



Published in final edited form as:

Oper Res. 2010 November 1; 58(6): 1577–1591. doi:10.1287/opre.1100.0877.

Optimal Breast Biopsy Decision-Making Based on Mammographic Features and Demographic Factors

Jagpreet Chhatwal,

Health Economic Statistics, Merck Research Laboratories, North Wales, Pennsylvania 19454, jagpreet_chhatwal@merck.com

Oguzhan Alagoz, and

Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706, alagoz@engr.wisc.edu

Elizabeth S. Burnside

Department of Radiology, University of Wisconsin–Madison, Madison, Wisconsin 53792, eburnside@uwhealth.org

Abstract

Breast cancer is the most common non-skin cancer affecting women in the United States, where every year more than 20 million mammograms are performed. Breast biopsy is commonly performed on the suspicious findings on mammograms to confirm the presence of cancer. Currently, 700,000 biopsies are performed annually in the U.S.; 55%–85% of these biopsies ultimately are found to be benign breast lesions, resulting in unnecessary treatments, patient anxiety, and expenditures. This paper addresses the decision problem faced by radiologists: When should a woman be sent for biopsy based on her mammographic features and demographic factors? This problem is formulated as a finite-horizon discrete-time Markov decision process. The optimal policy of our model shows that the decision to biopsy should take the age of patient into account; particularly, an older patient's risk threshold for biopsy should be higher than that of a younger patient. When applied to the clinical data, our model outperforms radiologists in the biopsy decision-making problem. This study also derives structural properties of the model, including sufficiency conditions that ensure the existence of a control-limit type policy and nondecreasing control-limits with age.

Subject classifications

Markov decision processes; dynamic programming; control-limit policy; service operations; breast cancer; mammography; breast biopsy; medical decision-making

1. Introduction

Breast cancer is the most common non-skin cancer affecting women in the United States. According to the American Cancer Society (ACS), in 2010, an estimated 207,090 women would be diagnosed with invasive breast cancer, and more than 40,000 die from this disease (American Cancer Society 2010). In recent years, various breast cancer treatment options (combination of surgery, radiation therapy, chemotherapy, and hormone therapy) have become available; nonetheless, successful treatment depends highly on early diagnosis (Fryback et al. 2006).

Screening mammography is the current standard practice for identifying cancer early in asymptomatic women. Randomized clinical trials have shown that the use of screening

mammography in the general population reduces breast cancer mortality by at least 24 percent (Smith et al. 2003, Nyström et al. 2002). Although the cost-effectiveness of screening mammography is well established for women between 50–69 years of age, there has been a debate for younger age groups (Kerlikowske et al. 1995, Fracheboud et al. 2005). Despite the controversy, the ACS recommends an annual mammogram for all women over 40 years of age (Smith et al. 2006). It is estimated that more than 20 million mammograms are performed annually in the United States, and approximately 70% of women over the age of 40 have had a mammogram in the last two years (Freid et al. 2003).

Mammogram reading involves two component—perception and interpretation. Radiologists look for abnormalities (such as microcalcifications or masses), which sometimes can be hard to detect. After identifying abnormalities, they determine the risk of cancer to decide on appropriate management. Mammogram interpretation requires radiologists to evaluate mammographic findings based on their previously acquired knowledge from training or experience.

Based on the mammographic findings, critical decisions must be made to detect cancer early while sparing healthy women unnecessary procedures. If a mammogram looks suspicious, then a biopsy (examination of the breast tissue removed using a needle or surgical excision) is required to decide whether an abnormality is in fact a breast cancer. In their seminal work, Tversky and Kahneman (2004) show that physicians often misestimate disease risk before and after diagnostic tests such as mammograms. All humans use heuristics that can severely bias decisions; however, in health care these systematic errors lead to incorrect judgments that are highly consequential.

There are two ways in which a mammogram interpretation can be wrong. First, a mammogram can be interpreted as normal when, in fact, a cancer is present. As a result, there is a delay in the diagnosis of cancer and threat to the patient's life. The false-negative rate of mammography lies in the range of 10%–25% (Destounis et al. 2004). While mammographically occult cancers represent a portion of the false-negative rate, cancers that are apparent in retrospect represent the larger portion (Baines and Dayan 1999, Bird et al. 1992). Second, a mammogram can be labeled positive when a finding, in fact, is noncancerous. This results in over-treatment and unnecessary anxiety to the patients. The false-positive rate of mammography lies in the range of 55%–85% (Elvecrog 1993, Meyer et al. 1990, Opie et al. 1993, Parker 1991). Both the false-negative and the false-positive rates reflect the complexity of mammogram interpretation.

The accuracy of mammogram interpretation varies with the radiologists' skills and training (Barlow et al. 2004, Beam et al. 1996). Subspecialty radiologists, who often have fellowship training in mammography and read these studies exclusively, perform generally better than the community radiologists, who read the majority of all mammograms in the context of a diverse general practice (Sickles et al. 2002). Furthermore, the United States as a whole appears to have different performance rates than other countries. Smith-Bindman et al. (2003) report that although cancer detection rates are identical in the U.S. and in the United Kingdom, radiologists in the United States declared many more mammogram results uncertain or suspicious compared with their British counterparts; as a result, American women with and without cancer underwent at least twice as many follow-up tests, such as biopsies.

To standardize mammography practice, the American College of Radiology (ACR) has developed a lexicon—the Breast Imaging Reporting and Data System (BI-RADS) (BI-RADS 1998, Liberman and Menell 2002). The BI-RADS lexicon, which includes descriptors that are the best predictors for a benign or malignant diagnosis, is intended to

guide radiologists and referring physicians in the breast cancer decision-making process and facilitate the management of patients. Each mammogram is described using standardized descriptors and then classified into one of the six final assessment categories (BI-RADS assessment codes), depending upon the interpretation of the observations made on the mammogram. Table 1 shows the relationship between the assessment codes and the recommended actions. BI-RADS 0 is associated with acquiring further imaging; BI-RADS 1 and 2 are recommended if there is either no or benign observation on the mammogram, suggesting no action and continue routine yearly screening; BI-RADS 3 is assigned to a mammogram if a radiologist observes something probably benign that does not warrant immediate biopsy but needs surveillance for a short interval (usually six months to two years); and BI-RADS 4 and 5 are assigned to suspicious and high-risk findings, recommending immediate biopsy to the patient.

In spite of the standardization of mammography reporting, several issues have not been addressed by the current usage of the BI-RADS lexicon. BI-RADS categories do not take age into consideration when making biopsy decisions, despite the fact that older age groups have unique characteristics that deserve particular attention. For instance, breast cancers tend to be less aggressive in older women, which might suggest that a higher probability threshold for action could be more appropriate (Fowble et al. 1994, Jayasinghe et al. 2005). The false-positive interpretation of mammography resulting in unnecessary invasive procedures might be more difficult or risky for older individuals with comorbidities. In general, these distinctive features of breast cancer in older women indicate that the probability thresholds for biopsy in older women might be different than those in younger women.

Computer-aided diagnostic (CADx) models have a potential to reduce inter-observer variability and improve decision-making for early diagnosis of breast cancer (Jiang et al. 2001). The output of these models is usually the probability of cancer, which is more informative than the six discrete BI-RADS categories. Several CADx models have been developed using various statistical and artificial intelligence techniques. Burnside et al. (2006, 2009) constructed a Bayesian network to predict risk of malignancy using mammographic features and demographic factors. Baker et al. (1995) and Ayer et al. (2010) used artificial neural networks, Jesneck et al. (2006) used decision fusion, and Chhatwal et al. (2009) used logistic regression for breast cancer risk prediction using mammographic features. These studies suggest that more accurate decisions could be made using the probability of cancer as an outcome measure. However, CADx models only partly address the problem because they do not provide an optimal threshold for the decision to biopsy.

No methods exist that enable radiologists to determine an optimal threshold over which to recommend biopsy for a given patient. Around 700,000 breast biopsies are performed annually in the United States. Due to the inherent limitations of mammography, 55%–85% of women who undergo biopsy turn out to have benign breast lesions. As a result, an estimated \$250 million is spent every year on the false-positive biopsies. In addition, a false-positive mammogram exposes the patient to unnecessary anxiety, pain, and possible complications (Steggles et al. 1998). A study by Maxwell et al. (2000) indicated that the stress level during the period of breast biopsy is substantial. Furthermore, biopsy can also introduce changes (such as distortion) on future mammograms, which could make later diagnosis more difficult.

This paper addresses the optimal breast biopsy decision for an individual patient. The decision of whether to perform a biopsy is a function of the patient's probability of cancer, which is estimated based on her mammographic findings and demographic factors. We also investigate how the optimal decision of whether to biopsy changes with the patient's age.

Past studies have considered the optimal interval of screening mammography (Maillart et al. 2008, Michaelson et al. 1999) and an optimal model for breast cancer screening and treatment (Ivy 2006). To the best of our knowledge, our study is the first one to consider the optimal decision-making for breast biopsy. In addition, we provide clinically intuitive conditions for the existence of structured optimal policies in mathematical models that have a similar structure to our problem (such as any diagnostic decision-making problem, e.g., lung biopsy, prostate biopsy, etc.). We solve an important diagnostic decision problem using clinical data and find policies that were not recognized by the decision-makers before.

Our model provides patient-specific recommendations for breast biopsy, in contrast to population-based guidelines observed in the current clinical practice. A biopsy decision based on individual characteristics can also help promote the concept of personalized care in the diagnosis of breast cancer, which is considered the future of health care (Williams et al. 2003). The biopsy decision in our model is based on the probability of cancer instead of the six BI-RADS codes. This probability gives the patient an opportunity to better understand her risk of cancer and engage in shared decision-making with the radiologist, which is of increasing interest in the radiology community with regard to screening tests (Chan 2005, Hillman 2005).

The rest of the paper is organized as follows. In §2, we describe the model formulation of the optimal biopsy decision problem, followed by its structural properties in §3. In §4, we present patient-specific optimal policies and perform sensitivity analysis on various model parameters. We conclude in §5 with the discussion of our results and future directions. An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

2. Model Formulation

The optimal biopsy decision problem is formulated as a finite-horizon discrete-time Markov decision process (MDP) (Puterman 1994), which is referred to as the optimal biopsy decision model (OBDM). We first describe various assumptions made throughout the model. At every decision-epoch, a woman undergoes a mammogram that is examined by a radiologist (the decision-maker), who has only two decision options: biopsy, or wait until the next annual mammogram. The decision to biopsy also includes follow-up procedures. This decision is made based on the woman's current risk of breast cancer, which could be estimated by a radiologist or a computerized risk prediction model such as the one described in §4.1. Once the biopsy is performed, the patient is assumed to leave the system, i.e., the decision process ends. We assume that patients adhere to the decisions made by the radiologists, i.e., they will get the next annual mammogram (or biopsy) with certainty if the action taken in the current decision epoch is to wait until the next annual mammogram (or biopsy). We assume a finite set of states that completely describe the state space. We also assume that both the patient and the radiologist are risk neutral. Throughout the OBDM, we refer to the woman as the “patient,” irrespective of her health condition, for the sake of consistency. Next, we describe the notation used to build our model.

- Decision epochs, $t = 0, 1, 2, \dots, T, T < \infty$. We define t as the number of years above the age of 40. The decision epochs start at age 40 because the ACS recommends regular mammography screening annually for all women over 40 years of age (Smith et al. 2006). We end our decision horizon at age 100 (i.e., $t = 60$), which is consistent with the U.S. life tables from the National Center for Health Statistics (NCHS) of Center for Disease Control and Prevention (CDC).
- s_t : State of the system at time t such that $s_t \in \{0, 1, \dots, S, S + 1, S + 2\}$, where $s_t \in \{0, 1, \dots, S\}$ represents the risk score, $\{S + 1\}$ represents the post-biopsy state, and

$\{S + 2\}$ represents death. The risk score (the current probability of cancer) can be estimated by either a CADx model (Ayer et al. 2010, Burnside et al. 2009, Chhatwal et al. 2009) or asking a radiologist. We use a mammography Bayesian network (MBN) described in §4.1 to estimate the current probability of breast cancer based on mammographic features and demographic factors. The risk score is obtained by discretizing the probability of breast cancer (outcome of the MBN) to a number between 0 and 100. For example, if the probability of breast cancer given by the MBN is 0.1538, then the corresponding risk score is 15.

- \mathbb{S} State space, $\mathbb{S} = \{0, 1, 2, \dots, S, S + 1, S + 2\}$.
- $w_t(s_t)$: Probability of cancer at time t when the current state is s_t . In our model, we assume $w_t(s_t) = s_t/100$.
- $\mathbb{A}(t)$: Action space at time t when the patient is in state s ; $\mathbb{A}(t) = \{\text{Biopsy (Bx)}, \text{annual mammogram (Am)}\}$.
- $p_t^{Am}(s' | s_t)$: Probability that the patient will be in state $s' \in \mathbb{S}$ at time $t + 1$, given that she is in state $s_t \in \mathbb{S}$ at time t , and the action is “annual mammogram.”

For $t = 1, 2, \dots, T$, we define $p_t^{Am}(S+1|S+1) = 1$, i.e., post-biopsy state is an absorbing state; $p_t^{Am}(S+1|s_t) = 0$ for $s_t \in \mathbb{S} + 1$, i.e., the patient cannot move to post-biopsy state in the next decision epoch when annual mammogram is recommended; $p_t^{Am}(S+2|S+2) = 1$, i.e., “death” is an absorbing state.

- $p_t^{ben}(S+2)$: Probability of death during the decision epoch t when the patient is cancer-free.
- $p_t^{mal}(S+2)$: Probability of death during the decision epoch t when the patient is diagnosed with breast cancer and has started receiving treatment.
- $p_t^{mal2}(S+2)$: Probability of death during the decision epoch t when the patient has breast cancer but has not started receiving any treatment. This is usually the case when the patient is unaware of her disease. We use $p_t^{mal2}(S+2)$ to estimate the transition probability of moving to the death state given that the patient's current risk of cancer is s_t , i.e., $p_t^{Am}(S+2|s_t)$. However, $p_t^{mal2}(S+2)$ cannot be observed in practice, hence, we use $p_t^{mal}(S+2)$ to estimate $p_t^{mal2}(S+2)$ as defined in §4.1.
- $p_t^{Am}(S+2|s_t)$: Probability of death before the decision epoch $t + 1$, given that the patient's current risk score is s_t at time t , and the action is “annual mammogram.”

$$p_t^{Am}(S+2|s_t) = w_t(s_t) * (p_t^{mal2}(S+2)) + (1 - w_t(s_t)) * (p_t^{ben}(S+2)). \quad (1)$$

- $p_t^{Bx}(s' | s_t)$ = Probability that the patient will be in state $s' \in \mathbb{S}$ at time $t + 1$, given that she is in state $s_t \in \mathbb{S}$ at time t , and the action is “biopsy.”

We define $p_t^{Bx}(s' | s_t) = 0$ for all $t, s_t \in \mathbb{S}, s' \in \mathbb{S}\{S + 1\}$; $p_t^{Bx}(S+1|s_t) = 1$ for all t and $s_t \in \mathbb{S}\{S + 1, S + 2\}$, $p_t^{Bx}(S+1|S+1) = 1$, and $p_t^{Bx}(S+2|S+2) = 1$ for all t .

- \mathbb{P}_t^{Am} : Transition probability matrix at time t when the action taken is “annual mammogram,” i.e., $\mathbb{P}_t^{Am} = [p_t^{Am}(s' | s_t)]$.

- \mathbb{P}_t^{Bx} : Transition probability matrix at time t when the action taken is “biopsy” i.e.,

$$\mathbb{P}_t^{Bx} = \left[p_t^{Bx}(s' | s_t) \right]$$
- $a_t(s_t)$: Action taken at time t in state $s_t \in \mathbb{S}\{S+1, S+2\}$, where $a_t(s_t) \in \{Am, Bx\}$. If the action taken is Bx , then the patient quits the process; otherwise, she waits for one more decision epoch until the next mammogram.
- $r_t(s_t, Am)$: Total expected intermediate reward accrued at time t , when the patient's state is s_t , and “annual mammogram” is chosen. Examples of reward include expected life in years, or expected quality-adjusted life years (QALYs). QALYs measure both the quality and the quantity of life years by incorporating risk-neutral utilities of health states in the expected life years (Pliskin et al. 1980), and are commonly used in medical decision making (Drummond 2005). In our model, we define reward as QALYs, which are assigned the intermediate reward of one year if the patient is alive in the current decision epoch, and one-half year if the patient dies in the current decision-epoch to account for the half-cycle correction (Sonnenberg and Beck 1993), i.e., to account for the fact that the patient could die in the first half or the second half of the decision epoch. Thus, the expected life in that decision epoch is given by

$$\begin{aligned} r_t(s_t, Am) &= 1 * P(\text{alive in the current decision epoch}) + 1/2 * P(\text{death in the current decision epoch}), \\ r_t(s_t, Am) &= 1 * (w_t(s_t) * (1 - P_t^{mal2}(S+2)) + (1 - w_t(s_t)) * (1 - P_t^{ben}(S+2))) + 1/2 * (w_t(s_t) * (1 - P_t^{mal2}(S+2)) + (1 - w_t(s_t)) * (P_t^{ben} \end{aligned} \tag{2}$$

We define $r_t(s_t, Am) = 0$ for $s_t \in \{S+1, S+2\}$.

- $r_t(s_t, Bx, -)$: Total expected discounted post-biopsy reward at time t when the patient's state is s_t and the biopsy outcome is negative (benign). In other words, $r_t(s_t, Bx, -)$ represents the life expectancy of a cancer-free woman.
- $r_t(s_t, Bx, +)$: Total expected discounted post-biopsy reward at time t when the patient's state is s_t and the biopsy outcome is positive (malignant).
- $q_t(- | s_t)$: Probability that the outcome of the biopsy is negative (benign) when the patient was in state s_t at time t .
- $(q_t(+ | s_t)$: Probability that the outcome of biopsy is positive (malignant) when the patient was in state s_t at time t .
- $dBx(t)$: Disutility of biopsy at time t . The disutility of biopsy exists because (i) patients associate biopsy decision with high chance of breast cancer, resulting in emotional distress; and (ii) potential surgical complications.
- $r'_t(s_t, Bx)$: Total expected discounted post-biopsy reward when the patient's state is s_t at time t . We assume that the patient receives a one-time lump-sum reward when she undergoes biopsy at time t and quits the process (i.e., she receives zero reward in the future decision epochs).

$$r'_t(s_t, Bx) = q_t(- | s_t) * r_t(s_t, Bx, -) + q_t(+ | s_t) * r_t(s_t, Bx, +). \tag{3}$$

- $r_t(s_t, Bx)$: Total expected net quality-adjusted post-biopsy life when the patient's state is s_t at time t .

$$r_t(s_t, Bx) = r'_t(s_t, Bx) - dBx(t). \quad (4)$$

We define $r_t(s_t, Bx) = 0$ for $s_t \in \{S+1, S+2\}$.

- λ : Annual discount factor ($0 \leq \lambda \leq 1$).
- $\vartheta_t(s_t)$: Maximum total expected quality-adjusted life that the patient can attain when her state is s_t at time t .

The optimal solution can be obtained by solving the following set of equations:

$$\vartheta_t(s_t) = \max \left\{ r_t(s_t, Bx), r_t(s_t, Am) + \lambda \sum_{s' \in \mathcal{S}} p_t^{Am}(s' | s_t) \vartheta_{t+1}(s') \right\}, t=1, 2, \dots, T-1. \quad (5)$$

For $t = T$, we add a boundary condition as follows:

$$\begin{aligned} r_T(s_T, Bx, S+1) &= r_T(s_T, Am) \\ &= V_T(s_T), \quad \text{for all } s_T \in \mathcal{S}. \end{aligned} \quad (6)$$

Note that at decision epoch T (age 100 in our model), patients are not assumed to die but instead are assigned a terminal reward, representing the conditional life expectancy of a woman at the end of decision horizon.

- $\psi_t(s_t, Am)$: Maximum quality-adjusted expected life when the patient's state is s_t at time t and the action is "annual mammogram."

$$v_t(s_t, Am) = r_t(s_t, Am) + \lambda \sum_{s' \in \mathcal{S}} p_t^{Am}(s' | s_t) \vartheta_{t+1}(s').$$

Figure 1 shows the state-transition diagram of the OBDM. A patient gets a mammogram every year. Based on her mammographic features and demographic factors, she is assigned a risk score (probability of cancer), which is estimated by a CADx model (such as the one in §4.1) or a radiologist. At each decision epoch, the radiologist has two options for each risk score: immediate biopsy (represented by Bx), or ask the patient to wait until the next annual mammogram (represented by Am). When the radiologist recommends the patient get a biopsy, she moves to the post-biopsy state, which is an absorbing state (i.e., the patient never leaves once she enters it). As a result, the patient receives a one-time lump-sum reward of $r_t(s_t, Bx)$ that represents the expected quality-adjusted post-biopsy life years if her current risk score is s_t at the time of biopsy. If the radiologist recommends the patient wait until the next decision epoch, then the patient's risk score next year changes according to the state transition probability matrix \mathbb{P}_t^{Am} . As a result, she receives a reward of $r_t(s_t, Am)$ that corresponds to her intermediate expected QALYs before the next decision epoch. All transitions occur randomly once the radiologist makes a decision. We assume that biopsy is perfect. Medical literature shows that the false-negative rate of biopsy is less than 3% (Lee et al. 1999, Liberman 2000). Most of the computer-aided diagnosis models in the medical literature also assume that biopsy is perfect and treat biopsy outcome as a gold standard (Baker et al. 1995, Jiang et al. 1999).

We formulate the biopsy decision problem as a finite-horizon MDP for the following two reasons. First, our rewards and state transition probabilities should depend on patient age to reflect clinical observations; i.e., as a woman gets older, her probability of death increases, her expected life reduces, and her natural history of breast cancer might change. Second, our policies should take age into consideration to answer the fundamental question in our problem: should the decision to biopsy depend on the patient's age? An alternate modeling approach would be to formulate this problem as an infinite-horizon MDP. To capture patient's age in the model, we would need to incorporate age into the state space, which would make the model numerically intractable and could be solved using approximate dynamic programming.

We also formulate this biopsy decision problem as a partially observable MDP (POMDP), where the decisionmaker has only partial information about the patient's true health state from a mammography observation. Appendix B in the electronic companion provides the model formulation of our POMDP model, compares it to our MDP model (OBDM), and describes why an MDP model is more suitable than a POMDP model for the biopsy decision-making problem.

3. Structural Properties

In this section, we investigate the structural properties of the OBDM with their clinical relevance. These properties provide insights to the decision-maker (radiologist, referring physician, and/or patient) on how various optimal decisions are made. We first make some assumptions that are used throughout this section.

ASSUMPTION 1. (As1) The function $r_t(s_t, Am)$ is non-increasing in s_t for all t , and in t . This implies that the patient's one year QALY does not increase with her risk of cancer or age.

ASSUMPTION 2. (As2) The function $r_t(s_t, Bx, -)$ is non-increasing in s_t for all t , and in t . This implies that the patient's expected post-biopsy QALYs, when the outcome of biopsy is benign, do not increase with her risk of cancer or age.

ASSUMPTION 3. (As3) The function $r_t(s_t, Bx, +)$ is non-increasing in s_t for all t , and in t . This implies that the patient's expected post-biopsy QALYs, when the outcome of biopsy is malignant, do not increase with her risk of cancer or age.

ASSUMPTION 4. (As4) The expected post-biopsy QALYs after a benign biopsy are never lower than that after a malignant biopsy, i.e., $r_t(s_t, Bx, -) \geq r_t(s_t, Bx, +)$.

ASSUMPTION 5. (As5) \mathbb{P}_t^{Am} satisfies the following:

$$\sum_{s'=j}^{S+2} p_t^{Am}(s'|i) \leq \sum_{s'=j}^{S+2} p_{t+1}^{Am}(s'|i)$$

for $i, j \in \mathbb{S}\{S+1\}$. It can be viewed as: the older the patient, the more probable that she will move to high-risk states, including death.

The above condition includes probability of death from cancer as well as other causes. Although the risk of cancer can decrease with a patient's age, this decrease might be outweighed by the increase in her probability of death from co-morbidities as she gets older (Schairer et al. 2004).

Definition 1 (Barlow and Proschan 1965). A Markov chain is said to be IFR (increasing failure rate) if its rows are in the increasing stochastic order; i.e.,

$$q(i) = \sum_{j=k}^{S+2} P(j|i)$$

is nondecreasing in i for all $k = 0, 1, \dots, S + 2$.

The IFR definition implies that as the patient's risk of cancer increases, her risk of further deterioration also increases. This definition is equivalent to the well-known notion of first-order stochastic dominance.

We first show the monotonicity of $\vartheta_t(s_t)$ in s_t and t . Proposition 1 provides a sufficiency condition under which $\vartheta_t(s_t)$ is nonincreasing in s_t , i.e., the patient's total expected quality-adjusted life years never increase with her risk score (probability of cancer). The proof of Proposition 1 as well as the proofs of all other results in this section are presented in Appendix C in the electronic companion.

PROPOSITION 1. If \mathbb{P}_t^{Am} is IFR for all $t = 1, 2, \dots, T$, then $\vartheta_t(s)$ is nonincreasing in s , for $s = 1, \dots, S$, and $t = 0, 1, 2, \dots, T - 1$.

In Proposition 2, we show that $\vartheta_t(s_t)$ is nonincreasing in t , i.e., the patient's total expected QALYs never increase with her age. We first present Lemma 1, which is used to prove Proposition 2.

LEMMA 1. If (As5) holds for $t = 1, 2, \dots, T - 1$, then for any $f(i)$ that is nonincreasing in i , the following holds:

$$\sum_{s' \in \mathcal{S}} p_t(s'|i) f(s') \geq \sum_{s' \in \mathcal{S}} p_{t+1}(s'|i) f(s').$$

PROPOSITION 2. The optimal value function, $\vartheta_t(s_t)$, is non-increasing in t for all $s_t \in \mathcal{S}$ i.e.,

$$\vartheta_t(s) \geq \vartheta_{t+1}(s).$$

In Theorem 1, we provide sufficiency conditions under which there exists an optimal control-limit type policy such that it is always optimal to biopsy if the patient's risk exceeds a risk score threshold, and wait until the next annual mammogram otherwise. We use the following lemma to prove Theorem 1.

LEMMA 2. Let $\mathbb{P} = [p_t(j|i)]$ for $i, j = 1, 2, \dots, N$ be an IFR transition probability matrix such that $\sum_{k=i+1}^{k^*} p_t(k|i) \geq \sum_{k=i+1}^{k^*} p_t(k|i+1)$ for $i < k^* \leq N$ and $t = 1, 2, \dots, T - 1$. If $f(i)$ is a nonincreasing function in i , then the following holds:

$$\begin{aligned} (a) \quad & \sum_{k=1}^i \{p_t(k|i) - p_t(k|i+1)\} f(k) \geq \sum_{k=1}^i \{p_t(k|i) - p_t(k|i+1)\} f(i), \\ (b) \quad & \sum_{k=i+1}^{k^*} \{p_t(k|i) - p_t(k|i+1)\} f(k) \geq \sum_{k=i+1}^{k^*} \{p_t(k|i) - p_t(k|i+1)\} f(i+1). \end{aligned}$$

Theorem 1 shows the existence of an optimal control-limit policy under Conditions (7) and (8). Inequality (7) implies that as the risk score increases, the percentage reduction in the

post-biopsy reward is less than the increase in the risk of waiting until the next decision epoch. Inequality (8) implies that the higher the risk score of the patient, the more likely she will move to a higher score in the future. Note that this condition is similar to the IFR condition. Figure 2 shows an example of the optimal policy in which there exists a control-limit type policy for all decision epochs.

THEOREM 1. If \mathbb{P}_t^{Am} is IFR, and \mathbb{P}_t^{Am} and $r_t(s_t, Bx)$ satisfy the following conditions:

$$\frac{r_t(s_t, Bx) - r_t(s_t+1, Bx)}{\lambda r_{t+1}(s_t+1, Bx)} \leq p_t^{Am}(S+2|s_t+1) - p_t^{Am}(S+2|s_t), \quad (7)$$

$$\sum_{s''=s_t+1}^S p_t^{Am}(s''|s_t+1) \geq \sum_{s''=s_t+1}^S p_t^{Am}(s''|s_t), \quad (8)$$

for all $s_t \in \mathbb{S}\{S+1\}$ and $t = 1, 2, \dots, T-1$, then there exists an optimal control-limit policy i.e., there exists $s_t^* \in \mathbb{S}\{S+1\}$ for $t = 0, 1, \dots, T-1$ such that

$$a^*(s_t) = \begin{cases} Bx & \text{if } s_t \geq s_t^*, \\ Am & \text{if } s_t < s_t^*. \end{cases}$$

DEFINITION 2 (FAN 1967). A function $f(x, y)$ is said to be superadditive if for $x_1 \leq x_2$ and $y_1 \leq y_2$,

$$f(x_1, y_1) + f(x_2, y_2) \geq f(x_1, y_2) + f(x_2, y_1);$$

and if the reverse inequality holds, then $f(x, y)$ is called subadditive.

Theorem 2 provides conditions under which the optimal value function is superadditive, i.e., the reduction in the QALYs of a patient as her current risk of cancer increases, reduces with her age. The clinical significance of this theorem is that it shows that the patient's expected life reduction with increase in her risk of cancer reduces with her age. We use Lemma 3 to prove Theorem 2. Inequality (10) in the following lemma implies that the expected benefit of biopsy in any state over waiting and then biopsy in the same state is nondecreasing with the patient's age. This means that as the patient gets older, her benefit of biopsy decreases. The intuitive explanation of Condition 4 in Theorem 2 is that the difference in the probability of moving to a higher risk score between two consecutive years decreases with the risk score.

LEMMA 3. Let \mathbb{P}_t^{Am} is IFR for all t , $\sum_{s'=j}^{S+2} p_{t+1}^{Am}(s'|i) \leq \sum_{s'=j}^{S+2} p_t^{Am}(s'|i)$, and \mathbb{P}_t^{Am} and $r_t(s_t, Bx)$ satisfy the following:

$$r_t(s, Am) - r_{t+1}(s, Am) \geq r_t(s+1, Am) - r_{t+1}(s+1, Am), \quad (9)$$

and

$$r_t(s, Bx) - \left\{ r_t(s, Am) + \sum_{s' \in \mathbb{S}} p_t^{Am}(s'|s) r_{t+1}(s', Bx) \right\} \leq r_{t+1}(s, Bx) - \left\{ r_{t+1}(s, Am) + \lambda \sum_{s' \in \mathbb{S}} p_t^{Am}(s'|s) r_{t+2}(s', Bx) \right\}, \quad (10)$$

then $\vartheta_t(s) - \vartheta_{t+1}(s) \geq r_t(s, Bx) - r_{t+1}(s, Bx)$ for all t , and $s \in \mathbb{S}\{S + 1\}$.

THEOREM 2. *If the following conditions are satisfied:*

1. The conditions of Theorem 1 hold such that there exists a control-limit type optimal policy for all t ,
2. $r_t(s, Am)$ is superadditive in s and t ,
3. $r_t(s, Bx)$ is superadditive in s and t ,
4. $\sum_{s'=k}^{S+2} p_t^{Am}(s'|s)$ is subadditive in s and t ,
5. Condition (10) holds then $\varphi_t(s)$ is superadditive in s and t , i.e.,

$$\vartheta_t(s) - \vartheta_t(s+1) \geq \vartheta_{t+1}(s) - \vartheta_{t+1}(s+1) \tag{11}$$

for all $s \in \mathbb{S}\{S + 1\}$ and $t \in 0, 1, \dots, T - 1$.

Theorem 3 provides a set of sufficiency conditions that ensure that the optimal control limit does not decrease with time. We first present Lemma 4 that provides an upper bound, $r_t^*(0)$ on $v_t(0, Am)$, where $r_t^*(0)$ is defined as the total expected life of a patient if her risk of cancer remains zero for the rest of her life and she never undergoes a biopsy. Inequality (12) implies that the expected post-biopsy reward in any state is greater than the upper bound on the total reward obtained by waiting for one more year and using the same risk score as the biopsy threshold. The clinical explanation of this condition is that the benefit of delaying biopsy decreases with time due to limited potential benefits obtained by biopsying. Figure 2 shows an example of an optimal control-limit policy in which the control limit does not decrease with time.

LEMMA 4. *Let \mathbb{P}_t^{Am} satisfies IFR assumption,*

$$r_t^*(0) = r_t(0, Am) + \sum_{i=t}^{T-1} \lambda^{i+1-t} \left(\prod_{j=t}^i (1 - p_j^{Am}(S+2|0)) \right) \cdot r_{i+1}(0, Am) \text{ and } r_t^*(0) \geq r_t(0, Bx)$$

then $\vartheta_t(0) \leq r_t^*(0)$ for all $t = 1, 2, \dots, T - 1$.

THEOREM 3. *A patient with a transition probability matrix, \mathbb{P}_t^{Am} satisfying IFR assumption, and having an optimal control-limit threshold for all ages, s_t^* , has a non-decreasing s_t^* in t if*

$$r_t(s, Bx) \geq r_t(s, Am) + \lambda \sum_{s' < s} p_t^{Am}(s'|s) r_{t+1}^*(0) + \lambda \sum_{s' \geq s} p_t^{Am}(s'|s) r_{t+1}(s', Bx) \tag{12}$$

for all $s \in \mathbb{S}$

COROLLARY 1. *If the patient's risk of breast cancer is non-decreasing in time (this is the case when the cancerous tumor is not detected or treated), i.e., $p_t^{Am}(s'|s_t) = 0$ for all $s' < s_t$, then the optimal control-limit threshold s_t^* is nondecreasing in t if*

$$r_t(s, Bx) \geq r_t(s, Am) + \lambda \sum_{s' \geq s^*} p_t^{Am}(s'|s) r_{t+1}(s', Bx) \tag{13}$$

for all $s \in \mathbb{S}$

DEFINITION 3 (ALAGOZ ET AL. 2004). Let $\mathbb{P}_1 = [p_1(j|i)], i, j = 1, 2, \dots, n$ and $\mathbb{P}_2 = [p_2(j|i)], i, k = 1, 2, \dots, n$ be any two transition probability matrices. We say that \mathbb{P}_1 dominates \mathbb{P}_2 i.e., $\mathbb{P}_1 \geq \mathbb{P}_2$, if $\sum_{j \leq k} p_1(j|i) \geq \sum_{j \leq k} p_2(j|i)$, for all $i, k = 1, 2, \dots, n$.

The clinical significance of Definition 3 is that if two patients \mathbb{W}_1 and \mathbb{W}_2 satisfy transition probability matrices \mathbb{P}_1 and \mathbb{P}_2 , respectively, as defined above, then patient \mathbb{W}_2 has a more aggressive growth of cancer, i.e., her risk of cancer increases faster than that of \mathbb{W}_1 .

Proposition 3 shows that a patient having more aggressive growth of cancer has a lower total expected QALYs and is more likely to get biopsy, compared to the patient having less aggressive cancer.

PROPOSITION 3. Let \mathbb{W}_1 and \mathbb{W}_2 be two patients for whom the optimal policy is of control-limit type with control limits κ_t^1 and κ_t^2 at decision epoch t , respectively, and transition probability matrices associated with “annual mammogram” \mathbb{P}_t^1 and \mathbb{P}_t^2 , respectively. If \mathbb{W}_1 and \mathbb{W}_2 have the same reward functions, $r_t(s, Am)$ and $r_t(s, Bx)$, and $\mathbb{P}_t^1 \geq \mathbb{P}_t^2$ for all $t = 1, 2, \dots, T$, then

$$\vartheta_t^1(s) \geq \vartheta_t^2(s) \quad \text{for all } s \in \mathbb{S} \text{ and } t=1, 2, \dots, T, \quad (14)$$

$$\kappa_t^1 \geq \kappa_t^2 \quad \text{for all } t=1, 2, \dots, T. \quad (15)$$

Proposition 4 shows that the decision to biopsy will be preferred more often if a biopsy technique having lower disutility (but equally effective) becomes available. Figures 4 and 5 provide examples for Proposition 4. That is, as the disutility of biopsy procedure decreases, the control limits become smaller.

PROPOSITION 4. Let \mathbb{B}_1 and \mathbb{B}_2 be the two biopsy procedures with disutilities of biopsy dB_1 and dB_2 , respectively, such that $dB_1 \leq dB_2$. If a patient satisfying the control-limit theorem has control limits κ_t^1 and κ_t^2 for \mathbb{B}_1 and \mathbb{B}_2 , respectively, then $\kappa_t^1 \leq \kappa_t^2$ for all $t = 1, 2, \dots, T$.

We next define the maximum and average violations of the assumptions and conditions described above. Maximum violation of (As5):

Maximum violation of (As5):

$$\epsilon_{\max}^1 = \max_{i,j,t} \left\{ \max \left\{ 0, \sum_{s'=j}^{S+2} p_t^{Am}(s'|i) - \sum_{s'=j}^{S+2} p_{t+1}^{Am}(s'|i) \right\} \right\}$$

for $i, j \in \mathbb{S}$ and $t = 1, 2, \dots, T - 1$.

Maximum violation of the IFR assumption:

$$\epsilon_{\max}^2 = \max_{i,j,t} \left\{ \max \left\{ 0, \sum_{s'=j}^{S+2} p_t^{Am}(s'|i) - \sum_{s'=j}^{S+2} p_t^{Am}(s'|i+1) \right\} \right\}$$

for $i, j \in \mathbb{S}$ and $t = 1, 2, \dots, T$.

Maximum violation of Condition (7):

$$\epsilon_{\max}^3 = \max_{s,t} \left\{ \max \left\{ 0, \frac{r_t(s, Bx) - r_t(s+1, Bx)}{\lambda r_{t+1}(s+1, Bx)} - \left\{ p_t^{Am}(S+2|s_t+1) - p_t^{Am}(S+2|s) \right\} \right\} \right\}$$

for $s = 0, 1, 2, \dots, S-1, t = 1, 2, \dots, T-1$.

Maximum violation of Condition (8):

$$\epsilon_{\max}^4 = \max_{s,t} \left\{ \max \left\{ 0, \sum_{s''=s+1}^S p_t^{Am}(s''|s) - \sum_{s''=s+1}^S p_t^{Am}(s''|s+1) \right\} \right\}$$

for $s = 0, 1, 2, \dots, S, S+2, t = 1, 2, \dots, T-1$.

Maximum violation of Condition (10):

$$\epsilon_{\max}^5 = \max_{s,t} \left\{ \max \left\{ 0, \{r_t(s, Bx) - r_{t+1}(s, Bx)\} - \{r_t(s, Am) - r_{t+1}(s, Am)\} - \lambda \sum_{s' \in S} p_t^{Am}(s'|s) \{r_{t+1}(s', Bx) - r_{t+2}(s', Bx)\} \right\} \right\}$$

for $s = 0, 1, 2, \dots, S, S+2, t = 1, 2, \dots, T-1$.

Maximum violation of superadditivity of $r_t(s, Am)$:

$$\epsilon_{\max}^6 = \max_{s,t} \{ \max \{ 0, \{r_t(s+1, Am) - r_{t+1}(s+1, Am)\} - \{r_t(s, Am) - r_{t+1}(s, Am)\} \} \}$$

for $s = 0, 1, 2, \dots, S, S+2, t = 1, 2, \dots, T-1$.

Maximum violation of superadditivity of $r_t(s, Bx)$:

$$\epsilon_{\max}^7 = \max_{s,t} \{ \max \{ 0, \{r_t(s+1, Bx) - r_{t+1}(s+1, Bx)\} - \{r_t(s, Bx) - r_{t+1}(s, Bx)\} \} \}$$

for $s = 0, 1, 2, \dots, S, S+2, t = 1, 2, \dots, T-1$.

Maximum violation of superadditivity of $\sum_{s'=j}^{S+2} p_t^{Am}(s'|i)$:

$$\epsilon_{\max}^8 = \max_{i,j,t} \left\{ \max \left\{ 0, \sum_{s'=j}^{S+2} \{p_t^{Am}(s'|i) - p_{t+1}^{Am}(s'|i)\} - \sum_{s'=j}^{S+2} \{p_t^{Am}(s'|i+1) - p_{t+1}^{Am}(s'|i+1)\} \right\} \right\}$$

for $i, j = 0, 1, 2, \dots, S, t = 1, 2, \dots, T-1$.

Maximum violation of Condition (12):

$$\epsilon_{\max}^9 = \max_{i,t} \left\{ \max \left\{ 0, r_t(i, Am) + \lambda \sum_{s' < i} p_t^{Am}(s'|i) r_{t+1}^*(0, Bx) + \lambda \sum_{s' \geq i} p_t^{Am}(s'|i) r_{t+1}(s', Bx) - r_t(i, Bx) \right\} \right\}$$

for all $i \in S$ and $t = 1, 2, \dots, T-1$.

The average violations are defined as follows:

Average violation of (As5):

$$\epsilon_{\text{avg}}^1 = \frac{1}{|S+3||S+3||T-1|} \sum_{t=1}^{T-1} \sum_{j \in \mathcal{S}} \sum_{i \in \mathcal{S}} \cdot \left\{ \max \left\{ 0, \sum_{s'=j}^{S+2} p_t^{Am}(s'|i) - \sum_{s'=j}^{S+2} p_{t+1}^{Am}(s'|i) \right\} \right\}.$$

Similarly, we define $\epsilon_{\text{avg}}^2, \dots, \epsilon_{\text{avg}}^9$.

4. Computational Experiments

The clinical data used for computational experiments came from 65,892 mammographic findings from 18,270 patients at Medical College of Wisconsin (MCW), Milwaukee, between April 5, 1999 and February 9, 2004. Reference standard outcomes for these data were obtained from breast biopsy as well as cancer registry match data, which provided a benign or malignant class labels for all findings.

In the rest of this section, we first summarize OBDM's parameter estimation including states, state transition probabilities, and rewards. Details of parameter estimation are presented in Appendix A in the electronic companion. Second, we estimate the violations of the assumptions and the conditions described in §3 using the available clinical data. Third, we perform extensive sensitivity analyses to check the robustness of our model. Last, we evaluate the performance of the OBDM's policies on mammography exams and compare it to radiologists' performance in practice.

4.1. Parameter Estimation

The *states* of the OBDM—defined as the risk score (probability of breast cancer)—are estimated using a mammography Bayesian network (MBN). Our MBN (Figure EC.2) estimates the current probability of breast cancer using patient demographic factors and mammographic features recorded in a National Mammography Database (NMD) format. We rounded the probabilities to map to the risk score of our OBDM. For example, the probability of cancer equal to 0.0382 will correspond to the risk score of 4%. Our MBN also incorporates features such as mass density and mass size (which may have changed from the previous mammogram) that incorporate information from previous mammograms to estimate the risk of cancer. The *state transition probabilities* are estimated by tracking changes in the risk score of consecutive mammographic examinations of each patient with time from the MCW data.

The *expected intermediate reward* ($r_t(s_t, Am)$), accrued in the current time period when the patient's state is s_t and annual mammogram action is taken, is assumed to be one year if the patient is alive in that decision-epoch, and one-half year otherwise. We assume quality-of-life (QOL) factor equals to 1 for all states when the action is Am , i.e., we use expected life as the reward function because we could not find appropriate studies in the literature that estimate QOL factor associated with the risk of cancer. The age-dependent lump-sum *post-biopsy rewards* of breast cancer associated with each risk score are estimated using age-specific probabilities of death from breast cancer (Jemal et al. 2007), and no-cancer from the 2003 U.S. life tables reported by the NCHS of CDC (Arias 2006). We differentiate between the probability of death from cancer within one year if the patient is not treated, $p_t^{mal2}(S+2)$, versus treated, $p_t^{mal}(S+2)$. We use probability of death from untreated cancer to estimate intermediate expected rewards (as in this case a cancer would not be diagnosed, hence no treatment would be given to the patient). We use a parametric model (Haybittle 1998) to

estimate $p_t^{mal2}(S+2)$. We estimate $p_t^{mal}(S+2)$ from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (Jemal et al. 2007).

4.2. Numerical Results

To obtain the optimal policies of the OBDM, we first assign values to OBDM's parameters from literature when available, and we assume otherwise. Next, we perform sensitivity analysis in §4.3 to check the model's robustness. The following values are assumed as base case values:

- Disutility of biopsy at age 40 ($dBx@40$) is two weeks (Velanovich 1995) and increases linearly with age. The disutility is assumed to increase because of increasing co-morbidities and biopsy complications associated with older ages.
- Disutility of biopsy factor at age 100 ($dBxFactor@100$), defined as the multiplication factor for the disutility of biopsy at age 100. For example, $dBxFactor@100$ of 4 implies that the disutility of biopsy at age 100, $dBx@100 = 4dBx@40$.
- The ratio of invasive cancer to all cancers (invasive and in-situ) = 0.75 (Jemal et al. 2007).
- "Treatment effectiveness factor," defined as the ratio of the probability of death from cancer when the patient is not treated to the probability of death when the patient is treated $(p_t^{mal2}(S+2) / (p_t^{mal}(S+2))) = 1.6$ (Haybittle 1998).

Figure 2 shows the optimal probability thresholds for biopsy decisions in different age groups when $\lambda = 1$. The optimal policy can be interpreted as follows: the radiologist should send a patient between the ages 40 and 42 for biopsy if her probability of cancer is 1% or higher. This probability threshold rises to 2% for a patient between the ages of 43 and 63, 3% for a patient between the ages of 64 and 82, and 7% for a 90-year-old patient. Note that the optimal policy is of control-limit type, and the optimal probability threshold for biopsy is nondecreasing with age. We prove the optimality of such a structured policy in Theorems 1 and 3. The optimal policy implies that as a patient gets older, she is less likely to be recommended for biopsy if QALYs are maximized. This could be because the aggressive biopsy in older ages might not significantly increase the total expected QALYs. For example, aggressive diagnosis of low grade or pre-invasive lesions in older women (women 65 years or older) might represent overdiagnosis that does not contribute to reduced morbidity or mortality because of limited life expectancy (Ernster et al. 1996,2002).

We also compare the optimal policies of two women of different age groups having identical disutility of biopsy. We observe that the higher the disutility of biopsy, the higher the optimal biopsy threshold. Also, the optimal biopsy threshold of an 80-year-old patient is never lower than that of a 50-year-old woman for the same disutility of biopsy (Figure 3). This trend again suggests age-dependent biopsy thresholds.

Next, we estimate the violations defined in §3 using clinical data available from MCW and other sources. Table 2 shows the maximum and average violations. Note that all violations are very small. Although the maximum violation of Condition (10) (denoted by ϵ_5) and Condition (12) (denoted by ϵ_9) are relatively high, their small average violations suggest that the maximum violations could have occurred at the boundary conditions.

4.3. Sensitivity Analysis

We check the robustness of the model by performing sensitivity analysis on the variables defined in §4.2. First, we change the disutility of biopsy at age 40. The higher the disutility

of biopsy, the less likely that the optimal action would be biopsy (Figure 4), which is proved by Proposition 4. If the disutility of biopsy is 0 for all ages, biopsy is the optimal decision for all states, which is expected because we do not consider any costs in this model. Note that the dotted line is not at 0 because of the rounding of the states (risk scores) in our model. At risk score 0, the expected life of choosing a biopsy or annual mammogram are equal when the disutility of biopsy is zero. We choose the optimal decision rule as annual mammogram when two decision rules yield the same rewards.

Next, we vary the disutility of biopsy at age 100 (Figure 5). As proved by Proposition 4, the higher the disutility of biopsy, the less likely that the optimal action would be biopsy. Note that even when the disutility of biopsy stays constant (multiplication factor is 1) with age or decreases (multiplication factor is 0.5) with age, the control limit is still increasing, which supports the clinical intuition that potential savings with biopsy are less for older women. Similarly, as the fraction of invasive cancer increases, the patient is more likely to opt for biopsy (Figure 6). This is because the life expectancy of a patient with invasive cancer is much lower than that of a patient with in-situ cancer. Next, as the treatment effectiveness factor increases, the total benefit of biopsy increases and as a result, the probability-threshold to biopsy decreases (Figure 7). Last, the optimal policy was insensitive to the change in discount rate values; particularly, no trend in the optimal policy was observed with the change in the discount rates to 0.97 and 0.95.

4.4. Comparison with Radiologists

We compare the decisions made by radiologists in real life to the optimal policies given by the OBDM using MCW data. We assume BI-RADS 1, 2, and 3 as negative, and BI-RADS 0, 4, and 5 as positive for a direct comparison between our model and radiologists. We use different data sets for estimating OBDM's parameters and testing OBDM's performance, and perform two-fold cross-validation (Stone 1974). Specifically, we divide our data set in two folds. We first use fold 1 to estimate the parameters of the OBDM, including the risk scores and transition probabilities, and compute an optimal policy. Then we assign the outcomes of the computed policy to the data in fold 2. Similarly, we use fold 2 to estimate the parameters of the OBDM, and evaluate the policies on fold 1. We used the Wisconsin Cancer Reporting System (cancer registry) to match findings recorded in our database to the actual outcomes. All newly diagnosed cancer cases are reported to the Wisconsin Cancer Reporting System. This registry collaborates with 17 other state agencies to collect a range of data including demographic information, tumor characteristics, treatment, and mortality. If a radiologist missed a cancer on the mammogram (assigned BI-RADS 1, 2, or 3), which was found later during the registry match, we label that case as “false negative.” On the other hand, if a radiologist recommended a lesion for biopsy (assigned BI-RADS 0, 4, or 5) that turned out to be cancerous, we labeled that finding as “true positive.”

Because the subjective disutilities of biopsies of individual patients in the MCW database are not available, we assume an equal disutility of biopsy for the whole population in the study and perform sensitivity analysis. We run nine scenarios by varying the subjective disutility of biopsy and compare the performance of the OBDM's policies to the radiologists' actions (Table 3). The number of false-positive cases of the OBDM is always lower than that of the radiologists. For six out of nine biopsy-disutility values, the OBDM's true-positive cases are comparable or greater than that of the radiologists.

5. Conclusions and Future Work

Mammography interpretation varies significantly with radiologists' experience and skills. No methods exist that enable radiologists to determine an optimal threshold over which to recommend biopsy for a given patient. Medical literature shows that 55%–85% of the breast

biopsies turn out as benign findings, resulting in over-treatment, unnecessary anxiety, and expenditures. In this paper, we address the decision problem faced by radiologists: when should a woman be sent for biopsy based on her mam-mographic features and demographic factors? To the best of our knowledge, this is the first quantitative study that addresses the problem of optimal breast biopsy decision-making. We formulate the optimal biopsy decision model as a finite-horizon discrete-time MDP, and we investigate the structural properties of the model to gain insights on how decision are made.

Our results provide patient-specific optimal policies for breast biopsy. The optimal policy is age-dependent, with older women having a higher biopsy threshold than younger women. This might appear counterintuitive; however, note that breast cancer tends to be less aggressive in older women, and an aggressive biopsy policy might not significantly increase the total expected QALYs for older women. In addition, false-positive interpretation of mam-mography leading to unnecessary invasive procedures can result in complications in older patients with comorbidities. It is important to note that in the current clinical practice, radiologists using BI-RADS assessment do not take age into consideration while making biopsy decisions. For younger patients, the optimal biopsy thresholds estimated by the OBDM are close to the current clinical practice of 2% threshold as recommended by the American College of Radiology. However, for older patients the optimal biopsy thresholds given by the OBDM are higher than those recommended by the current clinical practice.

The performance of our OBDM's optimal policies is better than the decisions made by radiologists on real-life mammography data. We believe that the suboptimal performance (higher false-positives and biopsies) of radiologists is probably attributable to their inability to accurately estimate the risk of breast cancer without any CADx model and BI-RADS guidelines not taking age into account. This is further exacerbated by the high penalty for missing a breast cancer, causing radiologists to unwittingly lower their threshold below what is recommended. While it is difficult to accurately calculate risk estimates to biopsy recommendations, it is equally difficult to tailor these recommendations to individual patients based on unique features like age. Our work addresses some of these challenges faced by radiologists.

Our study can also be used to solve other similar medical decision-making problems. For example, the framework of the OBDM can be applied to find the optimal timing of biopsy in prostate cancer diagnosis. In addition, the structural properties of our OBDM can be used to gain insights on the structure of the optimal policies (such as the existence of control-limit type optimal policies or trends in the control limits with patient's demographic factors) of similar medical problems.

Our model supports the idea of personalized care by providing patient-specific policies for breast biopsy in contrast to the current practice of overarching, non-tailored guidelines. In addition, the OBDM facilitates shared decision-making by providing optimal patient-specific biopsy policies as a function of the probability of cancer. Patients have a better understanding of their health condition from probabilities, which are intuitive to understand, as opposed to six discrete BI-RADS codes.

Next, we discuss our study's results with a recent controversy on optimal policy for mammography screening (Nelson et al. 2009, Partridge and Winer 2009, Murphy 2010). The U.S. Preventive Services Task Force (USPSTF) issued a new guideline suggesting starting routine mammog-raphy screening for low-risk women at age 50 as opposed to 40, whereas several influential medical organizations such as the American Cancer Society, American Medical Association, and American College of Radiology suggested continuing with the existing policy of annual screening starting from age 40. Our recommendations do

not contradict or support any screening guidelines. The USPSTF recommends less aggressive management of mammography screening in younger women, whereas our study recommends aggressive management of biopsy decisions in the same group. Our study does not imply that we will biopsy more young women than older ones, because the prevalence of breast cancer is higher in older women.

Our MDP model formulation assumes that we have complete information of our states. Alternatively, we could relax this assumption and formulate our problem using a POMDP model. However, there are several limitations with this approach in the context of biopsy decision-making problem. First, we cannot reliably estimate the observation probabilities. Second, the POMDP approach would use cancer incidence rates and the Bayesian update formula for estimating transitions between belief states, whereas an MDP approach could use clinical data and MBN for estimating these transitions. In conventional POMDPs, observation probabilities depend on current (or previous) core state and the action, not on the previous belief state. Therefore, such information (i.e., transition probability from a belief point to another one) cannot be incorporated into the POMDP framework directly.

One of the limitations of our study is that we assume that the decision ends if a biopsy is performed because the decision process for a patient who had already undergone biopsy might be different compared to women who have never undergone a biopsy. Patients who have a history of biopsy will have a higher risk profile and might experience scarring due to a biopsy, making future interpretation of mammograms more difficult. Such differences require the specific estimation of cancer risk, transition probabilities, and post-biopsy QALYs. Because we do not have any available data for this subset of women, we decided to end the decision process for women at the time of biopsy, which implies that our decision model applies only to women who never had a biopsy. Our modeling approach does not imply that a woman may not have multiple biopsies; it simply does not consider two or more biopsies explicitly. We plan this modification for future work. Note that incorporating the possibility of multiple biopsies is straightforward in our modeling framework; we only need to include a transition from post-biopsy benign state to the risk scores. Such a modeling change would impact most of our structural results; however, we do not expect our numerical results to change significantly.

We assume that the disutilities of both benign and malignant biopsy are equal. However, in practice, the disutility of biopsy might depend on the outcome of the biopsy. In one scenario, a woman could have a lower disutility of benign biopsy than that of malignant biopsy because she is relieved from the thought of having a cancer. On the other hand, a woman might have a higher disutility of benign biopsy than that of malignant biopsy if she thinks she had to go through unnecessary anxiety considering the possibility of a cancer. Therefore, we do not differentiate between the disutility of positive and negative biopsies but instead perform sensitivity analysis on a combined disutility of biopsy.

In our OBDM, we only use mammography features to estimate the risk of breast cancer. However, our model can easily be updated to incorporate additional imaging features such as ultrasound and MRI. This would be achieved by updating the risk of breast cancer based on information from these other imaging modalities. We assume that our probability estimates (risk of breast cancer) from our MBN are perfect. While we have a high performance of MBN as measured by an ROC curve (AUC = 0.961), we acknowledge that this is a limitation of our study.

There are several future research directions of this study. In our OBDM, we assume that decisions are made annually, and there are only two possible actions: biopsy, or wait until the next annual mammogram. In the future, we will add another action: short-interval

imaging follow-up, which radiologists recommend for patients if their risk of cancer is neither very high/imminent nor negligible. Adding follow-up as another action will change the structure of the OBDM and make parameter estimation from the clinical data more challenging; hence, it warrants the building of a different model. We also plan to extend the OBDM to explicitly incorporate multiple biopsies, where the patient's state will also include information of personal history of biopsy. In our present model, we do not consider the cost of biopsy or mammograms in the reward function and leave it as a future work. Another possible future direction is to include different biopsy procedures (core-needle, fine aspiration, etc.) in the action space, but a limited availability of relevant clinical data makes it challenging.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

Acknowledgments

This research was supported by the National Cancer Institute grants R21CA129393, R01CA127379, and K07CA114181. The authors thank Turgay Ayer, Fatih Erenay, and Karin Witte, as well as Stefanos Zenios, an anonymous associate editor, and three anonymous referees for their suggestions and insights, which improved this manuscript.

References

- Alagoz O, Maillart LM, Schaefer AJ, Roberts MS. The optimal timing of living-donor liver transplantation. *Management Sci* 2004;50(10):1420–1430.
- American Cancer Society. Breast cancer facts and figures: 2010. American Cancer Society; Atlanta: 2010.
- Arias E. United States life tables, 2003. *National Vital. Stat. Rep* 2006;54(14):1–40.
- Ayer T, Alagoz O, Chhatwal O, Shavlik JW, Kahn CE, Burnside ES. Breast cancer risk estimation with artificial neural networks revisited: Discrimination and calibration. *Cancer* 2010;116(14):3310–3321. [PubMed: 20564067]
- Baines C, Dayan JR. A tangled web: Factors likely to affect the efficacy of screening mammography. *J. National Cancer Inst* 1999;91(10):833–838.
- Baker JA, Kornguth PJ, Lo JY, Williford ME, Floyd CE Jr. Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* 1995;196(3):817–822. [PubMed: 7644649]
- Barlow, RE.; Proschan, F. *Mathematical Theory of Reliability*. Wiley; New York: 1965.
- Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, Hendrick RE, Elmore JG. Accuracy of screening mammography interpretation by characteristics of radiologists. *J. National Cancer Inst* 2004;96(24):1840–1850.
- Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by U.S. radiologists. Findings from a national sample. *Arch. Intern. Med* 1996;156(2):209–213. [PubMed: 8546556]
- BI-RADS. Breast Imaging Reporting and Data System (BI-RADS). 3rd ed. American College of Radiology; Reston, VA: 1998.
- Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology* 1992;184(3):613–617. [PubMed: 1509041]
- Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: Initial experience. *Radiology* 2006;240(3):666–673. [PubMed: 16926323]
- Burnside ES, Davis J, Chhatwal J, Alagoz O, Lindstrom MJ, Geller BM, Littenberg B, Shaffer KA, Kahn CE, Page CD. A probabilistic computer model developed from clinical data in the national

- mammography database format to classify mammography findings. *Radiology* 2009;251(3):663–672. [PubMed: 19366902]
- Chan EC. Promoting an ethical approach to unproven imaging tests. *J. Amer. Coll. Radiol* 2005;2(4):311–320. [PubMed: 17411824]
- Chhatwal J, Alagoz O, Lindstrom MJ, Shaffer KA, Kahn CE, Burnside ES. A logistic regression model based on the national mammography database format to aid breast cancer diagnosis. *Amer. J. Roentgenol* 2009;192(4):1117–1127. [PubMed: 19304723]
- Destounis SV, DiNitto P, Logan-Young W, Bonaccio E, Zuley ML, Willison KM. Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate? Initial experience. *Radiology* 2004;232(2):578–584. [PubMed: 15229350]
- Drummond, MF. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press; Oxford, UK: 2005.
- Elvecrog EL. Nonpalpable breast lesions: Correlation of stereotaxic large-core needle biopsy and surgical biopsy results. *Radiology* 1993;188(2):453–455. [PubMed: 8327696]
- Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson C. Incidence of and treatment for ductal carcinoma in situ of the breast. *JAMA* 1996;275(12):913–918. [PubMed: 8598618]
- Ernster VL, Ballard-Barbash R, Barlow WE, Zheng Y, Weaver DL, Cutter G, Yankaskas BC, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J. National Cancer Inst* 2002;94(20):1546–1554.
- Fan K. Subadditive functions on a distributive lattice and an extension of Szász's. *J. Math. Anal. Appl* 1967;18:262–268.
- Fowble BL, Schultz DJ, Overmoyer B, Solin LJ, Fox K, Jardines L, Orel S, Glick JH. The influence of young age on outcome in early stage breast cancer. *Internat. J. Radiat. Oncol. Biol. Phys* 1994;30(1):23–33.
- Fracheboud J, Groenewoud JH, Boer R, Draisma G, de Bruijn AE, Verbeek AL, de Koning HJ. Seventy-five years is an appropriate upper age limit for population-based mammography screening. *Internat. J. Cancer* 2005;118(8):2020–2025.
- Freid, VM.; Prager, K.; MacKay, AP.; Xia, H. *Health, United States, 2003*. National Center for Health Statistics; Hyattsville, MD: 2003. Chartbook on trends in the health of Americans.
- Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. Chapter 7: The Wisconsin breast cancer epidemiology simulation model. *J. National Cancer Inst. Mono* 2006;2006(36):37–47.
- Haybittle JL. Life expectancy as a measurement of the benefit shown by clinical trials of treatment for early breast cancer. *Clin. Oncol. (R. Coll. Radiol.)* 1998;10(2):92–94. [PubMed: 9610897]
- Hillman BJ. Informed and shared decision making: An alternative to the debate over unproven screening tests. *J. Amer. College Radiology* 2005;2(4):297–298.
- Ivy, JS. *Balancing patient and payer preferences: A maintenance-based model for breast cancer treatment and detection*. North Carolina State University; Raleigh: 2006. Working paper
- Jayasinghe UW, Taylor R, Boyages J. Is age at diagnosis an independent prognostic factor for survival following breast cancer? *ANZ J. Surgery* 2005;75(9):762–767.
- Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. *Cancer statistics, 2007*. *CA: A Cancer J. Clinicians* 2007;57(1):43–66.
- Jesneck JL, Nolte LW, Baker JA, Floyd CE, Lo JY. Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis. *Medical Phys* 2006;33(8):2945–2954.
- Jiang Y, Nishikawa RM, Schmidt RA, Toledano AY, Doi K. Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcification. *Radiology* 2001;220(3):787–794. [PubMed: 11526283]
- Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer assisted diagnosis. *Acad. Radiol* 1999;6(1):22–33. [PubMed: 9891149]
- Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. *JAMA* 1995;273(2):149–154. [PubMed: 7799496]

- Lee CH, Philpotts LE, Horvath LJ, Tocino I. Follow-up of breast lesions diagnosed as benign with stereotactic core-needle biopsy: Frequency of mammographic change and false-negative rate. *Radiology* 1999;212(1):189–194. [PubMed: 10405741]
- Liberian L. Centennial dissertation. Percutaneous imaging-guided core breast biopsy: State of the art at the millennium. *Amer. J. Roentgenol* 2000;174(5):1191–1199. [PubMed: 10789761]
- Liberian L, Menell JH. Breast imaging reporting and data system (BI-RADS). *Radiol. Clin. North Amer* 2002;40(3):409–430. [PubMed: 12117184]
- Maillart L, Ivy JS, Ransom S, Diehl K. Assessing dynamic breast cancer screening policies. *Oper. Res* 2008;56(6):1411–1427.
- Maxwell JR, Bugbee ME, Wellisch D, Shalmon A, Sayre J, Bassett LW. Imaging-guided core needle biopsy of the breast: Study of psychological outcomes. *Breast J* 2000;6(1):53–61. [PubMed: 11348335]
- Meyer JE, Eberlein TJ, Stomper PC, Sonnenfeld MR. Biopsy of occult breast lesions. Analysis of 1261 abnormalities. *JAMA* 1990;263(17):2341–2343. [PubMed: 2157903]
- Michaelson JS, Halpern E, Kopans DB. Breast cancer: Computer simulation method for estimating optimal intervals for screening. *Radiology* 1999;212(2):551–560. [PubMed: 10429717]
- Murphy AM. Mammography screening for breast cancer: A view from 2 worlds. *JAMA* 2010;303(2):166–167. [PubMed: 20068212]
- Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: An update for the U.S. preventive services task force. *Ann. Internal Medicine* 2009;151(10):727–737.
- Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. *Lancet* 2002;359(9310):909–919. [PubMed: 11918907]
- Opie H, Estes NC, Jewell WR, Chang CH, Thomas JA, Estes MA. Breast biopsy for nonpalpable lesions: A worthwhile endeavor? *Amer. Surg* 1993;59(8):490–493. [PubMed: 8393310]
- Parker SH. Nonpalpable breast lesions: Stereotactic automated large-core biopsies. *Radiology* 1991;180(2):403–407. [PubMed: 1648757]
- Partridge AH, Winer EP. On mammography—More agreement than disagreement. *New England J. Medicine* 2009;361(26):2499–2501.
- Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Oper. Res* 1980;28(1):206–224.
- Puterman, ML. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc.; New York: 1994.
- Schairer C, Mink PJ, Carroll L, Devesa SS. Probabilities of death from breast cancer and other causes among female breast cancer patients. *J. National Cancer Inst* 2004;96(17):1311–1321.
- Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: Specialist and general radiologists. *Radiology* 2002;224(3):861–869. [PubMed: 12202726]
- Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer. *CA: A Cancer J. Clinicians* 2006;56(1):11–25.
- Smith RA, Saslow D, Andrews Sawyer K, Burke W, Costanza ME, Evans WP, Foster RS, Hendrick E, Eyre HJ, Sener S. American Cancer Society guidelines for breast cancer screening: Update 2003. *CA: A Cancer J. Clinicians* 2003;53(3):141–169.
- Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA* 2003;290(16):2129–2137. [PubMed: 14570948]
- Sonnenberg FA, Beck JR. Markov models in medical decision making: A practical guide. *Medical Decision Making* 1993;13(4):322–338. [PubMed: 8246705]
- Steggles S, Lightfoot N, Sellick SM. Psychological distress associated with organized breast cancer screening. *Cancer Prev. Control* 1998;2(5):213–220. [PubMed: 10093635]
- Stone M. Cross-validation choice and assessment of statistical procedures. *J. Roy. Statist. Soc* 1974;36(2):111–147.

- Tversky, A.; Kahneman, D. Judgment under uncertainty: Heuristics and biases. In: Shafir, E., editor. Preference, Belief, and Similarity: Selected Writings. Amos Tversky. MIT Press; Cambridge, MA: 2004.
- Velanovich V. Immediate biopsy versus observation for abnormal findings on mammograms: An analysis of potential outcomes and costs. *Amer. J. Surg* 1995;170(4):327–332. [PubMed: 7573723]
- Williams RS, Willard HF, Snyderman R. Personalized health planning. *Science* 2003;300(5619):549. [PubMed: 12714710]

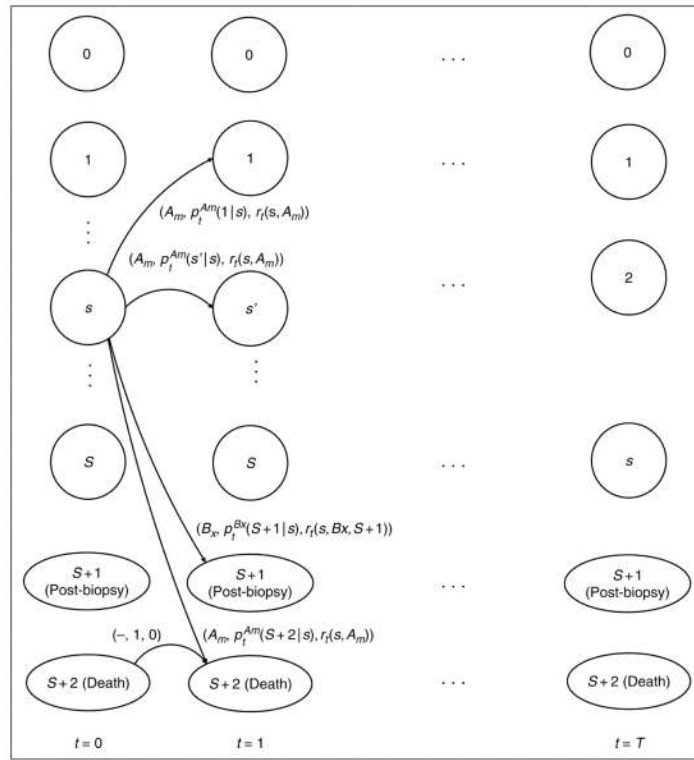


Figure 1.
State transition diagram of the OBDM.

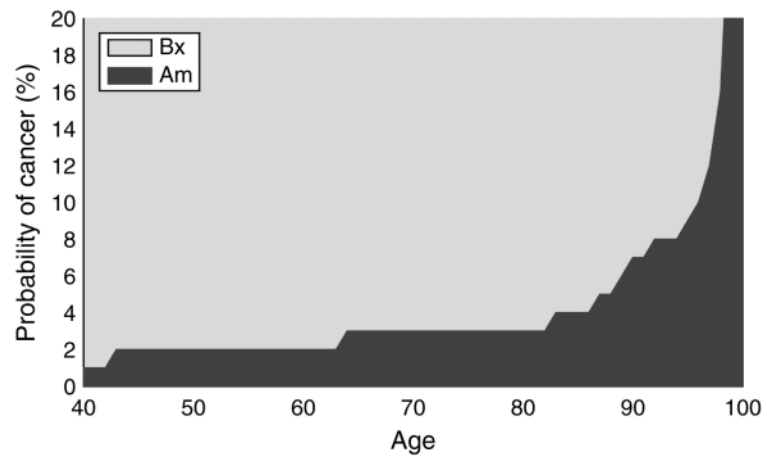


Figure 2.
Optimal age-dependent policy to perform biopsy.

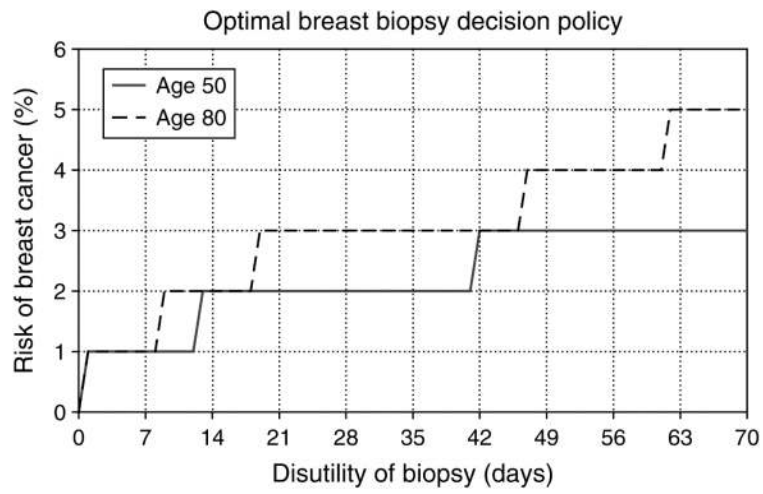


Figure 3. Optimal policy to perform biopsy for 50-and 80-year-old women based on their subjective disutility of biopsy.

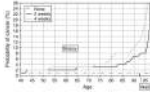


Figure 4.
Optimal probability threshold for biopsy changing with biopsy disutility at age 40.

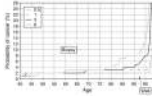


Figure 5. Optimal probability threshold for biopsy changing with the multiplication factor of biopsy disutility at age 100.

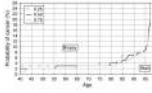


Figure 6.
Optimal probability threshold for biopsy changing with the ratio of invasive/in-situ cancer.

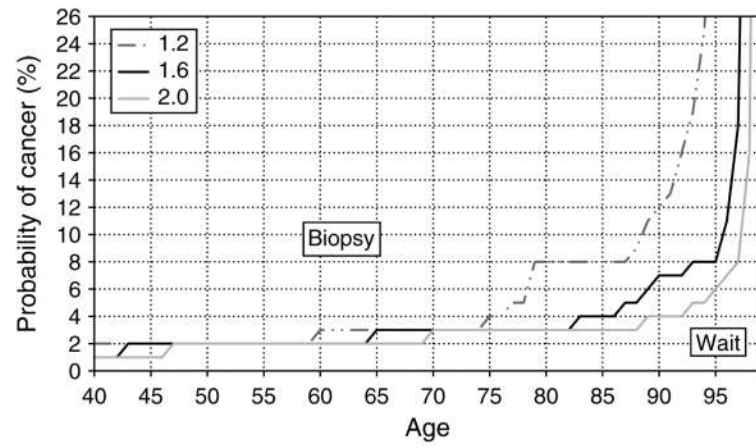


Figure 7. Optimal probability threshold for biopsy changing with the ratio of treatment effectiveness factor.

Table 1

BI-RADS final assessment codes with recommended actions.

Category	Definition	Recommended action
0	Need additional imaging evaluation and/or prior mammograms for comparison	Additional imaging evaluation
1	Negative finding	Wait (routine yearly screening)
2	Benign finding	Wait (routine yearly screening)
3	Probably benign finding (less than a 2% risk of malignancy)	Short-term follow-up (6 months)
4	Suspicious abnormality (risk of malignancy is between 2% and 95%)	Biopsy
5	Highly suggestive of malignancy (95% risk of malignancy)	Biopsy

Table 2

Error estimation.

Violation	Maximum	Average
ε^1	0.0391	0.0001
ε^2	0.0000	0.0000
ε^3	0.0031	0.0015
ε^4	0.0000	0.0000
ε^5	0.1629	0.0914
ε^6	0.0851	0.0092
ε^7	0.0363	0.0001
ε^8	0.0283	0.0000
ε^9	0.1677	0.0055

Table 3

Performance of the OBDM on the MCW data.

<i>dBx</i> @40	<i>dBx</i> @100	True positive	False positive	False negative	True negative
1 day	4 days	451	3,527	59	58,182
5 days	20 days	440	2,861	70	58,848
1 week	4 weeks	434	2,501	76	59,208
2 weeks	1 week	434	2,287	76	59,422
2 weeks	2 weeks	434	2,080	76	59,629
2 weeks	4 weeks	432	2,004	78	59,705
2 weeks	6 weeks	427	1,829	83	59,880
2 weeks	8 weeks	424	1,759	86	59,950
3 weeks	12 weeks	416	1,517	94	60,192
Radiologists		435	7,357	75	54,352