0.034‰	-0.034‰
	$P_t(UD)$	-0.384‰	-0.384‰	-0.384‰	N	+0.079‰	+0.079‰
80	d_t^M, d_t^B	N	N	N	N	N	-0.611‰
	$P_t(UD)$	-0.494‰	-0.494‰	-0.494‰	N	+0.152‰	+0.152‰

3.3.5. Comparison of Different Screening Policies

For women with an average risk of breast cancer, the ACS recommends that they undergo annual screening mammography starting at age 45. Women 55 years and older should either switch to biennial screenings or continue screening annually. We compare the outcome of using the

decisions generated by our optimal model with that of annual screenings starting at age 45, which is one of the screening scenarios recommended by the ACS. In addition, in order to validate the advantage of the tabu search-based risk model, a risk model using all of the breast cancer risk factors suggested by Barlow et al. (2006) as well as a model using predictors selected by a simple greedy search are included in the comparison. Likewise, the MCT is used to determine the optimal cut-off points of generating screening mammogram decisions for the Barlow model and the model selected by greedy search. We also use AUC as an alternative criterion in tabu search to demonstrate the superiority of using the H measure over traditional model selection criteria. Table 5 summarizes the decision policies used in the comparison.

Table 5 Summary of decision models and screening policy for comparison

Decision Policy	Search Strategy	Model Selection Criterion	Mammogram Decision rule
Optimal model	Tabu search	H measure	MCT threshold
Greedy search model	Greedy search	H measure	MCT threshold
AUC model	Tabu search	AUC	MCT threshold
Barlow model	N/A	N/A	MCT threshold
Annual screening from age 45	N/A	N/A	Annual mammogram

We calculate the average loss of life expectancy caused by a single screening decision at each age when following the decision policies in Table 5. The comparison focuses on the average impact of a single screening decision at a particular age on a woman's life expectancy. With the intention of validating the out-of-sample performance of our proposed model, the loss of life expectancy for each of the decision models are calculated based on the test set split from the original dataset. Since every woman in our dataset has in fact undergone a screening mammogram and been informed of the result, it is straightforward to determine each subject's actual loss of life expectancy using the calculated costs of false negative results and false positive results. Thus, for every age from 45 through 84 years old, we calculate the loss of life

expectancy for each woman in the age group when the woman adopts each of the five decision policies. Then we use each woman's risk characteristics to calculate her possibility of having breast cancer at her current age, and generate the optimal screening decision for her based on the optimal threshold. Since whether a woman has cancer is known, we are able to calculate her loss of life expectancy under different screening decision policies. Then we sum the total losses for all women in each age group and compute the average loss L_t, for every policy using

$$L_t = \frac{WC_t \times C_t^N + WN_t \times C_t^P}{n_t},$$

where WC_t is the number of cancerous observations which are incorrectly identified as non-cancerous by the given decision model; WN_t is the number of non-cancerous observations which are falsely identified as cancerous by the given decision model; and n_t is the total number of observations in age group *t*.

The age-specific average losses of the five decision policies are plotted in Figure 5. Our optimal decision model consistently outperforms the annual screening and the remaining four decision models. When following the optimal decisions, the average loss of life expectancy caused by a single decision ranges between 0.001062 and 0.001663 years. The annual screening policy results in a significantly higher range of loss (between 0.001699 and 0.001738 years) as compared with our optimal decisions. Adopting the ACS's biennial mammography screening after age 54 incurs higher loss of life expectancy (results not shown on Figure 5). In addition, the average loss of switching to biennial screening mammography after age 54 ranges from 0.01678 to 0.00521 years, which is three to nine times higher than the average loss when continuing annual mammography at the corresponding ages. The decision model based on the Barlow model shows slightly better outcomes compared to the annual screening policy. The performance of the model selected by a simple greedy search lies between the Barlow model and the optimal model.

The decision model selected by the AUC has the worst out-of-sample performance, which may be due to the ROC curve assigning the same weights to false positive and false negative errors.

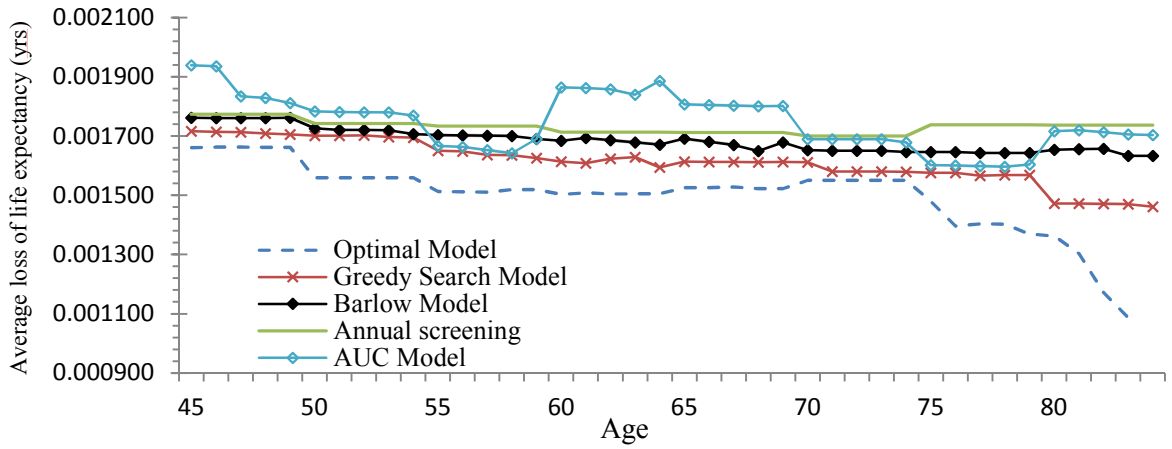


Figure 5 Comparing the optimal decisions with the annual screening policy and other models

Figure 6 presents the optimal cut-off probabilities associated with the four decision models starting at age 45. The annual screening policy does not make screening recommendations based on risk probabilities, and thus is not compared. The cut-off points from the greedy search model are closest to those of the optimal model. The cut-off points of the AUC model are significantly higher than the other three models, while the Barlow model has the lowest cut-off points among the four models. Since the Barlow model uses the same predictors for all age groups, it has a smoother curve than the other three models.

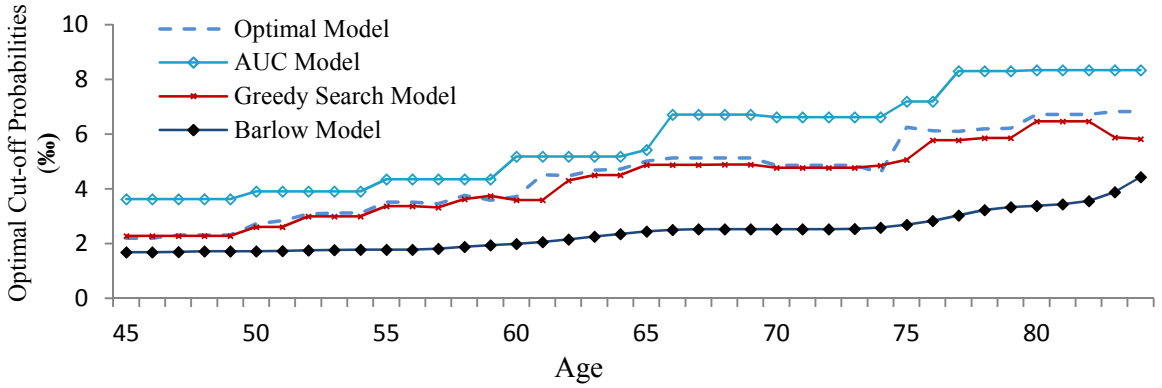


Figure 6 Comparing the optimal cut-off probabilities associated with the four decision models

We further compare the five screening policies in terms of the total number of screening mammograms recommended in the population using the test set. As shown in Table 6, our optimal model recommends the second lowest number of screening mammograms (331,504), which is only 86% of the total number of mammograms recommended in the annual screening policy (384,900). Therefore, using the optimal screening policy developed in this study not only decreases the loss of life expectancy of the general population, but also potentially reduces the amount of medical expenses on breast cancer screenings.

Table 6 Total number of mammograms under different policies

Decision Policy	Total Number of Mammograms	Ratio to the Annual Screening Policy
AUC model	310501	81%
Optimal model	331504	86%
Greedy search model	358603	93%
Barlow model	374564	97%
Annual screening from age 45	384900	100%

It is noteworthy that the comparison is based on the general population, which includes both low risk and high risk women, so that the average differences of loss between the optimal policies and other policies are low in terms of magnitude (less than one day). However, a lifetime cumulative comparison would make a more dramatic difference. In addition, if we were to apply the same comparison to a group of low risk women, instead of the general population or high risk women, we expect the differences to be sharper, as our optimal decision model propels low risk women to undergo fewer mammograms.

3.4 Discussion and Future Work

This paper presents a personalized decision-making model that aims to minimize the negative effects of current breast cancer screening practice. In Figure 7, we illustrate the process of using our personalized model to make screening mammogram decisions in practice. Whether a woman

should receive a screening mammogram is determined based on her risk level in the current year, our "on-line" screening policy is adaptive to a woman's health status. While most previous studies strive to improve the breast cancer screening guidelines by offering lifetime mammography screening schedules, our model proposes utilizing a dynamic decision aid to determine if breast cancer screening should be conducted in the current year. The personalized model does not attempt to answer when to start or end routine mammography screening and how often women should be screened, but instead uses a woman's personal risk of breast cancer and the net benefit of receiving a screening mammogram as two variable constraints to dynamically adjust the screening decision. As a uniform and static breast cancer screening policy is by no means perfectly applicable to different women or even the same woman at different time points, this research may inform the long-term controversies on breast cancer screening policies. Furthermore, the two-stage decision-making framework proposed in this study could also be applied to other disease prevention and treatment areas, such as prostate cancer screening, as long as there exists a binary decision-making process and the costs of two types of wrong decisions can be well quantified.

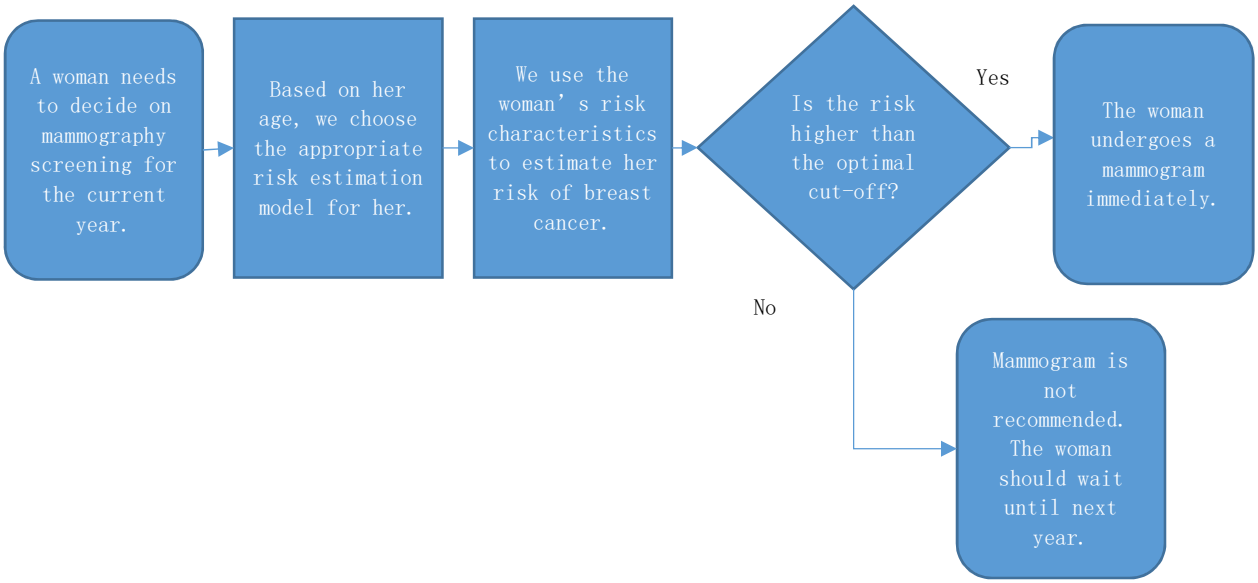


Figure 7 Flow chart of mammography decision-making process

There are methodological contributions in this study. Different from previous personalized optimization models that do not purposely select variables but adopt all available static risk factors, the determinants of the proposed risk estimation model are selected based on a particular focus, i.e., making an optimal decision that minimizes the expected cost of false prediction. This study combined the H measure, a novel prediction-cost-oriented criterion of classification and prediction problems, with an efficient heuristic search algorithm model selection to overcome the high dimensionality of explanatory variables and the large number of observations. We expect this combination can provide new insights into the improvement of efficiency and effectiveness when dealing with large scale problems, especially for medical decision making. Meanwhile, we incorporate interaction effects between different risk factors to the risk estimation model of breast cancer. Many prior studies revealed that different risk factors do not affect women's breast cancer risk independently (Mayberry and Stoddard-Wright 1992, Cleary and Maihle 1997, Clavel-Chapelon and Gerber 2002). However, earlier risk estimation studies fail to directly formulate these possible interaction effects in their model (Gail and Rimer 1998, Gail et al. 1989, Barlow et al. 2006). While stratifying risk models by age is commonly adopted, interaction effects between other risk factors have received little attention in previous screening mammography optimization studies. Our results show that interaction effects between risk factors play a critical role in the decision process.

A major limitation of this study is the estimation of the misclassification cost of a false negative result. Life expectancy calculation ($L_t(\text{ins}_1)$ and $L_t(\text{inv})$) is based on the assumption that a woman will strictly undergo annual screening mammograms in the future. In fact, whether the woman will receive mammograms in the subsequent years depends on her future risk factors. However, since the misclassification cost of a false negative result is considered as the difference

between the life expectancies of skipping a necessary screening mammogram and receiving a necessary screening mammogram, the subtraction greatly offsets the possible bias caused by the assumption. Another major limitation is that the model incorporates only the most recent mammogram result as a risk factor, rather than the full history of previous mammography screenings. Different numbers of prior negative and positive mammography results may indicate different levels of risk. Moreover, the interval between the last mammogram and the current time point may also make a difference in terms of cancer risk. Due to the fact that the BCSC dataset only contains one observation per woman, we are only able to consider the most recent screening mammogram regardless of the number or interval, which is a common practice in prior studies (Barlow et al. 2006, McCann et al. 2002). Lastly, some parameters in our model are based on the general population rather than being woman-specific. For instance, different women may have different cancer progression rates, but the progression rates used in this study are age-specific, which means all women in the same age group share the same cut-off probability regardless of their risk characteristics. To the best of our knowledge, there is no information on these patient-specific parameters reported in the literature. The model proposed in this study will be able to personalize both risk models and optimal cut-off points when such information becomes available in the future. Compared with Markovian models, our proposed method has a disadvantage in representing the abovementioned dynamics of the breast cancer screening problem. Our model requires a significant amount of data to properly incorporate these dynamics and such large data is unavailable in most cases. Although the proposed method is capable of processing dynamic risk factors, it primarily focuses on the current decision epoch and does not sufficiently model the women's screening decisions in the future, which may potentially impact the outcome (i.e., total life expectancy) of the current screening decision.

As a future research direction, a different model could be formulated to determine optimal decisions for women who have a personal history of breast cancer and who are undergoing surveillance mammography. The risk of cancer recurrence is not only affected by the identified breast cancer risk factors but also by type of treatment, number of years since treatment, and initial tumor characteristics (Buist et al. 2010). Using the methods developed in this study with data on women with a personal history of breast cancer may allow for the creation of another model that will contribute to our ability to generate personalized recommendations for prevention and control of breast cancer.

Acknowledgements

We appreciate the anonymous reviewers for their supportive and constructive comments.

Data collection and sharing was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040, HHSN261201100031C). A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: http://breastscreening.cancer.gov/

References

Alagoz, O., Ayer, T., & Erenay, F. S. (2011). Operations research models for cancer screening. Wiley Encyclopedia of Operations Research and Management Science.

American Cancer Society. (2016). Cancer Facts & Figures 2016. Atlanta: American Cancer Society;

Arias, E. (2011). United States life tables, 2007. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 59(9), 1-60.

Augugliaro, A., Dusonchet, L., & Sanseverino, E. R. (1999). Genetic, simulated annealing and tabu search algorithms: three heuristic methods for optimal reconfiguration and

compensation of distribution networks. European transactions on electrical power, 9(1), 35-41.

Ayer, T., Alagoz, O., & Stout, N. K. (2012). A POMDP Approach to Personalize Mammography Screening Decisions. *Operations Research*, 60(5), 1019-1034.

Ayvaci, M. U., Alagoz, O., & Burnside, E. S. (2012). The effect of budgetary restrictions on breast cancer diagnostic decisions. *Manufacturing & Service Operations Management*, 14(4), 600-617.

Beck, J. R., Pauker, S. G., Gottlieb, J. E., Klein, K., & Kassirer, J. P. (1982). A convenient approximation of life expectancy (the "DEALE"): II. Use in medical decision-making. *The American journal of medicine*, 73(6), 889-897.

Bajeh, A. O., & Abolarinwa, K. O. (2011). Optimization: A comparative study of genetic and tabu search algorithms. International Journal of Computer Applications (IJCA), 31(5).

Barlow, W. E., White, E., Ballard-Barbash, R., Vacek, P. M., Titus-Ernstoff, L., Carney, P. A., & Kerlikowske, K. (2006). Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute*, 98(17), 1204-1214.

Bellman, R. (1957). E. 1957. Dynamic Programming. Princeton University Press.

Bertsimas, D., O'Hair, A., Relyea, S., & Silberholz, J. (2016). An analytics approach to designing combination chemotherapy regimens for cancer. *Management Science*, 62(5), 1511-1531.

Breast Cancer Surveillance Consortium, (2009), Performance Measures for 1,960,150 Screening Mammography Examinations from 2002 to 2006 by Age. Retrieved from http://breastscreening.cancer.gov/data/performance/screening/2009/perf_age.html

Braithwaite, D., Zhu, W., Hubbard, R. A., O'Meara, E. S., Miglioretti, D. L., Geller, B., ... & Kerlikowske, K. (2013). Screening outcomes in older US women undergoing multiple mammograms in community practice: does interval, age, or comorbidity score affect tumor characteristics or false positive rates? *Journal of the National Cancer Institute*, *105*(5), 334-341.

Bruening, W., Schoelles, K., Treadwell, J., Launders, J., Fontanarosa, J., & Tipton, K. (2009). Comparative effectiveness of core-needle and open surgical biopsy for the diagnosis of breast lesions.

Buist, D. S., Abraham, L. A., Barlow, W. E., Krishnaraj, A., Holdridge, R. C., Sickles, E. A., & Geller, B. M. (2010). Diagnosis of second breast cancer events after initial diagnosis of early stage breast cancer. *Breast cancer research and treatment*, 124(3), 863-873.

Burnside, E. S., Chhatwal, J., & Alagoz, O. (2012). What is the optimal threshold at which to recommend breast biopsy? *PloS one*, 7(11), e48820.

Chhatwal, J., Alagoz, O., & Burnside, E. S. (2010). Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Operations research*, 58(6), 1577-1591.

Clavel-Chapelon, F., & Gerber, M. (2002). Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? *Breast cancer research and treatment*, 72(2), 107-115.

Cleary, M. P., & Maihle, N. J. (1997). The role of body mass index in the relative risk of developing premenopausal versus postmenopausal breast cancer. *Experimental Biology and Medicine*, 216(1), 28-43.

Cleves, M. (2008). An introduction to survival analysis using Stata. Stata Press.

Erenay, F. S., Alagoz, O., & Said, A. (2014). Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management*, 16(3), 381-400.

Fouskakis, D., & Draper, D. (2002). Stochastic optimization: a review. *International Statistical Review*, 70(3), 315-349.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mulvihill, J. J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute*, 81(24), 1879-1886.

Gail, M., & Rimer, B. (1998). Risk-based recommendations for mammographic screening for women in their forties. *Journal of clinical oncology*, 16(9), 3105-3114.

Glover, F. (1989). Tabu search-part I. *ORSA Journal on computing*, 1(3), 190-206.

Glover, F. (1990). Tabu search: A tutorial. *Interfaces*, 20(4), 74-94.

Grabowski, J., & Pempera, J. (2007). The permutation flow shop problem with blocking. A tabu search approach. Omega, 35(3), 302-311.

Greiner, M. (1996). Two-graph receiver operating characteristic (TG-ROC): update version supports optimisation of cut-off values that minimise overall misclassification costs. *Journal of immunological methods*, 191(1), 93-94.

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine learning*, 77(1), 103-123.

Hand, D. J. (2010). Evaluating diagnostic tests: the area under the ROC curve and the balance of errors. *Statistics in Medicine*, 29(14), 1502-1510.

Howlader, N., Noone, A. M., Krapcho, M., Neyman, N., Aminou, R., Altekruse, S. F., & Cronin, K. A. (2012). SEER cancer statistics review, 1975-2009 (vintage 2009 populations). Bethesda, MD: National Cancer Institute.

Hubbard, R. A., Kerlikowske, K., Flowers, C. I., Yankaskas, B. C., Zhu, W., & Miglioretti, D. L. (2011). Cumulative Probability of False-Positive Recall or Biopsy Recommendation After 10 Years of Screening MammographyA Cohort Study. *Annals of internal medicine*, 155(8), 481-492.

Jackson, R. R. P., McKay, V. L., & Malpas, J. S. (1981). Patient progress modelling and results in Wilm's tumour. Omega, 9(5), 527-536.

Jünger, M., Liebling, T. M., Naddef, D., Nemhauser, G. L., Pulleyblank, W. R., Reinelt, G., ... & Wolsey, L. A. (Eds.). (2009). 50 years of integer programming 1958-2008: From the early years to the state-of-the-art. Springer Science & Business Media.

Kadane, J. B., & Lazar, N. A. (2004). Methods and criteria for model selection. Journal of the American Statistical Association, 99(465), 279-290.

Kirch, R. L. A., & Klein, M. (1974). Surveillance schedules for medical examinations. Management Science, 20(10), 1403-1409.

Lenz, H. J., Schmid, W., & Wilrich, P. T. (2012). Frontiers in Statistical Quality Control 10. Physica-Verlag.

Maillart, L. M., Ivy, J. S., Ransom, S., & Diehl, K. (2008). Assessing dynamic breast cancer screening policies. *Operations Research*, 56(6), 1411-1427.

Mayberry, R. M., & Stoddard-Wright, C. (1992). Breast cancer risk factors among black women and white women: similarities and differences. *American journal of epidemiology*, 136(12), 1445-1456.

McCann, J., Stockton, D., & Godward, S. (2002). Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast cancer research and treatment*, 4(5), R11.

National Cancer Institute. (2004) Breast Cancer Surveillance Consortium: Evaluating Screening Performance in Practice. NIH Publication No. 04-5490. Bethesda, MD: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services, April 2004. Available at: http://breastscreening.cancer.gov/espp.pdf

Nelson, R. (2010). Mammography Guidelines Generate Confusion and Debate. AJN The American *Journal of Nursing*, 110(3), 14-15.

Miller, A. (2002). Subset selection in regression. CRC Press.

Misiunas, N., Oztekin, A., Chen, Y., & Chandra, K. (2016). DEANN: A healthcare analytics methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status. *Omega*, 58, 46-54.

Örkcü, H. (2013). Subset selection in multiple linear regression models: A hybrid of genetic and simulated annealing algorithms. *Applied Mathematics and Computation*, 219(23), 11018-11028.

Ozanne, E. M., Shieh, Y., Barnes, J., Bouzan, C., Hwang, E. S., & Esserman, L. J. (2011). Characterizing the impact of 25 years of DCIS treatment. *Breast cancer research and treatment*, 129(1), 165-173.

Özekici, S., & Pliska, S. R. (1991). Optimal scheduling of inspections: A delayed Markov model with false positives and negatives. *Operations Research*, 39(2), 261-273.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*(2), 365-375.

Recht, A., Rutgers, E. J., Fentiman, I. S., Kurtz, J. M., Mansel, R. E., & Sloane, J. P. (1998). The fourth EORTC DCIS Consensus meeting (Chateau Marquette, Heemskerk, The Netherlands, 23–24 January 1998)—conference report. *European Journal of Cancer*, 34(11), 1664-1669.

Rosenberg, R. D., Yankaskas, B. C., Abraham, L. A., Sickles, E. A., Lehman, C. D., Geller, B. M., ... & Ballard-Barbash, R. (2006). Performance Benchmarks for Screening Mammography 1. *Radiology*, 241(1), 55-66.

Rosset, S. (2004). Model selection via the AUC. In Proceedings of the twenty-first international conference on Machine learning (p. 89). ACM.

Saarni, S. I., Härkänen, T., Sintonen, H., Suvisaari, J., Koskinen, S., Aromaa, A., & Lönnqvist, J. (2006). The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. *Quality of Life Research*, 15(8), 1403-1414.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 289-310.

Sickles, E. A., D'Orsi, C. J., & Bassett, L. W. (2013). ACR BI-RADS mammography. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology.

Sivagaminathan, R. K., & Ramakrishnan, S. (2007). A hybrid approach for feature subset selection using neural networks and ant colony optimization. *Expert systems with applications*, 33(1), 49-60.

Tabar, L., Yen, M. F., Vitak, B., Chen, H. H. T., Smith, R. A., & Duffy, S. W. (2003). Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *The Lancet*, 361(9367), 1405-1410.

Tarnay, C. (2012). Care of Elder Women. Women's Health Review: A Clinical Update in Obstetrics-gynecology, 413.

Verkooijen, H. M., Fioretta, G. M., Rapiti, E., Bonnefoi, H., Vlastos, G., Kurtz, J., & Bouchardy, C. (2005). Patients' refusal of surgery strongly impairs breast cancer survival. *Annals of surgery*, 242(2), 276.

Wears, R. L., & Lewis, R. J. (1999). Statistical models and Occam's razor. *Academic Emergency Medicine*, 6(2), 93-94.

Weaver, D. L., Rosenberg, R. D., Barlow, W. E., Ichikawa, L., Carney, P. A., Kerlikowske, K., ... & Ballard-Barbash, R. (2006). Pathologic findings from the breast cancer surveillance consortium. *Cancer*, 106(4), 732-742.

Weedon-Fekjær, H., Romundstad, P. R., & Vatten, L. J. (2014). Modern mammography screening and breast cancer mortality: population study. *BmJ*, 348, g3701.

West, D., Mangiameli, P, Rampal, R, & West, V. (2005). Ensemble strategies for a medical diagnostic decision support system: A breast cancer diagnosis application. *European Journal of Operational Research*, 162(2), 532-551.

Wu, C. J., & Hamada, M. S. (2011). Experiments: planning, analysis, and optimization (Vol. 552). John Wiley & Sons.

Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. In Feature extraction, construction and selection (pp. 117-136). Springer US.

Yankaskas, B. C., Taplin, S. H., Ichikawa, L., Geller, B. M., Rosenberg, R. D., Carney, P. A., ... & Barlow, W. E. (2005). Association between Mammography Timing and Measures of Screening Performance in the United States 1. *Radiology*, 234(2), 363-373

Yen, M. F., Tabar, L., Vitak, B., Smith, R. A., Chen, H. H., & Duffy, S. W. (2003). Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *European Journal of Cancer*, 39(12), 1746-1754.

Zelen, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika*, 80(2), 279-293.

Zhang, S., (2011). Modeling the Complexity of Breast Cancer under Conditions of Uncertainty. PhD dissertation, North Carolina State University.

Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., & Inman, B. A. (2012). Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management*, 14(4), 529-547.

4 Personalized Modeling for Assessing Human Papillomavirus (HPV) Vaccination Policies for Women

Abstract

Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States. HPV can cause genital warts and multiple types of cancers on females, including cervical, anal, vaginal, vulvar, and oropharyngeal cancers. In order to prevent future HPV-related diseases, currently, all boys and girls ages 11 or 12 years are recommended to receive the three-dose HPV vaccine. Catch-up vaccines are recommended for males and females through ages 21 and 26, respectively, if they have not been vaccinated previously. In this research, we hypothesize that the cut off age of catch-up vaccine should be determined based on every single woman's risk characteristics, rather than a one-size-fits-all age 26. We model the impact of HPV vaccination at different ages on every individual and track her course of life to estimate the clinical outcomes that resulted from receiving vaccines. With the purpose of providing patient-specific HPV vaccination strategies, especially personalized catch-up vaccination policies, we develop a discrete-event simulation model to evaluate multiple clinical consequences after a woman is vaccinated based on a number of personal attributes. As our simulation model works at the individual level, we build an HPV risk estimation model reflecting every woman's HPV risk, which dynamically changes over time, to support the lifetime simulation model. We estimate the following patient-specific clinical consequences: life time risk of developing HPV-related cancers, life expectancy, life years saved by vaccines. Our study shows that catch-up vaccines still benefit all women after age 26 from the perspective of clinical outcomes. Especially, women facing high risk of HPV infection is expected to gain more health benefits compared with women with low HPV risk. The study reveals that the optimal HPV vaccination catch-up policy should be personalized for different women.

4.1 Introduction

Human papillomavirus (HPV) is the most common sexually transmitted disease in the United States. According to Satterwhite et al.'s estimation (2013), 14,100,000 people are newly infected with HPV each year. Among over 150 identified HPV subtypes, more than 40 HPV types can be transmitted via sexual contacts (Bernard et al. 2010). Although most HPV infections are asymptomatic and often regress on their own spontaneously, some HPV subtypes can cause genital warts and multiple types of cancers for females, including cervical, anal, vaginal, vulvar, and oropharyngeal cancers. In particular, HPV types 16 and 18, result in 70% of cervical cancers, which is the second largest cause of cancer deaths among women worldwide. (Arbyn et al.2011, Lowy and Schiller 2012).

While there is no effective treatment for HPV, HPV infections can be prevented through vaccination. Currently, three vaccines have been approved by the Food and Drug Administration (FDA), including Gardasil (bivalent vaccine), Cervarix (quadrivalent vaccine) and Gardasil 9 (9-valent vaccine). While bivalent HPV vaccine protects against HPV 16 and 18, quadrivalent vaccine protects against HPV 6 and 11 in addition to HPV 16 and 18. 9-valent vaccine, which was newly approved by FDA, provides protection against HPV types of 6, 11, 16, 18, 31, 33, 45, 52, and 58. Although the HPV subtypes covered by the three vaccines vary, all of them provide strong protection against HPV 16 and 18, which account for most HPV-related cancers. The American Cancer Society (ACS) recommends that girls and boys should receive HPV vaccination at age 11 to 12 (Saslow et al. 2016). For those who have not previously been vaccinated, catch-up HPV vaccination is also recommended for females through age 26 and for males through age 21.

Since the introduction and implementation of HPV vaccination, significant reductions in vaccine type infections among women have been reported. Markowitz et al. (2013) found a 56%

decrease in the prevalence of vaccine-type HPV among females aged 14 to 19 years in the 4 years after vaccine introduction. A clinic-based study reported substantial decrease in the prevalence of vaccine-type HPV in both vaccinated and unvaccinated women from a community 4 years after the quadrivalent HPV vaccine was introduced (Kahn et al. 2012). Previous studies also suggested substantial reduction in HPV-related diseases. Ali et al. (2013) reported that there was a 60% reduction in new cases of genital warts. Clinical trials showed that HPV vaccines have very high efficacy for prevention of vaccine-type pre-cancers. Specifically, the efficacy of bivalent and quadrivalent against cervical pre-cancer is higher than 93% (Dunne et al. 2014).

Previous studies evaluated HPV vaccination program at the population level. Most of these studies assessed the cost-effectiveness and clinical outcomes of HPV vaccination program based on simulation models. Sanders and Taira (2003) developed a Markov model to evaluate the cost-effectiveness of vaccinating adolescent girls with high-risk of HPV infections. They assumed a 75% probability of immunity against high-risk HPV infection and found that such a vaccine would result in a gain of 4 quality-adjusted life days at a cost of $246. If all 12-year-old girls in the United States were vaccinated, over 1,300 deaths from cervical cancer would be prevented. Elbasha et al. (2007) assessed the clinical outcomes and cost-effectiveness of quadrivalent vaccines and found that vaccinating girls before age 12 would reduce the incidences of genital warts and cervical cancers by 83% and 78%, respectively. Chesson et al. (2008, 2011) estimated the decrease in the health and economic burden of HPV-related diseases in males and females as a result of HPV vaccination. Their models incorporated cervical, vaginal, vulvar, anal, oropharyngeal, and penile cancers as well as genital warts. Guzzetta et al. (2014) estimated the impact of HPV vaccination on Italian women. Their study also found that catch-up vaccines of 25-year-old women can avert 9.6% of all cervical cancer cases expected in the scenario without

catch-up. Velde et al. (2011) investigated why different HPV vaccination evaluation studies usually produced very inconsistent results. They concluded that differences in the elements of model design, such as natural immunity, partnership duration, HPV types, and waning of vaccine protection, result in significant differences in the estimated effectiveness of HPV vaccines.

However, to the best of our knowledge, there is no study evaluating HPV vaccination at the individual level. As Velde et al. (2011) suggested, differences in some model parameters may result in very different outcomes. It is reasonable to expect that people with different levels of HPV risk have very different clinical outcomes after receiving vaccines. Previous studies identified a number of risk factors that determine the level of HPV risk on different women, including demographic attributes, personal life style and sexual behavior (Winer et al. 2003, Dempsey 2008, Shi et al. 2014). The HPV risk varies dramatically on different individuals. Thus, a personalized evaluation model incorporating individual HPV risk will precisely reveal the different impacts of HPV vaccination on different people.

In addition, although the current HPV vaccination guideline proposed by the ACS does not recommend vaccines for women older than 26, recent studies showed that women over the age of 26 still bear significant HPV risk (Velentzis et al. 2014). In some other countries, HPV vaccination recommendations go beyond age 26. Particularly, in Australia, HPV vaccines are registered for use in females up to 45 years old (Mazza et al. 2014). Some studies showed that HPV vaccines are still beneficial to older women (Muñoz et al. 2009, Schwarz et al. 2009, Westra et al. 2011). Those conclusions were drawn mainly from the efficacy perspective rather than from cost-effectiveness considerations.

In this study, we address the abovementioned problems with two objectives: (1) providing a personalized evaluation model of HPV vaccination, and (2) examining the clinical outcomes of

HPV vaccination for women older than 26. We also seek to synthesize these two problems and determine the cut-off age of HPV catch-up vaccine based on every single woman's risk characteristics. Our study models the impact of HPV vaccination at different ages on every individual, and tracks her course of life to estimate the clinical outcomes that has resulted from receiving vaccines. With the purpose of providing patient-specific HPV vaccination strategies, especially personalized catch-up vaccination policies, we develop a discrete-event simulation model to evaluate multiple clinical consequences after a woman receives vaccines based on a number of personal attributes. As our simulation model works at the individual level, we build an HPV risk estimation model incorporating every woman's HPV risk, which dynamically changes over time, to support the lifetime simulation model. We estimate the following patient-specific clinical consequences: life time risk of developing HPV-related cancers, life expectancy, life years saved by vaccines.

The contribution of our study is two-fold. Firstly, this is the only patient-specific simulation model for the evaluation of HPV vaccination, which provides more practical and accurate decision support for both individual woman and medical decision makers. The model proposes a new framework which is able to reflect the dynamic HPV risk of an individual during the simulated life course. Secondly, our study considers vaccination after age 26, which exceeds the recommended age of catch-up vaccines in the U.S. The results of the study would be instrumental for medical decision makers to rationally determine the catch-up vaccination and potentially amplify the public health benefits of HPV vaccines.

The remainder of the chapter proceeds as follows. Section 4.2 introduces the basic framework of the simulation model and the adjunct HPV risk model. Section 4.3 presents the numerical experiment performed. We introduce the model input, data processing and experiment

results under different scenarios in this section. Section 4.4 provides a discussion on the proposed model and concludes the significance of the study.

4.2 Methodologies

The HPV vaccination evaluation model consists of two sub-models. The main body of the model is a discrete event simulation model that keeps track of every individual woman's course of life, which involves multiple deterministic events and probabilistic events. An HPV risk model dynamically estimates the patient-specific HPV risk of every simulated woman's life course so as to update the likelihoods of probabilistic events while a simulation is running.

4.2.1. Estimation of Clinical Consequences Using Discrete Event Simulation

We use a discrete event simulation to model individual women's course of life in different vaccination scenarios (i.e. varying age at vaccination). The model simulates every woman's life course with the given risk characteristic repeatedly over a planning horizon from the starting age to age 100. The system clock is incremented by a fixed amount of time at each step of the simulation. In this study, we define the increment of time by 1 year. We use t to represent the current time epoch (i.e. a woman's age). We use five main states to represent a woman's health status.

- **Susceptible (S)**: the woman has no immunity against HPV and therefore is susceptible to HPV
- **Immune (I)**: the woman has been vaccinated and therefore is immune to HPV
- **HPV infection (H)**: the woman without immunity is infected with HPV.
- **Cancer (C)**: the woman has developed an HPV-related cancer.
- **Death (D)**: the woman dies from HPA-related cancers or other causes.

A woman's state at time t is denoted by s_t, where $s_t \in \{\mathrm{S}, \mathrm{I}, \mathrm{H}, \mathrm{C}, \mathrm{D}\}$. During the course of a woman's simulated life, her health status will switch between the five states until she enters death or reaches age 100. There are both deterministic events and probabilistic events that may change a woman's state. Table 1 summarizes the main deterministic events and probabilistic events in a woman's simulated life course. Figure 1 shows the five states and the specific events resulting in the transitions between these states. In particular, we assume all women complete 3-dose HPV and acquire full immunity against HPV. As the long-term clinical trials that follow up the HPV vaccines' protection duration are still ongoing and have reported almost persistent efficacy during the whole follow-up period (Romanowski et al. 2016), we also assume that vaccines provide women with lifetime immunity against HPV infection, which is a common practice in previous studies (Elbasha et al. 2007). Thus, once a woman gets vaccinated, she will never enter "HPV infection" or "Susceptible" states. At each step of the simulation, the system determines the occurrences of the events based on the woman's current state and the corresponding likelihoods of probabilistic events. The woman stays in one of the states until an event occurs and changes her state.

Table 1 Main events in the simulation model

Deterministic events	Probabilistic events
▪ The woman will get vaccinated at a specific age.	▪ The woman may or may not be infected with HPV.
▪ She will die at age 100 if not dying at an earlier age.	▪ After the woman is infected with HPV, the HPV may regress spontaneously.
	▪ After the woman is infected with HPV, the HPV may progress to an HPV-related cancer.
	▪ If the woman develops an HPV-related cancer, the cancer may result in death or be successfully treated.
	▪ The woman may die at any age.

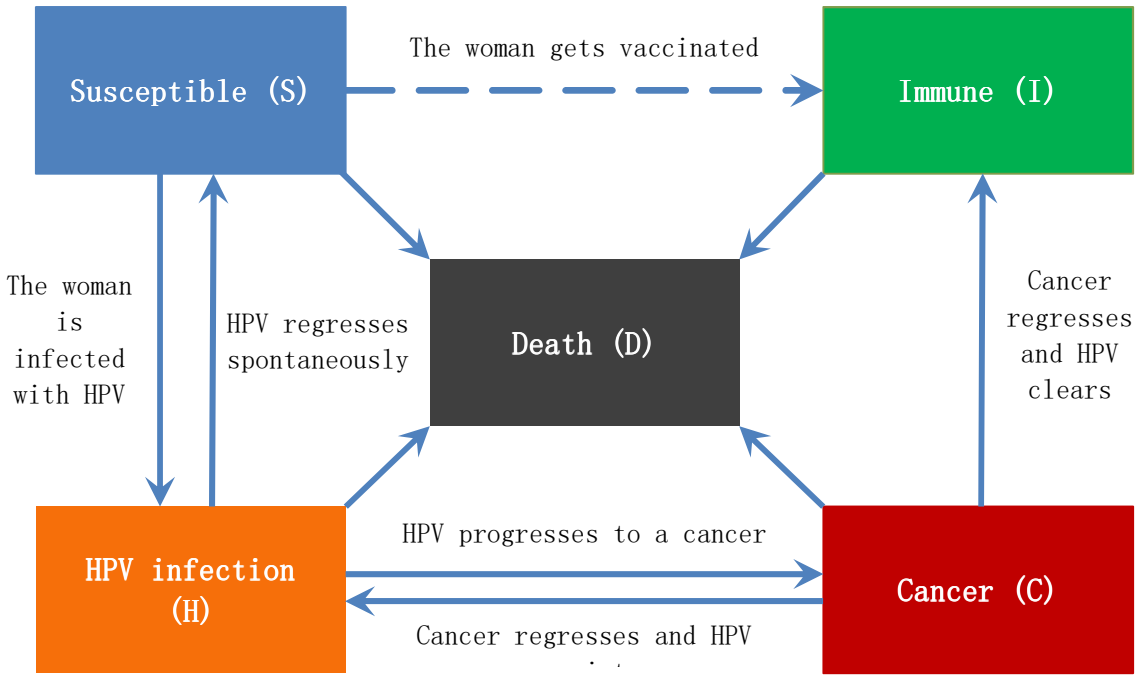


Figure 1 The main structure of the simulation model

Let $P_t(s_{t+1}|s_t)$ denote the transition probability from state s_t to state s_{t+1} for a woman at age *t*. Table 2 summarizes how the transition probabilities are determined.

Table 2 States transitions in the simulation model

Probability	Transition	Value
$P_t(\mathrm{H}\|\mathrm{S})$	The woman is infected with HPV.	Patient-specific annual HPV incidence rate
$P_t(\mathrm{S}\|\mathrm{H})$	HPV regresses spontaneously.	One-year regression rate of HPV
$P_t(\mathrm{D}\|\mathrm{H})$	The infected woman dies from reasons other than HPV-related cancers.	One-year death rate (excluding HPV-related cancers)
$P_t(\mathrm{C}\|\mathrm{H})$	HPV progresses to an HPV-related cancer.	Cancer-specific one-year progression rate of HPV
$P_t(\mathrm{D}\|\mathrm{C})$	The woman dies from any reasons, including HPV-related cancer.	Cancer-specific one-year death rate (including other reasons)
$P_t(\mathrm{D}\|\mathrm{I})$	The uninfected woman dies from reasons other than HPV-related cancers.	One-year death rate (excluding HPV-related cancers)
$P_t(\mathrm{D}\|\mathrm{S})$	The uninfected woman dies from reasons other than HPV-related cancers.	One-year death rate (excluding HPV-related cancers)
$P_t(\mathrm{I}\|\mathrm{S})$	The woman gets vaccinated.	100% at a specified age 0% at other ages
$P_t(\mathrm{H}\|\mathrm{C})$	The HPV-related cancer regresses. HPV persists.	HPV persistent rate after cancer treatment or cancer regression
$P_t(\mathrm{I}\|\mathrm{C})$	The HPV-related cancer regresses. HPV clears.	1 - HPV persistent rate (excluding death in one year)

Different from other states, “cancer” is a complex state that consists of several sub-states representing the development, diagnosis and treatment of HPV-related cancers. Figure 2 illustrates the subsequent process after a woman’s HPV develops to HPV-related cancers. We use multiple sub-states to differentiate detected and undetected HPV-related cancers as well as different stages of the cancers, as they have different death rates. The simulation clock is still incremented by one year at each step after a woman enters the sub-states of the “cancer” state. When a woman completes the transitions from “HPV infection” state to “cancer” state, she will immediately be assigned to “pre-cancer/ in situ cancer” state. Then the woman’s pre-cancer or in situ cancer may progress to a invasive cancer or be diagnosed. The state transitions in the sub-structure are similar to the transitions in the main structure of the simulation model.

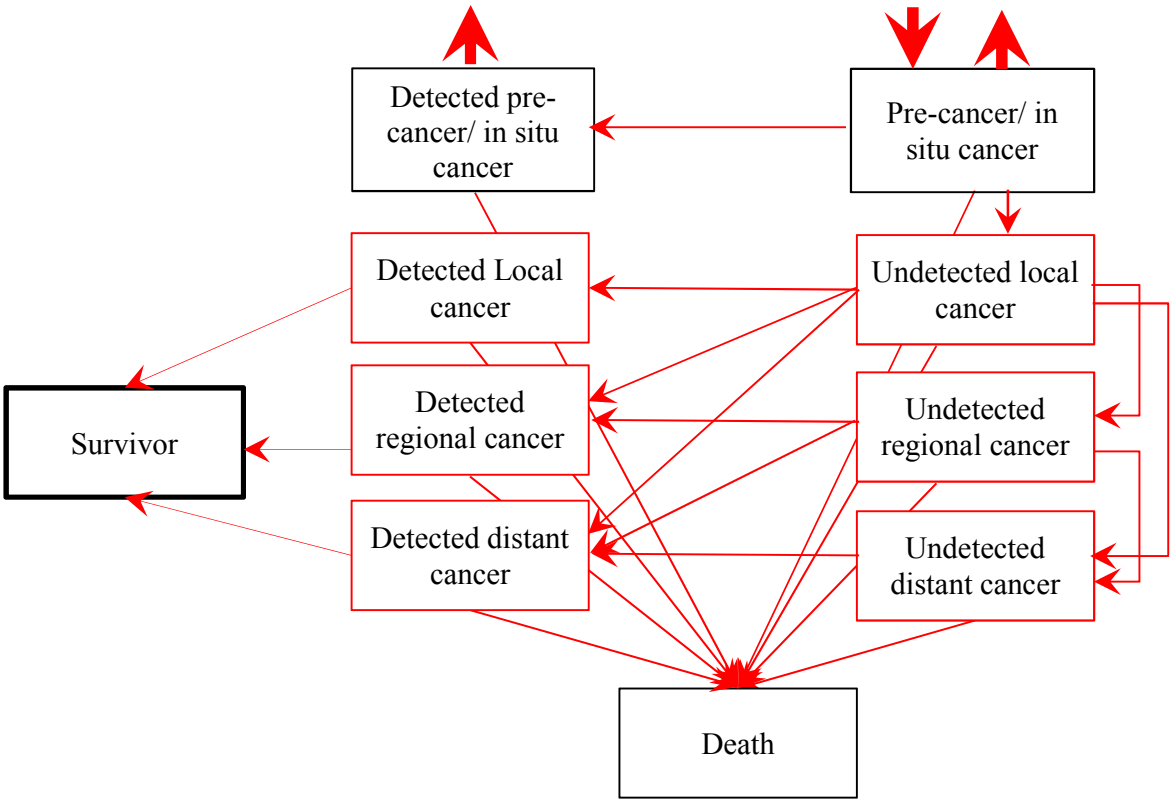


Figure 2 The sub-structure of the “cancer” state

Every year, a simulated woman's health status may stay at the same state or change to another state with a certain possibility. The woman's status switches between the different states until she enters an absorbing state (i.e. "survivor" or "death") or leave the sub-structure (i.e., two outgoing arrows from the "pre-cancer/ in situ cancer" states).

We take cervical cancer as an example to explain the detailed simulation process after a woman's HPV progresses to HPV-related cancer. Every year, an HPV infection may naturally progress to HPV-related cancers with a certain probability. The first status after the progression is the pre-cancer or in situ cancer stage. For cervical cancer, this status includes three states: cervical intraepithelial neoplasia (CIN) 1, CIN 2 and CIN 3, which are cervical cancer's non-malignant pre-cancer stages with the propensities of regression and progression (Schiffman et al. 1993). These pre-cancer stages may be detected by routine cervical cancer screening (i.e., Pap test) with a certain probability every year. Once detected, the CINs will be treated. Then the simulated woman will re-enter "HPV infection" state or "Immune" state in Figure 1, depending upon whether the HPV infection persists or not. During the year, the woman may also die from another causes other than cervical cancer. If the CINs are not detected, the woman's status may stay the same, regress to "HPV infection", regress to "Immune", or progress to a real cancer, i.e. local cervical cancer. When the woman enters "undetected local cancer" state, similarly, the local cancer has potentials of evolving to a regional cervical cancer and being detected. Since the woman has cervical cancer, there is a certain possibility for her to develop cancer-related symptoms, which finally results in diagnosis and treatment of the cancer. The combination of the symptom development rate and cervical cancer screening rate is the transition probability from undetected local cancer to detected local or regional cancer. If the cancer is detected, the woman will enter the "detected local cancer" state and end up with "survivor". The "survivor" state

makes the simulated woman quit the simulation and be assigned with a lump-sum life expectancy based on her specific cancer stage. A similar process is true for regional cancer and distant cancer.

In addition to cervical cancer, HPV has been identified as an important cause of four other cancers of women, including anal cancer, vaginal cancer, vulvar cancer and oropharyngeal cancer (Chaturvedi 2010). The simulation model takes into account the five HPV-related cancers. We assume that a woman will never develop multiple HPV-related cancers simultaneously, as the incidence of synchronous primary cancers of the female genital tract is very rare (Tong et al. 2008). An HPV infection may progress to one of the five cancers with a certain probability every year. Once a woman develops an HPV-related cancer, she is temporarily free from the other cancers.

The goal of the simulation is to estimate patient-specific clinical consequences in different vaccination scenarios (i.e. varying age at vaccination) and with given individual HPV risk characteristics. A woman's life course is simulated for multiple times to derive the average values of life time risk of developing HPV-related cancers and life expectancy. By changing the age at vaccination and performing the simulation on women with different HPV risk, we expect to receive dramatically different outcomes of these metrics.

4.2.2. HPV Risk Model

In the simulation model, individual HPV risk is embodied in the transition possibility $P_t(\mathrm{H}|\mathrm{S})$, which is from "Susceptible" to "HPV infection". Different women's possibilities of being infected with HPV at every age are estimated based on their personal risk characteristic. Previous studies have identified several behavioral and demographic risk factors of HPV (Moscicki et al. 1990, Winer et al. 2003, Dempsey 2008, Shi et al. 2014, Reiter and McRee 2016). We use a

number of identified HPV risk factors to build a penalized regression model to estimate the personal HPV risk for every individual, including demographic attributes, personal life style and sexual behaviors.

Table 3 Candidate HPV risk factors

Demographic attributes	Personal life style	Sexual behaviors
Age	Recent alcohol use	Age at first sex
Race	Recent tobacco use	Lifetime number of sex partners
Marital status		Number of recent sex partners
Ratio of family income to poverty		Sex orientation
Age at first menarche		
Parity		

Table 3 summarizes the twelve candidate variables that are used to build the HPV risk model. We use these risk factors to determine every individual's HPV risk at different ages. The risk factors are employed as the predictors of the logistic regression model to estimate the HPV risk. Since HPV 16 and 18 are responsible for most HPV-related cancers and preventable by the three approved vaccines, we only take these two HPV subtypes into account and treat the incidence of HPV 16 and 18 as a single binary variable. Thus, the response variable is whether a woman has HPV 16 or 18 at her current age.

However, it might not be necessary to include all available predictors in the regression analysis. Firstly, we need a parsimonious model – some risk factors are dynamic and therefore intractable to be tracked over time, such as number of sex partners. The change of a variable in a woman's life time needs to be fully modeled. Therefore, a simple model with as few predictors as possible would greatly reduce the difficulty of data preparation. Secondly, multicolinearity may exist in the model, as many variables are inherently correlated. For instance, a woman's marital status is naturally correlated with the number of her recent sex partners. Thirdly, incorporating too many variables may result in overfitting, which leads to a poor out-of-sample predictive power. Thus, we choose lasso regression to perform variable selection for the logistic

regression as it penalizes multicollinearity by removing correlated variables while pursuing a good predictive performance. The optimal tuning penalty parameter (lambda) is determined by a10-fold cross-validation. Optimality is evaluated based on the area under curve (AUC) since the data are highly unbalanced (i.e., observations with response variables equal to 0 make up the vast majority of the data).

In order to be easily tracked over time, all numerical candidate variables are converted to categorical variables, which means regular lasso regression may have undesirable results. Firstly, lasso regression makes the matrix of predictors sparse—every categorical variable is broken down to multiple dummy variables, resulting in very high column dimension. Secondly, lasso regression often partially selects dummy variables—it is not reasonable to select only a portion of dummy variables from one categorical predictor. Especially, lasso regression may not be able to reduce the difficulty of data collection. It may only remove part of the dummy variables derived from a predictor, which means the predictor still needs to be considered and fully prepared in the simulation model.

Therefore, group lasso (Yuan and Lin 2006, Meier et al. 2008) is an appropriate choice of penalization method for multicollinearity in this case. The group lasso does variable selection on predefined groups of variables in regression models, which allow us to group all dummy variables from a categorical variable and manipulate them at the same time. In a typical logistic regression set-up, assume that we have a binary response variable $y_i \in \{0,1\}$ and G independent variables $x_i = (x_{i,1}^{\mathrm{T}}, \ldots, x_{i,G}^{\mathrm{T}})^{\mathrm{T}}$ with $x_{i,g} \in \mathbb{R}^{\mathrm{df}_g}$, where $g = 1, \ldots, G$. And df_g is the degree of freedom. In the context of this study, the categorical variable with three levels has a $\mathrm{df}_g = 2$, while a numerical variable has $\mathrm{df}_g = 1$. The conditional probability $p_\beta(x_i) = p_\beta(Y = 1 | x_i)$ is modeled by

$$\log\left\{\frac{p_\beta(x_i)}{1-p_\beta(x_i)}\right\} = \eta_\beta(x_i), \quad (1)$$

with $\eta_\beta(x_i) = \beta_0 + \sum_{g=1}^{G} x_{i,G}^{\mathrm{T}}\,\beta_g,$

where β_0 is the intercept; β_g is the coefficient vector of the gth independent variable with a degree freedom of df_g. So the estimator $\hat{\beta}_\lambda$ is calculated by minimizing the convex function

$$S_\lambda(\beta) = -l(\beta) + \lambda \sum_{g=1}^{G} s(\mathrm{df}_g)\left\|\beta_g\right\|_2, \quad (2)$$

where $l(\beta) = \sum_{i=1}^{n} y_i \eta_\beta(x_i) - \log(1 + \mathrm{e}^{\eta_\beta(x_i)})$, which is the log-likelihood function of β (Meier et al. 2008). λ is a non-negative tuning parameter that controls the strength of penalization. $s(\mathrm{df}_g) = \sqrt{\mathrm{df}_g}$ is used to rescale the penalty with respect to the dimensionality of the coefficient vector β_g and ensure that the penalty is of the order of df_g. For a logistic regression with categorical variables, all dummy variables from the same categorical variable are grouped, so they are either entirely kept or discarded during the variable selection process.

We adopt the logistic regression penalized by the group lasso method to build the HPV risk model and use the HPV risk as the input for the transition possibility $P_t(\mathrm{H}|\mathrm{S})$, which represents the change from the "susceptible" state to the "HPV infection" state. Then the simulation is personalized for different women and generates patient-specific clinical outcomes.

4.3 Numerical Experiments

4.3.1. Parameter Estimation For Simulation Model

Table 4 summarizes the data sources that are used to estimate the transition probabilities between the main states in the simulation model. It is worth noting that our HPV risk model actually outputs the prevalence of HPV associated with people having specific risk characteristics. For initial state distribution of every individual woman at the beginning of the simulation, we use the

HPV risk model to estimate the prevalence of HPV and assign the woman to either "Susceptible" or "HPV infection" states according to the calculated prevalence. The epidemiologic equation $\text{Prevalence} = \text{Incidence} \times \text{Duration}$ (Aschengrau and Seage 2013) is used to calculate $P_t(H|S)$, which is the incidence associated with a specific risk characteristic. The 1.2 years duration of oncogenic HPV 16/18 infection is adopted (Elbasha et al. 2007).

Table 4 Sources of data input for transition probabilities

Transition Probability	Source for Parameter Estimation
$P_t(H|S)$	refer to the HPV risk model in Section 4.3.2
$P_t(S|H)$	Sanders and Taira 2003
$P_t(D|H)$	Arias et al. 2016
$P_t(C|H)$	Elbasha et al. 2007, Chesson et al. 2008
$P_t(D|C)$	Adriano et al. 2003, Elbasha et al. 2007, Arias et al. 2016
$P_t(D|I)$	Arias et al. 2016
$P_t(D|S)$	Arias et al. 2016
$P_t(I|S)$	100% at specified vaccination age, and 0% at other ages
$P_t(H|C)$	Elbasha et al. 2007
$P_t(I|C)$	Elbasha et al. 2007

This study considers five HPV-related cancers, including cervical cancer, anal cancer, vaginal cancer, vulvar cancer and oropharyngeal cancer. Therefore, there should be five different sub-structures of "cancer" state, each of which represent one of the five cancers. For the annual progression rates, annual regression rates, annual symptom development probabilities, annual screening rate and stage-specific annual death rates of cervical cancer, we use the data reported by Sanders and Taira (2003) as well as Elbasha et al. (2007). For the stage-specific lump-sum life expectancy of the absorbing state "survivor", we use the DEALE method based on the 5-year survival of invasive cervical cancer (Beck et al. 1986, Howlader et al. 2013).

However, unlike cervical cancer, there are relatively limited epidemiological data on the nature history, diagnosis and prognosis of anal cancer, vaginal cancer, vulvar cancer and oropharyngeal cancer. Due to the low incidences rates of these four cancers, most prior studies on HPV vaccination evaluation only took cervical cancer into consideration and focused on HPV

vaccination's health outcomes from the perspective of cervical cancer prevention (Sanders and Taira 2003, Elbasha et al. 2007, Ley-Chavez 2012, Guzzetta et al. 2014). Therefore, our study simplifies the four HPV-related cancers' impacts of the four HPV-related cancers on women's life expectancy by introducing the loss of life years associated with different cancers reported by Chesson et al. (2008), rather than use the sub-structures of "cancer" state to simulate the outcomes of these cancers. Once a woman develops one of the four cancers, her life expectancy is deducted by a certain amount of life years (Chesson et al. 2008), which are not only cancer-specific but also age-specific.

4.3.2. Parameter Estimation for HPV Risk Model

We use the National Health and Nutrition Examination Survey (NHANES) 2011-2012 data to build the individual HPV risk model. 13,431 people from 30 different study locations across the US participated in the survey. The publicly available datasets include data for people aged 18-59 years. The NHANES data report a large number of personal information of participants, including demographic, socioeconomic, dietary, and health-related information as well as medical laboratory exam results, which cover all the variables summarized in Table 3. Especially, the NHANES 2011-2012 include exam results of multiple HPV subtypes. It provides us an access to evaluate the risk of HPV infection in different groups.

However, one critical issue of NHANES data that needs to be addressed is its large number of missing data. Although 13,431 people were included in the survey, only 1,767 observations include HPV 16/18 data for participants. It is noteworthy that, among the 1,767 observations, only 247 subjects have complete data in all the twelve candidate variables. If we use the simple listwise deletion to deal with missing values (i.e. as long as there is at least one variable with missing value, the whole observation is removed), the deletion procedure would discard around

86% of the observations, which results in a considerable loss of data and potential bias. Following the previous studies that dealt with a large number of missing values in training data (Barlow et al. 2006, Shi et al. 2014), we treat the missing values in category variables as a single level "unknown" and only remove the observations with unknown or missing values in numerical variables (e.g. age).

The infection of HPV 16/18 (i.e. positive exam result) is modeled as the response variable in the regression analysis. We also exclude all women who have been vaccinated before they took the surveys. We conduct the group lasso variable selection on the twelve candidate variables and find the best combination of HPV risk factors.

Since the study simulates every woman's lifetime, our model takes into account the change of every individual's risk characteristic. Among the twelve candidate risk factors, eight of them are dynamic and change over time, including age, marital status, ratio of family income to poverty, recent alcohol use, recent drug use, parity, lifetime number of sex partners, and number of recent sex partners. Age is simply increased by 1 with every increment of the system clock. For the remaining seven dynamic variables, longitudinal data are needed to track their changes over a long period of time. Moreover, the relationships between HPV risk and some variables are not necessarily linear. In order to simplify the tracking process, we convert all numerical dynamic variables to categorical variables. Specifically, "age" is categorized by 5-year age groups, starting with group "20-24". We model "ratio of family income to poverty" as a binary variable with 1 indicating that the ratio is lower than or equal to 1, and 0 indicating that the ratio is higher than 1. "Parity" is also converted to a binary variable with 1 indicating that the woman has had a live birth, and 0 indicating that the woman has no live birth. As heavy drinking is defined as 8 or more drinks a week for women (U.S. Department of Health and Human Services

2015), we convert recent alcohol use to a binary variable with 1 indicating the woman drink more than one alcoholic beverages on average every day in the past 12 months, and 0 indicating not more than one drink every day. For "lifetime number of sex partners", we use five levels, "=0", "1","2-4","5-9",">=10", and "unknown", to categorize the variable. "Number of recent sex partner" is categorized by "0", "1", ">=2" and "unknown". In addition, although "age at first menarche" and "age at first sex" are not dynamics variables, in order to incorporate missing values, we also categorize them using five levels, "<=12", "13-15", "16-19",">=20" and "unknown".

Table 5 Model selection result

Candidate variable	Original data type	Property	Selected
Age	Numerical	Dynamic	Yes
Race	Categorical	Static	Yes
Marital status	Categorical	Dynamic	Yes
Ratio of family income to poverty	Numerical	Dynamic	No
Age at first menarche	Numerical	Static	No
Parity	Numerical	Dynamic	Yes
Recent alcohol use	Numerical	Dynamic	No
Recent tobacco use	Numerical	Dynamic	Yes
Age at first sex	Numerical	Static	Yes
Lifetime number of sex partners	Numerical	Dynamic	No
Number of recent sex partners	Numerical	Dynamic	Yes
Sex orientation	Categorical	Static	No

The entire analysis is performed in R 3.32. Using the group lasso logistic regression, seven variables are selected (Table 5). Among the seven selected variables, age, race, age at first menarche and age at first sex are either static or uniformly increased. In contrast, the values of marital status, recent tobacco use, parity and number of recent sex partners are fluctuating and need to be modeled throughout every woman's whole life. During the simulation, if one of these variables has any changes, the HPV risk is also updated accordingly. We use multiple data sources to model these dynamic variables as follows.

Marital status

We model women's changes of marital status based on the method proposed by Ley-Chavez (2012). Five statuses are considered in the model, including first marriage, first divorce, second marriage, second divorce and never-married. The duration of each status is modeled by a discrete distribution depending on age. The basic process of modeling marital status proceeds as follows. The simulation randomly assigns an age of first marriage to women based on the probability distribution estimated from the Survey of Income and Program Participation (U.S. Census Bureau 2015). The marriage will last for a certain amount of time or a lifetime, based on the probability distribution of 1st marriage duration estimated using the National Survey of Family Growth (Centers for Disease Control and Prevention data 2002) and the Survey of Income and Program Participation (U.S. Census Bureau (2015). Once a woman divorces, her length of divorce status is determined by a probability distribution reported by the Survey of Income and Program Participation. A divorced woman may get married for the second time. In addition, the duration of a woman's second marriage is determined by a probability distribution, using the data of the Survey of Income and Program Participation. We do not model third and higher-order marriages as they are very rare (Ley-Chavez 2012).

Recent tobacco use

For women's smoking behavior, women are classified into two groups: smokers or non-smokers. The likelihood of a non-smoker turning into a smoker is directly estimated using the NHANES data, which also report participants' ages at starting smoking cigarettes regularly. If a woman becomes a smoker, she may quit smoking at any age based on an age-specific probability of quitting smoking. We use the recent smoking cessation rate reported by the Centers for

Disease Control and Prevention (2011) to evaluate the yearly likelihood of a smoker to quit smoking.

Number of sex partners

The yearly changes of the number of sex partners is modeled based on the method proposed by Ley-Chavez (2012). The number of a woman's acquired new sex partner every year is categorized into three groups: 0, 1, 2 and "3 or more". Women in the category "3 or more" are assigned 3 partners in the simulation. In this study, unmarried women include the following status: "widowed", "divorced", "separated" and "never married", while married women include "married" and "living with partner". The probability of unmarried women having 0 to 3 or more new partners each year is modeled by an age-specific discrete distribution. So an unmarried woman is randomly assigned to one of the four categories according to the discrete distribution based on her age group. For married woman, we assume the probability of getting a new partner in one year is 0.06 (O'Dowd 2003).

Parity

The parity in this study is defined by a binary variable indicating if a woman has had live birth at her current age. For those who had no live birth at their current ages, they may expect a birth in the future. The birth rates of 1st child are used to probabilistically determine when a woman will have her first child (Martin et al. 2017). The birth rate is grouped by three factors: age, marital status and race. As the reported birth rates of 1st child do not distinguish marital statuses, we assume the percentage of 1st births to unmarried women (among all women) is the same as the percentage of all births to unmarried women, which is reported to be 40.3% (Martin et al. 2017). Then the birth rates of 1st child by age, marital status and race are estimated based

on the three parameters: the percentages of 1st births to unmarried women, the percentage of unmarried woman and overall birth rates of 1st child.

4.3.3 Design of Numerical Experiments

Based on the HPV risk model, two typical risk characteristics are identified using the selected variables: a high risk woman, who has the highest HPV incidence; and a low risk woman, who has the lowest HPV incidence (Table 6). We set multiple vaccination scenarios for the two women, that is, the two patients are vaccinated at age 20, 26, 30, 35, 40 and 45, respectively. The age at vaccination is also the starting age of the simulation. The simulation is run for the two women starting from age 20 under each of the scenarios for 10,000 times. Then the average values of the four metrics are reported for each woman under different scenarios, including life expectancy, lifetime risk of developing cervical cancer, and lifetime risk of developing an HPV-related cancer.

Table 6 Simulation scenarios

Candidate variable	High risk woman	Low risk woman
Age at vaccination	20, 26, 30, 35, 40, 45	20, 26, 30, 35, 40, 45
Race	Black	Asian
Marital status	Unmarried	Married
Parity	Yes	No
Recent tobacco use	Yes	No
Age at first sex	<=12	>=20
Number of recent sex partners	>=3	0

As a contrast, we also use the general female population's HPV incidences to run the simulation under the same 6 scenarios. The contrast demonstrates how the clinical outcomes differ between a high risk woman, a low risk woman and an average woman. When simulating an average woman, we do not use HPV risk model to dynamically update $P_t(\mathrm{H}|\mathrm{S})$, but use the age-specific incidences reported by Myers et al (2000). Table 7 summarizes the annual incidences of HPV among the three different women. The high risk and low risk women's incidence rates are directly estimated using the NHANES data.

Table 7 HPV incidences of three groups

Age at vaccination	High risk woman	Low risk woman	Average woman*
20	27.1%	4.0%	15.0%
26	24.2%	1.1%	5.0%
30	23.5%	0.4%	1.0%
35	23.4%	0.3%	1.0%
40	23.2%	0.1%	1.0%
45	23.2%	0.1%	1.0%

* Average woman's annual incidences are collected from Myers et al. (2000)

4.3.4 Results of Numerical Experiments

Since the incidences of the HPV-related cancer among teenagers are very low (Howlader et al. 2013), we assume no woman has HPV-related cancer before age 20 and only consider every woman's HPV-related cancer from age 20 to death. The simulation outputs the average values of the gain in life expectancy, lifetime risk of developing cervical cancer and lifetime risk of HPV-related cancers for 10,000 runs. The gain in life expectancy is equal to the difference of life expectancy between a woman receiving vaccines at a specific age and the same woman who will never receive the vaccine. The lifetime risk of HPV-related cancers reflects the impact from the HPV-infection before a woman gets vaccinated. Therefore, the later a woman receives vaccines, the higher risk of developing HPV-related cancers she faces.

Table 8 shows the three types of clinical outcomes of vaccinations at different ages. The clinical outcomes show dramatic differences among the women with different level of HPV risk. The gain in life expectancy for high risk patient is over 10 times higher than that for the low risk woman. However, no matter how high the risk is, receiving catch-up vaccines almost always benefit women. The high risk woman has the highest gain in life expectancy when receiving vaccines at age 20, which amounts to 64 days (0.176 year) on average. When getting vaccinated at the same age, the average woman improves her life expectancy by 12 days (0.032), while the low risk woman only gains 4 additional days in her life expectancy. The current HPV

vaccination policy only recommends catch-up vaccines for women up to age 26, but we still see significant improvements in life expectancy on women older than 26, especially the woman bearing a high HPV risk, which suggests that catch-up vaccine after age 26 should be deliberately considered.

Table 8 Clinical outcomes of HPV vaccination at different ages

Gain in life expectancy (in years)

Age at vaccination	**High risk woman**	**Low risk woman**	**Average woman**
20	0.176	0.012	0.032
26	0.113	0.004	0.019
30	0.075	0.002	0.015
35	0.043	0.001	0.014
40	0.040	0.001	0.012
45	0.031	0.000	0.011

Lifetime risk of developing cervical cancer

Age at vaccination	**High risk woman**	**Low risk woman**	**Average woman**
20	0.42%	0.00%	0.11%
26	0.76%	0.01%	0.23%
30	0.85%	0.01%	0.32%
35	0.90%	0.03%	0.37%
40	1.03%	0.03%	0.40%
45	1.10%	0.05%	0.45%

Lifetime risk of developing HPV-related cancer

Age at vaccination	**High risk woman**	**Low risk woman**	**Average woman**
20	0.91%	0.01%	0.24%
26	1.44%	0.02%	0.47%
30	1.73%	0.02%	0.65%
35	1.94%	0.06%	0.80%
40	2.12%	0.06%	0.82%
45	2.21%	0.09%	0.91%

Although the average gain in life expectancy from receiving catch-up vaccines after age 26 is numerically modest even for the high risk individual, the potentially averted HPV-related cancer cases can aggregate to substantial numbers on a group of women with the same risk profile. For instance, the high risk woman vaccinated at age 30 has a lifetime risk of developing cervical cancer equal to 0.85%. If the catch-up vaccine is delayed to age 45, the risk rises to 1.1%. The

increased risk (i.e. 0.26%) is equivalent to a significant number of cancer cases when applying to the population with the same high risk characteristics. For instance, assuming there are 100,000 women having the high risk profile, if we set the vaccination at age 45 as a baseline, the catch-up vaccination at age 26 would prevent 770 more cancer cases compared with the vaccination at age 45. However, the vaccination as late as age 30 still could prevent 260 cervical cancer cases compared with baseline age 45. In contrast, the increased risk on low risk woman is relatively low, implying the necessity of personalizing the catch-up vaccination policy.

4.4 Discussion

This study showcases a novel vaccination evaluation framework, which combines a discrete-event simulation with a dynamically updated regression-based risk model. Specifically, we show how to address dynamic risk factors and model these factors over every woman's lifespan. The modeling process involves a lot of data from a variety of sources and generates reasonable results. The same approach may be applied to the clinical outcome evaluations on the vaccines for other diseases such as hepatitis B.

What makes the study different from the prior studies on this topic is that the model works at the individual level rather than the population level. We evaluate the impact of HPV vaccines on different women and reveal that the catch-up vaccines after age 26 are still very beneficial for women, especially for those with high HPV risk. Our study finds that the gain in life expectancy for a high risk woman at age 45 is almost as high as that for an average woman if vaccinated at age 20. These results prove that, from a pure health outcomes perspective, the HPV vaccination policy should be personalized, rather than set at a one-size-fits-all cutoff age. Our results are consistent with many previous arguments that support providing catch-up HPV vaccines to older women (Muñoz et al. 2009, Schwarz et al. 2009, Westra et al. 2011). We also confirm that

Australian's HPV vaccination policy, which provides vaccines to women up to 45 years old, is also rational for some US women.

Table 9 Main assumptions in the study

Assumptions	Justifications and references
▪ All women complete the 3-dose HPV vaccination and acquire full immunity against HPV.	Chesson et al. 2008, Dunne et al. 2014
▪ Vaccines provide women with lifetime immunity against HPV infection.	Elbasha et al. 2007, Chesson et al. 2008
▪ A woman will never develop multiple HPV-related cancers simultaneously.	Tong et al. 2008
▪ Women who have not married by age 50 will never get married.	Ley-Chavez 2012
▪ We assume no woman has HPV-related cancers before age 20 and only consider every woman's HPV-related cancers from age 20 to death.	Howlader et al. 2013

Our analysis also has some limitations. Firstly, due to the data scarcity, the diagnosis and prognosis of the HPV-related cancers other than cervical cancer are not sufficiently modeled. The simplified method may result in some inaccuracy in our final outcomes. The model can be improved when the corresponding data are available. Secondly, we made several assumptions in the parameters estimation (see Table 9). For instance, we assume the vaccines' duration of protection against HPV is lifelong. Currently, the long-term clinical trials that follow up the HPV vaccines' protection duration are still ongoing and only report 10 years effective duration (Romanowski et al. 2016). Once the number is updated, the model may generate different outcomes. In addition, we also assume that the efficacy of HPV is 100%. However, the vaccines' efficacy range between 90% and 100% by vaccine types and number of vaccine shots (Dunne et al. 2014).

In the future, a sensitivity analysis on the HPV efficacy and protection duration will be conducted. Our study may also be extended to males and examines how men's vaccination ages impact their spouses or partners' health. Moreover, our current study only considers life-threating cancers related to HPV. In the future study, we may also incorporate HPV-related non-cancer diseases (e.g. genital warts) to evaluate HPV vaccinations policy's impact on quality-adjusted life years (QALYs). This extension will show how different types of vaccines, especially those against HPV subtypes other than 16 and 18, impact women's health.

Acknowledgements

This work was funded by the Women's Giving Circle of University of Arkansas.

References

Adriano, E., Jagoe, J. M., Harrison, T., Riffenburgh, R. H., & Johnstone, P. A. (2003). Survival of patients with untreated cervical carcinoma. American journal of clinical oncology, 26(4), 369-373.

Ali, H., Donovan, B., Wand, H., Read, T. R., Regan, D. G., Grulich, A. E., ... & Guy, R. J. (2013). Genital warts in young Australians five years into national human papillomavirus vaccination programme: national surveillance data. Bmj, 346, f2032.

Arbyn, M., Castellsagué, X., de Sanjosé, S., Bruni, L., Saraiya, M., Bray, F., & Ferlay, J. (2011). Worldwide burden of cervical cancer in 2008. Annals of oncology, 22(12), 2675-2686.

Arias, E., Heron, M., & Xu, J. (2016). United States Life Tables, 2012. National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 65(8), 1.

Aschengrau, A., & Seage, G. R. (2013). Essentials of epidemiology in public health. Jones & Bartlett Publishers.

Barlow, W. E., White, E., Ballard-Barbash, R., Vacek, P. M., Titus-Ernstoff, L., Carney, P. A., ... & Yankaskas, B. C. (2006). Prospective breast cancer risk prediction model for women undergoing screening mammography. Journal of the National Cancer Institute, 98(17), 1204-1214.

Beck, J. R., Pauker, S. G., Gottlieb, J. E., Klein, K., & Kassirer, J. P. (1982). A convenient approximation of life expectancy (the "DEALE"): II. Use in medical decision-making. The American journal of medicine, 73(6), 889-897.

Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 2010;401:70–9.

Bramlett, M. D., & Mosher, W. D. (2002). Cohabitation, marriage, divorce, and remarriage in the United States. Vital health statistics, 23(22), 1-32.

Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1979 cohort, 1979-2010 (rounds 1-24). Produced and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2012.

Centers for Disease Control and Prevention (CDC). (2002). National Survey of Family Growth.

Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data 2011-2012. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, [2016] [https://www.cdc.gov/nchs/nhanes/nhanes2011-2012/overview_g.htm].

Centers for Disease Control and Prevention (2011) Quitting smoking among adults—United States, 2001-2010. MMWR Morb Mortal Wkly Rep 60: 1513–1519.

Chaturvedi, A. K. (2010). Beyond cervical cancer: burden of other HPV-related cancers among men and women. Journal of Adolescent Health, 46(4), S20-S26.

Chesson, H. W., Ekwueme, D. U., Saraiya, M., & Markowitz, L. E. (2008). Cost-effectiveness of Human Papillomavirus Vaccination in the United States. Emerging Infectious Diseases, 14(2), 244-251.

Chesson, H. W., Ekwueme, D. U., Saraiya, M., Dunne, E. F., & Markowitz, L. E. (2011). The cost-effectiveness of male HPV vaccination in the United States. Vaccine, 29(46), 8443-8450.

Dempsey, A. F. (2008). Human papillomavirus: the usefulness of risk factors in determining who should get vaccinated. Reviews in Obstetrics and Gynecology, 1(3), 122.

Doorbar J, Quint W, Banks L, et al. The biology and life-cycle of human papillomaviruses. Vaccine 2012;30(Suppl 5):F55–70.

Dunne, E. F., Markowitz, L. E., Saraiya, M., Stokley, S., Middleman, A., Unger, E. R., ... & Iskander, J. (2014). CDC grand rounds: reducing the burden of HPV-associated cancer and disease. MMWR Morb Mortal Wkly Rep, 63(4), 69-72.

Elbasha, E. H., Dasbach, E. J., & Insinga, R. P. (2007). Model for assessing human papillomavirus vaccination strategies. Emerging infectious diseases, 13(1), 28.

Guzzetta, G., Faustini, L., Panatto, D., Gasparini, R., & Manfredi, P. (2014). The impact of HPV female immunization in Italy: model based predictions. PloS one, 9(3), e91698.

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2009). US Census Bureau.

Herriot, R. A., & Kasprzyk, D. (1984). Survey of Income and Program Participation.

Howlader, N., Noone, A. M., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S. F., ... & Cho, H. (2013). SEER cancer statistics review, 1975–2010. 2013. Bethesda, MD: National Cancer Institute.

Kahn, J. A., Brown, D. R., Ding, L., Widdice, L. E., Shew, M. L., Glynn, S., & Bernstein, D. I. (2012). Vaccine-type human papillomavirus and evidence of herd protection after vaccine introduction. Pediatrics, 130(2), e249-e256.

Lowy, D. R., & Schiller, J. T. (2012). Reducing HPV-associated cancer globally. Cancer Prevention Research, 5(1), 18-23.

Ley-Chavez, A. (2012). Quantitative Models to Design and Evaluate Risk-Specific Screening Strategies for Cervical Cancer Prevention (Doctoral dissertation, The Ohio State University).

Markowitz, L. E., Hariri, S., Lin, C., Dunne, E. F., Steinau, M., McQuillan, G., & Unger, E. R. (2013). Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003–2010. Journal of Infectious Diseases, jit192.

Martin, J. A., Hamilton, B. E., Osterman, M. J., Driscoll, A. K., & Mathews, T. J. (2017). Births: Final Data for 2015. National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, 66(1), 1.

Mazza, D., Petrovic, K., Grech, C., & Harris, N. (2014). HPV vaccination in women aged 27 to 45 years: what do general practitioners think?. BMC women's health, 14(1), 91.

Meier, L., Van De Geer, S., & Bühlmann, P. (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(1), 53-71.

Moscicki, Anna-Barbara, Joel Palefsky, John Gonzales, and Gary K. Schoolnik. "Human Papillomavirus Infection Sexually Active Adolescent Females & colon; Prevalence and Risk Factors." Pediatric Research 28, no. 5 (1990): 507-513.

Muñoz, N., Méndez, F., Posso, H., Molano, M., van den Brule, A. J., Ronderos, M., ... & Instituto Nacional de Cancerologia HPV Study Group. (2004). Incidence, duration, and determinants of cervical human papillomavirus infection in a cohort of Colombian women with normal cytological results. Journal of Infectious Diseases, 190(12), 2077-2087.

Muñoz, N., Manalastas, R., Pitisuttithum, P., Tresukosol, D., Monsonego, J., Ault, K., ... & Bautista, O. (2009). Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine in women aged 24–45 years: a randomised, double-blind trial. The Lancet, 373(9679), 1949-1957.

Myers, E. R., McCrory, D. C., Nanda, K., Bastian, L., & Matchar, D. B. (2000). Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. American journal of epidemiology, 151(12), 1158-1171.

O'Dowd, K. J. (2003). Sexual Risk Status and Behavior of New Jersey Adults: Results from the New Jersey Behavioral Risk Factor Survey, 1998-1999. New Jersey Department of Health and Senior Services Topics in Health Statistics. January 2003.

Reiter, P. L., & McRee, A. L. (2016). HPV Infection among Sexual Minority Women: Does it Matter How Sexual Orientation Is Measured?. Cancer Epidemiology and Prevention Biomarkers, 25(3), 559-560.

Romanowski, B., Schwarz, T. F., Ferguson, L., Peters, K., Dionne, M., Behre, U., ... & Struyf, F. (2016). Sustained immunogenicity of the HPV-16/18 AS04-adjuvanted vaccine administered as a two-dose schedule in adolescent girls: Five-year clinical data and modeling predictions from a randomized study. Human vaccines & immunotherapeutics, 12(1), 20-29.

Sanders, G. D., & Taira, A. V. (2003). Cost effectiveness of a potential vaccine for human papillomavirus. Emerging infectious diseases, 9(1), 37-48.

Saslow, D., Andrews, K. S., Manassaram-Baptiste, D., Loomer, L., Lam, K. E., Fisher-Borne, M., ... & Fontham, E. T. (2016). Human papillomavirus vaccination guideline update: American Cancer Society guideline endorsement. CA: A Cancer Journal for Clinicians, 66(5), 375-385.

Satterwhite, C. L., Torrone, E., Meites, E., Dunne, E. F., Mahajan, R., Ocfemia, M. C. B., ... & Weinstock, H. (2013). Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. Sexually transmitted diseases, 40(3), 187-193.

Schiffman, M. H., Bauer, H. M., Hoover, R. N., Glass, A. G., Cadell, D. M., Rush, B. B., ... & Stanton, C. K. (1993). Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. Journal of the National Cancer Institute, 85(12), 958-964.

Schwarz, T. F., Spaczynski, M., Schneider, A., Wysocki, J., Galaj, A., Perona, P., ... & Dubin, G. (2009). Immunogenicity and tolerability of an HPV-16/18 AS04-adjuvanted prophylactic cervical cancer vaccine in women aged 15–55 years. Vaccine, 27(4), 581-587.

Shi, R., Devarakonda, S., Liu, L., Taylor, H., & Mills, G. (2014). Factors associated with genital human papillomavirus infection among adult females in the United States, NHANES 2007–2010. BMC research notes, 7(1), 544.

Tong, S. Y., Lee, Y. S., Park, J. S., Bae, S. N., Lee, J. M., & Namkoong, S. E. (2008). Clinical analysis of synchronous primary neoplasms of the female reproductive tract. European Journal of Obstetrics & Gynecology and Reproductive Biology, 136(1), 78-82.

U.S. Census Bureau. (2015). Survey of Income and Program Participation.

US Department of Health and Human Services. (2015). 2015–2020 dietary guidelines for Americans. Washington (DC): USDA.

Velentzis, L. S., Sitas, F., O'Connell, D. L., Darlington-Brown, J., Egger, S., Sinha, R., ... & Canfell, K. (2014). Human papillomavirus 16/18 seroprevalence in unvaccinated women over 30 years with normal cytology and with high grade cervical abnormalities in Australia: results from an observational study. BMC infectious diseases, 14(1), 3861.

Westra, T. A., Rozenbaum, M. H., Rogoza, R. M., Nijman, H. W., Daemen, T., Postma, M. J., & Wilschut, J. C. (2011). Until which age should women be vaccinated against HPV infection? Recommendation based on cost-effectiveness analyses. Journal of Infectious Diseases, 204(3), 377-384.

Winer, R. L., Lee, S. K., Hughes, J. P., Adam, D. E., Kiviat, N. B., & Koutsky, L. A. (2003). Genital human papillomavirus infection: incidence and risk factors in a cohort of female university students. American journal of epidemiology, 157(3), 218-226.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.

5 Conclusion

This dissertation performs an in-depth exploration of personalized decision modeling of cancers intervention and prevention. We investigate the patient-specific prevention strategies for breast cancer and HPV-related cancers, representative of cancers prevented by screening and vaccination, respectively. Three popular healthcare analytics techniques, Markov models, regression-based forecasting models, and discrete-event simulation, are further developed in the context of personalized cancer medicine. Although widely used in quantities analyses of healthcare, utilizing these approaches to optimize personalized cancer medicine was rarely addressed in the previous literature. This dissertation explores multiple possibilities of incorporating patient-specific risk into personalized cancer prevention and intervention strategies and showcases three practical examples.

Chapter 2 presents a relatively simple method of personalizing cancer screening decisions. Individual breast cancer risk is reflected in the transition probabilities of a Markov decision process model to optimize biopsy referral decisions. The optimal biopsy referral policy can be considered as a lifetime mammography schedule, which informs a patient when she should skip the routine mammography in her lifetime. The study suggests that both screening mammography and surveillance mammography schedules should be determined on a patient-specific basis.

Chapter 3 differs from chapter 2 by providing an "on-line" screening policy adaptive to a woman's health status. Whether a woman should receive a mammogram is determined according to her risk level in the current year. This study emphasizes the risk model of breast cancer and makes decisions based on the tradeoff of type I errors and type II errors. While most previous studies optimized the breast cancer screening policy by providing lifetime mammography screening schedules, our proposed model provides the dynamic decision model updated with women's latest cancer risk.

Chapter 4 shift focus from personalized screening decisions to personalized vaccination strategy. This study does not provide optimal decisions but rather evaluates different vaccination scenarios' outcomes. We demonstrate a hybrid framework combining regression-based risk estimation with discrete-event simulation. This study discusses how to overcome the difficulty of modeling dynamic risk factors, which are excluded in Chapter 2 and Chapter 3.

Our findings suggest that, by receiving personalized screening and vaccination, patients are expected to have longer life expectancy and less possibility of dying from cancer. We also provide three practical solutions for patient-specific screening and vaccination strategies. Preventive screening and vaccination programs for other cancers or diseases, which have clearly identified risk factors and measurable risk, may all benefit from patient-specific policies.