Fall 11-2020

Stochastic Decision Modeling to Improve Breast Cancer Preventive Care

Sevda Molani
Louisiana Tech University

STOCHASTIC DECISION MODELING TO IMPROVE BREAST CANCER PREVENTIVE CARE

by

Sevda Molani, M.S.

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

COLLEGE OF ENGINEERING AND SCIENCE
LOUISIANA TECH UNIVERSITY

October 2020

LOUISIANA TECH UNIVERSITY

GRADUATE SCHOOL

October 1, 2020
Date of dissertation defense

We hereby recommend that the dissertation prepared by

Sevda Molani, M.S.

entitled **Stochastic Decision Modeling to Improve Breast Cancer Preventive Care**

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational Analysis & Modeling

Dr. Mahboubeh Madadi
Supervisor of Dissertation Research

Dr. Weizhong Dai
Head of Computational Analysis & Modeling

Doctoral Committee Members:
Dr. Katie Evans
Dr. Weizhong Dai
Dr. Eric Sherer
Dr. Pradeep Chowriappa

Approved:

Hisham Hegab
Dean of Engineering & Science

Approved:

Ramu Ramachandran
Dean of the Graduate School

ABSTRACT

Breast cancer is a leading cause of premature mortality among women in the United States. Breast cancer screening tests can help with detecting breast cancer in early stages and thereby reducing the breast cancer mortality risk. However, due to the imperfect nature of screening tests, there is always some associated overdiagnosis, false positives, and false negatives risks. Therefore, to improve breast cancer preventive care, we defined the focus of this dissertation on modeling breast cancer screening decisions.

Breast cancer overdiagnosis is the first issue that is addressed in this dissertation. Although overdiagnosis is known to be the major risk inherent in mammography screening; currently there is no way to distinguish between overdiagnosed cancers and the ones that would cause problems over a patient's lifetime. Overdiagnosis risk significantly depends on a patient's compliance with screening recommendations. In Chapter 2, we use a stochastic framework to perform a harm-benefit analysis to compare the overdiagnosis risk with the benefits that breast cancer screening provides. In addition, we estimate the lifetime mortality risk of breast cancer while considering the overdiagnosis risk and the uncertainty in a patient's adherence behavior. Our results show that, although overdiagnosis rate is relatively high in breast cancer screening, the benefits of breast cancer mammography screening outweigh the overdiagnosis risk.

The second issue that is addressed in this dissertation is false negative results caused by density of breast tissue. Breast density is known to increase breast cancer risk and decrease mammography screening sensitivity. Breast density notification laws, require physicians to inform women with high breast density of these potential risks. The laws usually require healthcare providers to notify patients of the possibility of using more sensitive supplemental screening tests (e.g., ultrasound). Since the enactment of the laws, there have been controversial debates over i) their implementations due to the potential radiologists bias in breast density classification of mammogram images and ii) the necessity of supplemental screenings for all patients with high breast density. Breast density is a dynamic risk factor. Therefore, in the third chapter, we apply a hidden Markov model (HMM) on a sparse unbalanced longitudinal data to quantify the yearly progression of breast density based on Breast Imaging Reporting and Data System (BI-RADs) classifications.

In Chapter 4, we use the results from previous chapter to investigate the effectiveness of supplemental screening and the impact of radiologists' bias on patients' outcomes under the breast density notification law. We consider the conditional probability of eventually detecting breast cancer in early states given that the patient develops breast cancer in her lifetime and the expected number of supplemental tests as patient's outcome. Our results indicate that referring patients to a supplemental test solely based on their breast density may not necessarily improve their health outcomes and other risk factors need to be considered when making such referrals. Additionally, average-skilled radiologists' performances are shown to be comparable with the performance of a perfect radiologist.

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author _____

Date _____

DEDICATION

To my beloved mom and dad, who encouraged and supported me every step of my life, and my dear Mitra, Afshin and Nila for all the love, support and motivation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my sincerest gratitude to my advisor, Dr. Mahboubeh Madadi, for the indispensable knowledge and support that she provided during the difficulties I faced while working on this dissertation. I am so thankful for her enthusiasm, motivation and patience throughout this process. I would also like to thank Dr. Katie Evans, Dr. Weizhong Dai, Dr. Eric Sherer and Dr. Pradeep Chowriappa for their enormous contribution and guidance to this dissertation as members of my graduate committee. I would also like to specially thank Dr. Collin Wick for providing me much needed support and guidance during studies.

I would like to thank all my friends at Louisiana Tech for their friendship, help, support, motivation, and love. Thank you to Kamal, Greta Vladeanu and Ghufran Aldawood for your precious friendship and for making this journey an enjoyable experience for me. I would like to thank Wesley Wilkes, Alhossin Alsadi, Mehdi Mofidian, Saleh Bajaba, Omar Ahmed, Hillary Husband and Shashank Reddy for their friendship and support.

Last but not least, I have to thank my parents for their love and support throughout my life. Thank you both for giving me strength to reach for the stars and chase my dreams. Grandma Akram, uncle Ali, Mitra, Afshin, Nila, Rima and Alp deserve my wholehearted thanks as well.

CHAPTER 1

INTRODUCTION

1.1 Background

Breast cancer is the most common non-cutaneous cancer among women in the U.S. [1]. It is currently estimated that a woman's lifetime risk of developing breast cancer is 1 in 8 [2]. In 2020, an estimated 276,480 new cases of invasive breast cancer, and 48,530 new cases of non-invasive (in situ) breast cancer will be diagnosed in the U.S. [1]. Moreover, breast cancer is one of the leading causes of death in women with a lifetime mortality risk of 1 in 38 [1]. Based on the estimates from the American Cancer Society (ACS), approximately 42,170 women will die as a result of breast cancer in 2020 [1]. Breast cancer screening tests, the most common of which is mammography, can help with detecting breast cancer in early stages and thereby reducing the breast cancer mortality risk by treating patients when they have a higher survival chance. However, due to the imperfect nature of mammography screening, there is always some associated overdiagnosis, false positives, and false negatives risks. Overdiagnosis should not be confused with false-positive results. In a false-positive test result, the disease is mistakenly believed to be present in the patient's body based on the initial test. A more accurate follow-up test (e.g., biopsy), however, falsifies the initial belief. In overdiagnosis, however, the disease is truly present in the patient's body, but it

would not cause any harm if remained undetected. False negative rates are especially higher in women with dense breasts due to the reduced sensitivity of mammography caused by the masking effect of high density breast tissue. These risks and benefits of screening tests are functions of each patient's features such as patient's adherence, age, breast density, and family history. Thus, a screening policy which is tailored to the different features of individuals are desirable.

1.2 Objectives

The fundamental purpose of decision modeling is to provide a methodology for comparing a set of choices or strategies by calculating the expected value of a specific outcome resulting from those strategies. One of the main application of decision analysis is in preventive healthcare modeling which provides a mechanism for evaluating different preventive strategies using multiple outcome criteria such as life expectancy, quality of life, and costs. Thus, the purpose of this dissertation is to evaluate the preventive healthcare decisions for breast cancer which includes evaluating cancer screening policies in order to diagnosis the disease before its effect on the patient's life is irreversible. For this comparison, we consider different factors such as patient's adherence, age, breast density, family history, and risk of overdiagnosis.

1.3 Dissertation Organization

A detailed outline of this dissertation is presented below.

Chapter 2 addresses the problem of overdiagnosis in breast cancer screening. Overdiagnosis risk significantly depends on a patient's compliance with screening

recommendations. Specifically, we use two partially observable Markov chains developed by Molani et al. [3] to perform a harm-benefit analysis in order to compare the overdiagnosis risk with the benefits that breast cancer screening provides. In this chapter, we also estimate some other criteria such as the expected number of cancers detected and overdiagnosed, and the expected number of lives saved through mammography screening. Additionally, we estimate the breast cancer lifetime mortality risk and compare the results with the lifetime mortality risk reported by American Cancer Society(ACS) for some in-practice policies.

In chapter 3, we review and apply some of the hidden Markov model parameter approximation methods to estimate breast density transition probabilities using mammography screening data from Louisiana Cancer Prevention and Control Programs [4]. Breast density is defined as the prevalence of fibroglandular tissue in the breast and is categorized into four categories. Higher breast density can significantly reduce the mass detection rate since the normal tissues in dense breasts appear as bright areas in mammography. Breast density is also associated with increased risk of breast cancer [5]. In addition, due to the lower sensitivity of screening mammography in women with dense breasts, the cancer is more likely to remain undetected. Breast density is a dynamic risk factor and typically decreases as a patient becomes older. Therefore, in this chapter, we use a frequentest method (Baum-Welch) to quantify the dynamics of breast density. The results of this chapter will be used in the next section of dissertation to evaluate breast cancer screening policies considering breast density as an important breast cancer risk factor.

Chapter 4 addresses the controversy over breast density notification law and its potential unintended consequences as well as quality of its implementation. These laws generally require physicians to inform women with high breast density of their potential increased risk of breast cancer and the impact of high breast density on the sensitivity of mammogram test. The laws usually require healthcare providers to notify patients of the possibility of using more sensitive supplemental screening tests such as magnetic resonance imaging (MRI) and ultrasound. In this study, we formulate a finite-horizon, discrete time partially observable Markov chain (POMC) to investigate the efficacy of supplemental screening and the impact of radiologist's behavior and expertise on patients' outcomes. The patients' outcome measures include the conditional probability of eventually detecting breast cancer in early states given that the patient develops breast cancer in her lifetime and the expected number of supplemental tests.

Chapter 5 presents some concluding remarks of the research presented in this dissertation, as well as future work for improving the models presented.

CHAPTER 2

HARM-BENEFIT ANALYSIS OF BREAST CANCER SCREENING CONSIDERING OVERDIAGNOSIS RISK AND PATIENT'S ADHERENCE

2.1 Introduction

Although mammography screening reduces breast cancer mortality risk, there has been increasing concerns that it unintentionally leads to overdiagnosis by identifying small, indolent, or regressive breast tumors that would not otherwise become clinically apparent. Overdiagnosis occurs when breast cancer is detected during a routine screening mammogram, yet the cancer would never have presented clinically in the absence of screening. Overdiagnosis results from the current inability of physicians to determine if a patient diagnosed with breast cancer will or will not eventually develop any breast cancer related complications or symptoms over the course of her lifetime. Therefore, the standard treatment is administered to all patients whose cancer has been detected. For the overdiagnosed patients, however, this treatment is unnecessary and results in patients' overtreatment. These patients go on to experience unnecessary medical intervention, financial costs, and psychological stress due to being overdiagnosed. The impact of overdiagnosis on patients' well-being and physical health is life-long and thus, overdiagnosis and its subsequent unnecessary treatment that comes along with different risks are the most important potential harms of mammogram screening [6].

Overdiagnosis should not be confused with false-positive results. In a false-positive test result, the disease is mistakenly believed to be present in the patient's body based on the initial test. A more accurate follow-up test (e.g., biopsy), however, falsifies the initial belief. In overdiagnosis, however, the disease is truly present in the patient's body, but it would not cause any harm if remained undetected.

Quantifying overdiagnosis risk is very challenging as overdiagnosis cannot be observed directly. The main reason is the administration of treatment upon cancer detection which makes it impossible to directly determine the risk. Therefore, there is a wide variation on the estimation of the extent at which breast cancer overdiagnosis occurs. Various clinical trials, each with different sets of assumptions, have been applied to indirectly estimate the overdiagnosis risk. Dominant research methods that have been used to measure overdiagnosis are cohort, and follow-up of randomized controlled trials (RCTs) studies [7], which require a long follow-up of the patients for accurate quantification of overdiagnosis. These studies, however, are prone to bias in estimation of overdiagnosis due to the limitation in their designs and assumptions. According to the independent UK panel on breast cancer screening [8], potential bias (e.g., suboptimal randomization), uncertainty in the relevance of old trials to current screening programs, and the unavailability of some key information in such trials are some of the main reasons that make these estimations unreliable. Lee and Etzioni [9] conclude that for generating reliable estimates for overdiagnosis, one needs to recognize limitations which include the problems of unknown counter-factual incidence, insufficient follow-up time, and trial design limitations.

There are a limited number of studies that use mathematical/statistical modeling for estimation of overdiagnosis. These studies use modeling frameworks for estimation of overdiagnosis to overcome the limitations in RCTs and cohort studies. These studies, however, are limited in their assumptions of patients' compliance with a screening policy. Adherence is an important, but often ignored, aspect of disease screening. The World Health Organization (WHO) defines adherence behavior as "the extent to which a person's behavior (e.g., taking medication, following a diet, executing lifestyle changes) corresponds with agreed recommendations from a health care provider" [10]. Currently, the WHO estimates the adherence rate of patients to all health care recommendations at roughly 50% in developed countries [10]. In terms of mammography screening, adherence behavior refers to a patient's level of compliance with a screening policy, i.e. showing up to prescribed mammography tests. A study by the Centers for Disease Control and Prevention (CDC) shows that in 2010, only 66.5% of women had a mammogram within the past two years [11]. In another study by the U.S. Preventive Services Task Force (USPSTF), approximately 66% of women aged 40 years and older have had a screening mammography within the past two years [12]. In addition, based on the 2018 ACS statistics, 50% of U.S. women 40 years of age and older reported having had a mammogram within the past year and 64% reported having had a mammogram in the past two years [13]. Given the current screening guidelines in the U.S. (i.e., annual or biennial mammogram screenings depending on the patient age and breast cancer risk), the reported compliance rates are considered relatively low. As undergoing a screening test is a necessary step in occurrence of overdiagnosis, screening compliance rate should not be ignored in the quantification of overdiagnosis risk.

Incorporating adherence behavior provides a more realistic estimation of overdiagnosis and helps the decision makers to better evaluate and compare screening policies.

Besides overdiagnosis and its subsequent unnecessary treatment, false positives, unnecessary biopsies, and radiation exposure are other risks associated with screening mammography. These potential risks have led to much debate regarding mammography screening recommendations. On the other hand, the impact of mammogram in detecting cancer in its early stages and subsequently increasing survival rates makes the judgment about the necessity and frequency of mammogram screening challenging. There are multiple health agencies in the U.S. including the American Cancer Society (ACS), the United States Preventive Services Task Force (USPSTF), and the American College of Radiology (ACR), which recommend different breast cancer screening guidelines. Currently, there is not even an established consensus regarding the starting age, the stopping age, or the interval between two consecutive mammography screenings due to the uncertainty regarding the balance in benefits and harms associated with mammography screening.

In this chapter, we use a stochastic modeling framework developed by Molani et al. [3] to perform a harm-benefit analysis for overdiagnosis risk and estimate the ratio of detected cancers that are overdiagnosed, the ratio of lives saved per each overdiagnosed case, and other measures of interest for several screening policies.

Molani et al. [3] quantified the overdiagnosis risk associated with different breast cancer screening policies while incorporating the uncertainty in women's adherence behaviors. Often, calculation of overdiagnosis is based on the excess incidence (EI) in a

screened population compared to an unscreened reference population during the screening period. Overdiagnosis is usually reported as the proportion of screen-detected cancers which do not cause harm if left undetected. In this study, they introduced different measures of overdiagnosis risk. In addition to the previously commonly used measure, i.e., the ratio of number of overdiagnosed cases to the number of screen-detected cancers, they estimated the lifetime overdiagnosis risk. The lifetime overdiagnosis risk measures the risk of overdiagnosis imposed on patients if they follow a screening policy over their lifetime. Additionally, they estimated breast cancer stage-specific and age-specific risk of overdiagnosis, that gives the conditional risk of overdiagnosis given that a patient is diagnosed at a specific age and/or a specific cancer stage. Unlike cohort and RCT studies which are prone to different biases (as discussed above) and are limited to assess the overdiagnosis risk for only one policy that is applied to the screened arm, the proposed framework can assess different overdiagnosis risks for any given policy. In addition, the proposed model takes into consideration the uncertainty in patients' compliance to provide a realistic estimation of the risk. In this study, two partially observable Markov chain models are developed to estimate overdiagnosis measures of interest. The first partially observable Markov chain model represents women's adherence behaviors, and the second one represents natural history of breast cancer.

The remainder of this chapter is as follows. In Section 2.2, we present our proposed model for estimation of breast cancer mortality risk and a harm-benefit analysis framework. Section 2.3 presents the parameter estimations, and the design of our numerical studies. In Section 2.4, we present the results. We summarize and conclude in Section 4.6.

2.2 Model Formulation

In this section, we review the partially observable Markov chains developed by Molani et al. [3] to model patients' adherence behaviors and breast cancer natural history. The modeling of adherence behavior, which is detailed in Section 2.2.1, provides a framework to calculate age-specific probabilities of patients' compliance with a prescribed mammogram screening. Using the adherence model and the breast cancer natural history model, we formulate a framework to estimate the breast cancer mortality risk considering overdiagnosis and a harm-benefit framework to analyze the harm-benefit trade-off under different breast cancer screening policies.

2.2.1 Adherence Behavior

To model the uncertainty in adherence behavior, patients are classified into two distinct groups based on the CISNET definition: regular screeners and irregular screeners. A *regular screener* is a patient whose mean time between two consecutive screenings is less than or equal to 2 years. An *irregular screener* is a patient whose mean time between two consecutive screenings is greater than 2 years. However, a patient's classification is not static as adherence behavior can change over time. It has been shown that a patient's age, past screening behavior, education level, income level, and perceived risk are all influential factors in mammography screening adherence [14]. Therefore, a patient's adherence behavior is subject to change as these influential factors change.

A discrete-time partially observable Markov chain was developed to model a patient's adherence behavior. Figure 2.1 shows the state transition diagram of

the proposed Markov chain modeling a patient's adherence behavior. A partially observable model is used due to the uncertainty in characterizing a patient's compliance with a screening recommendation. From this model, we can determine the probability that a patient complies with a policy prescribed screening at any age. The probability of compliance with a policy's prescribed mammography is then used in Section 2.2.2 to calculate adherence-adjusted overdiagnosis risks. The following is the list of notations used in the model.

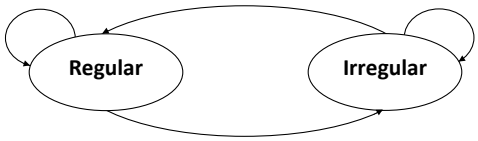


Figure 2.1: State transition diagram of the underlying adherence behavior Markov model

- t: Time period, $t = 0, 1, \cdots, T$. We model a patient's adherence behavior starting at age 40 ($t = 0$). We account for possible behavioral changes every six months, and the modeling ends at age 100 ($t = T = 120$).
- s_b: Core adherence behavior state, $s_b \in S_b = \{\mathcal{R}, \mathcal{I}\}$. The two adherence core states correspond to the two types of screeners: regular ($\mathcal{R}$) and irregular ($\mathcal{I}$). The states are partially observable as the decision maker does not have full information about a patient's adherence behavior.
- $T_t(s_b'|s_b)$: Core adherence behavior state transition probabilities, that is, the probability of a patient being in state s_b' at time $t+1$ given that she was in state s_b at time t.

- a_t: Prescribed action at time t, $a_t \in A = \{M, W\}$, where M and W represent action mammogram and wait, respectively.
- O^{a_t}: Observation space, which includes observations seen upon taking action a_t at time t. If $a_t = M$, the observation at time t is $o_t \in O^M = \{c, \bar{c}\}$, where c represents compliance with the prescribed mammography, and $\bar{c}$ represents failure to comply. If $a_t = W$, then no observation will be received and $O^W = \emptyset$.
- $Q_t(o_t|a_t, s_b)$: Observation probabilities, which represent the probability of receiving observation o_t given that the patient is in state s_b and action a_t is taken at time t.
- h_t: Screening attendance history up to time t, $h_t = (a_1, o_1, a_2, o_2, \ldots, a_{t-1}, o_{t-1}) \in H_t$, where H_t is the set of all possible screening attendance sample paths.
- κ_{h_t}: History dependent adherence behavior belief state, $\kappa_{h_t} = [\,\kappa_{h_t}(\mathcal{R}), \kappa_{h_t}(\mathcal{I})]\,$, where $\kappa_{h_t}(\mathcal{R})$ and $\kappa_{h_t}(\mathcal{I})$ represent the probability of being in core adherence state $\mathcal{R}$ and $\mathcal{I}$, respectively, given the patient history of compliance to recommended screenings is h_t.
- β_t: Expected adherence behavior belief state at time t, $\beta_t = [\,\beta_t(\mathcal{R}), \beta_t(\mathcal{I})]$, where $\beta_t(\mathcal{R})$ and $\beta_t(\mathcal{I})$ represent the probability of being in core adherence state $\mathcal{R}$ and $\mathcal{I}$, respectively.

A patient's adherence behavior belief is updated based on the observations received at the time epochs where the prescribed action is to undergo a mammogram and a patient screening history. If the screening policy does not prescribe a mammogram,

the adherence behavior belief state is updated based on the core adherence state transition probabilities to account for adherence behavior dynamics. Given the adherence behavior belief state distribution associated with sample path h_t at time t is κ_{h_t}, Equation (2.1) calculates the updated adherence behavior belief state when the prescribed action is a_t and observation o_t is received. More specifically, $\kappa_{h_{t+1}}(s'_b|\kappa_{h_t}, a_t, o_t)$ gives the probability that the patient is in adherence state s'_b at time $t+1$, given her adherence belief up to time t is κ_{h_t}, action a_t is taken and observation o_t is received at time t.

$$\kappa_{h_{t+1}}(s'_b|\kappa_{h_t}, a_t, o_t) = \begin{cases} \frac{\sum_{s_b=\mathcal{R},\mathcal{I}} \kappa_{h_t}(s_b)\cdot Q_t(o_t|M,s_b)\cdot T_t(s'_b|s_b)}{\sum_{s_b=\mathcal{R},\mathcal{I}} \kappa_{h_t}(s_b)\cdot Q_t(o_t|M,s_b)}, & a_t = M, \quad o_t \in O^M, \\ \sum_{s_b=\mathcal{R},\mathcal{I}} \kappa_{h_t}(s_b)\cdot T_t(s'_b|s_b)\,, & a_t = W. \end{cases} \tag{2.1}$$

The first line in Equation (2.1) represents the case when the recommended action is a mammogram test. In such case, we use Bayes rule to update the belief about the patient being a regular or irregular screener based on the received observation o_t at time t. The second line in Equation (2.1) represents the case when the prescribed action is to wait, in which case the dynamics of adherence behavior is used to update the adherence behavior belief states.

Let $P(h_t)$ denote the probability associated with the screening attendance sample path h_t. The probability of sample path h_t can be calculated recursively as follows:

$$P(h_t) = P(h_{t-1}) \cdot Q_t(o_t|a_t), \tag{2.2}$$

where $P(h_0) = 1$ since there is no uncertainty associated with a sample path when no action is taken, yet. In addition, $Q_t(o_t|a_t) = \sum_{s_b=\mathcal{R},\mathcal{I}} \kappa_{h_t}(s_b) \cdot Q_t(o_t|a_t, s_b)$.

The expected adherence behavior belief state (the average over all possible screening attendance sample paths) at time t and the probability that a patient complies with a policy prescribed screening at time t can be calculated according to Equations (2.3) and (2.4), respectively.

$$\beta_t = \sum_{h_t \in H_t} \kappa_{h_t} \cdot P(h_t), \tag{2.3}$$

$$C_t = \sum_{s_b=\mathcal{R},\mathcal{I}} Q_t(c|a_t, s_b) \cdot \beta_t(s_b), \tag{2.4}$$

where β_t is calculated by taking into consideration all possible screening attendance sample paths up to time t, and C_t is calculated based on the prescribed action at time t using the total probability rule.

2.2.2 Overdiagnosis Risk Estimation

A discrete-time partially observable Markov chain model was developed to characterize the natural history of breast cancer. A partially observable model is developed because mammography screening is imperfect and provides only partial information about the core health state of a patient. In this model, three breast cancer health states are included: early stage, advanced stage, and symptomatic breast cancer. This Markov model was proposed by Maillart et al. [15], which is based on the American Joint Committee on Cancer (AJCC) classification, with some minor changes to develop our POMC model. We consider early stage breast cancer to be the grouping of stage 0, stage I, and stage II without lymph node involvement. We

consolidate stage II with lymph node involvement, stage III, and stage IV breast cancer into the single health state of advanced breast cancer. Further, we assume that the manifestation of symptoms is only possible from the advanced stage of the disease. Based on this model, cancers take at least one year from their onset to grow to a symptomatic size (given the time interval between two epochs is six months). This is in line with previous studies and reports since even the cancer with the highest progression rates cannot grow from a cell to a symptomatic size in six months [16].

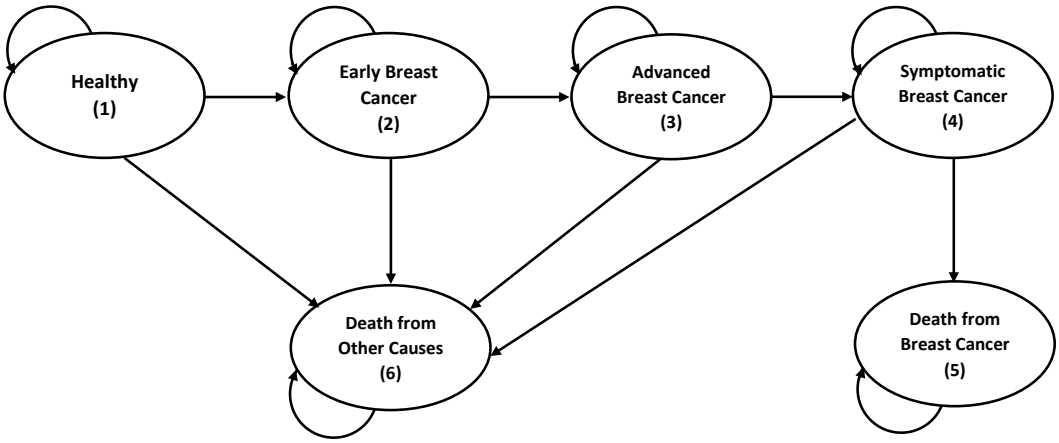


Figure 2.2: State transition of the underlying health Markov model representing the natural history of breast cancer

The following is the list of the notations used in this section. Note that notation t and a_t are defined previously in Section 2.2.1 and will be used here.

- s_h: Core health state, $s_h \in S_h = \{1, 2, \cdots, 6\}$, where the state space S_h is composed of six states: healthy (1), early breast cancer (2), advanced breast cancer (3), symptomatic breast cancer (4), death from breast cancer (5), and death from other causes (6). Due to the imperfect nature of mammography

screening, states 1, 2, and 3 are partially observable. All other states are fully observable.

- $P_t(s'_h|s_h)$: Core health state transition probability, which represents the probability that a patient is in state s'_h at time $t+1$ given she was in state s_h at time t.
- Θ^{a_t}: Observation space when the prescribed action is a_t. At each time period, an action dependent observation will be received. If $a_t = M$, then the observation at time t will be $\theta_t \in \Theta^M = \{M^+, M^-\}$, where M^+ and M^- represent positive and negative mammogram result, respectively. If $a_t = W$, then no observation will be received, and $\Theta^W = \emptyset$.
- $K_t(\theta_t|a_t, s_h)$: Observation probabilities which represent the probability of receiving observation θ_t given action a_t is taken, and the patient is in state s_h at time t. Note that $K_t(M^+|M, s_h)$ is the sensitivity of mammography screening for early stage ($s_h = 2$) and advanced stage ($s_h = 3$) breast cancer at time t. Further, $K_t(M^-|M, s_h = 1)$ is the specificity of mammography test at time t.
- χ_t: History of screening results up to time t, $\chi_t = (a_1, \theta_1, \ldots, a_{t-1}, \theta_{t-1}) \in X_t$, where X_t is the set of all possible screening results' sample paths.
- η_{χ_t}: History dependent health belief state, which represents the probability distribution of a patient being in the partially observable core health states given the screening results history χ_t, i.e., $\eta_{\chi t} = [\eta_{\chi_t}(1), \eta_{\chi_t}(2), \eta_{\chi_t}(3)]$. Note that $\eta_{\chi_t}(s_h)$ represents the probability of being in health state s_h given the screening results history χ_t.

- π_t: Expected health belief state, $\pi_t = [\pi_t(1), \pi_t(2), \pi_t(3)]$, where $\pi_t(s_h)$ represents the probability of being in the core health state s_h at time t.

At each time epoch t, one of the two possible actions of screening mammogram or wait is prescribed by a policy. When the prescribed action is a mammogram, patients may follow the prescribed mammogram or skip it (i.e., wait). We assume actions are taken at the beginning of each period. Regardless of the action taken, we account for possible incidence and progression of breast cancer. If a patient undergoes a mammogram and receives a positive result, we assume a biopsy will be performed to confirm the presence of cancer. Upon cancer detection, the patient will not receive screening any longer. We assume no treatment is administered upon cancer detection to estimate overdiagnosis. However, we follow up the patient up to time T to determine if the patient dies from breast cancer or other causes to calculate the overdiagnosis risk.

The health belief state of a patient at each time epoch t is updated based on the action taken and the observation received. Let $\tau[\eta_{\chi_t}, a_t, \theta_t]$ denote the updated health belief state for sample path χ_{t+1} given that at time t the health belief state is η_{χ_t}, action a_t is taken and observation θ_t is received. We have

$$\eta_{\chi_{t+1}}(s_h') = \tau[\eta_{\chi_t}, a_t, \theta_t](s_h') = \begin{cases} \frac{1}{L_t} \frac{\sum_{s_h=1}^{3} \eta_{\chi_t}(s_h) \cdot K_t(M^-|M, s_h) \cdot P_t(s_h'|s_h)}{\sum_{s_h=1}^{3} \eta_{\chi_t}(s_h) \cdot K_t(M^-|M, s_h)}, & a_t = M,\ \theta_t = M^-\,, \\ \frac{1}{L_t} P_t(s_h'|s_h = 1), & a_t = M,\ \theta_t = M^+, s_h = 1, \\ \frac{1}{L_t} \sum_{s_h=1}^{3} \eta_{\chi_t}(s_h) \cdot P_t(s_h'|s_h)\,, & a_t = W, \end{cases} \tag{2.5}$$

where L_t is a normalizing factor and represents the probability that the patient survives time t. The first line in Equation (2.5) represents the case when the patient undergoes a prescribed mammogram and receives a negative result. In such case, we use Bayes rule to update the patient's health belief state based on the received result. The second case represents the situations in which the patient receives a false positive mammogram result (that is, the patient receives a negative biopsy after receiving a positive mammogram result). In this case, the patient's belief state is updated by considering possible cancer development from the cancer-free state at time t (i.e., $s_h = 1$ at time t). When the prescribed action is to wait, we use the dynamics of breast cancer natural history to update our belief about the patient's health status. This is represented in the third case of Equation (2.5).

Similar to the adherence behavior model, the probability of each sample path can be calculated as $P(\chi_{t+1}) = P(\chi_t) \cdot K(\theta_t|a_t)$, where $K(\theta_t|a_t) = \sum_{s_h=1}^{3} K_t(\theta_t|a_t, s_h) \cdot \eta_{\chi_t}(s_h)$. Therefore, the expected health belief state can be calculated as follows.

$$\pi_t = \sum_{\chi_t \in X_t} \eta_{\chi_t} \cdot P(\chi_t). \tag{2.6}$$

For an overdiagnosis to happen, a cancer needs to be detected first. This requires that (i) the patient is in one of the cancer states, (ii) the patient shows up to a prescribed mammogram screening, and (iii) the screening mammography detects the cancer. The probabilities of events (i) and (iii) depend on the patient's cancer stage. The probability of event (ii), however, is independent of the cancer stage as the patient has no information about the presence of cancer in her body and decides independently about attending a prescribed screening. Let $D_t(s_h, a_t)$ denote

the probability of detecting a cancer in state $s_h, (s_h = 2, 3)$ at time t when action a_t is taken. Equation (2.7) calculates this probability.

$$D_t(s_h, a_t) = \begin{cases} \pi_t(s_h) \cdot C_t \cdot K_t(M^+|M, s_h), & a_t = M, \\ 0, & a_t = W. \end{cases} \tag{2.7}$$

Moreover, let $\omega_t(s_h)$ denote the conditional probability that a patient with a breast cancer in state s_h $(s_h = 2, 3)$ at time t eventually dies from other causes without developing symptomatic breast cancer. Equation (2.8) calculates this probability for early stage breast cancer and comes from the following logic. Given that the patient has early breast cancer at time t, she can die from other causes in time period t, or she can survive time t and eventually die from other causes. Note that if the patient survives time t, she either stays in early breast cancer or progresses to advanced breast cancer at time $t + 1$.

$$\omega_t(2) = P_t(6|2) + \sum_{s_h=2,3} P_t(s_h|2) \cdot \omega_{t+1}(s_h). \tag{2.8}$$

Following the same logic, Equation (2.9) calculates the probability that a patient who is in advanced breast cancer at time t eventually dies from other causes without developing symptomatic breast cancer.

$$\omega_t(3) = P_t(6|3) + P_t(3|3) \cdot \omega_{t+1}(3). \tag{2.9}$$

At the final age, we assume a patient eventually dies from other causes with certainty due to age-related diseases and co-morbidities. Therefore, $\omega_T(2) = \omega_T(3) = 1$. Note that we assume $T = 120$, which represents the age of 100.

Next, we estimate age-specific and stage-specific overdiagnosis risk, $\Omega_t(s_h)$, which calculates the probability of overdiagnosis if a patient is diagnosed with breast cancer in stage s_h at time t. This includes the probability of both detecting breast cancer at time t and the probability of eventually dying from other causes before developing symptomatic breast cancer; that is

$$\Omega_t(s_h) = D_t(s_h, a_t) \cdot \omega_t(s_h). \tag{2.10}$$

To compute age-specific overdiagnosis risk, we note that detection of early breast cancer and detection of advanced breast cancer are mutually exclusive events. Thus the probability that a patient, who is diagnosed with breast cancer at time t is overdiagnosed can be calculated as follows:

$$\Omega_t = \sum_{s_h=2}^{3} \Omega_t(s_h)\,. \tag{2.11}$$

Over her lifetime, a patient's breast cancer can be diagnosed at any age. However, screening stops once breast cancer has been detected. Therefore, a patient cannot be diagnosed at two distinct ages. Hence, the lifetime overdiagnosis risk of a screening policy can be obtained as follows:

$$\Omega = \sum_{t=0}^{T} \Omega_t. \tag{2.12}$$

2.2.3 Harm-Benefit Analysis

A harm-benefit analysis framework is presented to quantify the harms' and benefits' trade-off for different screening policies. More specifically, we investigate the associated harms and benefits of applying a screening policy to the U.S. female population

and measure the expected number of cancers detected and overdiagnosed, as well as the expected number of lives saved through mammography screening. Let n_{α_t} denote the U.S. female population size at age $\alpha_t, \alpha_t = 40, 40.5, \ldots, 100$. Note that patient age and time period t are related through $\alpha_t = 0.5t+40, (t = 0, \ldots, 120)$. The expected number of screen-detected cancers and the expected number of detected cancers that lead to overdiagnosis are calculated as $\sum_{t=0}^{120} n_{\alpha_t} \cdot \sum_{s_h=2,3} D_t(s_h, a_t)$ and $\sum_{t=0}^{120} n_{\alpha_t} \cdot \Omega_t$, respectively. The expected number of screen-detected cases who survive breast cancer is calculated as $\sum_{t=0}^{120} n_{\alpha_t} \cdot SR_{\alpha_t}$, where SR_{α_t} is the probability of detecting a cancer through screening at age α_t that will be cured with treatment and is calculated as follows:

$$SR_{\alpha_t} = \sum_{s_h=2,3} D_t(s_h, M) \cdot \gamma_{\alpha_t}(s_h), \tag{2.13}$$

where $\gamma_{\alpha_t}(s_h)$ is the probability that a patient diagnosed with cancer stage s_h at age α_t survives breast cancer and eventually dies of a competing cause given that treatment is administered upon cancer detection.

2.2.4 Lifetime Breast Cancer Mortality Risk

In this section, we present a framework to estimate the breast cancer lifetime mortality risk derived from the proposed model. Let $\psi_t^a(\eta_{\chi_t})$ denote the probability that a patient with a screening history χ_t and expected health belief state η_{χ_t} eventually dies from breast cancer when action a_t is taken at time t. If the recommended action at time t is to wait (W), the patient risk of eventually dying from breast cancer is calculated as follows:

$$\psi_t^W(\eta_{\chi_t}) = \eta_{\chi_t}(3) \cdot P_t(4|3) \cdot r_{t+1}(4) + \Big[1 - \sum_{s_h=1}^{3} \eta_{\chi_t}(s_h) \cdot P_t(6|s_h)\Big] \cdot \psi_{t+1}^a(\tau[\eta_{\chi_t}, W, .]), \tag{2.14}$$

where $r_{t+1}(4)$ is the probability that a patient diagnosed with symptomatic cancer ($s_h = 4$) at time $t+1$ eventually dies of breast cancer. The logic of Equation (2.14) is as follows. If the patient is in advanced cancer state at time t, she may proceed to symptomatic cancer within period t with probability $P_t(4|3)$ and eventually die from breast cancer with probability $r_{t+1}(4)$, or she may survive to the next period in which case her belief state for time $t+1$ is updated and her probability of eventually dying from breast cancer is calculated recursively. The updated expected health belief state, $\eta_{\chi_{t+1}}$, can be calculated based on the third case in Equation (2.5).

If the prescribed action at time t is a mammogram screening, the probability that the patient eventually dies of breast cancer is

$$\begin{aligned}
\psi_t^M(\eta_{\chi_t}) = {} & \eta_{\chi_t}(1) \cdot K(M^+|M, s_h = 1) \cdot (1 - P_t(6|1)) \cdot \psi_{t+1}^a(\tau[\eta_{\chi_t}, M, M^+]) \\
& + \sum_{s_h=2,3} \eta_{\chi_t}(s_h) \cdot K(M^+|M, s_h) \cdot r_t(s_h) \\
& + \eta_{\chi_t}(3) \cdot K(M^-|M, s_h = 3) \cdot P_t(4|3) \cdot r_{t+1}(4) \\
& + \Bigg[\sum_{s_h=1,2} \eta_{\chi_t}(s_h) \cdot K(M^-|M, s_h) \cdot (1 - P_t(6|s_h)) \\
& \qquad + \eta_{\chi_t}(3) \cdot K(M^-|M, s_h = 3) \cdot P_t(3|3) \Bigg] \cdot \psi_{t+1}^a(\tau[\eta_{\chi_t}, M, M^-]). \qquad (2.15)
\end{aligned}$$

The logic of Equation (2.15) is as follows. Assume that the mammography result is positive. Two cases are possible: (i) The true health state of the patient is cancer-free ($s_h = 1$), in which case the follow-up biopsy reveals that the initial mammogram positive result was inaccurate. In this case, the patient survives period t with probability $(1 - P_t(6|1))$, her health belief is updated using the second case in Equation (2.5), and her risk of dying from breast cancer in the future is calculated

recursively. (ii) The patient is in one of the cancer states, in which case the patient starts treatment upon receiving a positive result, and her risk of eventually dying from breast cancer would be $r_t(s_h)$ depending on the cancer status ($s_h = 2, 3$). Note that we do not model the treatment explicitly. We assume that the patient leaves the model when cancer is detected. If the mammogram screening result is negative, two cases are possible: (i) the patient develops some symptoms when in advanced breast cancer state ($s_h = 3$), in which case she proceeds to treatment and her risk of eventually dying from breast cancer would be $r_{t+1}(4)$. (ii) She survives time epoch t with probability $(1 - P_t(6|s_h))$ when in states $s_h = 1, 2$ and with probability $P_t(3|3)$ when in advanced breast cancer ($s_h = 3$). In such case, the patient health belief is updated based on Equation (2.5), and her mortality risk is calculated recursively to account for the future risk of dying from breast cancer.

2.3 Numerical Analysis

2.3.1 Model Input

The data sources that are used in the model parameters' estimation are listed in Table 2.1. Estimation of the core health state transition probabilities is the most challenging due to the lack of a single data source that includes all the age-specific dynamics of untreated, unscreened natural history of breast cancer. We, therefore, use several sources to estimate the breast cancer dynamics parameters [15, 17, 18]. Our primary source for the core health transition probabilities estimations is Maillart et al. [15]. However, since the breast cancer natural history model presented here is more detailed and includes symptomatic cancer, we use some additional sources for

parameters' estimation. The age-specific non-breast cancer induced death probabilities, that is $P_t(6|s_h)$ for $s_h \in S_h \setminus \{5,6\}$, are calculated based on the 2010 report on annual mortality rate from all causes of death by the CDC [19]. We use the same approach proposed by Maillart et al. [15] to estimate these probabilities. The transition probabilities for the healthy state, $P_t(s_h|1)$ for $s_h = 1, 2$, are also adopted from Maillart et al. [15]. The probability that an untreated early breast cancer remains in the early breast cancer state in six months is calculated based on a study by Wu et al. [17], which reports the likelihood that untreated pre-clinical local breast cancer (PL) remains pre-clinical local breast cancer over the course of one year, i.e., $\sqrt{PL} = P_t(2|2)$. The probability of progression of untreated breast cancer from early stage to advanced stage, $P_t(3|2)$, is obtained using the normalizing condition, i.e., $P_t(3|2) = 1 - P_t(2|2) - P_t(6|2)$. The six-month transition probability from advanced breast cancer to symptomatic breast cancer, $P_t(4|3)$, is obtained from Bloom et al. [18]. Similarly, the survival probability of symptomatic breast cancer in six months is estimated based on the five-year survival probabilities from the onset of symptoms (SF) reported by Bloom et al. [18], i.e., $\sqrt[10]{SF} = P_t(4|4)$. Although outdated, the Bloom et al. [18] study is a unique source for late stage breast cancer progression modeling. The probability that an untreated advanced breast cancer remains in advanced breast cancer in six months, $P_t(3|3)$, and the probability of dying from breast cancer in six month when in symptomatic breast cancer state, $P_t(5|4)$, are calculated based on normalizing conditions, i.e., $P_t(3|3) = 1 - P_t(4|3) - P_t(6|3)$, and $P_t(5|4) = 1 - P_t(4|4) - P_t(6|4)$, respectively.

Table 2.1 Input data sources for parameters estimation

Model Parameter	Parameter Values	Source
Adherence state transition probabilities	Table A.1	National Health Interview Survey [20]
Adherence rates	Table A.2	HINTS [21], and Madadi et al. [22]
Initial adherence belief	Table A.3	National Health Interview Survey [20]
Health state transition probabilities	Table A.4	Maillart et al. [15], Wu et al. [17], and Bloom et al. [18]
Mammography sensitivity	Table A.5	Maillart et al. [15]
Mammography specificity	0.89300	Ayer et al. [23]
Initial health belief	Table A.6	Gail risk model [24]
Survival rate for screen-detected breast cancer by state	Table A.7	Maillart et al. [15]
Survival rate for symptomatic breast cancer	0.84427	Allgood et al. [25]
U.S. age composition	Table A.8	U.S. Census Bureau [26]
Breast cancer treatment costs	Table A.9	Mariotto et al. [27], and Ong and Mandl [28]

The transition probabilities of the adherence behavior model, $T_t(s'_b|s_b)$, are estimated using the CDC 2015 National Health Interview Survey (NHIS) data. NHIS dataset includes mammogram participation report of $10,245$ cases over six years. Adherence to mammography recommendations is shown to be correlated with age [29, 30]. To calculate the age-specific estimates of adherence transition probabilities, the survey participants are first filtered by age into one of three age groups: $[40-50)$, $[50-65)$, or 65+. These age groups are adopted from previous studies [29, 30]. We use the first four years of NHIS data to classify survey participants as regular or irregular screeners based on the screening mammography history. Our analysis shows that 52%, 69%, and 63% of cases in age groups $[40-50)$, $[50-65)$, and 65+ are regular screeners, respectively. We then calculate the proportion of subjects who maintain their classification (stay a regular (irregular) screener if first identified as a regular (irregular) screener) based on the remaining data from the last two years. Let F_R and F_I denote the proportion of patients who maintained their classifications as a regular screener and irregular screener in the last two years, respectively. F_R and F_I can be interpreted as the two-year transition probabilities from regular to regular and irregular to irregular screener, respectively. To estimate the six-month transition

probabilities, we account for all possible sample paths a patient can take. Equations (2.16) and (2.17) represent all possible six-month transitions sample paths and their relations to F_R and F_I. Using Equations (2.16) and (2.17) along with the normalizing conditions $T(R|R) + T(I|R) = 1$, and $T(R|I) + T(I|I) = 1$, we can calculate all the four transition probabilities. Note that we assume that patients in each age group have similar adherence behavior dynamics. Therefore, we drop the time subscripts for the transition probabilities for the sake of notational simplicity.

$$\begin{aligned} F_R = {} & T(R|R)^4 \; + \; 3 \cdot T(R|R)^2 \cdot T(I|R) \cdot T(R|I) \\ & + \; 2 \cdot T(R|R) \cdot T(I|I) \cdot T(I|R) \cdot T(R|I) \\ & + \; T(I|I)^2 \cdot T(I|R) \cdot T(R|I) \; + \; T(I|R)^2 \cdot T(R|I)^2. \end{aligned} \tag{2.16}$$

$$\begin{aligned} F_I = {} & T(I|I)^4 \; + \; 3 \cdot T(I|I)^2 \cdot T(R|I) \cdot T(I|R) \\ & + \; 2 \cdot T(I|I) \cdot T(R|R) \cdot T(R|I) \cdot T(I|R) \\ & + \; T(R|R)^2 \cdot T(R|I) \cdot T(I|R) \; + \; T(R|I)^2 \cdot T(I|R)^2. \end{aligned} \tag{2.17}$$

We estimate F_R for the age groups [40-50), [50,65), and 65+ as 0.9361, 0.9169, and 0.9121, respectively. In addition, we estimate F_I for these age groups to be 0.4340, 0.6063, and 0.7644, respectively. The age-specific observation probabilities for the breast cancer model, $K_t(\theta_t|a_t, s_h)$, are calculated based on the mammography test sensitivity and specificity reported in previous studies [15, 23]. The age-specific observation probabilities for the adherence behavior model, $Q_t(o_t|a_t, s_b)$, are adopted from Madadi et al. [29]. Initial health belief state (risks of early and invasive cancers) for women aged 40 are estimated using the Gail model [24]. For women at an older

age, we use Bayesian updating to estimate the health belief state. Initial adherence belief state at age 40 is estimated using the NHIS data [20]. For the harm-benefit analysis, we use the U.S. Census Bureau population composition data to calculate the expected number of cancers detected and the expected number of overdiagnosis for each age group if screening policies applied to the U.S. female population. The associated breast cancer treatment costs are also adopted from the literature [27, 28] to estimate the extent of overdiagnosis and unnecessary overtreatment costs.

2.3.2 Experimental Design

In our numerical studies, we consider 203 distinct screening policies, including the in-practice policies in the United States as well as some alternative policies with different starting age, stopping age, and screening intervals. The in-practice policies include those recommended by the ACS, USPSTF, and ACR. Based on the ACS guideline, women are recommended to undergo annual screening between the ages of 45 and 55. After age 55 it is up to the patient to continue the annual screening or switch to biennial screening. The ACR guideline recommends women undergo annual mammography screening starting at age 40. The ACS and ACR guidelines do not specify a stopping age and recommend women undergo mammography screening as long as they are in good health. Therefore, in this study, we consider different stopping ages ranging from 70 to 100 with 5-year increments for the ACS and ACR policies. The USPSTF recommends biennial screening for women aged 50 to 74.

We consider two types of policies in our numerical analysis: *static* and *dynamic* policies. A *static* policy has a fixed screening interval over the patients' lifetime and

is represented as (a_b, i_1, a_e), where a_b, i_1, and a_e are screening starting age, screening interval length, and screening stopping age, respectively. For example, the USPSTF policy, which is a static policy, is represented as (50,2,74). In a *dynamic* policy, however, the screening interval length changes over the patient's lifetime to account for change in the dynamics of breast cancer. A *dynamic* policy is represented as $(a_b, i_1, a_s, i_2, a_e)$, where a_s, i_1, and i_2 are the switching age, first and second screening interval length, respectively, and a_b and a_e are similar to that of the *static* policies. For example, (45,1,55,2,70) is a version of the ACS guideline, which recommends annual screening between the ages of 45 and 55 and biennial screening between ages of 55 and 70. Table 2.2 shows different screening parameters considered in this study.

Table 2.2 Screening policies considered in the numerical analysis

Policy	**Start age**	**1^{st} interval**	**Switching age**	**2^{nd} interval**	**Stopping age**
Static	40-60	1,2	-	-	70-100 (5 year increments)
Dynamic	40-60	1,2	45-65	1,2	70-100

2.4 Results

In this section, we present the harm-benefit analysis results and lifetime breast cancer mortality risk for in-practice and some alternative policies.

2.4.1 Harm-Benefit Analysis Results

In this section, the results of the harm-benefit analysis are presented for some in-practice policies. Specifically, we analyze the harms and benefits associated with different variations of the ACS, ACR and USPSTF policies. We consider the cohort of U.S. women aged 40 to 100 and assume that women are screened under each policy. The expected number of cancers detected, the expected number of overdiagnosed cases, the

expected number of screen-detected patients who survive breast cancer, and the ratio of the number of lives saved per each overdiagnosed case are calculated for each policy. We also analyze the results in terms of the associated unnecessary costs that could be avoided if patients were identified to be overdiagnosed. The overtreatment cost considered here includes the initial treatment costs spent in the first year of treatment upon cancer detection. We compare the estimated overtreatment costs with the reported U.S. annual breast cancer care cost. The proportion of annual breast cancer care cost spent on overtreatment is reported for each policy. The annual cost of breast cancer care in 2010 was estimated to be $16.50 billion [28]. Note that although we report the results for one year of screening, we assume patients undergo each policy over their entire lifetime; that is the overdiagnosis and survival rates are calculated over the patients' lifetimes.

Tables 2.3 and 2.4 present the harm-benefit analysis results for the adherence-adjusted and perfect adherence case, respectively. Based on the results, about 10% to 22% of screen-detected cancers for both imperfect and perfect adherence cases are overdiagnosed which is in-line with the ratios reported by [28]. For both adherence-adjusted and perfect adherence case, the ratio of overdiagnosed cancers to screen-detected cancers increases as the screening policy becomes more invasive at older ages, with the ACS policy with changing screening intervals and stopping age of 70 having the lowest and the ACR policy with stopping age of 100 having the highest rate of overdiagnosed cancers to screen-detected cases. In addition, the variation of the ACS and ACR policies with stopping age of 70 have the highest numbers of lives saved per overdiagnosed cases. In general, the results support the implementation of screening for breast cancer preventive care since the expected numbers of lives saved are larger

than the number of overdiagnosed cases for all the policies considered here. The results also show that the benefits of screening decrease significantly as the patients become older. Therefore, organized screening programs are more effective when applied to women younger than 70 since this group of women has a higher life expectancy and thus lower risk of being overdiagnosed. The ratios of total annual cost disbursed on overtreatment are estimated for both annual estimated cost of $16.50 billion reported in the literature [28] and the expected annual breast cancer care calculated by the model, which includes both screening and treatment costs. In terms of the associated overtreatment costs, the ACS and ACR policies with stopping age of 70 and 100 have the lowest and the highest ratios of overtreatment cost to the total breast cancer care costs, respectively, which supports stopping screening at the age of 70.

Moreover, comparing the results for the perfect adherence and adherence-adjusted case shows an increase in the expected number of screen-detected breast cancers and consequently an increase in the expected number of overdiagnosed cases when assuming perfect adherence. However, for policies with lower stopping age (70 and 74), this increase in the expected number of screen-detected breast cancers mostly contributes to the number of cases that are not overdiagnosed. Thus, for these policies when assuming perfect adherence, we have a lower ratio of overdiagnosed cases to the expected number of detected cancers as well as a higher ratio of lives saved per overdiagnosed cases, which imply encouraging women to comply with the prescribed screening policies at younger ages. However, when the policy recommends stopping screening at an older age, perfect adherence case yields higher ratios of overdiagnosed cases to screen-detected cases as well as lower expected numbers of lives saved per

overdiagnosis when compared to the adherence-adjusted case. This happens as the probability of overdiagnosis in older women is higher. This implies the negative effects of intense mammography screening at older ages. Moreover, the increase in the expected number of overdiagnosed cases causes an increase in the overall overtreatment costs in the perfect adherence case compared to the adherence-adjusted case.

Table 2.3 Harm-benefit analysis of in-practice policies implemented to the U.S. female population for the adjusted-adherence case. (BC represents "breast cancer")

Policy	Expected number of screen-detected BC	Expected number of overdiagnosis	Ratio of overdiagnosed cases to screen-detected cases (%)	Expected number of lives saved per overdiagnosis	Ratio of Overtreatment to BC care cost (%)	Ratio of Overtreatment to BC care cost (model-based) (%)
ACR: (40,1,70)	81491.208	8904.079	10.926	6.694	1.958	1.715
ACR: (40,1,80)	102829.724	16471.229	16.018	4.669	3.612	3.039
ACR: (40,1,90)	113777.342	22849.362	20.082	3.756	5.000	4.121
ACR: (40,1,100)	115302.337	23956.860	20.777	3.634	5.238	4.307
ACS: (45,1,70)	80103.400	8891.599	11.100	6.519	1.965	1.723
ACS: (45,1,80)	101441.917	16458.749	16.225	4.573	3.620	3.047
ACS: (45,1,90)	112389.535	22836.882	20.319	3.686	5.006	4.129
ACS: (45,1,100)	113914.530	23944.380	21.020	3.567	5.246	4.315
ACS: (45,1,55,2,70)	76139.111	8018.493	10.531	6.518	1.901	1.666
ACS: (45,1,55,2,80)	97399.951	15034.767	15.436	4.551	3.579	3.000
ACS: (45,1,55,2,90)	109090.892	21490.948	19.700	3.600	5.108	4.179
ACS: (45,1,55,2,100)	110604.081	22531.952	20.372	3.486	5.352	4.365
USPSTF: (50,2,74)	81629.827	11033.332	13.516	4.983	2.673	2.297

Table 2.4 Harm-benefit analysis of in-practice policies implemented to the U.S. female population for the case of perfect adherence. (BC represents "breast cancer")

Policy	Expected number of screen-detected BC	Expected number of overdiagnosis	Ratio of overdiagnosed cases to screen-detected cases (%)	Expected number of lives saved per overdiagnosis	Ratio of Overtreatment to BC care cost (%)	Ratio of Overtreatment to BC care cost (model-based) (%)
ACR: (40,1,70)	160585.013	17457.084	10.871	6.718	3.840	2.906
ACR: (40,1,80)	202362.361	32272.335	15.947	4.683	7.079	5.004
ACR: (40,1,90)	226447.085	46452.839	20.514	3.676	10.160	6.921
ACR: (40,1,100)	230694.572	49537.494	21.473	3.517	10.826	7.329
ACS: (45,1,70)	156870.832	17409.910	11.098	6.520	3.848	2.921
ACS: (45,1,80)	198648.181	32225.160	16.222	4.573	7.087	5.024
ACS: (45,1,90)	222732.904	46405.664	20.835	3.599	10.168	6.946
ACS: (45,1,100)	226980.393	49490.319	21.803	3.444	10.834	7.354
ACS: (45,1,55,2,70)	149109.398	15700.510	10.529	6.520	3.722	2.822
ACS: (45,1,55,2,80)	190734.669	29437.234	15.434	4.551	7.007	4.928
ACS: (45,1,55,2,90)	215999.260	43540.736	20.158	3.521	10.344	6.973
ACS: (45,1,55,2,100)	220213.866	46440.190	21.089	3.372	11.022	7.380
USPSTF: (50,2,74)	159834.618	21602.270	13.515	4.983	5.233	3.841

2.4.2 Lifetime Breast Cancer Mortality Risk Results

Due to the wide variation in the overdiagnosis risk estimations in the literature, we choose to use a different measure of interest, rather than overdiagnosis, to validate our model and make sure that the proposed model represents the breast cancer natural history in the U.S. female population. To do so, we estimate the breast cancer lifetime mortality risk derived from the proposed model and compare the results with the lifetime mortality risk reported by ACS for some in-practice policies. We also compare our results with the estimates provided by two other studies in the literature [15, 22].

The ACS reports a lifetime mortality risk of 1 in 39 (2.57%) [31], which is calculated based on the mortality data adopted from the U.S. National Cancer Institute's Surveillance Epidemiology and End Results (SEER) database from 2010 through 2012. We estimate the lifetime mortality risks for ACR and different variations of the ACS policies. Table 2.5 shows the results. Note that the screening policy recommended by the ACS prior to 2015 was similar to the ACR policy (that is, annual screening starting at age 40). Our estimated lifetime breast cancer mortality risk for this policy is 3.23%, which is comparable to the ACS report. The 0.54% difference with the ACS report can be attributed to the fact that our model does not include other forms of cancer detection such as clinical breast exam (CBE) and breast self-examination (BSE). Please note that incorporation of CBE and BSE does not affect the mammography-induced overdiagnosis risk estimation, which is the focus of this study. This estimate is also comparable with 3.57% estimate reported by Maillart et al. [15], and 2.56% reported by Madadi et al. [22].

Table 2.5 Lifetime breast cancer mortality risk for some in-practice policies

Policy	Lifetime breast cancer mortality risk
ACR: (40,1,100)	0.0323
ACS: (45,1,80)	0.0359
ACS: (45,1,90)	0.0349
ACS: (45,1,100)	0.0342
ACS: (45,1,55,2,80)	0.0417
ACS: (45,1,55,2,90)	0.0413
ACS: (45,1,55,2,100)	0.0407

2.5 Conclusion

In this study, we estimate different measures of overdiagnosis risk of mammography screening while incorporating uncertainty in patients' adherence behaviors. Given the low rate of compliance with in-practice mammography screening recommendation, adherence behavior is a necessary factor to include. The measures of overdiagnosis risk investigated in this study include the common measure of estimation previously reported in other studies; that is, the proportion of detected cancers that are overdiagnosed in a screened population, mortality risk considering overdiagnosis, and overtreatment costs.

We analyze the harm-benefit trade-off of some in-practice policies by measuring the number of lives that are saved per each overdiagnosed case. We also estimate the associated proportion of overtreatment cost to breast cancer care cost for each policy. Our results suggest that policies with stopping age of 70 have higher numbers of lives saved per overdiagnosed cases. In addition, policies with stopping age 70 have a lower ratio of overtreatment costs to the total breast cancer care cost, which supports stopping of mammography screening at age 70.

There are several directions for future work. Given the disparity in incidence and mortality among different races, future work will include calibrating the models to determine race-specific overdiagnosis risk. Previous studies on optimal breast cancer screening did not incorporate the possibility of overdiagnosis. A possible future direction could be developing an optimization model to derive optimal screening policy that controls the risk of overdiagnosis.

CHAPTER 3

ESTIMATING TRANSITION PROBABILITY MATRIX IN A SPARSE UNBALANCED MARKOV CHAIN

3.1 Introduction

Multi-state Markov chain models have been receiving increasing attention in medical and public health research where health status, responses to treatment, or dynamic characteristics of a patient are represented by several states. The longitudinal data collected on patients over time can be modeled as a stochastic process. In medical longitudinal studies, it is inevitable to encounter missing data because of patient's missed scheduled visits, different testing times, and long time interval between observations. The missingness in longitudinal data could be categorized into two types: dropouts, where the patient doesn't show up for a scheduled visit and intermittent missingness, where patients usually show up only at intermittent visits [32].

The process that governs the likelihood of missingness is called the missing data mechanism. There are three mechanisms causing missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). In a typical longitudinal data with several individuals, each individual should have the same number of observations and we record the η^{th} patient's observation in a $T \times 1$ vector, where T denotes the number of scheduled observations. Let $\mathcal{Y} = (\mathcal{Y}^{obs}; \mathcal{Y}^{miss})$

denote a $T \times \Gamma$ matrix of T observations for Γ number of patients, where $\mathcal{Y}^{obs}$ is observed values and $\mathcal{Y}^{miss}$ is missing observations. In addition, let R denote a $T \times \Gamma$ matrix indicating whether the t^{th} observation of patient η is missing ($r_{t,\eta} = 1$) and 0 otherwise. Assume Υ is a vector of parameters describing the relationship between missingness, R and the dataset, $\mathcal{Y}$. Based on the definition presented in Laird [33] and Rubin [34], if the probability of an observation being missing is independent of the responses, then the data are said to be missing completely at random (MCAR),

$$P(R|\mathcal{Y}^{obs}, \mathcal{Y}^{miss}, \Upsilon) = P(R|\Upsilon). \tag{3.1}$$

If the probability of missing data only depends on the observed data, then the missing mechanism is missing at random,

$$P(R|\mathcal{Y}^{obs}, \mathcal{Y}^{miss}, \Upsilon) = P(R|\mathcal{Y}^{obs}, \Upsilon). \tag{3.2}$$

Finally, if the probability of missingness depends on both the missing and observed values, the mechanism is missing not at random (MNAR),

$$P(R|\mathcal{Y}^{obs}, \mathcal{Y}^{miss}, \Upsilon) = P(R|\mathcal{Y}^{obs}, \mathcal{Y}^{miss}, \Upsilon). \tag{3.3}$$

One important concept in analyzing the mechanisms of missing data is ignorability. Ignorability refers to the effect of missingness to the validity of the statistical inferences. In the MAR and MCAR mechanisms, we do not need to model the missing data mechanism as a part of the estimation process. On the other hand, non-ignorable MNAR missingness requires modeling the missing data mechanism to get good estimates of the parameters [35].

Missing observations and their effects on estimation of parameters have been well-studied using two general model-based approaches: Maximum Likelihood and Bayesian Inference. Maximum likelihood estimation (MLE) which is frequently used in the literature, obtains parameter estimates by maximizing the likelihood function of the incomplete data. On the other side, Bayesian inferences on parameters are based on the posterior distribution of the parameters. To derive a posterior distribution, prior distributions must initially be assumed for the parameters by sampling the missing variables through a sampler [36].

Expectation-Maximization (EM) algorithm is a general-purpose approach to calculate maximum likelihood estimates from incomplete data and has been proposed by Dempster et al. [37]. This method has been used for different missingness mechanisms. For example, for ignorable missing mechanism, Sherlaw-Johnson et al. [38] described expectation-maximization technique for finding the maximum likelihood estimates for a transition matrix when a system is observed at irregular time intervals. Craig and Sendi [39] summarized the maximum likelihood estimate of the transition matrix when the observation intervals have varying length or do not coincide with the cycle length. They used bootstrap in order to assess the uncertainty of the maximum likelihood estimate and to construct confidence interval for the transition matrix. Yeh et al. [32] considered a complete-case analysis using the observed one-step transitions, a non data-augmentation method (NL) by solving nonlinear equations, and a data-augmentation method (EM algorithm) for modeling a discrete-time Markov chain transition probabilities when multiple successive observations are missing at random between two observed outcomes.

For the non-ignorable missing mechanism, Troxel et al. [40] proposed a likelihood method to analyze continuous longitudinal data. They applied the method to a breast cancer dataset to confirm the non-ignorable missingness mechanism in this dataset. Albert [41] developed a transitional model for longitudinal binary data and proposed an EM algorithm for parameter estimation. Chen et al. [42] analyzed the incomplete data from progressive multi-state disease processes in which individuals are scheduled to be seen at periodic pre-scheduled assessment times using Maximum likelihood estimation via an EM algorithm. On the other paper, Chen and Zhou [43] developed methods for non-homogeneous Markov processes through time scale transformation when observation times are pre-planned with some observations missing. They have used Maximum likelihood estimation via the EM algorithm to derive parameter estimates. Van Den Hout and Matthews [44] formulated a continuous time three-state model with time-dependent transition intensities to describe transitions between healthy and unhealthy states before death. To deal with possible non-ignorable missing states, a maximum-likelihood model is proposed for the joint distribution of both the state and whether or not the state is observed. Yeh et al. [45], conducted a simulation study to examine the impact of ignorable and non-ignorable intermittent missing observations on the parameter estimates of HMM.

The other method that has been used in the literature to estimate Markov model parameters is Bayesian inference. Assoudou et al. [46] described a Bayesian estimation of the transition probabilities of a binary Markov chain observed from heterogeneous individuals. Pasanisi et al. [47] focused on Monte Carlo Markov Chain (MCMC) algorithms to perform Bayesian inference and evaluate posterior distributions

of the transition probabilities with missing-data framework. Efthimiou et al. [48] adopted a Bayesian framework in order to use a multi-state Markov model for the analysis of incomplete individual patient data for a dichotomous outcome reported over a period of time. The model accounts for patients dropping out of the study and also for patients relapsing.

Some paper chose to review both methods to analyze the incomplete data. Ghahramani and Jordan [49] reviewed the problem of learning from incomplete data from two statistical perspectives: the likelihood-based and the Bayesian. They have described a set of algorithms, derived from the likelihood-based framework, that handle clustering, classification, and function approximation from incomplete data. These algorithms are based on mixture modeling and make two distinct appeals to the expectation-maximization principle, both for the estimation of mixture components and for coping with the missing data. Deltour et al. [50] described two algorithms for estimating Markov chain models in the case of intermittent missing data in longitudinal studies, a stochastic EM algorithm and the Gibbs sampler, which is used for a full Bayesian inference. Ma et al. [51] focused on analyzing data with missing at random values within discrete-time Markov chain models. The naive method, nonlinear (NL) method, Expectation-Maximization (EM) algorithm, and a Bayesian framework, using an adjusted rejection algorithm to sample the posterior distribution, and estimating the transition probabilities with a Monte Carlo algorithm are discussed in this paper.

In this chapter, we review and apply some of the hidden Markov model parameter approximation methods to estimate breast density transition probabilities using mammography screening data from Louisiana Cancer Prevention and Control

Programs [4]. The results of this chapter will be used in the next section of dissertation to evaluate the breast cancer screening policies considering breast density as an important breast cancer risk factor.

The remainder of this chapter is as follows. In Section 3.2, we review some of the well-known methods for estimating parameters of Markov models from an incomplete data with ignorable missingness mechanism. In Section 3.3, we talk about the computational analysis of a simulated data using described methods and a description of our main dataset and the results of the parameter estimation. We summarize and conclude in Section 3.4.

3.2 A Review of the Methodologies in the Literature

In this section, we review different methods for estimating Markov model parameters for an incomplete longitudinal set of data, where different characteristics of a group of patients are scheduled to be observed at equal time intervals. We discuss methods for an ignorable mechanism of missing values. Generally, there are three categories of methods for handling incomplete data with ignorable missingness mechanism: deletion, imputation, and augmentation. Data deletion is the most simple and common method. However it causes bias in parameter estimation. Data imputation and data augmentation methods are comparatively similar methods, however, in data imputation methods missing observations substitute with imputed values but in data augmentation, parameter estimation is augmented by the information gained from assuming certain probability models from observed data [52]. Data augmentation procedures included maximum likelihood (ML) and Bayesian Inference.

In this section, we discuss each of these methods separately. The following is the list of notations used in these models.

- t: Time periods, $t = 0, 1, \cdots, T$, where T denotes the number of scheduled observations for each patient.
- Γ: Number of patients, Note that we use subscript $\eta \in \{1, ..., \Gamma\}$ in order to represent η^{th} patient.
- S: Patient's core state space, where $s_t \in S$ represents patient's state at time t. Core state space will be defined based on the type of the data. For example in analyzing breast density, we assume $S = \{d_I, d_{II}, d_{III}, d_{IV}\}$, which is the set of partially observable breast density states including almost entirely fatty tissue, scattered fibroglandular densities, heterogeneously dense class, and extremely dense class.
- $\mathfrak{S}_\eta$: State sequence of patient η, where $\mathfrak{S}_\eta = s_{\eta,1}, s_{\eta,2}, ..., s_{\eta,T}$ represents the actual states of patient η at each time point.
- Y: Observation space, where $y_t \in Y$ represents the observation that the patient gets at time t. In hidden Markov models, the state of patient is not directly observable and it would be determine through the imperfect interpretation of screening results.
- $\mathcal{Y}_\eta$: Observation sequence for patient η, where $\mathcal{Y}_\eta = \{y_{\eta,1}, y_{\eta,2}, ..., y_{\eta,T}\}$ represents the observation for patient η at time t. Note that we assume the data is not complete and some of observations are missing. In other words, we have $\mathcal{Y}_\eta = \{\mathcal{Y}_\eta^{obs}, \mathcal{Y}_\eta^{miss}\}$.
- m: Number of missing values between two observed values.

- n_{ij}: Number of one-step transitions between state i and state j among all Γ patients when the data is complete.
- $n_{i(m)j}$: Number of transitions from state i to state j with m missing observations between these two observations among all Γ patients.
- π_{s_t}: Expected patient belief state at time t that represents the occupancy distribution over core states.
- A: State transition probability matrix, where $a_{ij} \in A$ represents the probability that a patient will be in state j at time $t+1$, given that s/he is in state i at time t.
- $a_{ij}^{[m]}$: m-step transition probability from state i to state j. Note that $a_{ij}^{[1]} = a_{ij}$.
- B: Observation probability matrix, where $b_{s_t}(y_t) \in B$ represents the probability of receiving observation y_t given that patient is in state s_t at time t.
- λ: Vector of model parameters, where $\lambda = (\pi, A, B)$.

3.2.1 Data Deletion Method

Listwise Deletion

The most common approach to handle the missing data is to exclude the cases with missing values and analyze the remaining data. Listwise method can produce unbiased estimates, if the sample is large enough and the missingness mechanism is MCAR.

Pairwise Deletion

Despite the listwise method that omits the whole case if it has a missing value, pairwise method excludes only the variable of the case with missing value. Note that

pairwise method still uses the case when analyzing other variables with non-missing values. This method has the best use of data and hence is more preferable compared to listwise deletion. However, this method use different number of cases for each analyze and it causes an overestimation or underestimation of standard errors [53].

3.2.2 Data Imputation Methods

In this section, we review some of the well-known single and multiple imputation methods for handling missing values.

LOCF/NOCB Methods

One of the well-known single imputation methods for longitudinal data is Last Observation Carried Forward (LOCF). This imputation method carries the last observed non-missing value to fill in missing values at a later point. Therefore, the response remains constant at the last observed value. A similar approach named Next Observation Carried Backward (NOCB) works in the opposite direction by taking the first observation after the missing value and carrying it backward. These methods can introduce a positive or negative bias and because all the missing values for an individual are replaced with the same numbers, the within-subject variability is reduced [53].

Mean/Mode Substitution

Mean substitution method is another single imputation procedure where the missing value of a variable can be replaced with the average of known values for that variable. For categorical data, mode of the previous data can be used as substitute value for imputation. This method's estimation may lead to inconsistent bias, if the

missing values are not random or the number of missing values for different variables is not equal [53].

Regression Imputation

This approach estimates and imputes the missing data using a regression model and other relevant variables in the dataset. Therefore, it maintains all the cases and does not change the standard deviation or the shape of the distribution. However, in this method other variables in the dataset are used to impute missing observations without adding any new information. This causes an increase in sample size and consequently a decrease in the standard error [54].

K-Nearest Neighbors Imputation Method (KNN)

The KNN algorithm stores all the training cases and when it encounters a new input vector, it performs a prediction by considering its K closest training cases according to a given distance metric. A well-known distance function is the Euclidean distance which is used to identify the k nearest neighbors of each case with two inputs, $\mathcal{Y}_A$ and $\mathcal{Y}_B$,

$$d(\mathcal{Y}_A, \mathcal{Y}_B) = \sqrt{\sum_{\vartheta} d_{\vartheta}(\mathcal{Y}_{A\vartheta}, \mathcal{Y}_{B\vartheta})^2}, \tag{3.4}$$

where $d_{\vartheta}(\mathcal{Y}_{A\vartheta}, \mathcal{Y}_{B\vartheta})$ is the distance between the two cases on its ϑ^{th} attribute and it can be calculated as follow,

$$d_\vartheta(\mathcal{Y}_{A\vartheta}, \mathcal{Y}_{B\vartheta}) = \begin{cases} 1, & \text{if } \mathcal{Y}_\vartheta \text{ is missing in } \mathcal{Y}_A \text{ or } \mathcal{Y}_B, \\ ED, & \text{if data is continuous,} \\ 1, & \text{if data is categorical and } \mathcal{Y}_{A\vartheta} \neq \mathcal{Y}_{B\vartheta}, \\ 0, & \text{if data is categorical and } \mathcal{Y}_{A\vartheta} = \mathcal{Y}_{B\vartheta}, \end{cases} \tag{3.5}$$

where ED is the normalized Euclidean distance and can be calculated as follow,

$$ED = \frac{\mid \mathcal{Y}_{A\vartheta} - \mathcal{Y}_{B\vartheta} \mid}{\max(\mathcal{Y}_\vartheta) - \min(\mathcal{Y}_\vartheta)}. \tag{3.6}$$

Given an incomplete vector of dataset, for each missing value in $\mathcal{Y}$, the KNN imputation method finds the corresponding set of k closest training cases with observed values in the incomplete feature to be imputed. Then, the unknown value $\mathcal{Y}_\vartheta$ can be estimated by the mean or mode of k nearest neighbors [55] and the parameter can be estimated using the imputed complete dataset.

3.2.3 Data Augmentation Methods

In this section, we talk about methods that use maximum likelihood and Bayes theorem as a basic for calculation of the parameters. In order to show the calculation burden of the maximum likelihood methods that do not use data augmentation, first we talk about a nonlinear method introduced by Yeh et al. [32] and then we talk about expectation-maximization and Viterbi algorithms that use data augmentation to decrease the complexity of calculations. Later in this section, we will talk about Bayesian inference assumptions and algorithm.

Nonlinear Maximum Likelihood Method

Nonlinear method uses all the one and multi-step transitions to compute the parameters. For example, the $(m+1)$-step transition probability from state i to state j can be computed by summing over all possible routes,

$$a_{ij}^{[m+1]} = \sum_{l_1=1}^{S} a_{il_1} a_{l_1 j}^{[m]} = \sum_{l_1=1}^{S} \sum_{l_2=1}^{S} a_{il_1} a_{l_1 l_2} a_{l_2 j}^{[m-1]} = \sum_{l_1=1}^{S} \sum_{l_2=1}^{S} \dots \sum_{l_m=1}^{S} a_{il_1} a_{l_1 l_2} \dots a_{l_{m-1} l_m} a_{l_m j}. \tag{3.7}$$

The log likelihood function for this equation can be expressed as,

$$L(\lambda|Y) = \sum_{\eta=1}^{\Gamma} \sum_{i=1}^{S} \sum_{j=1}^{S} \sum_{m=0}^{T-2} n_{i(m)j} \log(a_{ij}^{[m+1]}). \tag{3.8}$$

The MLE of the transition probabilities can be computed by solving $\frac{\Delta L(\lambda|Y)}{\Delta \lambda} = 0$. This MLE function is nonlinear due to the summation inside the logarithm and should be solved with nonlinear procedures.

Expectation-Maximization Algorithm

Expectation-Maximization(EM) algorithm is an iterative approach for a broad range of Markov model parameter estimation. On each iteration of EM algorithm there is an expectation step and a maximization step. In the expectation step (E-Step), a log-likelihood function is evaluated using the current estimate for the parameters. The second step (the M-step) of the EM algorithm is to maximize the expectation function we computed in the E-step. To define the maximum-likelihood function, we assume that data $\mathcal{Y}^{obs}$ is observed and is generated by some distribution. We assume that a complete dataset exists, $\mathcal{Y} = (\mathcal{Y}^{obs}; \mathcal{Y}^{miss})$ and also assume a joint density function as,

$$P(\mathcal{Y}|\lambda) = P(\mathcal{Y}^{obs}, \mathcal{Y}^{miss}|\lambda) = P(\mathcal{Y}^{miss}|\mathcal{Y}^{obs}, \lambda) P(\mathcal{Y}^{obs}|\lambda). \tag{3.9}$$

In the first step, EM algorithm finds the expected value of the complete data log-likelihood with respect to the unknown data $\mathcal{Y}^{miss}$ given the observed data $\mathcal{Y}^{obs}$ and the current parameter estimates,

$$Q(\lambda, \lambda^{\nu-1}) = E[\log P(\mathcal{Y}^{obs}, \mathcal{Y}^{miss}|\lambda)|\mathcal{Y}^{obs}, \lambda^{\nu-1}], \tag{3.10}$$

where $\lambda^{\nu-1}$ is the set of current parameters estimates that we use to evaluate the expectation function and λ is the set of new parameters that we optimize to increase Q . The second step (M-step) of the EM algorithm is to maximize the expectation function we computed in the first step. That is,

$$\lambda^{\nu} = \arg\max_{\lambda} Q(\lambda, \lambda^{\nu-1}). \tag{3.11}$$

These two steps are repeated until convergence. Each iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function.

If the Markov model is fully observable, Equation (3.10) can be written as,

$$Q(\lambda, \lambda^{\nu-1}) = E[\log P(\mathcal{Y}^{obs}, \mathcal{Y}^{miss}|\lambda)|\mathcal{Y}^{obs}, \lambda^{\nu-1}] = \sum_{i=1}^{S}\sum_{j=1}^{S} n_{ij}^{\nu} \log(a_{ij}), \tag{3.12}$$

where $n_{ij}^{\nu} = n_{i(0)j} + m_{ij}^{\nu}$ is the ν^{th} iterate of the expected number of transitions from state i to state j for the whole data. Note that $n_{i(0)j}$ is the number of one-step transitions and m_{ij}^{ν} is the number of multi-step transitions from state i to state j at ν^{th} iteration.

In each iteration, M-step maximizes the Q function in Equation (3.12). The $(\nu+1)^{th}$ iteration of the estimates of transition from state i to state j is,

$$\hat{a}_{ij}^{\nu+1} = \frac{n_{ij}^{\nu}}{n_{i+}^{\nu}}, \quad \text{where} \quad n_{i+}^{\nu} = \sum_{\iota=1}^{S} n_{i\iota}^{\nu}. \tag{3.13}$$

The estimated transition probabilities $a_{ij}^{\nu+1}$ are substituted into the Q function, and the steps E and M were repeated until the Q function converges.

In order to learn the parameters of an HMM model, the EM algorithm has been derived for finding the maximum-likelihood estimate of the parameters of a hidden Markov model given a set of observed feature vectors. This algorithm is known as the Baum-Welch algorithm. There are three basic problems that EM algorithm tries to solve in each iteration and finally estimate the parameters: evaluation problem, finding optimal state sequence problem, and optimization problem [56, 57]. Here we will talk about each problem and the required steps for Baum-Welch algorithm.

In Evaluation problem, we compute the probability of the observation sequence $\mathcal{Y}_\eta = y_{\eta,1}, y_{\eta,2}, ..., y_{\eta,T}$ for patient η given the parameters of a model, $\lambda = (A, B, \pi)$. One way to calculate this probability is through enumerating every possible state sequence of length T. Consider fixed state sequence of $\mathfrak{S}_\eta = s_{\eta,1}, s_{\eta,2}, ..., s_{\eta,T}$. The probability of observation sequence of $\mathcal{Y}_\eta$ for $\mathfrak{S}_\eta$ state sequence is,

$$P(\mathcal{Y}_\eta|\mathfrak{S}_\eta, \lambda) = \prod_{t=1}^{T} P(y_{\eta,t}|\mathfrak{S}_\eta, \lambda) = b_{s_{\eta,1}}(y_{\eta,1}) b_{s_{\eta,2}}(y_{\eta,2}) ... b_{s_{\eta,T}}(y_{\eta,T}). \tag{3.14}$$

In addition, the probability of state sequence $\mathfrak{S}_\eta$ can be written,

$$P(\mathfrak{S}_\eta|\lambda) = \pi_{s_{\eta,1}} a_{s_{\eta,1},s_{\eta,2}} a_{s_{\eta,2},s_{\eta,3}} ... a_{s_{\eta,T-1},s_{\eta,T}}. \tag{3.15}$$

The probability of $\mathcal{Y}_\eta$ given the model parameters can be obtained by summing the joint probability over all possible state sequences:

$$P(\mathcal{Y}_\eta|\lambda) = \sum_{\mathfrak{S}_\eta} P(\mathcal{Y}_\eta|\mathfrak{S}_\eta, \lambda)P(\mathfrak{S}_\eta|\lambda)$$

$$= \sum_{\mathfrak{S}_\eta} \pi_{s_{\eta,1}} b_{s_{\eta,1}}(y_{\eta,1}) a_{s_{\eta,1},s_{\eta,2}} b_{s_{\eta,2}}(y_{\eta,2}) ... a_{s_{\eta,T-1},s_{\eta,T}} b_{s_{\eta,T}}(y_{\eta,T}). \tag{3.16}$$

The logic of this equation is as follow. Let's assume patient η is in state $s_{\eta,1}$ with probability $\pi_{s_{\eta,1}}$ and observes observation $y_{\eta,1}$ with probability of $b_{s_{\eta,1}}(y_{\eta,1})$ at time 1. Then, the patient makes a transition from state $s_{\eta,1}$ to state $s_{\eta,2}$ at time 2 with probability $a_{s_{\eta,1},s_{\eta,2}}$ and observe observation $y_{\eta,2}$ with probability $b_{s_{\eta,2}}(y_{\eta,2})$ until s/he gets to time T. An efficient procedure named forward-backward can be used to solve this problem. The forward-backward algorithm computes the marginal probability of a given state at a time point. In this procedure, the forward variable, α, is defined as the probability of seeing the partial sequence $(y_{\eta,1}, ..., y_{\eta,t})$ for patient η and ending up in state i at time t given the model parameters.

$$\alpha_{\eta,t}(i) =$$

$$P(\mathcal{Y}_{\eta,1} = y_{\eta,1}, ..., \mathcal{Y}_{\eta,t} = y_{\eta,t}, s_{\eta,t} = i|\lambda), \quad t = 1, 2, ..., T, \eta = 1, 2, .., \Gamma. \tag{3.17}$$

where,

$$\alpha_{\eta,1}(i) = \pi_i b_i(y_{\eta,1}), \tag{3.18}$$

$$\alpha_{\eta,t+1}(j) = \Big[\sum_{i=1}^{S} \alpha_{\eta,t}(i) a_{i,j}\Big] b_j(y_{\eta,t+1}), \tag{3.19}$$

$$P(\mathcal{Y}_\eta|\lambda) = \sum_{i=1}^{S} \alpha_{\eta,T}(i). \tag{3.20}$$

Similarly, we can calculate backward variable $\beta_t(i)$ defined as,

$$\beta_{\eta,t}(i) =$$

$$P(\mathcal{Y}_{\eta,t+1} = y_{\eta,t+1}, ...\mathcal{Y}_{\eta,T} = y_{\eta,T}|s_{\eta,t} = i, \lambda), \quad t = 1, 2, ..., T-1, \eta = 1, 2, ..., \Gamma. \tag{3.21}$$

Backward variable is the probability of ending sequence $(y_{\eta,t+1}, ..., y_{\eta,T})$ for patient η, given that we started at state i at time t and the model parameters. Note that backward variable will be used in solution for optimization problem. We can solve Equation (3.21) using two steps:

$$\beta_{\eta,T}(i) = 1 \qquad 1 \leq i \leq S, \tag{3.22}$$

$$\beta_{\eta,t}(i) = \sum_{j=1}^{S} a_{i,j} b_j(y_{\eta,t+1}) \beta_{\eta,t+1}(j). \tag{3.23}$$

In the second problem, we try to compute the probability of a state sequence $\mathfrak{S}_\eta = s_{\eta,1}, s_{\eta,2}, ..., s_{\eta,T}$ for patient η given the observation sequence, $\mathcal{Y}_\eta = y_{\eta,1}, y_{\eta,2}, ..., y_{\eta,T}$ and the parameters of a hidden Markov model, $\lambda = (\pi, A, B)$ to find the optimal state sequence. The optimality criteria that has been used for this section is expected to maximize the expected number of correct individual states. For this purpose we define,

$$\gamma_{\eta,t}(i) = P(s_{\eta,t} = i|\mathcal{Y}_\eta, \lambda) = \frac{P(s_{\eta,t} = i, \mathcal{Y}_\eta|\lambda)}{P(\mathcal{Y}_\eta|\lambda)} = \frac{\alpha_{\eta,t}(i)\beta_{\eta,t}(i)}{\sum_{j=1}^{S} \alpha_{\eta,T}(j)}, \tag{3.24}$$

where $\gamma_{\eta,t}(i)$ is the probability that patient η is in state i at time t, given the observation sequence of $\mathcal{Y}_\eta$ and the model parameters of λ. Using $\gamma_t(i)$, we can find the most

likely sequence of states for an observation sequence,

$$s_{\eta,t} = \underset{1\leq i\leq S}{\arg\max}[\gamma_{\eta,t}(i)]. \tag{3.25}$$

In optimization problem, we implement an iterative procedure to estimate the parameters of the model. In this procedure, HMM parameters will be updated and improved iteratively. For this purpose, first we define $\xi_{\eta,t}(i,j)$ as the probability of patient η being in state i at time t and in state j at time $t+1$, given the parameters of the model and observation sequence.

$$\begin{aligned}\xi_{\eta,t}(i,j) = P(s_{\eta,t}=i, s_{\eta,t+1}=j|\mathcal{Y}_\eta,\lambda) &= \frac{P(s_{\eta,t}=i, s_{\eta,t+1}=j, \mathcal{Y}_\eta|\lambda)}{P(\mathcal{Y}_\eta|\lambda)} \\ &= \frac{\alpha_{\eta,t}(i)a_{ij}b_j(y_{\eta,t+1})\beta_{\eta,t+1}(j)}{\sum_{i=1}^{S}\alpha_{\eta,T}(i)}\,.\end{aligned} \tag{3.26}$$

Note that based on our definition for $\gamma_{\eta,t}(i)$ as the probability of patient η being in state i at time t, given the observation sequence and the model, we can relate $\gamma_{\eta,t}(i)$ to $\xi_{\eta,t}(i,j)$ by summing over j, giving,

$$\gamma_{\eta,t}(i) = \sum_{j=1}^{S}\xi_{\eta,t}(i,j). \tag{3.27}$$

In addition, summing $\gamma_{\eta,t}(i)$ over t, we get the expected number of transitions from state i. Similarly, summation of $\xi_{\eta,t}(i,j)$ over t can be interpreted as the expected number of transitions from state i to state j.

Using the above formulas, we can estimate HMM parameters in each iteration.

$$\begin{aligned}\pi_i^\nu &: \text{Expected probability of being in state } i \text{ at time } 1 \\ &= \frac{\sum_{\eta=1}^{\Gamma}\gamma_{\eta,1}^\nu(i)}{\sum_{\eta=1}^{\Gamma}\sum_{i=1}^{S}\gamma_{\eta,1}^\nu(i)},\end{aligned} \tag{3.28}$$

$$a_{ij}^{\nu} : \frac{\text{Expected number of transitions from state } i \text{ to } j}{\text{Expected number of transitions from state } i} = \frac{\sum_{\eta=1}^{\Gamma} \sum_{t=1}^{T-1} \xi_{\eta,t}^{\nu}(i,j)}{\sum_{\eta=1}^{\Gamma} \sum_{t=1}^{T-1} \gamma_{\eta,t}^{\nu}(i)}, \tag{3.29}$$

$$b_i(j)^{\nu} : \frac{\text{Expected number of times in state } i \text{ and observing } j}{\text{Expected number of times in state } i} = \frac{\sum_{\eta=1}^{\Gamma} \sum_{t=1}^{T} \gamma_{\eta,t}^{\nu}(i)\delta(y_{\eta,t}=j)}{\sum_{\eta=1}^{\Gamma} \sum_{t=1}^{T} \gamma_{\eta,t}^{\nu}(i)\delta(y_{\eta,t}\neq .)}, \tag{3.30}$$

where $\delta(y_{\eta,t} = j)$ is the index function that equals 1 if the η^{th} patient at time t is observed to be in state j, and zero otherwise.

Viterbi Algorithm

The Viterbi algorithm estimates the state sequence of a discrete-time finite-state Markov process by recursively taking the most probable path that could lead to each cell. Based on Rabiner [57] to implement this algorithm first we define,

$$\zeta_{\eta,t}(j) = \max_{s_{\eta,1},s_{\eta,2},s_{\eta,t-1}} P(y_{\eta,1}, y_{\eta,2}, ..., y_{\eta,t}, s_{\eta,1}, s_{\eta,2}, ..., s_{\eta,t-1}, s_{\eta,t} = j|\lambda). \tag{3.31}$$

Note that $\zeta_{\eta,t}$ represents the most probable path by taking the maximum over all possible previous state sequences given that we had already computed the probability of being in every state at time $t-1$. For a given state s_j at time t and for patient η, the value $\zeta_{\eta,t}(j)$ can also be computed as,

$$\zeta_{\eta,t}(j) = \max_{i=1}^{S} \zeta_{\eta,t-1}(i) a_{ij} b_j(y_{\eta,t}). \tag{3.32}$$

The actual state sequence is retrieved by tracking the transitions that maximize the $\zeta_{\eta,t}(j)$ scores for each patient, time point t and state j using $\psi_{\eta,t}(j)$.

Modified Viterbi algorithm can be used when the dataset contains some missing values [58]. The steps required for this algorithm is

1. Detect the most probable path at each time over all possible previous state sequences.

   If $t = 1$,

$$\zeta_{\eta,1}(i) = \begin{cases} \pi_i, & y_{\eta,1} \text{ is missing,} \\ \pi_i b_i(y_{\eta,1}), & \text{otherwise,} \end{cases} \tag{3.33}$$

$$\psi_{\eta,1}(i) = 0.$$

   If $t = 2, \cdots, T$,

$$\zeta_{\eta,t}(j) = \begin{cases} \max_{1\le j\le S}[\zeta_{\eta,t-1}(i)a_{ij}], & y_{\eta,1}\text{is missing,} \\ \max_{1\le j\le S}[\zeta_{\eta,t-1}(i)a_{ij}b_j(y_{\eta,t})], & \text{otherwise,} \end{cases} \tag{3.34}$$

$$\psi_{\eta,t}(j) = \arg\max_{1\le i\le S}[\zeta_{\eta,t-1}(i)a_{ij}].$$

2. Update the terminal state,

$$s_T = \arg\max_{1\le i\le S}[\zeta_{\eta,T}(i)]. \tag{3.35}$$

3. State sequence backtracking

$$s_{\eta,t} = \psi_{\eta,t+1}(s_{\eta,t+1}). \tag{3.36}$$

In each iteration, the algorithm generates a new set of parameters based on the most probable state sequence and previous estimation of parameters. The Viterbi algorithm is expensive, both in terms of memory and compute time [59].

Bayesian Inference (MCMC)

In this section, we review Gibbs sampler method derived from Vidotto et al. [60] and investigate the performance of mixture hidden models for missing categorical longitudinal data. This model is implemented using Bayesian inference method which requires defining the prior distribution of the model parameters in order to obtain the posterior distribution of the model's unknown parameters given the observed data ($\mathcal{Y}^{obs}$). In this method, the V sets of imputations are obtained from the posterior predictive distribution of the missing data. To implement this method, first, we sample parameter values, $\lambda^{\nu}(\nu = 1, ..., V)$, from $P(\lambda|\mathcal{Y}^{obs})$ and then the imputations are drawn from $P(\mathcal{Y}^{mis}|\lambda^{\nu})$.

To estimate the parameters of hidden Markov model, a Multinomial distribution is defined for each of the categorical variables as follow,

- $(s_{\eta,1}) \propto$ Multinomial(π), where $\pi = (\pi_1, ..., \pi_S)$ is the initial state probabilities vector.
- $(s_{\eta,t+1}|s_{\eta,t} = i) \propto$ Multinomial$(a_i), \forall_t$, where a_i is the transition probability vector starting from state i.
- $(y_{\eta,t}|s_{\eta,t} = i) \propto$ Multinomial(b_i), where $b_i = (b_i(1), ..., b_i(Y))$ is the observation probability vector given we are in state i.

The core of Bayesian inference is to combine two different likelihood and prior distributions into one posterior distribution to estimate the parameters. Since the variables are categorical and we have multinomial distribution as likelihood distribution, if we choose a conjugate prior distribution, instead of multiplying the likelihood with the prior distribution, the posterior distribution can be updated easily using the

prior parameters. In view of the fact that Dirichlet distribution is the conjugate of multinomial distribution, we define the prior distribution of parameters as follow,

- $\pi \sim \text{Dirichlet}(\omega)$ with $\omega = (\omega_1, ..., \omega_S), \quad \omega_s > 0, \quad \forall_s.$
- $a_i \sim \text{Dirichlet}(\theta)$, with $\theta = (\theta_1, ..., \theta_S), \quad \theta_s > 0, \quad \forall_s.$
- $b_i \sim \text{Dirichlet}(\varphi_i)$, with $\varphi_i = (\varphi_i(1), ..., \varphi_i(Y)), \quad \varphi_i > 0, \quad \forall_i.$

To choose the model parameters for each these Dirichlet distribution, if we have no previous knowledge about the imputation of model parameters, a symmetric Dirichlet priors can be chosen (Dirichlet$(c_1, c_2, ..., c_s)$ where $c_1 = c_2 = ... = c_s$).

The Gibbs sampler has been used in this section for model estimation and imputation. The algorithm steps are as follow,

1. Sample hidden states for each patient $\eta = 1, ..., \Gamma$ and for all time points $t = 1, ..., T$ from a conditional categorical distribution defined by the following probabilities.

$$P(s^{\nu}_{\eta,t} | \lambda^{\nu-1}, y_{\eta,t}).$$

At $t = 1$, for all patients $\eta = 1, ..., \Gamma$, and for all states $j = 1, ..., S$,

$$P(s^{\nu}_{\eta,1} = j | \lambda^{\nu-1}, y_{\eta,1}) = \frac{\pi_j^{\nu-1} b_j(y_{\eta,1})^{\nu-1}}{\sum_c \pi_c^{\nu-1} b_c(y_{\eta,1})^{\nu-1}}. \tag{3.37}$$

For $t = 2, ..., T$ and for patients $\eta = 1, ..., \Gamma$ we have,

$$P(s^{\nu}_{\eta,t} = j | \lambda^{\nu-1}, y_{\eta,t}) = \frac{\sum_i P(s^{\nu}_{\eta,t-1} = i | \lambda^{\nu-1}, y_{\eta,t-1}) a_{ij}^{\nu-1} b_j(y_{\eta,t})^{\nu-1}}{\sum_c P(s^{\nu}_{\eta,t} = c | \lambda^{\nu-1}, y_{\eta,t-1}) b_c(y_{\eta,t})^{\nu-1}}. \tag{3.38}$$

2. Update each parameter values as follow,

   For initial state probabilities $(t = 1)$ $\forall_\eta$,

$$\pi^\nu | s^\nu_{\eta,1} \sim$$

$$\text{Dirichlet}(\omega_1 + \sum_\eta I_{\eta,1}(s^\nu_{\eta,1} = 1), ..., \omega_S + \sum_\eta I_{\eta,1}(s^\nu_{\eta,1} = S)), \tag{3.39}$$

   where $I_{\eta,t}(s^\nu_{\eta,t} = s) = 1$ if for patient η, $s^\nu_{\eta,t} = s$ and 0 otherwise.

   For transition probabilities and for $i = 1, ..., S$ and $\forall_{\eta,t\geq 1}$,

$$a^\nu_i | s^\nu_{\eta,t}, s^\nu_{\eta,t+1} \sim$$

$$\text{Dirichlet}(\theta_1 + \sum_{\eta, s^\nu_{\eta,t}=i} I_{\eta,t+1}(s^\nu_{\eta,t+1} = 1), ..., \theta_S + \sum_{\eta, s^\nu_{\eta,t}=i} I_{\eta,t+1}(s^\nu_{\eta,t+1} = S)). \tag{3.40}$$

   For the emission probabilities and for $j = 1, ..., S$ and $\forall_{\eta,t}$,

$$b^\nu_j | s^\nu_{\eta,t}, Y^{obs} \sim$$

$$\text{Dirichlet}(\varphi_j(1) + \sum_{\eta, s^\nu_{\eta,t}=j} I_{\eta,t}(y_{\eta,t} = 1), ..., \varphi_j(Y) + \sum_{\eta, s^\nu_{\eta,t}=j} I_{\eta,t}(y_{\eta,t} = Y), \tag{3.41}$$

   where $I_{\eta,t}(y_{\eta,t}) = 1$ if $y_{\eta,t} = y$ and $y_{\eta,t} \in Y^{obs}$ and 0 otherwise.

3.3 Computational Analysis

In this section, first, we conduct a simulation to study the accuracy of the Markov model parameter estimates obtained from Baum-Welch and Bayesian Inference algorithms. The simulated data will have the same number of states, patients, and missing mechanism as our main dataset. In addition, we present the estimated Markov

parameters and the bootstrapping standard errors for breast density longitudinal dataset in Section 3.3.2.

3.3.1 Monte Carlo Simulation

In order to understand the impact of missing values and their uncertainty in estimating model parameters, we construct a Mont Carlo simulation. The objective of this simulation is to approximate the sampling distribution of an estimator by generating large number of random independent datasets from a known values of the parameters and computing the estimator for each dataset using different algorithms. The sample mean of the estimates over all datasets is an estimate of the mean of the sampling distribution of the estimator and the standard deviation of the estimates over all datasets is an estimate of the standard deviation of the sampling distribution. The following are the basics of this simulation model,

- Generate $\mathcal{D}$ datasets from true statistical parameters with true value of λ.
- For each dataset, estimate λ using each of the algorithms.
- Obtain bias, mean square error, and standard error for λ.

Let $\hat{\lambda}$ be the estimator of true value of parameters and $\hat{\lambda}_{\int}$ be the estimate obtained from the $\int^{th}$ dataset, where $\int = 1, ..., \mathcal{D}$. We would want to calculate an unbiased estimator for the parameters. Thus, we would hope that the mean of the sampling distribution is close to the true value of λ with only small bias. To assess this difference of values, we calculate the Monte Carlo bias for estimator $\hat{\lambda}$ from the sampling distribution as follow,

$$\text{Bias} = \frac{1}{\mathcal{D}} \sum_{\int=1}^{\mathcal{D}} \hat{\lambda}_{\int} - \lambda. \tag{3.42}$$

To compare the precision of two estimators based on the $\mathcal{D}$ estimates of each, we could compare the Monte Carlo mean square error (MSE) for each estimator. The estimated MSE based on the $\mathcal{D}$ estimates $\hat{\lambda}_\jmath$ is defined as,

$$\text{MSE} = \frac{1}{\mathcal{D}} \sum_{\jmath=1}^{\mathcal{D}} (\hat{\lambda}_\jmath - \lambda)^2 = \frac{1}{\mathcal{D}} \sum_{\jmath=1}^{\mathcal{D}} (\hat{\lambda}_\jmath - \bar{\lambda})^2 + (\bar{\lambda} - \lambda)^2, \tag{3.43}$$

where $\bar{\lambda} = \frac{1}{\mathcal{D}} \sum_{\jmath=1}^{\mathcal{D}} \hat{\lambda}_\jmath$. Note that second part of Equation (3.43) includes square values of the bias and standard error of estimator based on $\mathcal{D}$ observations.

For our purpose, we generated simulated datasets with the same number of states, patients, missing mechanism, and missing value percentage as our main dataset which will be discussed in Section 3.3.2. The simulation was repeated 1000 runs for each method under each scenario. We estimated the transition parameters and compared the estimates to the true parameter values in terms of bias, standard errors, and mean square errors (MSE). Based on simulation results, when the model is fully observable both algorithms estimate parameters with high accuracy. However, for a hidden state sequence of observations, Baum-Welch algorithm has lower bias and lower required calculation time.

3.3.2 Main Data Description and Bootstrapping

This section describes a longitudinal dataset that is used to quantify the dynamics of breast density which will be used in our last chapter for evaluation of breast screening policies considering patients' breast density. This longitudinal data has been gathered from 436 patients since 2016 at Louisiana Cancer Prevention and Control Center [4]. The records of patients give us information about the patient's date of birth, date of screening, BI-RADS health and density states, future recommendation,

and radiologist name. This dataset contains missing values which comes from the missed scheduled visits of patients for evaluation of their health and breast density status. We assume the missingness mechanism in this dataset is ignorable since the probability of a patient shows up for breast screening mostly depends on their previous screening results rather than current observation. This comes from the fact that the symptomatic breast cancer usually happens at the very advanced stage of a cancer, which means that patients who show up for screening are mostly because of previous suspicious results or to follow-up the screening policy recommendation. Furthermore, we divided observations based on the age of patients into two groups $(40-55/55+)$ in order to consider the age and menopause effect on breast density dynamics in our analysis. The summary counts for observed breast density status and transitions for each group are shown in Table 3.1.

Table 3.1 Transition counts between different states in breast density dataset

		40+				40-55				55+			
		d_I [1]	d_{II} [2]	d_{III} [3]	d_{IV} [4]	d_I	d_{II}	d_{III}	d_{IV}	d_I	d_{II}	d_{III}	d_{IV}
one-step	d_I	41	8	0	0	11	5	0	0	26	2	0	0
	d_{II}	11	200	10	0	4	61	4	0	7	136	6	0
	d_{III}	0	29	126	2	0	13	61	2	0	15	60	0
	d_{IV}	0	0	3	2	0	0	1	1	0	0	1	1
two-step	d_I	8	1	0	0	4	0	0	0	5	1	0	0
	d_{II}	5	52	3	0	2	16	1	0	1	27	2	0
	d_{III}	0	12	33	0	0	3	17	0	0	8	15	0
	d_{IV}	0	0	2	2	0	0	1	1	0	0	1	1

[1] Almost entirely fatty
[2] Scattered fibroglandular densities
[3] Heterogeneously dens
[4] Extremely dense

Due to small sample size for each age group and in order to assess the uncertainty of estimation, we also conduct a bootstrap that can be described as follow; we generate

a bootstrap set, by randomly sampling $n_{\mathfrak{B}}$ patients with replacement from the main dataset. Based on Efron's recommended bootstrap ([61]), the sample size is assumed to be the same as the original sample size and the sampling distribution is uniform, which means that each of the patients in dataset has the same probability of being selected. We repeat this sampling τ times to generate τ bootstrap sets, $\mathfrak{B}_1$, $\mathfrak{B}_2$,..., $\mathfrak{B}_\tau$. Finally, we construct a sampling distribution with these τ bootstrap statistics.

The initial values for the baseline distribution with starting ages 40 and 55 were chosen based on population distribution in Mandelblatt et al. [62] as:

$$\pi^{(40)} = \begin{bmatrix} 0.046 & 0.338 & 0.472 & 0.144 \end{bmatrix}$$

$$\pi^{(55)} = \begin{bmatrix} 0.098 & 0.471 & 0.373 & 0.058 \end{bmatrix}$$

We also chose the initial information matrix arbitrarily as,

$$B = \begin{bmatrix} 0.85 & 0.12 & 0.02 & 0.01 \\ 0.07 & 0.75 & 0.15 & 0.03 \\ 0.02 & 0.1 & 0.8 & 0.08 \\ 0.01 & 0.04 & 0.15 & 0.8 \end{bmatrix}$$

For the initial transition probabilities, we used the normalized one-step observation counts in each dataset. Table 3.2 shows the parameter estimator values, standard error and bootstrapping 95% confidence interval. Based on the estimates for yearly transitions, the successive observations are more likely to recur, which means that the amount of fibroglandular tissue in a patient's breast is most likely to stay

similar as the previous year for all four classes. However, the results show a higher probability of decline and lower probability of increased breast density per year as the patient's age increases. In addition, higher breast tissue density shows a higher decline each year. For example there is 28.6% probability of transitioning from state 4 to state 3 compared to 12.9% and 3.4% probability of transitioning from state 3 to state 2 and from state 2 to state 1, respectively. Based on the results for all age groups, the probability of transitioning between two nonconsecutive states is almost impossible. This means that if the state of patient's breast density is 3, the probability of changing state to 1 in one year is zero.

Table 3.2 HMM estimates, standard errors, and 95% confidence intervals for breast density dataset

		Transition to			
Age Groups	Transition from	d_I	d_{II}	d_{III}	d_{IV}
40+	d_I	0.947 (0.030) (0.874,0.995)	0.053 (0.030) (0.005,0.126)	0 (0) (0,0)	0 (0) (0,0)
	d_{II}	0.034 (0.012) (0.012,0.06)	0.960 (0.013) (0.934,0.983)	0.006 (0.004) (0.001,0.016)	0 (0) (0,0)
	d_{III}	0 (0) (0,0)	0.129 (0.026) (0.08,0.179)	0.869 (0.026) (0.819,0.918)	0.002 (0.001) (0,0.005)
	d_{IV}	0 (0) (0,0)	0 (0) (0,0)	0.286 (0.109) (0.093,0.518)	0.714 (0.109) (0.482,0.907)
40-55	d_I	0.900 (0.075) (0.722,0.999)	0.100 (0.075) (0.001,0.278)	0 (0) (0,0)	0 (0) (0,0)
	d_{II}	0.032 (0.021) (0.001,0.081)	0.960 (0.021) (0.911,0.993)	0.009 (0.007) (0,0.027)	0 (0) (0,0)
	d_{III}	0 (0) (0,0)	0.107 (0.033) (0.047,0.174)	0.885 (0.034) (0.817,0.946)	0.008 (0.006) (0.002,0.024)
	d_{IV}	0 (0) (0,0)	0 (0) (0,0)	0.232 (0.102) (0.068,0.468)	0.768 (0.102) (0.532,0.932)
55 +	d_I	0.979 (0.015) (0.942,0.999)	0.021 (0.015) (0.001,0.058)	0 (0) (0,0)	0 (0) (0,0)
	d_{II}	0.026 (0.011) (0.008,0.052)	0.962 (0.012) (0.938,0.984)	0.012 (0.005) (0.003,0.024)	0 (0) (0,0)
	d_{III}	0 (0) (0,0)	0.139 (0.036) (0.079,0.221)	0.861 (0.036) (0.779,0.921)	0 (0) (0,0)
	d_{IV}	0 (0) (0,0)	0 (0) (0,0)	0.338 (0.1) (0.162,0.436)	0.662 (0.1) (0.564,0.838)

3.4 Conclusion

In this section, we reviewed several methods to estimate Markov model parameters using a set of data with intermittent missing values. We discussed three categories of methods for datasets with ignorable missingness mechanism including deletion, imputation, and augmentation. Moreover, we executed these methods on several simulated data to measure the bias of the results from each algorithm. Based on the simulation results, algorithms tend to have a very small bias if the Markov model is fully observable. However, the uncertainty in the accuracy of the observations adds bias to these estimates. Finally we implemented Baum-Welch algorithm on a set of incomplete dataset consist of hidden observations for breast density of a group of patients at Louisiana Cancer Prevention and Control center. Due to the small size of dataset, we used the bootstrapping method to reduce the bias of estimations. Based on our results, estimated breast density transition matrices specify a relatively high probability that successive observations recur. This implies that future outcome has a high dependency on the current outcome. The results of breast density transition probability estimations will be used as the basic parameters in section 4 to evaluate the efficacy of supplemental screening tests for high breast density patients.

One of the limitation of this study is the small number of longitudinal observations for each patient. One possible direction for future study could be defined as implementing these methods on a larger dataset with more number of observations for each patient.

CHAPTER 4

INVESTIGATING THE EFFECTIVENESS OF BREAST DENSITY NOTIFICATION LAW CONSIDERING RADIOLOGISTS BIAS

4.1 Introduction

Breast density is defined as the prevalence of fibroglandular tissue in the breast. The Breast Imaging Reporting and Data System (BI-RADS) classification system classifies breast tissue density into four categories: almost entirely fatty which includes less than 25 percent glandular tissue (type 1), scattered fibroglandular densities which includes approximately 25-50 percent glandular tissue (type 2), heterogeneously dense class which includes approximately 51-75 percent glandular tissue (type 3), and extremely dense class which includes more than 75 percent glandular tissue (type 4) [63]. Higher breast density can significantly reduce the mass detection rate since the normal tissues in dense breasts appear as bright areas in mammography. Breast density is also associated with increased risk of breast cancer [5]. It is well-established that tumors in dense breasts may progress more rapidly than those in fatty breasts [5]. In addition, due to the lower sensitivity of screening mammography in women with dense breasts, the cancer is more likely to remain undetected. Keefer [64] showed that the relative risk associated with breast density is substantially higher than other relative risks such as a family history of breast cancer, and menstrual and reproductive

risk factors. The reported odds ratio for developing breast cancer for the most dense compared with the least dense breast tissue categories ranges from 1.46 [65] to 6.0 [66].

Breast density is a dynamic risk factor and typically decreases as a patient becomes older [67, 68, 69]. Younger women (especially those in premenopausal status) are more likely to have dense breasts [70]. According to Mandelblatt et al. [62], 58.8% of women aged 40-49 have highly dense breasts, while this percentage decreases to 42.7% and 31.1% for women aged 50-64 and 65-74, respectively.

Breast density notification laws have been enacted in 38 states in the U.S. (as of August 2020) to mitigate the increased breast cancer risk in women with high breast density which is partially caused by the masking effect of dense breast in screening mammography. These laws generally require physicians to notify patients with high breast density of their increased risk of breast cancer compared to low breast density women [63]. Moreover, in some states, the breast density notification law requires physicians to inform women with high breast density that adjunctive screening tests such as breast ultrasonography (US) and magnetic resonance imaging (MRI) may benefit them. Breast ultrasound uses high frequency sound waves to make an image of breast tissues and as a result has higher sensitivity than mammography in women with dense breasts. MRI uses intravenous contrast solution injection in order to produce 3-dimensional images of breast tissue.

Since the emergence of breast density notification laws, there have been a lot of controversy on its potential unintended consequences as well as quality of its implementation. It is believed that supplemental screening may result in an increased number of unnecessary biopsies and patients' overdiagnosis (that is,

detection of a cancer that would not have become clinically apparent over the patient's lifetime if left undetected). In addition, inter and intra-variabilities in breast density classification by radiologists raised some concerns since it results in patients' breast density misclassification (e.g., classification of a patient to a breast density category different from her true BI-RADS breast density class) [71]. According to Bahl et al. [72], the percentage of mammogram images reported as dense decreased after the enactment of breast density notification law. This reduction happens as radiologists may downgrade their assessment of density to avoid reporting requirements. On the other hand, there have been controversies that radiologists may upgrade their assessments so that supplemental screening can be ordered and their liability is minimized [72].

Currently, there is not a consensus among different health agencies in the U.S. regarding the necessity of supplemental screening in early breast cancer detection for women with high breast density. The American College of Radiology (ACR) advocates the use of ultrasound as an adjunctive screening test in women with dense breast tissues [73]. However, according to the ACS report, there is not enough evidence to make a recommendation for or against supplemental screening in addition to mammograms for women with dense breasts [74]. The U.S. Preventive Services Task Force (USPSTF) and the American College of Physicians (ACP) state that the current evidence is not sufficient to support recommendation of supplemental screenings [75, 76].

4.2 Relevant literature

In breast cancer related studies, Markov models have been previously used to evaluate/optimize breast cancer screening and treatment strategies. Nohdurft et al. [77] formulated a Markov decision process (MDP) model to derive optimal surgery decisions for women with breast cancer. Chhatwal et al. [78] developed a finite-horizon discrete-time MDP to provide patient-specific recommendations for breast biopsy based on the patient's mammographic features. Alagoz et al. [79] formulated a finite-horizon discrete-time MDP to optimize the post-mammography diagnostic decisions (choosing between biopsy or short-interval follow up mammogram) based on mammogram test findings. Ayvaci et al. [80] developed an MDP model to optimize the risk-sensitive diagnostic decisions after a mammography exam. In their study, the radiologist can select from biopsy, short-term follow-up, and routine mammography while considering the patient's preferences to maximize the quality-adjusted survival duration. In another study, Ayvaci et al. [81] investigated the impact of budgetary restrictions on breast biopsy decisions by developing a finite-horizon discrete-time constrained MDP. Çağlayan et al. [82] developed a Markov model to capture the breast cancer progression in women with certain risk factors such as gene mutations and family history of breast and ovarian cancer. They identified the optimal and most cost-effective population screening strategies.

As mammography is not perfect and may not reflect the true health status of a patient, some studies used partially observable Markov models in assessing/optimizing breast cancer screening policies. Maillart et al. [15] formulated a partially observable Markov chain (POMC) model to compare different policies in terms of lifetime breast

cancer mortality risk and the expected number of mammograms a woman should undergo under each screening policy. Ayer et al. [83] formulated a finite-horizon, partially observable Markov decision process (POMDP) to determine optimal personalized mammography screening policies based on patients' different risk characteristics. In another study, Ayer et al. [23] developed a POMDP framework to analyze the importance of heterogeneity in women's adherence to mammography screening policies. Madadi et al. [22] developed a discrete-time POMC model to evaluate mammography screening policies in terms of the expected QALYs and lifetime breast cancer mortality risk while incorporating the uncertainty in women's adherence behaviors. Molani et al. [3] developed two POMCs to quantify the age and stage-specific overdiagnosis risks while considering the uncertainty in a patient's adherence behavior. Cevik et al. [84] proposed a POMDP model to maximize the total expected QALYs of a patient when there is a constraint on the number of mammograms the patient can undergo. Sandikci et al. [85] formulated a POMDP model to determine the optimal breast cancer screening policies considering patients' breast density as risk factor. Otten et al. [86] formulated a finite horizon discrete-time POMDP to optimize and personalize breast cancer follow-up.

In this chapter, we develop a POMC model to investigate the impact of the radiologist bias on patients' health outcomes under the breast density notification law. The patients' outcomes include probability of detecting breast cancer in early and advanced cancer states and the expected number of supplemental screening a patient undergoes in her lifetime. To the best of our knowledge, Sandikci et al. [85] work, is the only study that investigate breast density notification law and explicitly considers breast density as a risk factor. Our study, however, is different from Sandikci et al.'s

work from several aspects: 1) We consider the conditional probability of detecting breast cancer in early and advanced states given the patient develops cancer in her lifetime as the main patient's outcome. This is the first study in the literature to consider these probabilities as patients' outcomes. Note that the main purpose of cancer screening is to detect the cancer in early states where it is more likely to be cured. 2) We investigate the impact of radiologists' bias in density classification of mammogram images on patients' outcomes. This is done by modeling breast density as a partially observable variable. In Sandikci et al.'s work, however, breast density is assumed to be fully observable (i.e., radiologist's evaluation of breast density perfectly correlates with the patient's actual density). 3) In this study, we use sequential mammography screening data of 436 patients from Louisiana Cancer Prevention and Control Programs [4] to estimate the dynamics of breast density to better capture the breast cancer risk dynamic caused by change in breast density.

The remainder of this chapter is as follows. In Section 4.3, we present our proposed POMC model and calculate the probability of detecting cancer in early and advanced states and the expected number of screening tests a patient undergoes in her lifetime under a screening strategy. Section 4.4 presents parameter estimations and model validation. Numerical results and sensitivity analyses are presented in Section 4.5. We summarize and conclude in Section 4.6.

4.3 Model formulation

A discrete-time finite-horizon partially observable Markov chain (POMC) is developed to model the breast cancer natural history and breast density dynamics.

A POMC is used as the imperfect nature of mammography tests (i.e., possibility of receiving false positives and false negatives) as well as the possibility of breast density misclassification by radiologists prevent the patient's true state to be fully observable to the decision maker.

The patients' outcome measures include the lifetime conditional probability of detecting cancer in early and advanced states. That is, we focus on the population of patients who would eventually develop breast cancer at some point in their lifetimes and their cancer eventually becomes symptomatic if not detected through screening tests. The latter assumption is made to rule out the over-diagnosed cases as for these cases, detection of cancer is not favourable. Note that early detection of cancers which will eventually become problematic is the main incentive of screening programs.

Obviously, the more aggressive a screening strategy is (more frequent and multiple screening modalities), the higher is the chance of detecting the cancer in early state when it is more likely to be treated. However, there are disutilities associated with screening tests which adversely impacts patient quality of life. Therefore, there is a trade-off between detecting a cancer in early state and the discomfort of undergoing aggressive frequent screening. As such, we consider the expected number of supplemental screening tests a patient would undergo in her lifetime as another patients' outcome to investigate the trade-ff.

We estimate the probability of eventually detecting breast cancer in early and advanced state in Section 4.3.1 and the expected number of supplemental screening tests a patient may undergo in her lifetime in Section 4.3.2. The following is the list

of notation used in the proposed model. Note that vectors and matrices notations are in bold.

- t: Time periods, $t = 0, 1, 2 \cdots, T$.
- s: Patient's core state; Specifically, $s = (h, d) \in \Omega = H \bigtimes D$ represents the patient's underlying state where $h \in H$ and $d \in D$ denote the patient's core health and breast density states, respectively. The health state set H includes three partially observable states of cancer free (state 0), early breast cancer (state 1), and advanced breast cancer (state 2) and one fully observable state of death due to breast cancer of other causes (state 3). Specifically, we refer to the partially and fully observable health state sets as H_1, and H_2, respectively, i.e., $H = H_1 \cup H_2$. We denote the subsets of patient's core states for which $h \in H_1$ and $h \in H_2$ by Ω_1 and Ω_2, respectively. Moreover, set D includes four BI-RADS density classes as discussed in Section 4.1, i.e., $D = \{1, 2, 3, 4\}$.
- $\boldsymbol{\beta}_t$: A vector of length $|\Omega_1|$ representing the patient's belief state at the beginning of period t. Specifically, $\boldsymbol{\beta}_t(s)$ denotes the probability that the patient is in partially observable state $s = (h, d), h \in H_1$ at the beginning of period t.
- $\boldsymbol{P}_t$: Underlying transition probability matrix capturing the natural history of breast cancer and breast density dynamics. That is, $\boldsymbol{P}_t(s'|s)$ represents the probability that a patient will be in state $s' = (h', d')$ at time $t + 1$, given that she is in state $s = (h, d)$ at time t.
- a_t: Prescribed action at time t, where possible actions include *wait* and *mammography*, denoted by W and M, respectively. A patient classified as high density may undergo a supplemental screening following a negative mammogram, for

which case the action is denoted by B. Let A denote set of all possible actions; we have $a_t \in A = \{W, M, B\}$

- o_t: Observation received at time t which includes both breast density classification by the radiologist and screening result. Specifically, $o_t = (\theta, \delta)$, where $\theta \in \Theta = D$ and $\delta \in \Delta^a$ respectively denote the assigned breast density class and screening result. For notation brevity, we define density observation subsets $\underline{\Theta} = \{1, 2\}$ and $\bar{\Theta} = \{3, 4\}$. Clearly, we have $\Theta = \underline{\Theta} \cup \bar{\Theta}$. Observations are received only if the prescribed action is to undergo a mammogram. Specifically, when patient undergoes only a *mammography*, the possible test results are negative and positive mammogram denoted by M^- and M^+, respectively. That is, $\Delta^M = \{M^-, M^+\}$. When a mammogram is *accompanied with a supplemental* test (i.e., $a_t = B$), possible observations are $M^-\&S^-$ and $M^-\&S^+$ which respectively represent a negative and a positive supplemental test followed by a negative mammogram, i.e., $\Delta^B = \{M^-\&S^-, M^-\&S^+\}$. If the action is *wait*, no observations will be received.
- $\boldsymbol{Q}_t^M$: Breast density information matrix, where $\boldsymbol{Q}_t^M(\theta|d)$ denotes the probability of a patient with true density d be classified in density class θ upon action $a_t = M$.
- $\boldsymbol{K}_t^a$: Health information matrix, where $\boldsymbol{K}_t^a(\delta|s)$ represents the probability of observation $\delta \in \Delta^a$ when action $a \in \{M, B\}$ is taken and the patient's true state is $s = (h, d)$ at time t. Note that health observation probability matrix for the case that the action is a mammogram, is a function of breast density due to the masking effect of high breast density on mammogram sensitivity.

- $\boldsymbol{\eta}_t$: A vector of length $|\Omega_1|$ representing the probabilities of a patient showing symptoms in period t. Specifically, $\boldsymbol{\eta}_t(s)$ is the probability of showing symptoms in state s at time t.

The one-period sample path for the breast cancer detection process under the notification law is presented in Figure 4.1. At each period, depending on the prescribed action and possible subsequent observations, patient takes a different path. We assume that after receiving a positive screening result (either a mammogram or a supplemental test), the patient undergoes a biopsy test. Biopsy is assumed to be perfect as its true positives rate is very close to 1 [87]. According to the U.S. Department of Health & Human Services report, surgical biopsy could be considered a test without measurement error [88]. We assume that probability of both developing cancer and showing symptoms in one period (one year) is zero. That is, cancers can only show symptoms in a period when the patient is in a cancer state at the beginning of that period. Moreover, we assume that breast cancer cannot spontaneously (without treatment) regress [89].

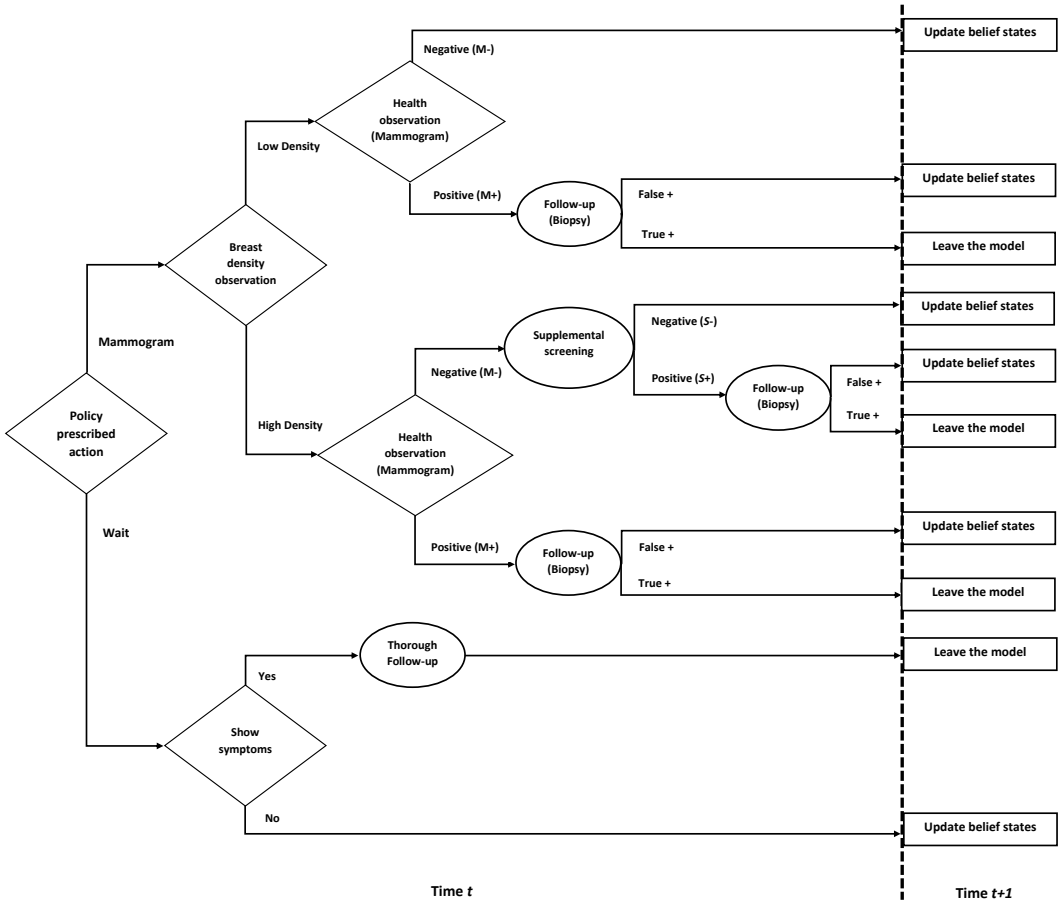


Figure 4.1: One-period sample path of the breast cancer detection process when supplemental screening is administered. Note that under action *wait*, symptoms can happen when the patient is in a cancer state

At each period, the patient's belief state is updated based on the action and possible observations received. Under action a_t, observation o_t and assuming that the patient belief at the beginning of period t is $\boldsymbol{\beta}_t$, Equation (4.1) calculates the patient's updated belief state ($\boldsymbol{\nu}$) at time $t+1$:

$$\boldsymbol{\nu}[\boldsymbol{\beta}_t, a_t, o_t](z) =$$

$$\begin{cases} \dfrac{\sum_{s\in\Omega_1}\boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^-|s)\boldsymbol{P_t}(z|s)}{\sum_{s'\in\Omega_1}\sum_{s\in\Omega_1}\boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^-|s)\boldsymbol{P_t}(s'|s)}, & \text{if } a_t = M,\, o_t = (\theta, M^-), \theta \in \underline{\Theta}, \\ \dfrac{\sum_{s\in\Omega_1}\boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^B(M^-\&S^-|s)\boldsymbol{P_t}(z|s)}{\sum_{s'\in\Omega_1}\sum_{s\in\Omega_1}\boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^B(M^-\&S^-|s)\boldsymbol{P_t}(s'|s)}, & \text{if } a_t = B,\, o_t = (\theta, M^-\&S^-), \theta \in \bar{\Theta}, \\ \dfrac{\boldsymbol{P_t}(z|s)}{1-\sum_{s'\in\Omega_2}\boldsymbol{P_t}(s'|s)}, & \text{if } s = (0,d), a_t = M,\, o_t = (\theta, \delta), \\ & \quad \delta \in \{M^+, M^-\&S^+\}, \\ \dfrac{\sum_{s\in\Omega_1}\boldsymbol{\beta}_t(s)\boldsymbol{P_t}(z|s)}{\sum_{s\in\Omega_1}\boldsymbol{\beta}_t(s)\big(1-\sum_{s'\in\Omega_2}\boldsymbol{P_t}(s'|s)\big)}, & \text{if } a_t = W. \end{cases} \tag{4.1}$$

The first and second case in Equation (4.1) represents the case when the patient undergoes a prescribed mammogram, receives a negative result and is classified into the low and high density class by the radiologist, respectively. In the latter case, the patient undergoes a supplemental screening test and receives a negative result. We use Bayes rule to update the patient's belief state in these two cases. In the second case where the patient undergoes both mammography and supplemental tests, the joint information from both tests is used to update the patient's belief state. The third case represents a false positive and consists two different situations: 1) a false positive mammogram result, and 2) a negative mammogram followed by a false positive supplemental test. In these cases, the patient's belief is updated by accounting for possible cancer development from the cancer-free state ($s = (0, d)$). The fourth case represent the situation where the action is *wait*. In this case, no observation is received and the dynamics of breast cancer natural history and breast density is used to update the patient's belief state. The term $1 - \sum_{s=(h',d'),h'\in H_2}\boldsymbol{P_t}(s'|s)$ in the third and fourth cases represents the probability that a patient in state s survives period t.

4.3.1 Probability of detecting cancer in early and advanced states

Let $\mathcal{E}_t^a(\boldsymbol{\beta}_t)$ and $\mathcal{A}_t^a(\boldsymbol{\beta}_t)$ denote the probability of eventually detecting a cancer in early and advanced cancer states when the patient belief state in period t is $\boldsymbol{\beta}_t$ and action a is taken, respectively. Note that we only consider the cancer population for which the cancer will eventually show symptom. That is, we exclude the over-diagnosed cases whose cancer may never show symptoms or cause any problem in the patient's lifetime. Equations (4.2) and (4.3) calculate these probabilities for the case when the prescribed action in epoch t is *wait*:

$$\begin{aligned}\mathcal{E}_t^W(\boldsymbol{\beta}_t) = &\sum_{s=(0,d)} \boldsymbol{\beta}_t(s)\Big[0 + \rho_t^0(s)\mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]) + (1-\rho_t^0(s))\cdot 0\Big] \\ &+ \sum_{s=(1,d)} \boldsymbol{\beta}_t(s)\Big[\boldsymbol{\eta}_t(s) + (1-\boldsymbol{\eta}_t(s))\Big(\rho_t^1(s)\mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]) + (1-\rho_t^1(s))\cdot 0\Big)\Big] \\ &+ \sum_{s=(2,d)} \boldsymbol{\beta}_t(s)\cdot 0, \end{aligned} \tag{4.2}$$

where $\rho_t^0(s) = \frac{\sum_{z=(h',d'):h'\in\{0,1\}} \boldsymbol{P}_t(z|s)}{1-\sum_{s'\in\Omega_2} \boldsymbol{P}_t(s'|s)}$ is the probability of remaining in the healthy states or transitioning to early cancer states in period t given that the patient is healthy (i.e., $s=(0,d)$) at the beginning of period t and survives the current period. Additionally, $\rho_t^1(s) = \frac{\sum_{z=(1,d')} \boldsymbol{P}_t(z|s)}{1-\sum_{s'\in\Omega_2} \boldsymbol{P}_t(s'|s)}$ is the probability of remaining in early cancer state (i.e., $s=(1,d)$) in period t given that the patient survives the current period.

The logic of Equation (4.2) is as follows: If the patient is in a healthy state at time t, the probability of cancer detection in the current period is zero. The future probability of eventually detecting cancer in early state is conditioned on patient's surviving the current period. In addition, the patient should stay healthy or if developed cancer, she has to be in an early cancer state in order to be possible to

detect the cancer in an early stage as we assume that cancers cannot spontaneously regress. The probability of such event is $\rho_t^0(s)$ and in such a case, the future probability of cancer detection in early state is $\mathcal{E}_{t+1}^a\Big(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]\Big)$. Note that $\boldsymbol{\nu}[\beta_t, W, \cdot]$ is the patient updated belief under action *wait* and is calculated using Equation (4.1). If the patient transitions to an advanced cancer state (with probability $1 - \rho_t^0(s)$), the probability that the cancer is eventually be detected in an early state is zero. If the patient is in an early cancer state at the beginning of period t, the cancer may show symptoms with probability $\boldsymbol{\eta}_t(s)$. In this case, the follow-up tests will reveal the cancer and the patient leaves the model. If the cancer remains undetected in the current period (which happens with probability $1 - \boldsymbol{\eta}_t(s), s = (1, d)$) and the patient remains in early cancer state (which happens with probability $\rho_t^1(s)$), the cancer might be eventually detected in an early state with probability $\mathcal{E}_{t+1}^a\Big(\boldsymbol{\nu}[\beta_t, W, \cdot]\Big)$. If the patient is in an advanced cancer state at the beginning of period t, the cancer can never be detected in an early state since we assume that no cancer regression can occur.

$$\begin{aligned}\mathcal{A}_t^W(\boldsymbol{\beta}_t) = &\sum_{s=(0,d)} \boldsymbol{\beta}_t(s)\Big[0 + \rho_t^0(s)\mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]) + (1 - \rho_t^0(s)) \cdot 1\Big] \\ &+ \sum_{s=(1,d)} \boldsymbol{\beta}_t(s)\Big[0 + (1 - \boldsymbol{\eta}_t(s))\Big(\rho_t^1(s)\mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]) + (1 - \rho_t^1(s)) \cdot 1\Big)\Big] \\ &+ \sum_{s=(2,d)} \boldsymbol{\beta}_t(s) \cdot 1. \end{aligned} \tag{4.3}$$

The logic of Equation (4.3) is as follows: If the patient is in a healthy state at the beginning of period t, the immediate probability of detecting cancer in an advanced state is zero. If the patient stays in a healthy or transitions to an early cancer state conditioning that she has survived the current period (which occurs with

probability $\rho_t^0(s)$), she might eventually be detected in an advanced cancer state with probability $\mathcal{A}_{t+1}^a\Big(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]\Big)$. However, if she transitions to an advanced cancer state, with certainty she will eventually be detected in an advanced cancer state. If the patient is in an early cancer state at the beginning of period t, her cancer needs to remain undetected in the current period (which happens with probability $1 - \boldsymbol{\eta}_t(s)$) in order to be detected later in an advanced state. In such a case, if she remains in the early cancer state, the future probability of detecting cancer in an advanced state is $\mathcal{A}_{t+1}^a\Big(\boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot]\Big)$, and if she transitions to an advanced state, the corresponding probability is one. Finally, if the patient is in an advanced cancer state, she will eventually be detected in an advanced state with certainty.

Equations (4.4) and (4.5) respectively present the probability of eventually detecting cancer in early and advanced cancer states starting from belief state $\boldsymbol{\beta}_t$ at the beginning of period t when the prescribed action is a screening mammogram with a possible subsequent supplemental test. Note that in compliance with the breast density notification laws, when the prescribed action is a *mammogram*, the patient may take different paths depending on their observed breast density class.

$$
\begin{aligned}
\mathcal{E}_t^M(\boldsymbol{\beta}_t) = & \sum_{s=(0,d)} \sum_{\theta \in \underline{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \\
& \Big[0 + \boldsymbol{K}_t^M(M^-|s) \Big(\rho_t^0(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-)]) + (1 - \rho_t^0(s)) \cdot 0 \Big) \\
& + \boldsymbol{K}_t^M(M^+|s) \Big(\rho_t^0(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^+)]) + (1 - \rho_t^0(s)) \cdot 0 \Big) \Big] \\
+ & \sum_{s=(0,d)} \sum_{\theta \in \bar{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \\
& \Big[0 + \boldsymbol{K}_t^B(M^- \& S^-|s) \Big(\rho_t^0(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^- \& S^-)]) + (1 - \rho_t^0(s)) \cdot 0 \Big) \\
& + \boldsymbol{K}_t^B(M^- \& S^+|s) \Big(\rho_t^0(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^- \& S^+)]) + (1 - \rho_t^0(s)) \cdot 0 \Big) \\
& + \boldsymbol{K}_t^M(M^+|s) \Big(\rho_t^0(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^+)]) + (1 - \rho_t^0(s)) \cdot 0 \Big) \Big] \\
+ & \sum_{s=(1,d)} \sum_{\theta \in \underline{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \Big[\boldsymbol{K}_t^M(M^+|s) \\
& + \boldsymbol{K}_t^M(M^-|s) \Big(\rho_t^1(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-)]) + (1 - \rho_t^1(s)) \cdot 0 \Big) \Big] \\
+ & \sum_{s=(1,d)} \sum_{\theta \in \bar{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \Big[\Big(\boldsymbol{K}_t^M(M^+|s) + \boldsymbol{K}_t^B(M^- \& S^+|s) \Big) \\
& + \boldsymbol{K}_t^B(M^- \& S^-|s) \Big(\rho_t^1(s) \mathcal{E}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^- \& S^-)]) + (1 - \rho_t^1(s)) \cdot 0 \Big) \Big] \\
+ & \sum_{s=(2,d)} \boldsymbol{\beta}_t(s) \cdot 0.
\end{aligned}
\tag{4.4}
$$

Equation (4.4) emerges from the following logic: If the patient is healthy, the immediate probability of cancer detection is zero. If she is classified in the low density class, which occurs with probability $\sum_{\theta \in \underline{\Theta}} \boldsymbol{Q}_t^M(\theta|d)$, she only receives a mammogram test. The mammogram test result might be a true negative or a false positive with corresponding probabilities of $\boldsymbol{K}_t^M(M^-|s)$ and $\boldsymbol{K}_t^M(M^+|s), s = (0, d)$. In either case, if the

patient remains in the healthy state or proceeds to early cancer state, her belief state is updated and the future probability of cancer being detected in early state is calculated. Note that we assume in case of a false positive, further examination (i.e, biopsy) reveals that the patient is healthy. If the patient transitions to an advanced cancer state, her future probability of being detected in an early cancer state is zero. When the patient is classified into a high breast density class (with probability $\sum_{\theta\in\bar{\Theta}} \boldsymbol{Q}_t^M(\theta|d)$), she may undergo a supplemental screening if the mammogram result is negative. Possible outcomes in such case are negative mammogram followed by a negative supplemental test (true negative with probability $\boldsymbol{K}_t^M(M^-\&S^-|s)$), negative mammogram followed by a positive supplemental test (false positive with probability $\boldsymbol{K}_t^M(M^-\&S^+|s)$), or a positive mammogram (false positive with probability $\boldsymbol{K}_t^M(M^+|s)$). In any of these cases, if the patient does not proceed to advanced cancer state, her belief state is updated based on the received observations and her future probability of being detected in early state is calculated. However, if the patient proceeds to advanced cancer state, the cancer will never be detected in early state. If the patient is in an early cancer states at the beginning of period t, her cancer may be detected in the current period through screening tests. Specifically, if the patient is classified as a low and high breast density patient, the cancer may be detected in the current period with probability $\boldsymbol{K}_t^M(M^+|s)$ and $\boldsymbol{K}_t^M(M^+|s) + \boldsymbol{K}_t^M(M^-\&S^+|s), s = (1, d)$, respectively, in which case the patient leaves the model. However, if the screening does not reveal the cancer, which occurs with probabilities $\boldsymbol{K}_t^M(M^-|s)$ and $\boldsymbol{K}_t^M(M^-\&S^-|s), s = (1, d)$ when the patient is classified as low and high density class, respectively, the patient belief is updated based on the sensitivity of screening test(s) that the patient has undergone and the future probability of the cancer being detected in early state is calculated.

Finally, if the patient is in an advanced cancer state at the beginning of period t, the probability that she eventually be detected in an early cancer state is zero.

$$
\begin{aligned}
\mathcal{A}_t^M(\boldsymbol{\beta}_t) = & \sum_{s=(0,d)} \sum_{\theta \in \underline{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \\
& \Big[0 + \boldsymbol{K}_t^M(M^-|s) \Big(\rho_t^0(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-)]) + (1 - \rho_t^0(s)) \cdot 1 \Big) \\
& + \boldsymbol{K}_t^M(M^+|s) \Big(\rho_t^0(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^+)]) + (1 - \rho_t^0(s)) \cdot 1 \Big) \Big] \\
+ \sum_{s=(0,d)} & \sum_{\theta \in \bar{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \\
& \Big[0 + \boldsymbol{K}_t^B(M^- \& S^-|s) \Big(\rho_t^0(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^- \& S^-)]) + (1 - \rho_t^0(s)) \cdot 1 \Big) \\
& + \boldsymbol{K}_t^B(M^- \& S^+) \Big(\rho_t^0(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^- \& S^+)]) + (1 - \rho_t^0(s)) \cdot 1 \Big) \\
& + \boldsymbol{K}_t^M(M^+|s) \Big(\rho_t^0(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^+)]) + (1 - \rho_t^0(s)) \cdot 1 \Big) \Big] \\
+ \sum_{s=(1,d)} & \sum_{\theta \in \underline{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \\
& \Big[0 + \boldsymbol{K}_t^M(M^-|s) \Big(\rho_t^1(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-)]) + (1 - \rho_t^1(s)) \cdot 1 \Big) \Big] \\
+ \sum_{s=(1,d)} & \sum_{\theta \in \bar{\Theta}} \boldsymbol{\beta}_t(s) \boldsymbol{Q}_t^M(\theta|d) \\
& \Big[0 + \boldsymbol{K}_t^B(M^- \& S^-|s) \Big(\rho_t^1(s) \mathcal{A}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^- \& S^-)]) + (1 - \rho_t^1(s)) \cdot 1 \Big) \Big] \\
+ \sum_{s=(2,d)} & \boldsymbol{\beta}_t(s) \cdot 1.
\end{aligned}
\tag{4.5}
$$

Equation (4.5) follows a logic similar to that of Equations (4.3) and (4.4) and thus is omitted for brevity.

For the boundary conditions, a healthy patient or a patient in early cancer states can be eventually detected in either early or advanced states. The probability

of such events are estimated using cancer progression rates and probability of showing symptoms after period T. For a patient in advanced cancer states at time T, the cancer will eventually be detected in the advanced states. Let $\gamma_E(s)$ and $\gamma_A(s)$ respectively denote the probability of eventually detecting cancer in early and advanced states for a patient in state s at time T. We have

$$\gamma_E\Big(s=(0,d)\Big) > \gamma_E\Big(s=(1,d)\Big), \quad \gamma_E\Big(s=(2,d)\Big)=0, \quad d\in D. \tag{4.6}$$

The probabilities $\gamma_A(s)$ are calculated using the fact that $\gamma_E(s)$ and $\gamma_A(s)$ are complementary in order to exclude overdiagnosed cases.

4.3.2 Expected number of supplemental screenings

Let $\mathcal{V}_t^a(\boldsymbol{\beta}_t)$ denote the expected number of supplemental screenings a patient undergoes in her remaining life years when she is in belief state $\boldsymbol{\beta}_t$ and the prescribed action in epoch t is a. Equations (4.7) and (4.8) calculate $\mathcal{V}_t^a(\boldsymbol{\beta}_t)$ for actions *wait* and *mammogram*, respectively. When the prescribed action is wait, the patient may undergo a supplemental screening when the cancer shows symptoms. Specifically, when the patient is in state $s=(h,d)\in\Omega_1$, she may develop symptoms with probability $\boldsymbol{\eta}_t(s)$. In the follow-up mammogram if she is classified as a high density and receives a negative result, she will undergo supplement test. However, if she does not show any symptoms (with probability $1-\boldsymbol{\eta}_t(s)$), she proceeds to the next period with probability $\sum_{z\in\Omega_1}\boldsymbol{P_t}(z|s)$ and the expected number of screening is $\mathcal{V}_{t+1}^W(\boldsymbol{\nu}[\boldsymbol{\beta}_t,W,.])$.

$$\mathcal{V}_t^W(\boldsymbol{\beta}_t)=\sum_{s\in\Omega_1}\sum_{\theta\in\bar{\Theta}}\boldsymbol{\beta}_t(s)\Big[\boldsymbol{\eta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^-|s)\cdot 1 + (1-\boldsymbol{\eta}_t(s))\sum_{z\in\Omega_1}\boldsymbol{P_t}(z|s)\mathcal{V}_{t+1}^W(\boldsymbol{\nu}[\boldsymbol{\beta}_t,W,.])\Big]. \tag{4.7}$$

When the prescribed action is a mammogram, the patient receives a supplemental screening only if she is classified as a high density patient in the mammogram screening and the mammogram result is negative. In such cases, she receives an immediate cost of 1. If the supplemental test result returns positive and the patient is actually in a cancer state, she leaves the model. In any other cases, the patient belief state is updated based on the screening test(s) and corresponding observation she has received and her future cost-to-go is calculated.

$$
\begin{aligned}
\mathcal{V}_t^M(\boldsymbol{\beta}_t) = & \sum_{s\in\Omega_1}\sum_{\theta\in\underline{\Theta}} \boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^-|s) \\
& \qquad \left[0 + \sum_{z\in\Omega_1} \boldsymbol{P_t}(z|s)\mathcal{V}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-)])\right] \\
& + \sum_{s\in\Omega_1}\sum_{\theta\in\bar{\Theta}} \boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^-\&S^-|s) \\
& \qquad \left[1 + \sum_{z\in\Omega_1} \boldsymbol{P_t}(z|s)\mathcal{V}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-\&S^-)])\right] \\
& + \sum_{\substack{s=(h,d)\\h=0}}\sum_{\theta\in\bar{\Theta}} \boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^-\&S^+|s) \\
& \qquad \left[1 + \sum_{z=(2,d)\in\Omega_1} \boldsymbol{P_t}(z|s)\mathcal{V}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^-\&S^+)])\right] \\
& + \sum_{\substack{s=(h,d)\\h=0}}\sum_{\theta\in\Theta} \boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\theta|d)\boldsymbol{K}_t^M(M^+|s) \\
& \qquad \left[0 + \sum_{z\in\Omega_1} \boldsymbol{P_t}(z|s)\mathcal{V}_{t+1}^a(\boldsymbol{\nu}[\boldsymbol{\beta}_t, M, (\theta, M^+)])\right] \\
& + \sum_{\substack{s=(h,d)\\h\in\{1,2\}}} \boldsymbol{\beta}_t(s)\boldsymbol{Q}_t^M(\bar{\theta}|d)\boldsymbol{K}_t^M(M^-\&S^+|s)\cdot 1. \qquad (4.8)
\end{aligned}
$$

4.4 Parameters estimation and model validation

The data sources used to estimate the parameters of the proposed model are presented in Table 4.1. Following the recommended screening policies in the U.S., we use age 45 and 75 (corresponding to $t = 0$ and $t = T = 30$) as the earliest and latest ages that a patient undergoes a breast cancer screening test. The start age of 45 is considered based on the new ACS policy recommendation and the fact that the risk of developing breast cancer is very small in women younger than 45 [74]. Moreover, we assume no screening is administered after age 75 since the risks associated with breast cancer screening outweigh its benefits in women older than 75 [90].

Table 4.1 Source of model inputs and parameters estimation

Model Parameter	**Parameter Values**	**Source**
Breast density state transition probabilities	Table 3.2	Molani [91]
Breast density observation probability matrix	Table B.1	Østerås et al. [92]
Age-specific health state transition probabilities	Table A.4	Maillart et al. [15], Duffy et al. [65]
Age and density-specific mammography specificity	Table B.2	Stout et al. [93]
Age-specific mammography sensitivity	Table B.3	von Euler-Chelpin et al. [94]
Joint Mammogram/ultrasound sensitivity	0.885	Devolli-Disha et al. [95]
Joint Mammogram/MRI specificity	0.77	Group et al. [96]
Joint Mammogram/MRI sensitivity	0.95	Çağlayan et al. [82]
Initial density belief state	Table B.4	Mandelblatt et al. [62]
Initial health belief state	Table B.5	BCSC model[97]

4.4.1 Transition probabilities

As discussed earlier, previous studies have shown that mammographic breast density is associated with increased breast cancer risk. We estimate the age-specific and density-specific core transition probabilities of the breast cancer natural history model by adjusting previously estimated transition probabilities by Maillart et al. [15] using different odds ratios. In our baseline analysis, we use odds ratio of 3.73 which is the midpoint of the odds ratio interval, reported in the literature ([65], [66]). We

will perform a sensitivity analysis to investigate the impact of odds ratio on results in Section 4.5.1. To calculate the density-specific health transition probabilities, we adjust disease development and progression probabilities using the odds ratios of breast cancer risk comparing high and low breast density patients and the proportion of women in low and high breast density class. Let p_{α_t} be the proportion of women in low breast density at age α_t associated with time period t. Let I_t denote the general breast cancer incidence probability (i.e., probability of going from cancer-free state to early breast cancer state) at time t. Moreover, let I_t^d represent the incidence probability for patients in breast density class d at time t. We calculate the incidence probability for low and high breast density at time t, denoted by $I_t^{\underline{d}}$ and $I_t^{\bar{d}}$ respectively, using the following set of equations.

$$I_t = p_{\alpha_t} I_t^{\underline{d}} + (1 - p_{\alpha_t}) I_t^{\bar{d}}, \tag{4.9a}$$

$$\text{odds ratio} = \frac{\frac{I_t^{\bar{d}}}{1-I_t^{\bar{d}}}}{\frac{I_t^{\underline{d}}}{1-I_t^{\underline{d}}}}. \tag{4.9b}$$

Therefore, the core transition probability $\boldsymbol{P_t}\Big(s' = (1, d') | s = (0, d)\Big)$ is calculated as $Pr(\text{transition from density state } d \text{ to } d') \cdot I_t^d$, where breast density transition probabilities are adopted from previous chapter of this dissertation. The cancer progression probabilities is calculated using a similar approach.

We estimate breast density transition probabilities using mammography screening data from Louisiana Cancer Prevention and Control Programs [4]. Patients in the dataset are grouped into two different age categories of 40-54 and 55+ to capture the impact of age and menopausal status on breast density. Previous studies have shown

a significant dependency between the menopausal status and breast tissue density [98]. The dataset contains 436 patients with longitudinal breast density assessments between 2016 and 2020. Assuming that breast density is partially observable and missing observations are ignorable, we estimated the transition probabilities using the Baum-Welch method. Note that in the ignorable missingness mechanism, the probability of missingness depends only on the observed values and not the missing values [34].

4.4.2 Observations probabilities

We estimate breast density information matrix for an average-skilled radiologist using a previous study by Østerås et al. [92]. In their study, a number of radiologists interpreted 537 mammogram images and reported their density classifications. The radiologists classification results were then compared with the volumetric breast density obtained from a commercially available software (Quantra). They reported that in 87% of the cases the clinical interpretation agreed with radiologist reports. We also considered information matrices reflecting a perfect radiologist, and radiologists who always downgrade and upgrade density classifications.

The health information matrix for each screening test are estimated using their associated sensitivity and specificity. Specificity is defined as the probability of receiving a negative result when the patient is in cancer free stage (i.e., true negative), and sensitivity is the probability of receiving a positive result when the patient is in a cancer state (i.e., true positive). Specifically, let $sens_t(a|s = (h, d))$ denote the sensitivity of action a ($a \in \{M, B\}$) when the patient is in a cancer state (i.e., $h \in \{1, 2\}$) and density

class d, and $spec_t(a|d)$ denote the specificity of action a when the patient is in breast density class d at time t. The health information matrix can be calculated as follows.

$$\boldsymbol{K}_t^a(a^-|s=(h=0,d)) = spec_t(a|d),$$

$$\boldsymbol{K}_t^a(a^+|s=(h,d)) = sens_t(a|s=(h,d)), \;\; h=1,2,$$

$$\boldsymbol{K}_t^a(a^+|s=(h=0,d)) = 1 - spec_t(a|d),$$

$$\boldsymbol{K}_t^a(a^-|s=(h,d)) = 1 - sens_t(a|s=(h,d)), \;\; h=1,2,$$

We use the cancer stage and density specific sensitivity and specificity of mammography provided in von Euler-Chelpin et al. [94] and Stout et al. [93], respectively. The sensitivity of joint mammogram/ultrasound and mammogram/MRI are adopted from Devolli-Disha et al. [95] and Çağlayan et al. [82], respectively. We use the specificity of joint mammogram/MRI from Group et al. [96].

4.4.3 Initial belief state

Initial health belief state is estimated using the Breast Cancer Surveillance Consortium (BCSC) risk model [97]. which estimates advanced breast cancer risk based on age, ethnicity, family history of breast cancer, history of a breast biopsy, and BI-RADS breast density [97]. To estimate the early breast cancer risk, we use the breast cancer stage distribution by race reported by the ACS [13], and the race distribution in the U.S. [99]. The ratio of early to advanced breast cancer cases among women in the U.S. is estimated as 1.78. The initial breast density distribution for the general population are adopted from Mandelblatt et al. [62]. We combine health and density initial belief to calculate the patient's initial belief state,

$$\boldsymbol{\beta}_0(s) = \quad \text{P(Health state } h \mid \text{Density state } d) \cdot \text{ P(Density state } d).$$

4.4.4 Model validation

To validate the estimated health transition probabilities, we calculate i) lifetime risk of developing breast cancer, ii) five-year and ten-year risks of developing breast cancer from the proposed model, and iii) lifetime mortality risk of breast cancer with some adjustments to the model proposed by Molani et al. [3]. The derived lifetime risk of developing breast cancer and mortality risk of breast cancer are compared with the corresponding risks reported by the ACS. The derived five-year and ten-year breast cancer risks are compared with the associated risks obtained from the BCSC risk assessment tool. Our estimation of the lifetime breast cancer risk (12.37% for general population) is close to the reported ACS risk of 1 to 8 women (12.5%). More specifically, we estimate the lifetime risk of developing breast cancer as 7.56% , 17.18%, and 12.37% for women with low and high breast density, and the general population, respectively. The estimated five-year risk of breast cancer using the BCSC risk assessment tool at age 60 and 70 are 1.68% and 2.00%, which are comparable with our estimations of 1.59% and 2.37%. Moreover, the ten-year breast cancer risk estimated using the BCSC risk tool are 2.51%, 3.45%, and 3.80% at age 50, 60 and 70. These are comparable with our corresponding estimated risks of 2.17%, 3.76%, and 4.03%, respectively. In addition, we estimate the lifetime breast cancer mortality risk under several screening policies (see Table 4.2) for the general population (average-risk). The average of lifetime mortality risks across all policy equals 2.88% which is comparable with the reported ACS risk of 1 to 38 women (2.63%) [100]. Note that we report the average lifetime breast cancer mortality risk across different policies (screening frequencies) to implicitly account for the variation in the breast cancer screening frequencies for women in the U.S.

4.5 Numerical analyses

In this section, we evaluate the efficacy of supplemental screening and the impact of radiologists bias on patients outcomes for some of the in-practice screening policies. Table 4.2 presents the policies, denoted by P1 through P7, evaluated in this study. The screening policies recommended by the two major U.S. health agencies, the ACS and USPSTF, along with some screening guidelines in the European countries are evaluated. These policies differ in terms of recommended screening starting and stopping ages and the screening interval. Biennial and triennial screening with starting and stopping age of 45 and 75, respectively, are also assessed. Additionally, we consider *do nothing* (DN) policy with no recommended screening test in a patient lifetime.

Table 4.2 Screening policies considered in the numerical analysis

Policy ID	Institution/Policy	Start age	End age	Screening Interval			
				40-44	45-49	50-54	55+
P1	The annual option of the ACS policy (2015)	45	75	NA	1	1	1
P2	The ACS policy with switching interval (2015)	45	75	NA	1	1	2
P3	Biennial screening between age 45 and 75	45	75	NA	2	2	2
P4	Triennial screening between age 45 and 75	45	75	NA	3	3	3
P5	USPSTF (2016), AAFP (2016), France and Netherlands	50	74	NA	NA	2	2
P6	Belgium, Denmark, Finland, Germany, Ireland, Poland and Spain	50	69	NA	NA	2	2
P7	United Kingdom	50	70	NA	NA	3	3

In our numerical analyses, patients are classified into two classes of low and high density where the former includes BI-RADs density classification of almost entirely fatty and scattered fibroglandular and the latter includes heterogeneously dense and extremely dense classes. This is because per the breast density notification law, patients follow the same guideline weather they are in breast density class 3 or 4. Classifying patients into two density groups also reduces the computational complexity, especially for policies with high number of prescribed screenings.

We consider 4 different radiologist types: radiologist with minimizing reporting requirement behavior (type 1), average-skilled radiologist (type 2), perfect radiologist (type 3), and radiologist with minimizing liability behavior (type 4). Radiologists type 1 and 4, always downgrad and upgrade patients breast density categories, respectively. Note that under radiologist type 1, the patient never undergoes a supplemental test while under radiologist type 4, all screening mammograms are followed by a supplemental test. Radiologist type 3 (perfect radiologist) classifies breast density with 100% accuracy and radiologist type 2 (average-skilled) has 13% missclasisfication probability [92], as discussed in Section 4.4.2.

Four patient cases differing in breast cancer risk characteristics including race, breast density, breast cancer family history, and biopsy history are considered. The initial belief for these patients are calculated using the BCSC risk model, described in Section 4.4.3. These cases are as follows:

Case 1: A 45-year-old white woman with no breast cancer family history or prior biopsy. It is assumed that this case is in density class 1 at age 45. This patient is considered to be a low risk case with initial (at age 45) estimated early and advanced breast cancer risks of 0.28% and 0.16% .

Case 2: All risk factors for this case are similar to *Case 1*, except for the initial breast density which is assumed extremely dense. This patient's risks of being in early and advanced breast cancer states at age 45 are estimated as 1.19% and 0.67%, respectively.

Case 3: A 45-year-old white woman with a family history of breast cancer and prior biopsy and breast density class of 4. The patient's estimated risks of early and advanced breast cancer at age 45 are estimated as 16.38% and 9.2%, respectively.

Case 4: This case represents the general population. The estimated risks of early and advanced breast cancer for average-risk population at age 45 are 1.67% and 0.94%, respectively [101]. Using the breast density distribution of the women population in the U.S. provided by [62], we estimate the initial belief state for this case.

Figures 4.2 and 4.3 present the probability of detecting cancer in early and advanced states and the expected number of supplemental screening tests for different radiologist types when ultrasound and MRI are used as supplemental screening tests, respectively. Note that under radiologists with minimizing reporting requirement behavior, patients undergo only mammogram tests. Obviously, more aggressive screening strategies are more likely to detect the cancer in early states, where there is a higher chance of survival. That is, 1) the ACS policy with fixed screening intervals has the highest probability of detecting cancer in early states for all four cases and all radiologist types. 2) Under radiologist type 4 where patients always undergo supplemental screenings, patients receive the highest probability of being detected in early cancer state.

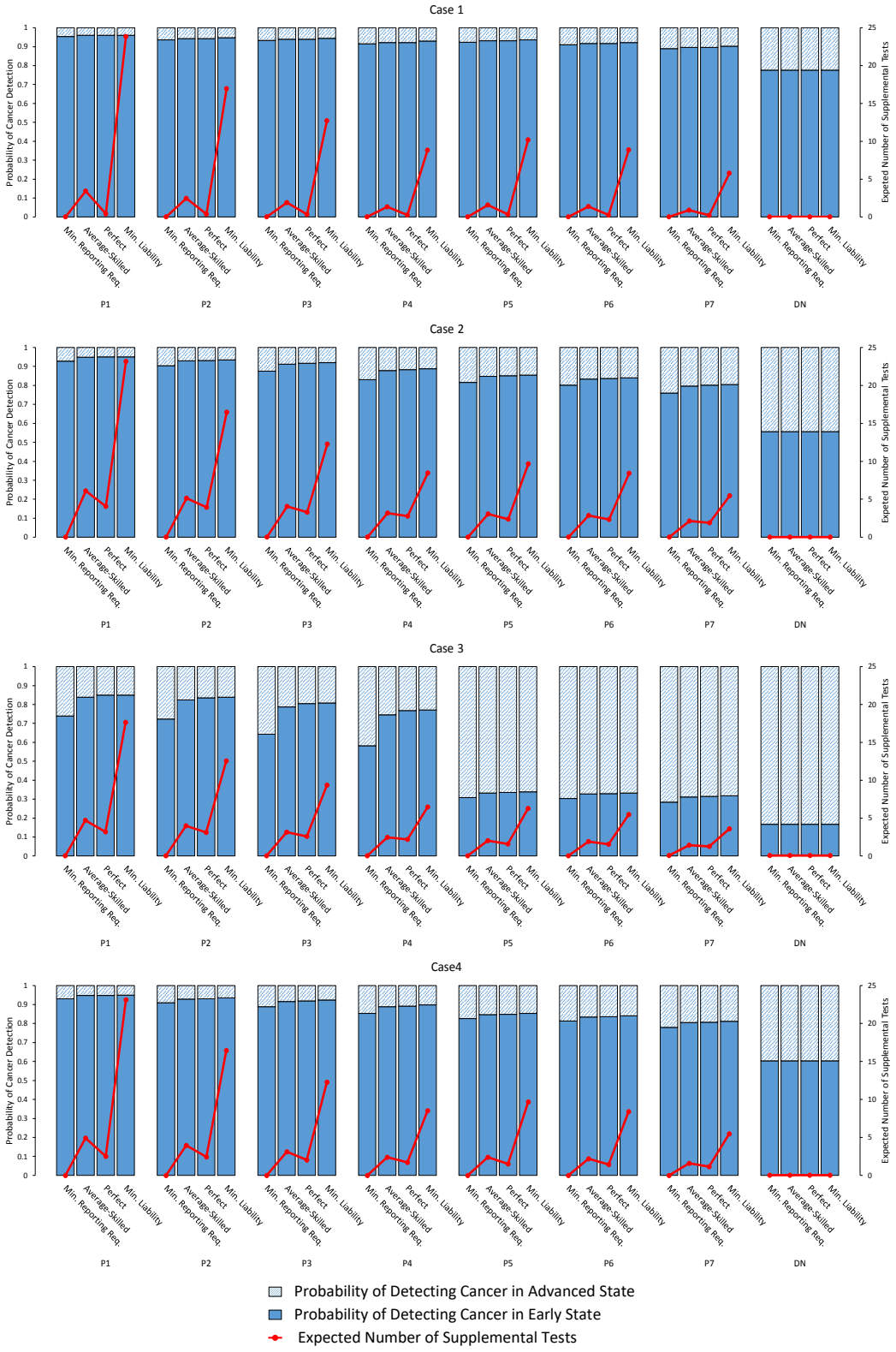


Figure 4.2: Lifetime probabilities of detecting cancer in early and advanced states and expected number of supplemental tests under different radiologist types–supplemental test: *ultrasound*

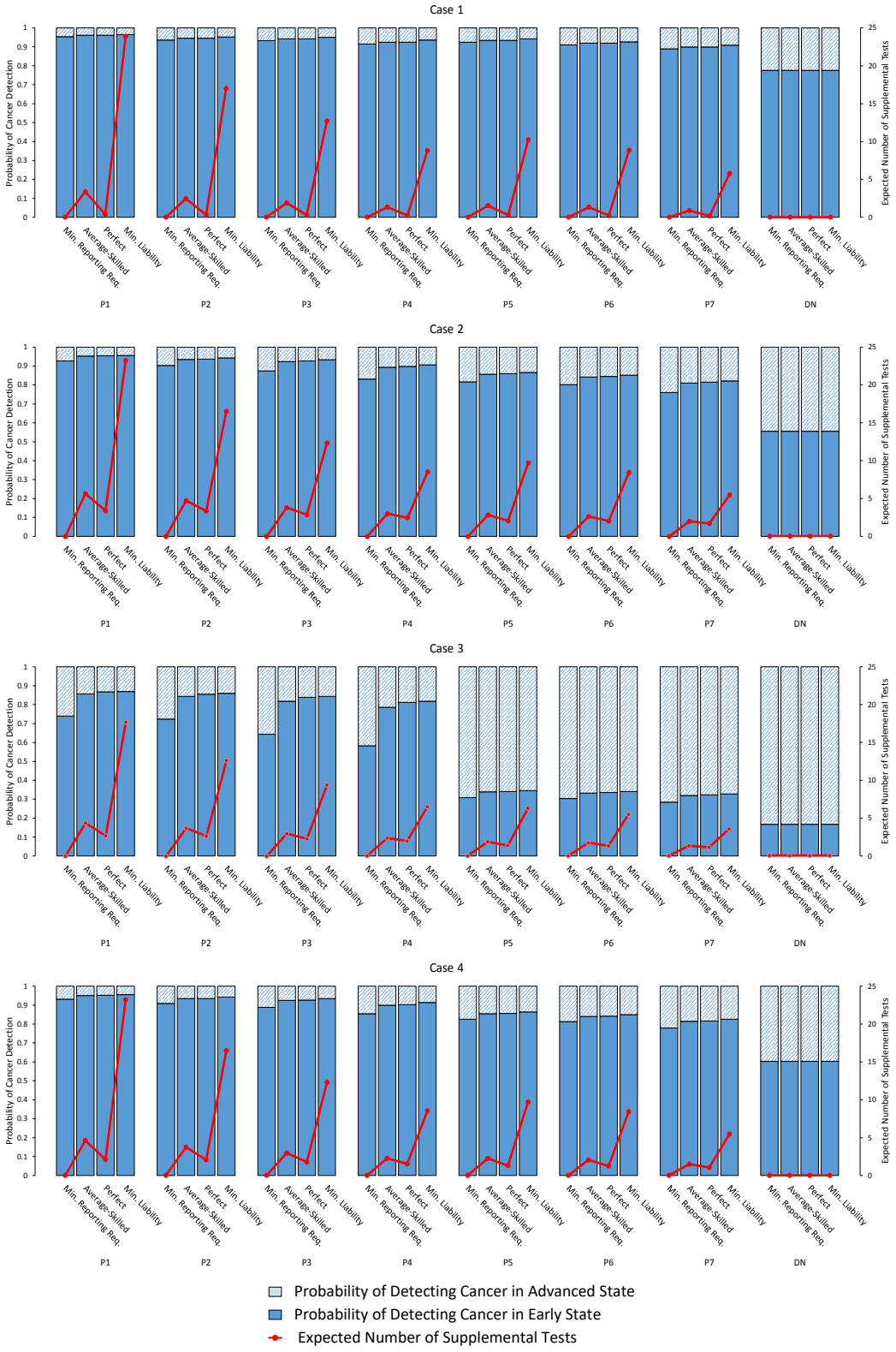


Figure 4.3: Lifetime probabilities of detecting cancer in early and advanced states and expected number of supplemental tests under different radiologist types–supplemental test: *MRI*

Moreover, the results suggest that in terms of the probability of detecting cancer in early states, the difference in performance of perfect and average-skilled radiologists is very small. In fact, the performances of perfect and average-skilled radiologists are very close to the performance of radiologist type 4. However, note that the radiologist type 4 impose a significantly higher number of supplemental screening tests on patients. For example, for *Case 1*, who would not really benefit from supplemental screening since she most likely remains in low breast density in her lifetime, the expected number of supplemental tests that she undergoes are 23.85 and 16.97 under the two ACS policies. Under the perfect radiologists, the expected number of supplemental screening tests are 0.41 and 0.36. Note that this implies that the patient undergoes 23.44 and 16.61 unnecessary supplemental tests under radiologist type 4 which adversely affect her quality of life. Under the average-skilled radiologist, the expected number of supplemental tests are 3.41 and 2.48 for the two ACS policies, which suggests that the unnecessary number of supplemental tests are 3.00 and 2.12, respectively. The expected number of unnecessary supplemental tests are smaller under the other screening policies as they are less aggressive. This implies that average-skilled radiologists performance is very close to perfect radiologists performance when comparing the probability of detecting cancer in early state. However, in terms of the expected number of supplemental tests, the difference might be significant (depending on the aggressiveness of screening polices).

The differences in the performance of radiologists become more evident as the patient's risk increases. That is, for *Case 1* and *Case 3* the differences are at their lowest and highest level, respectively. Specifically, for *Case 1* and under *ultrasound* as

the supplemental test, the probability of detecting the cancer in early state increases by only 0.63%, 1.11% for the ACS policies (P1 and P2) when going from radiologist type 1 to type 4. For *Case 3*, under radiologist type 4, the probability of detecting breast cancer in early states for the two ACS policies are 84.95% and 83.83%, as compared with 73.91% and 72.30% corresponding to radiologist type 1 (a difference of 11.04% and 11.52%, respectively). *Case 2* and *4* fall in between *Case 1* and *3*, with a slightly higher increase in probability of detecting cancer in early state when going from radiologist type 1 to type 4 (i.e., an increase between 1.83% and 3.15%).

This also implies that the efficacy of supplemental screening highly depend on the patients' overall breast cancer risk and not only their breast density. For patients with a lower risk (e.g., *Case 1*), the benefits of undergoing supplemental screening is minimal, as discussed above. For *Case 2* with all risk factors similar to *Case 1* but breast density, we observe an increase in early state probability detection when she undergoes supplemental test. For instance, under mammogram only policy (radiologist type 1), the early detection probability is 92.75% for policy P1 and this probability increases to 94.89% and 95.00% under radiologist type 2 and 3 who recommend patient undergo ultrasound test as needed, per breast density notification law. For *Case 3*, however, undergoing ultrasound screening provide a significantly higher benefit, as discussed above.

Comparing MRI and ultrasound, we observe that MRI always results in a higher probability of detection in early state as it is more sensitive than ultrasound. The difference, although, is negligible, especially for low risk cases. The highest difference in performance of MRI and ultrasound occurs for *Case 3*. For this case,

under the average-killed radiologist and biennial and triennial screening policies, using MRI results in corresponding early detection probabilities of 81.85% and 78.56%, as compared with 78.76% and 74.61% when undergoing ultrasound. On the other hand, MRI imposes slightly lower expected number of supplemental test compared to ultrasound. Note that MRI is a more aggressive test and therefore it might be beneficial to be used only for the cases with higher risk such as *Case 3*.

The results show that in general, policies that recommend starting screening at age 45 outperform those with starting age of 50. That is because breast cancer is more aggressive at younger ages. The impact of starting age in particular and screening policy in general on detecting cancer in early states is specifically very evident for *Case 3*. This implies that the probability of detecting cancer in early states is more impacted by the policy type and patient risk than the radiologist type.

All in all, the results show that 1) breast density is not a sufficient factor when administering supplemental screening as the increases in probabilities of detecting cancer in early states in all cases expect for *Case 3* are very negligible among different radiologist types for any given policy. Other risk factors must be taken into account when recommending a supplemental screening (as shown in *Case 3*). 2) Average and perfect radiologists performance are very similar and comparable to the performance of radiologist type 4. Additionally, if the patient is not a high risk case, radiologist type 1 performance is also comparable to the other radiologist types. This implies that radiologist type impact is not as significant as other factors such as the patient risk factors and screening frequency.

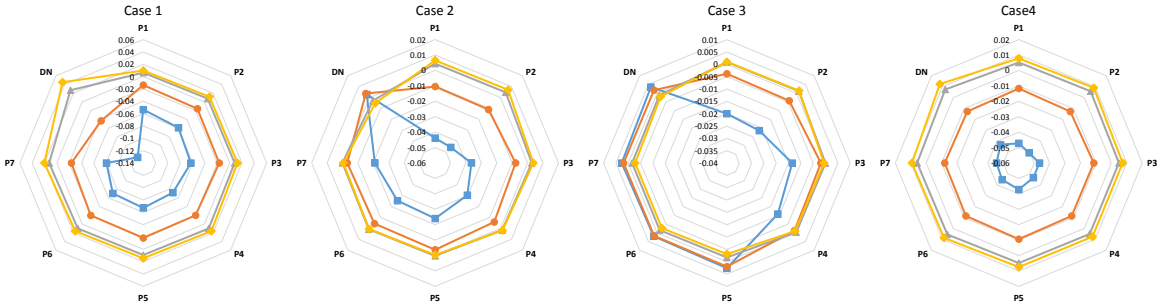
4.5.1 Sensitivity analyses

In this section, we conduct sensitivity analyses on 1) odds ratio (OR) of breast cancer risk comparing women in different density classes and 2) sensitivity and specificity of supplemental screening and probability of cancer showing symptoms. These parameters are selected due to the variability in their reported values in the literature. We consider 4 different ORs, excluding the baseline OR. In the second part of sensitivity analyses, we consider 18 different combinations for sensitivity and specificity of the supplemental test and the probability of the cancer showing symptoms. In total, for each patient case, we evaluate the outcomes under 832 settings.

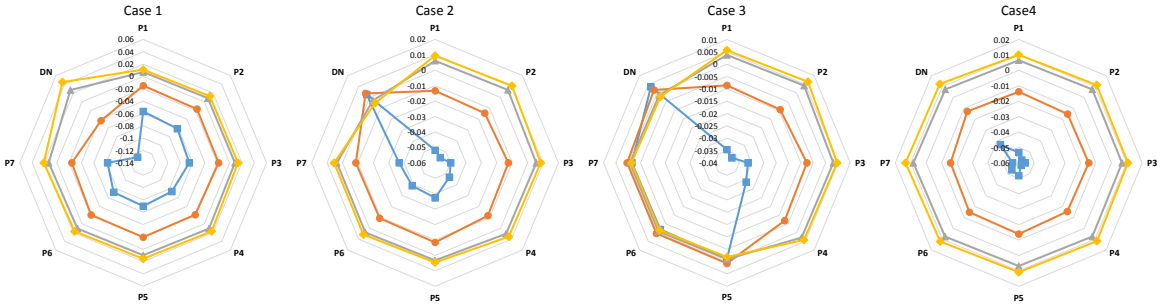
Odds ratios

Based on the previous studies, the odds ratio for developing breast cancer for the most dense compared with the least dense breast tissue categories ranges from 1.46 to 6.0 [65, 66]. We use the midpoint value (OR=3.73) in our baseline analyses, presented in Figures 4.2 and 4.3. Here, we consider OR values of 1.46, 2.595, 4.865, and 6. These ORs are selected to include minimum and maximum values reported in the literature, as well as the midpoints values of the intervals formed by these values and the baseline OR. Note that as the OR increases, the breast cancer risk gap between women with high and low breast density increases.

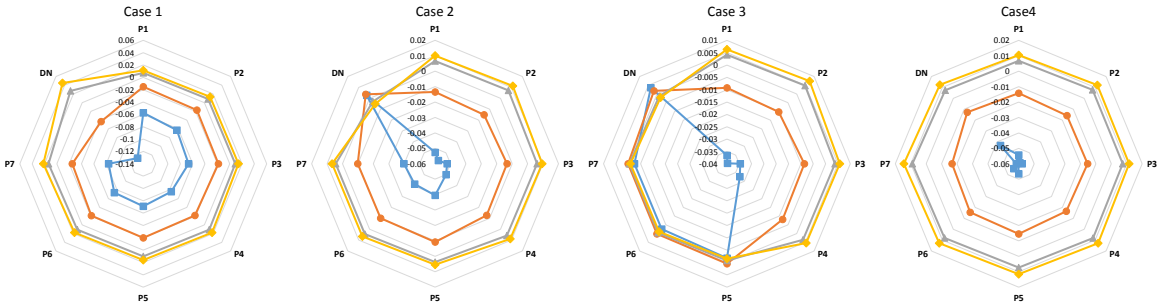
Figure 4.4 presents the change in the early detection probability caused by a change in OR values, when compared to the baseline. Note that negative and positive changes present a decreased and an increased probability of early detection, respectively.



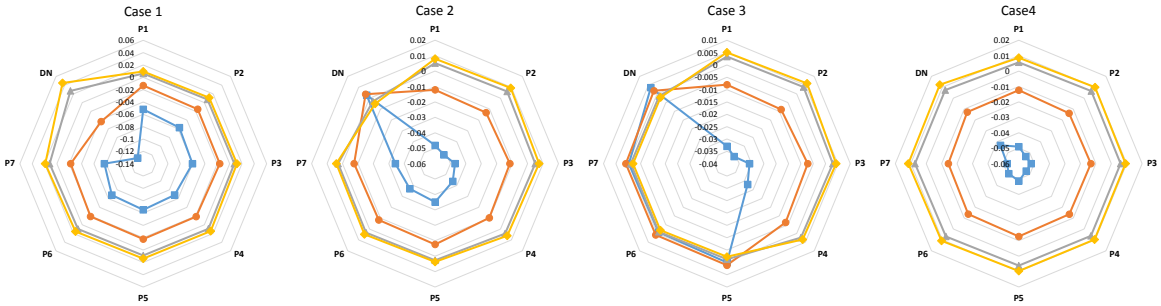
(a) Min. Reporting Req. Radiologist



(b) Average-Skilled Radiologist



(c) Perfect Radiologist



(d) Min. Liability Radiologist

Figure 4.4: Results of sensitivity analyses on odds ratio of breast cancer comparing fatty and extremely dense classes. Note that negative/positive values imply decreased/increased probability of detecting cancer in early states compared with the baseline ($OR = 3.73$)

Under the maximum OR value ($OR = 6$), we observe an average absolute change of 1.08% across all cases, radiologist types and policies, with a maximum of 4.5% (under *Case 1*, radiologist type 1, and *do nothing* policy). Under increased OR, patients with low breast density (e.g., *Case 1* at age 45) carry lower breast cancer incidence and progression rates compared to the baseline. On the other hand, patients with high breast density (e.g., *Case 2* at age 45) carry higher breast cancer risk. Interestingly, the results suggest that for both of these cases, the probability of early detection increases (except for *Case 2* under *do nothing* policy.) This is expected for *Case 1* as this case starts and most likely remains in density class 1 in her lifetime. For *Case 2*, the increase in detection probability seems counter-intuitive since this case has an increased breast cancer risk due to occupying density class 4. However, note that the increase in early detection probability is very negligible and due to probable transition of this patient to a low density class over the course of few epochs. Note that, based on our data and previous studies, breast density stochastically decreases over time [67, 68, 69]. For the other two cases, the changes are very negligible, especially for *Case 3*. Obviously, under OR value of 4.865, the changes are smaller but follow similar patterns.

With a decreased OR, we observe a higher changes in probability of early detection. Specifically, under OR value of 1.46, we observe a maximum change of 12.72% (for *Case 1* under *do nothing* policy). In general, we observe a decrease in probability of early detection for *Case 1*, *2*, and *4* (except for *do nothing* policy for *Case 2*). Note that with a decreased OR, patients with lower breast density carry higher risk compared with the baseline which results in a decrease in probability of early detection. Under *do nothing* policy for *Case 2*, we observe a very negligible

change in early detection probability. Note that the small change in this case is mainly contributed by the patient's belief of being in cancer state. The patient's cancer belief under this policy is generally higher due to lack of screening tests and consequent less informative risk adjustments of the patient. This is also true for *Case 3* whose change in early detection probability is very negligible for *do nothing* policy and screening policies starting at age 50.

Generally, with increased/decreased OR we observe an increase/decrease in early detection probability. The magnitude of changes, however, vary across different patients and screening policies. The results, in general, are consistent and prompt comparable conclusions with those derived under the baseline OR.

Supplemental test accuracy and probability of cancer symptoms

We consider 3 different levels of sensitivity and 3 different levels of specificity for supplemental screening test. We also consider 2 different levels for the probability of showing symptoms. Using a full factorial design, we have 18 different combinations for these parameters. We consider 1) joint sensitivity of mammogram and ultrasound decreased by 5%, 2) joint sensitivity of mammogram and MRI increased by 5%, and 3) the midpoint of the interval formed by the joint mammogram/ultrasound and mammogram/MRI sensitivity values in the baseline. Similarly, we calculate three levels for supplemental screening specificity. We consider 5% increase and decrease in the baseline probability of showing symptoms.

Figure 4.5 presents the change in the probability of detecting cancer in early state for *Case 3* under average-skilled radiologist. We present the results only for

Case 3 for brevity; however, note that in general, sensitivities to parameters changes are smaller for the other cases. For example, the maximum change in the probability of detecting cancer in early states for *Case 1* is only 1.57% when compared with the baseline. The maximum change for *Case 3* is 8.44%.

Obviously, the changes in sensitivity of testing has the highest impact on the probabilities of early detection. As the test sensitivity increases, the probability of detecting cancer in early state increases. The increase in early detection probability is smaller for policies with more frequent screening policies as more frequent screenings compensate for possible false negatives. The increase is at the highest for policy P4, due to the longer intervals between subsequent recommended screening tests (3 years). Note that for policy P7, although following the same screening frequency, the change is very small due to the delayed screening start age. That is, it is very likely that the cancer remains undetected and grows to advanced state between age 45 and 50 which consequently decreases the chance of detecting the cancer in early states. Recall that *Case 3* is a high risk case and has an aggressive breast cancer.

Similarly, increasing the probability of showing symptoms results in an increase in the probability of detecting cancer in early states. The highest impact is associated with less aggressive policies. For instance, for *Case 3* and under *do nothing* policy, we observe an increase of 3.86% in the probability of early detection when increasing the probability of showing symptoms. On the other hand, for the ACS policy with fixed screening interval (P1), we observe the lowest sensitivity to the probability of showing symptoms, with an average increase of 1.43% across different sensitivity analyses combinations considered here.

Specificity of screening tests has an opposite impact on the probability of detecting cancer in early states. That is, increased joint specificity causes a decrease in the probability of early detection. This happens as with increased joint specificity, the number of false positive observations, resulting in consequent biopsies, decreases. Recall that biopsies are perfect and determine with certainty that the patient is cancer free. This causes an overall decrease in the belief that the patient is in cancer states and the probability of detecting cancer in early state.

In summary, the sensitivity of joint mammogram and supplemental test has the largest and the specificity of joint mammogram and supplemental test has the smallest impact on the probability of early detection. The conclusions in the baseline analyses still hold. That is, as long as a supplemental test is administered for a patient (radiologist types 2–4), a bias in radiologist's classification has a negligible impact on the probability of detecting cancer in early states. That is, under the same policy, the difference in the probabilities of early detection across radiologists type 2 through 4 is minimal. Moreover, breast density should not be the sole determining factor as whether a patient should be referred to supplemental screening tests. Other breast cancer risk factors and frequency of screening tests should be considered when making such referrals.

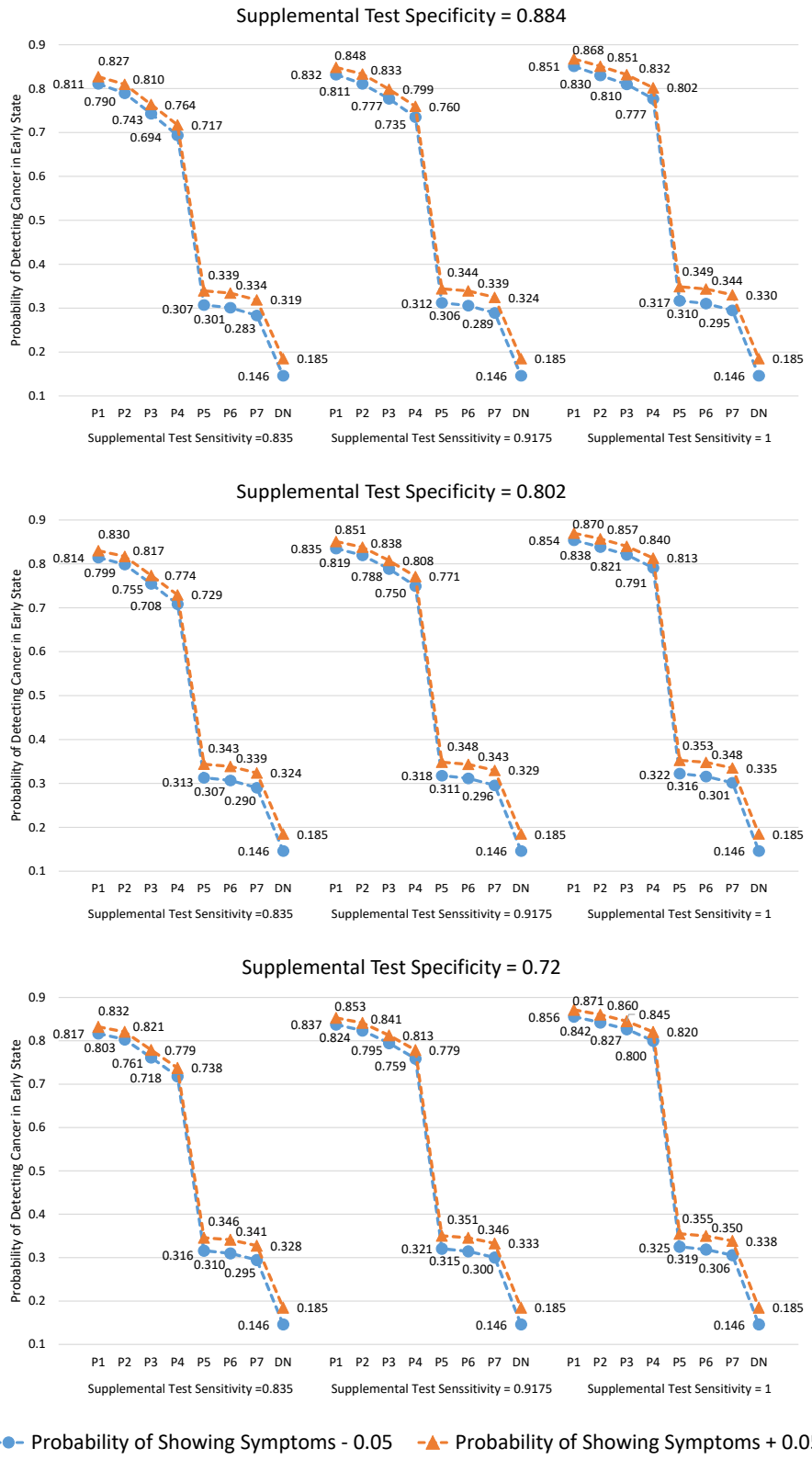


Figure 4.5: Results of sensitivity analyses on joint mammogram and supplemental test sensitivity and specificity and the probability of showing symptoms: *Case 3* and average-skilled radiologist

4.6 Conclusion

Breast density is associated with increased breast cancer risk and decreased mammography screening sensitivity. To promote early breast cancer detection in women with high breast density, breast density notification laws have been enacted in 38 states. The laws, however, have caused controversial debates on 1) whether supplemental screening improves patients' outcomes and 2) the impact of radiologists bias in breast density classification on patient's outcomes.

In this study, we develop a POMC model, incorporating both patients' health and breast density dynamics, to investigate the impact of supplemental screening tests and role of radiologists' bias on a patients' health outcomes. Specifically, we consider the conditional probability of detecting breast cancer in early and advanced states given the patient develops cancer in her lifetime. We consider the expected number of supplemental tests a patient undergoes in her lifetime as another outcome.

Our results indicate that 1) breast density should not be the only risk factor when referring a patient to supplemental screening and 2) the radiologist's bias may affect the efficacy of supplemental screening. Specifically, patient outcomes may be significantly affected under radiologists who consistently upgrade or downgrad patient's breast density. However, the bias introduced by an average-skilled radiologist may not significantly affect the patients' outcomes.

Given that screening technologies are continuously advancing, a possible future research direction would be to analyze the impact of emerging technologies (e.g., tomosynthesis) on the necessity of supplemental screening tests for women with high breast density. Moreover, patients' adherence is a very influential factor on the

effectiveness of a screening policy and patients' outcomes that is not considered in this study. A possible future work is to incorporate this factor.

CHAPTER 5

CONCLUSIONS

This dissertation proposed and investigated the use of stochastic decision models in breast cancer preventive care. We addressed several controversial issues in breast cancer screening programs, including overdiagnosis risk and the implications of use of supplemental screening test in breast ccancer preventive care.

In chapter 2, we use two stochastic frameworks for the patient's adherence and the breast cancer natural history to estimate different measures of overdiagnosis risk of mammography screening while incorporating uncertainty in patients' adherence behaviors. The measures of overdiagnosis risk investigated in this study include the proportion of detected cancers that are overdiagnosed in a screened population, mortality risk considering overdiagnosis, and overtreatment costs. We analyze the harm-benefit trade-off of some in-practice policies by measuring the number of lives that are saved per each overdiagnosed case. We also estimate the associated proportion of overtreatment cost to breast cancer care cost for each policy. Our results show that, although overdiagnosis rate is relatively high in breast cancer screening, the benefits of breast cancer mammography screening outweigh the overdiagnosis risk.

In chapter 3, we reviewed several methods to estimate Markov model parameters using a set of data with intermittent missing values. We discussed three

categories of methods for datasets with ignorable missingness mechanism including deletion, imputation, and augmentation. Moreover, we executed a frequentist setting (Expectation-Maximization algorithm) and a Bayesian framework (Gibbs Sampler) on several simulated data to measure the bias of the results from each algorithm. Finally, we used Baum-Welch algorithm on a set of sparse unbalanced data consist of hidden observations for breast density states of a group of patients at Louisiana Cancer Prevention and Control center. Estimated breast density transition matrices specify a relatively high probability that successive observations recur. This implies that future outcome has a high dependency on the current outcome. The results of this chapter were used in chapter 4 where we developed a POMC model to investigate the impact of the radiologists' expertise level and behavior on the implementation of the breast density notification law and their implications on patients' health outcomes. The efficacy of ultrasound and MRI as supplemental screening methods are also studied under implementation of the density notification laws. The patients' outcome measures investigated in this chapter include the lifetime probability of detecting cancer in early and advanced cancer states in the population of patients who would eventually develop cancer at some point in their lifetime and expected number of required supplemental screening for each policy under radiologists with different level of expertise and different supplemental screening. Based on our results, referring patients to a supplemental test solely based on their breast density may not significantly improve patient's outcomes and other risk factor might be considered when making such decision. Additionally, average-skilled radiologists performance are comparable with a perfect radiologist

performance. However, significant bias (i.e., consistent upgrading and downgrading of breast density classes) can negatively impact a patient health outcomes.

Some possible future works include:

- Development of models to determine race-specific overdiagnosis risk given the disparity in incidence and mortality among different races.
- Development of an optimization model to derive optimal screening policy that controls the risk of overdiagnosis.
- Implementation of Markov model parameter estimation methods on a larger breast density dataset with more number of observations for each patient to reduce the bias of these estimations.
- Analyzing the impact of emerging technologies (e.g., tomosynthesis) on the necessity of supplemental screening tests for women with high breast density.
- Considering patients' adherence as an influential factor on the effectiveness of a screening policy and patients' outcomes in analyzing the effectiveness of breast density notification law.

APPENDIX A

PARAMETER VALUES FOR CHAPTER 2

Table A.1 Adherence state transition probabilities

	Time t+1					
	40-49 years		50-64 years		65+ years	
Time t	Regular	Irregular	Regular	Irregular	Regular	Irregular
Regular	0.7627	0.2372	0.7299	0.2700	0.7429	0.2570
Irregular	0.3837	0.6162	0.2356	0.7643	0.1440	0.8559

Table A.2 Adherence rates

	40-49 years	50-84 years	85-100 years
Regular screening	0.9014	0.9551	0.8924
Irregular screening	0.2914	0.2619	0.0603

Table A.3 Initial adherence belief

	Regular screening	Irregular screening
Initial adherence belief	0.2309	0.7691

Table A.4 Health state transition probabilities

	Time $t+1$					
	s_1	s_2	s_3	s_4	s_5	s_6
Time t	40-44 years					
s_1	0.99784	0.000975	0	0	0	0.00118
s_2	0	0.7309	0.26792	0	0	0.00118
s_3	0	0	0.928548	0.070272	0	0.00118
s_4	0	0	0	0.8442	0.15462	0.00118
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	45-49 years					
s_1	0.99703	0.001281	0	0	0	0.001691
s_2	0	0.7309	0.267409	0	0	0.001691
s_3	0	0	0.928037	0.070272	0	0.001691
s_4	0	0	0	0.8442	0.154109	0.001691
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	50-54 years					
s_1	0.99563	0.001682	0	0	0	0.00269
s_2	0	0.7309	0.26641	0	0	0.00269
s_3	0	0	0.927038	0.070272	0	0.00269
s_4	0	0	0	0.8442	0.15311	0.00269
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	55-59 years					
s_1	0.99371	0.00198	0	0	0	0.004314
s_2	0	0.7309	0.264786	0	0	0.004314
s_3	0	0	0.925414	0.070272	0	0.004314
s_4	0	0	0	0.8442	0.151486	0.004314
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	60-64 years					
s_1	0.99086	0.002214	0	0	0	0.006927
s_2	0	0.7309	0.262173	0	0	0.006927
s_3	0	0	0.922801	0.070272	0	0.006927
s_4	0	0	0	0.8442	0.148873	0.006927
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1

Continued on next page

Table A.4 *Health state transition probabilities - Countinued ...*

	s_1	s_2	s_3	s_4	s_5	s_6
	65-69 years					
s_1	0.98656	0.002377	0	0	0	0.011064
s_2	0	0.7309	0.258036	0	0	0.011064
s_3	0	0	0.918664	0.070272	0	0.011064
s_4	0	0	0	0.8442	0.144736	0.011064
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	70-74 years					
s_1	0.97945	0.002509	0	0	0	0.018045
s_2	0	0.7309	0.251055	0	0	0.018045
s_3	0	0	0.911683	0.070272	0	0.018045
s_4	0	0	0	0.8442	0.137755	0.018045
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	75-79 years					
s_1	0.96702	0.0024	0	0	0	0.030581
s_2	0	0.7309	0.238519	0	0	0.030581
s_3	0	0	0.899147	0.070272	0	0.030581
s_4	0	0	0	0.8442	0.125219	0.030581
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	80-84years					
s_1	0.92519	0.002075	0	0	0	0.07274
s_2	0	0.7309	0.19636	0	0	0.07274
s_3	0	0	0.856988	0.070272	0	0.07274
s_4	0	0	0	0.8442	0.08306	0.07274
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	85-90 years					
s_1	0.92519	0.002075	0	0	0	0.07274
s_2	0	0.7309	0.19636	0	0	0.07274
s_3	0	0	0.856988	0.070272	0	0.07274
s_4	0	0	0	0.8442	0.08306	0.07274
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1

Continued on next page

Table A.4 *Health state transition probabilities - Countinued ...*

	s_1	s_2	s_3	s_4	s_5	s_6
	90-94					
s_1	0.92519	0.002075	0	0	0	0.07274
s_2	0	0.7309	0.19636	0	0	0.07274
s_3	0	0	0.856988	0.070272	0	0.07274
s_4	0	0	0	0.8442	0.08306	0.07274
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1
	95-100					
s_1	0.92519	0.002075	0	0	0	0.07274
s_2	0	0.7309	0.19636	0	0	0.07274
s_3	0	0	0.856988	0.070272	0	0.07274
s_4	0	0	0	0.8442	0.08306	0.07274
s_5	0	0	0	0	1	0
s_6	0	0	0	0	0	1

Table A.5 Mammography sensitivity

	Early breast cancer		Advance breast cancer	
	40-49	50+	40-49	50+
Mammography sensitivity	0.75033	0.85449	0.81860	0.93224

Table A.6 Initial health belief

Healthy	Early breast cancer	Advance breast cancer
0.997	0.0006	0.0024

Table A.7 Survival rate for screen-detected breast cancer by state

State	Survival probabilities
Healthy to Healthy	0.99
Healthy to Early	0.004
Early to Early	0.7309
Early to Advanced	0.263
Advanced to Symptomatic	0.0703
Advanced to Advanced	0.923

Table A.8 U.S. age composition

Age groups	Population by age	Age groups	Population by age
40-44 years	10,496,987	70-74 years	5,034,194
45-49 years	11,499,506	75-79 years	4,135,407
50-54 years	11,364,851	80-84 years	3,448,953
55-59 years	10,141,157	85-89 years	2,346,592
60-64 years	8,740,424	90-94 years	1,023,979
65-69 years	6,582,716	95-99 years	288,981

Table A.9 Breast cancer treatment costs

	Cost
Early breast cancer treatment	$ 32103
Advance breast cancer treatment	$ 51837
Screening	$ 102

APPENDIX B

PARAMETER VALUES FOR CHAPTER 4

Table B.1 Breast density observation probability matrices

	Observed state	
	Low density	High density
True state	Radiologist with Min. Reporting Requirement	
Low density	1	0
High density	1	0
	Average-Skilled Radiologist	
Low density	0.87	0.13
High density	0.13	0.87
	Perfect-skilled Radiologist	
Low density	1	0
High density	0	1
	Radiologist with Min. Liability	
Low density	0	1
High density	0	1

Table B.2 Age and density-specific mammography specificity

	Age groups						
Breast density	45-49	50-54	55-60	60-64	65-69	70-74	75-79
Low density	0.901563	0.916036	0.916889	0.917585	0.917988	0.918083	0.918201
High density	0.866351	0.889283	0.888364	0.887937	0.887579	0.88737	0.887407

Table B.3 Density-specific mammography sensitivity

	Low breast density	High breast density
Mammography sensitivity	0.765	0.58

Table B.4 Initial density belief state

	Age groups							
Breast density	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79
d_I	0.046	0.055	0.075	0.098	0.117	0.13	0.136	0.139
d_{II}	0.338	0.364	0.422	0.471	0.5	0.521	0.537	0.539
d_{III}	0.472	0.458	0.416	0.373	0.338	0.313	0.296	0.291
d_{IV}	0.144	0.123	0.086	0.058	0.045	0.036	0.031	0.031

Table B.5 Initial health belief state

	Health state		
Patient	Cancer free	Early breast cancer	Advance breast cancer
Case 1	0.995552	0.002848	0.0016
Case 2	0.981374	0.011926	0.0067
Case 3	0.74424	0.16376	0.092
Case 4	0.973868	0.016732	0.0094

APPENDIX C

DATA PRE-PROCESSING

In Chapter 3, we applied several hidden Markov model parameter approximation methods to estimate breast density transition probabilities using mammography screening data from Louisiana Cancer Prevention and Control Programs [4]. This longitudinal data has been gathered from 436 patients since 2016 and sent as the physical reports. These reports contained information such as the age of the patient, BI-RADS health and breast density state of the patient, date of screening, radiologist name, and the future recommendation for the patient. The first data preprocessing step was to convert these text files to excel worksheets manually.

In addition, the dataset contains missing values which comes from the missed scheduled visits of patients for evaluation of their health and breast density status. We assume the missingness mechanism in this dataset is ignorable since the probability of a patient shows up for breast screening mostly depends on their previous screening results rather than current observation. This comes from the fact that the symptomatic breast cancer usually happens at the very advanced stage of a cancer, which means that patients who show up for screening are mostly because of previous suspicious results or to follow-up the screening policy recommendation. In the next step, we divided observations based on the age of patients into two groups $(40-55/55+)$ in

order to consider the age and menopause effects on breast density dynamics in our analysis. A small section of this dataset is presented below.

Table C.1 A sample of dataset for estimation of breast density dynamics

ID	Date of birth	Date of screening	Age	Density state	Health state	Future recommendation	Radiologist
332398	11/27/1953	04/05/2016	62	d_{II}	2	Screening in 1 year	MBM
332398	11/27/1953	11/06/2017	64	d_{II}	2	Screening in 1 year	MBM
424266	10/23/1956	01/22/2016	59	d_{I}	1	Screening in 1 year	MBM
424266	10/23/1956	11/07/2017	61	d_{I}	1	Screening in 1 year	MBM
415831	08/08/1951	01/20/2015	63	d_{III}	2	Screening in 1 year	MBM
415831	08/08/1951	06/19/2015	64	d_{III}	2	Screening in 1 year	LFSR
415831	08/08/1951	08/26/2016	65	d_{III}	2	Screening in 1 year	BLM
415831	08/08/1951	11/07/2017	66	d_{II}	2	Screening in 1 year	MJM
624400	06/28/1970	11/01/2016	46	d_{III}	2	Screening in 1 year	BLM
624400	06/28/1970	11/07/2017	47	d_{III}	2	Screening in 1 year	MJM
298288	11/23/1950	10/12/2016	66	d_{II}	2	Screening in 1 year	BLM
298288	11/23/1950	11/08/2017	67	d_{I}	2	Screening in 1 year	MJM
390921	08/12/1960	11/04/2016	56	d_{II}	1	Screening in 1 year	BLM
390921	08/12/1960	11/08/2017	57	d_{II}	1	Screening in 1 year	MJM

APPENDIX D

PSEUDO-CODES

D.1 Codes for Chapter 2

Algorithm 1: Calculate patient's adherence-health belief state

$t \leftarrow 0$;
while $t <= T$ **do**
  $t \leftarrow t + 1$;
  if $a_t == W$ **then**
    $\kappa_{h_{t+1}}(s'_b) \leftarrow \kappa[\kappa_{h_t}, W, \cdot](s'_b)$;
    $\eta_{\chi_{t+1}}(s'_h) \leftarrow \tau[\eta_{\chi_t}, W, \cdot](s'_h)$;
  else
    $\kappa_{h_{t+1}}(s'_b) \leftarrow \kappa[\kappa_{h_t}, M, o_t](s'_b)$;
    $\eta_{\chi_{t+1}}(s'_h) \leftarrow \tau[\eta_{\chi_t}, M, \theta_t](s'_h)$;
  end
end

Algorithm 2: Calculate patient's outcomes in Chapter 2

$t \leftarrow T$;
while $t => 0$ **do**
  $t \leftarrow t - 1$;
  if $a_t == W$ **then**
    Calculate $\psi_t^W(\eta_{\chi_t})$ using Equation 2.14 recursively;
  else
    Calculate $\psi_t^M(\eta_{\chi_t})$ using Equation 2.15 recursively;
  end
end

D.2 Codes for Chapter 3

Algorithm 3: Baum-Welch Algorithm

Initialize;

for $\nu = 0$ **do**

  Select $\lambda^{(0)} = (\pi^{(0)}, B^{(0)}, A^{(0)})$;

end

Iterative calculation ;

for $\nu = 1, 2, \cdots$ **do**

$$\pi_i^{\nu} \leftarrow \frac{\sum_{\eta=1}^{\Gamma} \gamma_{\eta,1}^{\nu}(i)}{\sum_{\eta=1}^{\Gamma}\sum_{i=1}^{S} \gamma_{\eta,1}^{\nu}(i)};$$

$$a_{ij}^{\nu} \leftarrow \frac{\sum_{\eta=1}^{\Gamma}\sum_{t=1}^{T-1} \xi_{\eta,t}^{\nu}(i,j)}{\sum_{\eta=1}^{\Gamma}\sum_{t=1}^{T-1} \gamma_{\eta,t}^{\nu}(i)};$$

$$b_i(j)^{\nu} \leftarrow \frac{\sum_{\eta=1}^{\Gamma}\sum_{t=1}^{T} \gamma_{\eta,t}^{\nu}(i)\delta(y_{\eta,t}=j)}{\sum_{\eta=1}^{\Gamma}\sum_{t=1}^{T} \gamma_{\eta,t}^{\nu}(i)\delta(y_{\eta,t}\neq .)};$$

where,

$$\alpha_{\eta,t}(i) \leftarrow P(\mathcal{Y}_{\eta,1} = y_{\eta,1}, ..., \mathcal{Y}_{\eta,t} = y_{\eta,t}, s_{\eta,t} = i|\lambda)$$

$$\beta_{\eta,t}(i) \leftarrow P(\mathcal{Y}_{\eta,t+1} = y_{\eta,t+1}, ...\mathcal{Y}_{\eta,T} = y_{\eta,T}|s_{\eta,t} = i, \lambda)$$

$$\gamma_{\eta,t}(i) \leftarrow \frac{\alpha_{\eta,t}(i)\beta_{\eta,t}(i)}{\sum_{j=1}^{S} \alpha_{\eta,T}(j)}$$

$$\xi_{\eta,t}(i,j) \leftarrow \frac{\alpha_{\eta,t}(i)a_{ij}b_j(y_{\eta,t+1})\beta_{\eta,t+1}(j)}{\sum_{i=1}^{S} \alpha_{\eta,T}(i)}$$

end

Termination;

Obtain $\lambda^{\nu+1} = (\pi^{\nu+1}, B^{\nu+1}, A^{\nu+1})$

D.3 Codes for Chapter 4

Algorithm 4: Calculate patient's density-health belief state

$t \leftarrow 0$;
while $t <= T$ **do**
  $t \leftarrow t + 1$;
  if $a_t == W$ **then**
    $\boldsymbol{\beta_{t+1}} \leftarrow \boldsymbol{\nu}[\boldsymbol{\beta}_t, W, \cdot](z)$;
  else
    $\boldsymbol{\beta_{t+1}} \leftarrow \boldsymbol{\nu}[\boldsymbol{\beta}_t, M, o_t](z)$;
  end
end

Algorithm 5: Calculate patient's outcomes in Chapter 4

$t \leftarrow T$;
while $t => 0$ **do**
  $t \leftarrow t - 1$;
  if $t == T$ **then**
    Calculate $\gamma_E(s)$ and $\gamma_A(s)$ using cancer progression rates and probability of showing symptoms after period T
  else
    if $a_t == W$ **then**
      Calculate $\mathcal{E}_t^W(\boldsymbol{\beta}_t)$ using Equation 4.2 recursively;
      Calculate $\mathcal{A}_t^W(\boldsymbol{\beta}_t)$ using Equation 4.3 recursively;
      Calculate $\mathcal{V}_t^W(\boldsymbol{\beta}_t)$ using Equation 4.7 recursively;
    else
      Calculate $\mathcal{E}_t^M(\boldsymbol{\beta}_t)$ using Equation 4.4 recursively;
      Calculate $\mathcal{A}_t^M(\boldsymbol{\beta}_t)$ using Equation 4.5 recursively;
      Calculate $\mathcal{V}_t^M(\boldsymbol{\beta}_t)$ using Equation 4.8 recursively;
    end
  end
end

BIBLIOGRAPHY

[1] American Cancer Society. How Common is Breast Cancer, 2020. (Accessed 1-August-2020).

[2] American Cancer Society. Breast Cancer Facts and Figures 2019-2020, 2019.

[3] Sevda Molani, Mahboubeh Madadi, and Wesley Wilkes. A partially observable markov chain framework to estimate overdiagnosis risk in breast cancer screening: Incorporating uncertainty in patients adherence behaviors. *Omega*, 89:40–53, 2019.

[4] Health Sciences Center LSU. Louisiana cancer prevention and control programs, 2016-20. `https://louisianacancer.org/`.

[5] Svjetlana Mujagic. The influence of breast density on the sensitivity and specificity of ultrasound and mammography in breast cancer diagnosis. *Acta Medica Academica*, 40(2):132–139, 2011. ISSN 18401848. doi: 10.5644/ama2006-124.16.

[6] Elizabeth Morris, Stephen A Feig, Madeline Drexler, and Constance Lehman. Implications of overdiagnosis: impact on screening mammography practices. *Population health management*, 18(S1):S–3, 2015.

[7] Jamie L Carter, Russell J Coletti, and Russell P Harris. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ*, 350:g7773, 2015.

[8] Independent UK Panel On Breast Cancer Screening and others. The benefits and harms of breast cancer screening: an independent review. *The Lancet*, 380 (9855):1778–1786, 2012.

[9] Christoph I Lee and Ruth Etzioni. Missteps in current estimates of cancer overdiagnosis. *Academic radiology*, 24(2):226–229, 2017.

[10] Eduardo Sabaté. *Adherence to long-term therapies: evidence for action*. World Health Organization, 2003.

[11] Center for Disease Control and Prevention. `https://www.cdc.gov/nchs/data/hus/2013/083.pdf`, 2014. (Accessed 1-August-2020).

[12] R Edward Hendrick and Mark A Helvie. United states preventive services task force screening mammography recommendations: science ignored. *American*

Journal of Roentgenology, 196(2):W112–W116, 2011.

[13] American Cancer Society. Breast Cancer Facts and Figures 2017-2018, 2017.

[14] Kristin M Schueler, Philip W Chu, and Rebecca Smith-Bindman. Factors associated with mammography utilization: a systematic quantitative review of the literature. *Journal of women's health*, 17(9):1477–1498, 2008.

[15] Lisa M Maillart, Julie Simmons Ivy, Scott Ransom, and Kathleen Diehl. Assessing dynamic breast cancer screening policies. *Operations Research*, 56(6): 1411–1427, 2008.

[16] Stephen W Duffy, Hsiu-Hsi Chen, Laszlo Tabar, and Nicholas E Day. Estimation of mean sojourn time in breast cancer screening using a markov chain model of both entry to and exit from the preclinical detectable phase. *Statistics in medicine*, 14(14):1531–1543, 1995.

[17] Jenny Chia-Yun Wu, Matti Hakama, Ahti Anttila, Amy Ming-Fang Yen, Nea Malila, Tytti Sarkeala, Anssi Auvinen, Sherry Yueh-Hsia Chiu, and Hsiu-Hsi Chen. Estimation of natural history parameters of breast cancer based on non-randomized organized screening data: subsidiary analysis of effects of inter-screening interval, sensitivity, and attendance rate on reduction of advanced cancer. *Breast cancer research and treatment*, 122(2):553–566, 2010.

[18] Harris Julian Gaster Bloom, WW Richardson, and EJ Harries. Natural history of untreated breast cancer (1805-1933). *British medical journal*, 2(5299):213, 1962.

[19] Jiaquan Xu, Kenneth D Kochanek, Sherry L Murphy, Betzaida Tejada-Vera, et al. National vital statistics reports. *National vital statistics reports*, 58(19), 2010.

[20] Centers for Disease Control and Prevention. National Health Interview Survey, 2015. (Accessed 1-August-2020).

[21] David Nelson, Gary Kreps, Bradford Hesse, Robert Croyle, Gordon Willis, Neeraj Arora, Barbara Rimer, K Vish Viswanath, Neil Weinstein, and Sara Alden. The health information national trends survey (hints): development, design, and dissemination. *Journal of health communication*, 9(5):443–460, 2004.

[22] Mahboubeh Madadi, Shengfan Zhang, and Louise M Henderson. Evaluation of breast cancer mammography screening policies considering adherence behavior. *European Journal of Operational Research*, 247(2):630–640, 2015.

[23] Turgay Ayer, Oguzhan Alagoz, Natasha K Stout, and Elizabeth S Burnside. Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science*, 62(5):1339–1362, 2015.

[24] National Cancer Institute. Gail Risk Model, 2011. (Accessed 1-August-2020).

[25] PC Allgood, SW Duffy, O Kearins, E O'sullivan, N Tappenden, MG Wallis, and G Lawrence. Explaining the difference in prognosis between screen-detected and symptomatic breast cancers. *British journal of cancer*, 104(11):1680–1685, 2011.

[26] US Census Bureau. Age and Sex Composition: 2010, 2011. (Accessed 1-August-2020).

[27] Angela B Mariotto, K Robin Yabroff, Yongwu Shao, Eric J Feuer, and Martin L Brown. Projections of the cost of cancer care in the united states: 2010–2020. *Journal of the National Cancer Institute*, 103(2):117–128, 2011.

[28] Mei-Sing Ong and Kenneth D Mandl. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. *Health affairs*, 34(4):576–583, 2015.

[29] Mahboubeh Madadi, Shengfan Zhang, Karen H Kim Yeary, and Louise M Henderson. Analyzing factors associated with women's attitudes and behaviors toward screening mammography using design-based logistic regression. *Breast cancer research and treatment*, 144(1):193–204, 2014.

[30] Claudine Isaacs, Beth N Peshkin, Marc Schwartz, Tiffani A DeMarco, David Main, and Caryn Lerman. Breast and ovarian cancer screening practices in healthy women with a strong family history of breast or ovarian cancer. *Breast Cancer Research and Treatment*, 71(2):103–112, 2002.

[31] American Cancer Society. Lifetime Risk of Developing or Dying From Cancer , 2020. (Accessed 1-August-2020).

[32] Hung-Wen Yeh, Wenyaw Chan, Elaine Symanski, and Barry R Davis. Estimating transition probabilities for ignorable intermittent missing data in a discrete-time markov chain. *Communications in Statistics—Simulation and Computation®*, 39(2):433–448, 2010.

[33] Nan M Laird. Missing data in longitudinal studies. *Statistics in medicine*, 7 (1-2):305–315, 1988.

[34] Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

[35] F. Arteaga and A.J. Ferrer-Riquelme. 3.06 - missing data. In Steven D. Brown, Romá Tauler, and Beata Walczak, editors, *Comprehensive Chemometrics*, pages 285 – 314. Elsevier, Oxford, 2009. ISBN 978-0-444-52701-1. doi: https://doi.org/10.1016/B978-044452701-1.00125-3. URL `http://www.sciencedirect.com/science/article/pii/B9780444527011001253`.

[36] Muhamad Rashid Ahmed. An investigation of methods for missing data in hierarchical models for discrete data. 2011.

[37] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical*

Society: Series B (Methodological), 39(1):1–22, 1977.

[38] Chris Sherlaw-Johnson, Steve Gallivan, and Jim Burridge. Estimating a markov transition matrix from observational data. *Journal of the Operational Research Society*, 46(3):405–410, 1995.

[39] Bruce A Craig and Peter P Sendi. Estimation of the transition matrix of a discrete-time markov chain. *Health economics*, 11(1):33–42, 2002.

[40] Andrea B Troxel, David P Harrington, and Stuart R Lipsitz. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):425–438, 1998.

[41] Paul S Albert. A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, 56(2):602–608, 2000.

[42] Baojiang Chen, Grace Y Yi, and Richard J Cook. Analysis of interval-censored disease progression data via multi-state models under a nonignorable inspection process. *Statistics in Medicine*, 29(11):1175–1189, 2010.

[43] Baojiang Chen and Xiao-Hua Zhou. Non-homogeneous markov process models with informative observations with an application to alzheimer's disease. *Biometrical Journal*, 53(3):444–463, 2011.

[44] Ardo Van Den Hout and Fiona E Matthews. Estimating stroke-free and total life expectancy in the presence of non-ignorable missing values. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2):331–349, 2010.

[45] Hung-Wen Yeh, Wenyaw Chan, and Elaine Symanski. Intermittent missing observations in discrete-time hidden markov models. *Communications in Statistics-Simulation and Computation*, 41(2):167–181, 2012.

[46] Souad Assoudou, Belkheir Essebbar, et al. A bayesian model for binary markov chains. *International Journal of Mathematics and Mathematical Sciences*, 2004 (8):421–429, 2004.

[47] Alberto Pasanisi, Shuai Fu, and Nicolas Bousquet. Estimating discrete markov models from various incomplete data schemes. *Computational Statistics & Data Analysis*, 56(9):2609–2625, 2012.

[48] Orestis Efthimiou, Nicky Welton, Myrto Samara, Stefan Leucht, Georgia Salanti, and GetReal Work Package 4. A markov model for longitudinal studies with incomplete dichotomous outcomes. *Pharmaceutical statistics*, 16(2):122–132, 2017.

[49] Zoubin Ghahramani and Michael I Jordan. Learning from incomplete data. 1995.

[50] Isabelle Deltour, Sylvia Richardson, and Jean-Yves Le Hesran. Stochastic algorithms for markov models estimation with intermittent missing data.

Biometrics, 55(2):565–573, 1999.

[51] Junsheng Ma, Xiaoying Yu, Elaine Symanski, Rachelle Doody, and Wenyaw Chan. A bayesian approach in estimating transition probabilities of a discrete-time markov chain for ignorable intermittent missing data. *Communications in Statistics-Simulation and Computation*, 45(7):2598–2616, 2016.

[52] Shinichi Nakagawa and Robert P Freckleton. Missing inaction: the dangers of ignoring missing data. *Trends in ecology & evolution*, 23(11):592–596, 2008.

[53] Hyun Kang. The prevention and handling of the missing data. *Korean journal of anesthesiology*, 64(5):402, 2013.

[54] Carol M Musil, Camille B Warner, Piyanee Klainin Yobas, and Susan L Jones. A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, 24(7):815–829, 2002.

[55] Pedro J García-Laencina, Pedro Henriques Abreu, Miguel Henriques Abreu, and Noémia Afonoso. Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Computers in biology and medicine*, 59:125–133, 2015.

[56] Jeff A Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.

[57] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[58] VE Uvarov, AA Popov, and TA Gultyaeva. Imputation of incomplete motion data using hidden markov models. In *Journal of Physics: Conference Series*, volume 1210, page 012151. IOP Publishing, 2019.

[59] Shaunak Chatterjee and Stuart Russell. A temporally abstracted viterbi algorithm. *arXiv preprint arXiv:1202.3707*, 2012.

[60] Davide Vidotto, Jeroen K Vermunt, and Katrijn van Deun. Bayesian multilevel latent class models for the multiple imputation of nested categorical data. *Journal of Educational and Behavioral Statistics*, 43(5):511–539, 2018.

[61] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[62] Jeanne S Mandelblatt, Natasha K Stout, Clyde B Schechter, Jeroen J Van Den Broek, Diana L Miglioretti, Martin Krapcho, Amy Trentham-Dietz, Diego Munoz, Sandra J Lee, Donald A Berry, et al. Collaborative modeling of the benefits and harms associated with different us breast cancer screening strategiesbenefits and harms of us breast cancer screening strategies. *Annals of*

internal medicine, 164(4):215–225, 2016.

[63] American College of Radiology. ACR BI-RADS Atlas 5th Edition, 2013. `https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/Mammography-Reporting.pdf`, (Accessed 11-August-2020).

[64] R Keefer. Shedding light on breast density. *ACR Bull*, 13, 2012.

[65] Stephen W Duffy, Oliver WE Morrish, Prue C Allgood, Richard Black, Maureen GC Gillan, Paula Willsher, Julie Cooke, Karen A Duncan, Michael J Michell, Hilary M Dobson, et al. Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. *European Journal of Cancer*, 88:48–56, 2018.

[66] Jennifer A Harvey. Quantitative assessment of percent breast density: analog versus digital acquisition. *Technology in cancer research & treatment*, 3(6): 611–616, 2004.

[67] Mariëtte Lokate, Rebecca K Stellato, Wouter B Veldhuis, Petra HM Peeters, and Carla H van Gils. Age-related changes in mammographic density and breast cancer risk. *American journal of epidemiology*, 178(1):101–109, 2013.

[68] Celine Vachon, V Shane Pankratz, Christopher G Scott, Shaun D Maloney, Karthik Ghosh, Kathleen Brandt, Tia Milanese, Michael J Carston, and Thomas A Sellers. Longitudinal trends in mammographic percent density and breast cancer risk. *Cancer Epidemiology and Prevention Biomarkers*, 16(5): 921–928, 2007.

[69] Meghan E Work, Laura L Reimers, Anne S Quante, Katherine D Crew, Amy Whiffen, and Mary Beth Terry. Changes in mammographic density over time in breast cancer cases and women at high risk for breast cancer. *International journal of cancer*, 135(7):1740–1744, 2014.

[70] E.J. Aiello, D.S. Buist, E. White, and P.L. Porter. Association between mammographic breast density and breast cancer tumor characteristics. *Cancer Epidemiol.Biomarkers Prev.*, 14(1055-9965 (Print)):662–668, 2005.

[71] Charlotte C Gard, Erin J Aiello Bowles, Diana L Miglioretti, Stephen H Taplin, and Carolyn M Rutter. Misclassification of breast imaging reporting and data system (BI-RADS) mammographic density and implications for breast density reporting legislation. *The breast journal*, 21(5):481–489, 2015.

[72] Manisha Bahl, Jay A Baker, Mythreyi Bhargavan-Chatfield, Eugenia K Brandt, and Sujata V Ghate. Impact of breast density notification legislation on radiologists' practices of reporting breast density: a multi-state study. *Radiology*, 280(3):701–706, 2016.

[73] Carol H Lee, D David Dershaw, Daniel Kopans, Phil Evans, Barbara Monsees, Debra Monticciolo, R James Brenner, Lawrence Bassett, Wendie Berg, Stephen

Feig, et al. Breast cancer screening with imaging: recommendations from the society of breast imaging and the acr on the use of mammography, breast mri, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *Journal of the American college of radiology*, 7(1):18–27, 2010.

[74] American Cancer Society. ACS Breast Cancer Screening Guideline, 2016. `https://www.cancer.org/latest-news/special-coverage/american-cancer-society-breast-cancer-screening-guidelines.html` (Accessed 11-August-2020).

[75] U.S Preventive Services Task Force. Final Recommendation Statement, 2016. `https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/breast-cancer-screening1` (Accessed 1-August-2020).

[76] Timothy J Wilt, Russell P Harris, and Amir Qaseem. Screening for cancer: Advice for high-value care from the american college of physicianssscreening for cancer: Advice for high-value care from the acp. *Annals of internal medicine*, 162(10):718–725, 2015.

[77] Eike Nohdurft, Elisa Long, and Stefan Spinler. Was Angelina Jolie right? optimizing cancer prevention strategies among BRCA mutation carriers. *Decision Analysis*, 14(3):139–169, 2017.

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

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

[80] Mehmet Ulvi Saygi Ayvaci, Oguzhan Alagoz, Mehmet Eren Ahsen, and Elizabeth S Burnside. Preference-sensitive management of post-mammography decisions in breast cancer diagnosis. *Production and Operations Management*, 2017.

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

[82] Çağlar Çağlayan, Turgay Ayer, and Donatus U Ekwueme. Assessing multi-modality breast cancer screening strategies for brca 1/2 gene mutation carriers and other high-risk populations. *Available at SSRN 3139779*, 2018.

[83] Turgay Ayer, Oguzhan Alagoz, and Natasha K Stout. Or forum—a pomdp approach to personalize mammography screening decisions. *Operations Research*, 60(5):1019–1034, 2012.

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

[85] Burhaneddin Sandikci, Mucahit Cevik, and David Schacht. Screening for breast cancer: The role of supplemental tests and breast density information. *Chicago Booth Research Paper*, (18-03), 2018.

[86] Maarten Otten, Judith Timmer, and Annemieke Witteveen. Stratified breast cancer follow-up using a continuous state partially observable markov decision process. *European journal of operational research*, 281(2):464–474, 2020.

[87] Steve H Parker, Fred Burbank, Roger J Jackman, Charles J Aucreman, Gilda Cardenosa, Thomas M Cink, John L Coscia Jr, GW Eklund, WP Evans 3rd, and Paul R Garver. Percutaneous large-core breast biopsy: a multi-institutional study. *Radiology*, 193(2):359–364, 1994.

[88] Issa J Dahabreh, Lisa Susan Wieland, Gaelen P Adam, Christopher Halladay, Joseph Lau, and Thomas A Trikalinos. Core needle and open surgical biopsy for diagnosis of breast lesions. 2014.

[89] R Edward Hendrick. Obligate overdiagnosis due to mammographic screening: a direct estimate for us women. *Radiology*, 287(2):391–397, 2018.

[90] Mahboubeh Madadi, Mohammadhossein Heydari, Shengfan Zhang, Edward Pohl, Chase Rainwater, and Donna L Williams. Analyzing overdiagnosis risk in cancer screening: A case of screening mammography for breast cancer. *IISE Transactions on Healthcare Systems Engineering*, 8(1):2–20, 2018.

[91] Sevda. Molani. *Stochastic Decision Modeling to Improve Breast Cancer Preventive Care*. PhD thesis, Louisiana Tech University, 2020.

[92] Bjørn Helge Østerås, Anne Catrine T Martinsen, Siri Helene B Brandal, Khalida Nasreen Chaudhry, Ellen Eben, Unni Haakenaasen, Ragnhild Sørum Falk, and Per Skaane. Classification of fatty and dense breast parenchyma: comparison of automatic volumetric density measurement and radiologists' classification and their inter-observer variation. *Acta Radiologica*, 57(10):1178–1185, 2016.

[93] Natasha K. Stout, Sandra J. Lee, Clyde B. Schechter, Karla Kerlikowske, Oguzhan Alagoz, Donald Berry, Diana S M Buist, Mucahit Cevik, Gary Chisholm, Harry J. De Koning, Hui Huang, Rebecca A. Hubbard, Diana L. Miglioretti, Mark F. Munsell, Amy Trentham-Dietz, Nicolien T. Van Ravesteyn, Anna N A Tosteson, and Jeanne S. Mandelblatt. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography. *Journal of the National Cancer Institute*, 106(6), 2014. ISSN 14602105. doi: 10.1093/jnci/dju092.

[94] My von Euler-Chelpin, Martin Lillholm, Ilse Vejborg, Mads Nielsen, and Elsebeth Lynge. Sensitivity of screening mammography by density and texture: a cohort study from a population-based screening program in denmark. *Breast Cancer Research*, 21(1):111, 2019.

[95] Emine Devolli-Disha, Suzana Manxhuka-Kërliu, Halit Ymeri, and Arben Kutllovci. Comparative accuracy of mammography and ultrasound in women with breast symptoms according to age and breast density. *Bosnian journal of basic medical sciences*, 9(2):131, 2009.

[96] MARIBS Study Group et al. Screening with magnetic resonance imaging and mammography of a uk population at high familial risk of breast cancer: a prospective multicentre cohort study (maribs). *The Lancet*, 365(9473):1769–1778, 2005.

[97] Breast Cancer Surveillance Consortium (BCSC). BCSC BC Risk Calculator, 2016. `https://tools.bcsc-scc.org/BC5yearRisk/intro.htm`, (Accessed 12-January-2018).

[98] Natalie J Engmann, Christopher Scott, Matthew R Jensen, Stacey J Winham, Lin Ma, Kathleen R Brandt, Amir Mahmoudzadeh, Dana H Whaley, Carrie B Hruska, Fang-Fang Wu, et al. Longitudinal changes in volumetric breast density in healthy women across the menopausal transition. *Cancer Epidemiology and Prevention Biomarkers*, 28(8):1324–1330, 2019.

[99] The United states Census Bureau. US Population by Race, 2017. `https://www.census.gov/quickfacts/fact/table/US/RHI125217#viewtop`.

[100] American Cancer Society. Breast Cancer; How Common Is Breast Cancer? 2019. `https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html#written_by`, (Accessed 11-August-2020).

[101] National Cancer Institute. Breast Cancer Risk in American Women, 2012. `https://www.cancer.gov/types/breast/risk-fact-sheet` (Accessed 15-April-2018).