MANAGEMENT SCIENCE

https://pubsonline.informs.org/journal/mnsc

Vol. 00, No. 0, Xxxxx 0000, pp. 000–000

To be submitted to *Management Science*

ISSN 0025-1909, EISSN 1526-5501

Proposal: Personalized Breast Cancer Prevention by Combining Screening and Diagnostic Decisions

Hamed Milani

Haskayne School of Business, University of Calgary, hamed.milani@calgary.ca

Abstract. Mammography screening has been shown to be the most effective method for reducing breast cancer mortality. However, it still faces significant inefficiencies, primarily due to high false-positive rates and substantial unnecessary follow-up examinations, such as biopsies. In this research, we aim to address these two inefficiencies by combining screening and post-mammography diagnostic decisions into a unified framework to derive personalized policies for managing patients throughout the breast cancer screening pathway. Although prior research has advanced personalization in breast cancer care, existing studies typically treat screening decisions and post-mammography diagnostic decisions as separate problems. We hypothesize that screening and diagnostic decisions are inherently interconnected, and that addressing them separately could result in suboptimal and less personalized screening pathways for patients. We formulate this problem as a partially observable Markov decision process (POMDP), with two decision variables: (1) determining the optimal timing for mammography (screening decisions), and (2) selecting the appropriate follow-up examination method (diagnostic decisions).

Key words: partially observable Markov decision processes, dynamic programming, decision analysis, medical decision making, breast cancer, mammography screening, personalized screening

1. Introduction

Breast cancer is the first and most prevalent cancer diagnosed among women in Australia, with approximately 19,031 cases reported in 2023 (Australian Institute of Health and Welfare (AIHW 2024)). Breast cancer is also the second leading cause of cancer death among women in Australia, accounting for approximately 3,259 deaths in 2023 (AIHW 2024). It is estimated that 20,336 new cases of breast cancer will be diagnosed, with 3,353 related deaths in 2025 (AIHW 2024). This means that about 1 in 8 women will be diagnosed with breast cancer, and 1 in 57 women will die because of it. Early diagnosis of breast cancer is critically important, as it significantly improves survival rates by allowing interventions at stages when the cancer is more localized (Eby et al. 2022,

Helvie et al. 2014, Pocobelli and Weiss 2015). This impact is evident in the dramatic increase in the 5-year survival rate, which rises from 23% at stage 4 to 92% at stage 2 (Ellison and Saint-Jacques 2023).

Mammography screening remains the most effective method for early breast cancer detection and reducing its mortality rate at the population level, which can detect abnormalities in breast tissue 1.5 to 2.3 years before symptoms appear or are noticeable by patients (Løberg et al. 2015, Tabár 2018, Duffy 2020). In Australia, the age-standardized mortality rate dropped from 36.9 per 100,000 women in 1982 to 22.4 in 2024, largely attributed to increased screening and improved treatments (AIHW 2024). This trend, along with the age-standardized incidence rate, is illustrated in Figure 3. Although the breast cancer incidence rate rose in the 1980s-90s due to increased mammography use and hormonal factors, it has since been stabilized. While several randomized control trials (RCTs) have demonstrated the benefit of mammography screening programs in reducing breast cancer mortality, accounting for approximately 20% reduction in mortality rate (Marmot et al. 2013), several meta-analyses and studies show that these programs still face significant inefficiencies, primarily due to two sources: high rate of false-positive and high rate of unnecessary follow-up examinations (Løberg et al. 2015). We aim to build our work focusing on these two sources of inefficiency.

False-positive mammograms, the primary source of inefficiency in screening programs, occur when a suspicious finding is incorrectly interpreted as cancerous, prompting additional diagnostic tests that ultimately confirm no cancer. False positives arise in 2.5–14% of screening mammograms, translating to about 392,000 false positives among 2.8 million mammograms in Canada in 2018 (Løberg et al. 2015, CCS 2023). Several factors contribute to these errors. Dense breast tissue can obscure imaging, making benign and malignant patterns harder to distinguish, an issue more common in younger women (Corsetti et al. 2008). Benign conditions such as cysts, fibroadenomas, or calcifications may also appear abnormal. The expertise and risk behavior of radiologists further influence outcomes, as variability in interpretations can produce false positives in borderline cases (Alberdi et al. 2011). These errors cause substantial patient anxiety (Brewer et al. 2007), lead to unnecessary follow-up tests and overtreatment, and contribute to population-level over-screening.

Unnecessary follow-up examinations are the second major source of inefficiency. After mammography, patients either return to routine screening or undergo additional procedures, such as diagnostic mammograms, biopsies, ultrasounds, or MRIs, often scheduled immediately or within 3, 6, or 12 months. To standardize reporting and management, radiologists use the BI-RADS system (Magny et al. 2023), which classifies findings from 1 to 6 with corresponding recommendations

(Table 1). However, its subjectivity, shaped by radiologists' expertise and risk tolerance, can lead to overestimation of the cancer risk and many false-positive assessments (Alberdi et al. 2011). Its broad guidance is particularly problematic for category 4, which recommends biopsy across a very wide risk range (2–98%). Most BI-RADS 4 findings (70–80%) are later found to be benign after biopsy (Liu et al. 2024), underscoring the prevalence of unnecessary biopsies and the need for more personalized follow-up strategies.

The trade-off between the benefits and risks of mammography screening is a nuanced issue. While more frequent screening improves early detection and reduces mortality, it also raises the likelihood of false positives and unnecessary follow-up procedures. Public guidelines typically specify the recommended starting and ending ages and the screening frequency (annual, biennial, or triennial). As shown in Table 2, these guidelines differ notably across programs, especially in the recommended starting age. Since the launch of public screening programs in the 1980s, several organizations have revised their recommendations. For example, the U.S. Preventive Services Task Force (USPSTF) lowered the starting age from 50 to 40 in 2023 (USPSTF 2024). However, disagreement persists among public healthcare organizations regarding when to begin and end screening and how often to screen. For instance, unlike most U.S. programs, Australia's national BreastScreen program does not actively invite women aged under 50, and the Canadian Task Force on Preventive Health Care (CTFPHC) recommends against starting screening at age 40 (Department of Health and Aged Care 2025, CTFPHC 2024).

Supporters of earlier screening argue that breast cancer is more aggressive in younger women (Jayasinghe et al. 2005). Tumor doubling times illustrate this pattern, with a median of 80 days for women under 50, compared with 157 days for ages 50–70 and 188 days for women over 70 (Peer et al. 1993, Michaelson et al. 1999). Critics, however, emphasize the higher false-positive rates in younger women due to denser breast tissue and the need for more cautious guidelines (Kerlikowske et al. 2000). Faster tumor progression in younger women may also increase the likelihood of interval cancers, reducing screening effectiveness. This debate has received substantial media coverage (NYTimes 2024, CBC 2024), including ethical discussions that highlight the tension between saving lives and exposing many to unnecessary stress, anxiety, and medical procedures (Prevention 2016).

An emerging approach to this controversy is the personalization of screening and diagnostic policies, as numerous studies emphasize tailoring guidelines based on factors beyond age, such as demographics, family history, breast density, prior mammograms, and genetics (Gail and Rimer

1998, Ayer et al. 2012). Since breast cancer risk varies within the same age group, personalized policies can improve life-saving outcomes for high-risk women while reducing unnecessary screening and diagnostic procedures for low-risk women. Although prior research has advanced personalization in breast cancer screening, most studies treat personalized screening schedules and post-mammography diagnostic decisions as separate problems. In practice, however, these decisions are sequentially dependent: more frequent screenings can increase false positives and downstream diagnostics, while diagnostic outcomes influence subsequent screening choices and timing. We hypothesize that these decisions are inherently interconnected and that analyzing them in isolation may lead to suboptimal and less personalized screening pathways.

In this research, we aim to combine screening decisions and post-mammography diagnostic decisions into a unified framework for designing personalized policies to manage patients throughout the breast cancer screening pathway. We formulate this problem as a partially observable Markov decision process (POMDP), with two decision variables: (1) determining the optimal timing for mammography (screening decisions), and (2) selecting the appropriate follow-up examination method (diagnostic decisions). Building on this framework, we address the following core research questions:

- **RQ1:** How do screening frequency and post-mammography diagnostic choices interact dynamically within breast cancer screening pathway?
- **RQ2:** To what extent does combining screening and post-mammography diagnostic decisions improve clinical outcomes and resource efficiency for different risk subgroups?
- **RQ3:** What is the impact of integrating genetic testing into the breast cancer screening pathway on the reduction of unnecessary biopsies and overall healthcare costs?

The remainder of this proposal is organized as follows: In §2, we review the related literature, highlighting existing gaps and our contributions to this body of research. In §3 and §4, we present the problem definition and the POMDP formulation respectively.

2. Literature Review

Cancer screening and diagnostic decisions have been extensively studied in the Operations Research (OR) literature in the past 40 years (see Alagoz et al. (2011), and Ivy (2009) for a detailed review). Established screening programs are in place for five cancers: prostate, cervical, colorectal, lung, and breast cancer Alagoz et al. (2011). Among these, breast cancer screening received a lot of attention in OR community since it is relatively complex due to the variability in mammography

interpretation and the subsequent diagnostic follow-ups required, which introduces challenges in both clinical practice and modeling efforts. Studies within the breast cancer setting can broadly be categorized into population-level and personal-level modeling approaches. Population-level studies typically focus on optimizing screening policies for large cohorts based on average risk factors such as age (see Koleva-Kolarova et al. (2015) and Henderson et al. (2024) for comprehensive reviews of these studies). In contrast, personal-level studies emphasize tailoring screening and diagnostic decisions to individual characteristics beyond age, such as race, family history, breast density, gene mutations, and previous test results, offering a more individualized approach to cancer prevention.

2.1. Personal-Level Studies

In the Operations Research literature, two distinct research streams explore personalization in breast cancer setting. One stream focuses on optimizing personalized screening schedules for healthy women, without addressing diagnostic decision-making processes. The other stream concentrates on developing optimal policies for personalized post-mammography diagnostic decisions within regular public screening intervals, overlooking the potential for more personalized screening schedules.

2.1.1. Personalized Screening Decisions The seminal work by Ayer et al. (2012) was the first to incorporate both dynamic factors, such as previous screening results and family history, and static factors, like age and race, into a model that optimizes personalized screening schedules. Their proposed POMDP model generates personalized screening schedules that adapt over time and are not necessarily fixed to a regular pattern. For instance, a healthy woman who has undergone annual screening for five years with all negative results may be recommended to extend her next screening interval to over two years. These intervals continue to adjust based on changes in her health status. Building on this work, Ayer et al. (2015) addressed the same problem but incorporated non-perfect patient adherence behavior. While adherence behavior is widely studied in medical and behavioral science literature, few studies in healthcare operations management examine its impact. Ayer et al. (2015) showed that accounting for non-perfect adherence can significantly affect the optimal solution. In particular, they found that when screening strategies are optimized based on average adherence, the effect on patients with low adherence is relatively small, but patients with high adherence may be adversely affected. In both papers, mammography outcomes are defined only as positive or negative, disregarding the BI-RADS assessments. Additionally, if the mammogram is positive, the patient is assumed to proceed directly to biopsy, with no decision-making process regarding follow-up diagnostic examinations.

2.1.2. Personalized Post-Mammography Diagnostic Decisions In the second stream of studies on personalized post-mammography diagnostic decisions, Chhatwal et al. (2010) were the first to consider tailoring diagnostic decisions based on both mammographic and demographic features of patients. Their decision model considers biopsy as the only follow-up examination within an annual screening regime. They assume that all women undergo annual screenings, and after each mammogram, the decision-maker must choose between performing an immediate biopsy or waiting until the next screening. They propose a MDP formulation, deriving optimal dynamic policies based on age and cancer risk probabilities. However, their model's action set is binary (wait or biopsy), omitting other common diagnostic options such as short-term follow-up, ultrasound, or MRI.

Alagoz et al. (2013) extended Chhatwal et al. (2010)'s model by introducing an additional follow-up option of short-term diagnostic mammography. They derive dynamic, threshold-based policies based on age and cancer risk for choosing between three options after screening mammography. Tunç et al. (2022) further advanced this research by proposing a novel MDP model with the same set of diagnostic actions but a different approach to defining patient states. They defined the patient state using a tuple that specifies the risk of being in each cancer stage, resulting in a large-scale MDP. To manage the model's complexity, they applied dimension reduction techniques combined with a divide-and-search algorithm to determine feasible upper bounds on the optimal decision thresholds. Their findings demonstrate that incorporating cancer stages into diagnostic decisions can significantly reduce overdiagnosis. However, all these studies assume unlimited resources.

Ayvaci et al. (2012) is the only study in this stream that incorporates a cost-effectiveness analysis, considering budget constraints when determining optimal dynamic policies for diagnostic decisions. Their model derives dynamic, threshold-based policies based on age, cancer risk, and budget levels, specifying optimal choices between three possible options after screening mammography. Despite these advancements, all of these studies assume perfect adherence to follow-up recommendations and risk-neutral behavior on the part of patients.

The only study that accounts for patients' strategic behavior is Ayvaci et al. (2018), which considers the varying preferences of patients regarding health-related decisions. In their framework, different patient groups may have different valuations of the benefits and risks of follow-up diagnostic examinations, represented by distinct utility functions. They show that while incorporating patients' preferences into medical decision-making may result in a welfare loss at the population level in terms of survival duration, it can make care delivery more effective and potentially more

efficient. Across all of these studies, the models could be adapted to biennial or triennial screening intervals. However, they all assume fixed, non-personalized screening intervals.

2.2. Existing Gap

Screening and diagnostic decisions are inherently interconnected, and analyzing them separately may yield suboptimal policies for two main reasons. First, there is a sequential dependency between screening and diagnostic decisions. For instance, screening decisions can impact diagnostic outcomes; more frequent screenings may lead to an increased number of false positives, resulting in unnecessary biopsies or other diagnostic procedures. As noted by (Ayer et al. 2012), the risk of false-positive mammograms rises with more frequent screenings. Conversely, diagnostic decisions influence future screening choices, as the results of diagnostic tests, particularly when confirming a false positive, can be critical inputs for subsequent screenings. Additionally, following a biopsy, the presence of surgical scars may require adjustments to the screening schedule to allow for proper healing. Second, disjointed models for screening and diagnostic decisions can overlook the importance of feedback loops in sequential decision-making processes. These feedback loops are crucial in managing the underlying uncertainty in such problems. Separating the two decision processes could result in ignoring this feedback mechanism, which may lead to suboptimal outcomes. Moreover, without the feedback loop, diagnostic decisions might fail to fully account for the patient's evolving health status and the history of previous screening results, further contributing to less effective decision-making.

2.3. Our Contributions

From the application standpoint: (1) Rather than analyzing screening and diagnostic decisions in isolation, we propose a holistic framework that examines the interdependent effects of these decisions on each other. (2) Our model incorporates an additional diagnostic option, gene testing, that also requires follow-up actions. To the best of our knowledge, gene testing has not been studied extensively in cancer screening within the operations research literature, largely due to historically high costs and limited affordability. However, recent technological advances and increasing adoption in national screening programs have made it an important component of modern cancer prevention strategies (Thapa et al. 2025). Motivated by this shift, we include gene testing in our framework to assess both its operational impact and its potential cost-effectiveness in reducing unnecessary follow-ups and overall healthcare costs. By including this option, we aim to more accurately model the screening pathway, reflecting the broader range of diagnostic choices available in practice

beyond the commonly used options of short-term follow-up mammography, biopsy, and regular screening. This added flexibility allows us to capture more complex pathways following suspicious mammography results. (3) While much of the operations research (OR) literature focuses on maximizing Quality Adjusted Life Years (QALYs) as the objective function, we adopt a broader approach. Our objective function includes the number of false positives, false negatives, benign biopsies, as well as the costs associated with each testing method. By using this function, decision-makers can adjust their valuation of these objectives, recognizing that some are conflicting and require trade-offs. This allows for the creation of trade-off curves to inform screening and diagnostic decisions.

On the theoretical side, we propose a novel POMDP model that differs from conventional POMDPs in healthcare operations management. In our model, the time intervals before observing the outcomes of different types of decisions vary. For example, after choosing short-term follow-up mammography, the outcome is observed six months later, while the outcomes of a biopsy or gene test are typically available almost immediately. As a result, the model features different time periods, with transition epochs following each decision. This approach could be beneficial in modeling healthcare contexts where the timing of outcomes varies. For instance, in diabetes management, adjusting a patient's medication based on blood glucose levels may take days or weeks to show effects, while immediate interventions such as administering rapid-acting insulin can affect blood glucose levels almost immediately (within hours). Finally, we aim to analyze the structural characteristics of this model to derive simple threshold-based policies that are easy to understand and implement.

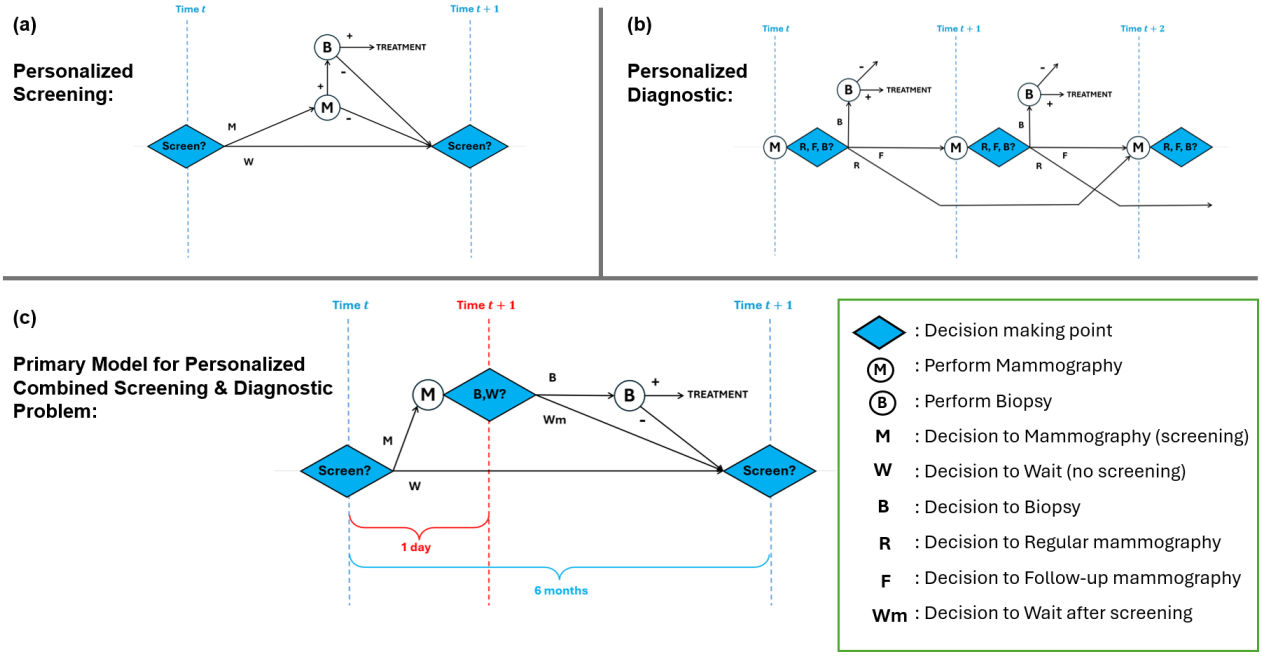
3. Problem Definition

3.1. Breast Cancer Screening Problem

As discussed in §1, public mammography screening programs typically recommend only the start and end ages for screening and the frequency of screenings. Women meeting the eligibility criteria, asymptomatic women within the prescribed age range and not having had a mammogram within the screening interval, receive invitation letters, and can undergo mammography without requiring a doctor's visit or referral. For high-risk women, consulting a physician is often recommended to determine the appropriate timing and method of screening.

In the context of personalized screening policies, a sequential decision-making process is employed (Ayer et al. 2012, 2015). Figure 1a shows a simplified process diagram for these models. In this process, at each decision point, the decision-maker must decide whether to refer an

Figure 1 Simplified Primary Models for Separate and Combined Screening and Diagnostic Problems



asymptomatic woman for a screening mammogram or wait until the next decision point, six months later. As seen in Figure 1a, these models assume binary outcomes for mammograms, with positive mammogram results leading to immediate biopsies. If the biopsy is positive, the patient exits the decision process and begins treatment; otherwise, the decision process continues. For the sake of simplicity, we omit death and self-diagnosis transitions from the diagrams of Figure 1. This process accounts for dynamic factors such as the most recent test results and family history, so the resulting schedules can be more flexible and adjusted in response to mammogram outcomes rather than public guidelines' fixed periodic intervals; for example, consecutive negative results may lead to longer future intervals between screenings.

3.2. Post-Mammography Diagnostic Problem

The BI-RADS system is a widely used framework for managing follow-up diagnostic examinations and assisting radiologists in their diagnostic decision-making. However, as discussed in §1, this system is often too generalized, requiring radiologists to rely on their own experience in many cases, which may introduce bias.

In personalized diagnostic decision models, a sequential decision-making process occurs after each regular screening mammogram (Chhatwal et al. 2010, Ayvaci et al. 2012, Alagoz et al. 2013, Tunç et al. 2022). Existing models for personalized post-mammography decisions typically consider three alternative decisions/actions after assessing a screening mammogram: (1) continue regular

screening (12 months later); (2) recommend short-term follow-up diagnostic mammography (6 months later); or (3) perform an immediate biopsy. Figure 1b presents a simplified process diagram for these models, in which patients exit the decision-making process after a biopsy, regardless of the result, and no death transitions are considered.

As per discussions with our clinical collaborator, a Medical Oncologist at the University of Calgary, the real-world practices are more complex than these models reflect. For example, follow-up diagnostic mammograms may be recommended within a 1- to 6-month window, rather than a fixed six-month period. Additionally, other diagnostic methods, such as breast-dedicated ultrasound and MRI, are frequently used to confirm abnormalities. Even after negative biopsy results, physicians may recommend further follow-up diagnostic tests to verify the negative findings. Moreover, in practice, post-diagnostic examinations often involve a concordance check, where new findings are compared with previous ones to ensure consistency. Discussions with the Cancer Centre at the University of Calgary also highlighted the expanding role of genetic testing in guiding screening pathways. Currently used for high-risk women, genetic testing is expected to become more affordable and accessible to broader sub-populations in the near future. Our model incorporates genetic testing as an option within the screening pathway, and we aim to conduct a cost-effectiveness analysis to assess its potential for reducing unnecessary biopsies and overall healthcare costs.

3.3. Primary Model for Combining Screening and Diagnostic Problems

When considering both screening and diagnostic problems together, a key challenge arises in defining decision epochs, necessitating a different modeling approach from the existing literature. Figure 1c presents a simplified version of our proposed combined screening and diagnostic problem, where, in addition to the screening decision, only two types of diagnostic actions are available: (1) regular screening and; (2) biopsy. At the start of each six-month epoch, the decision-maker must choose between performing a mammogram or waiting until the next six-month period. If the model decides to screen, after assessing the mammogram results and the patient's individual characteristics, it must then decide whether to wait until the next six-month epoch or proceed with a biopsy. If the biopsy result is positive, the patient exits the decision process and begins treatment; otherwise, the decision process continues. The key structural difference between the disjointed and combined models is as follows: As shown in Figure 1a and Figure 1b, existing disjointed models typically assume six-month decision epochs with only one decision point in each epoch. However, as illustrated in Figure 1c, in the combined problem, two decision points occur within each six-month interval. Combining screening and diagnostic decisions introduces an additional type of decision

period, lasting approximately one day, to account for the time required to process mammogram or other test results. Thus, if a screening mammogram is performed, a new decision point emerges the following day. This complexity necessitates defining two types of decision epochs with different lengths, leading to a fundamentally different modeling approach. In essence, nested decisions are required, where certain actions trigger a subsequent decision-making process with distinct decision options.

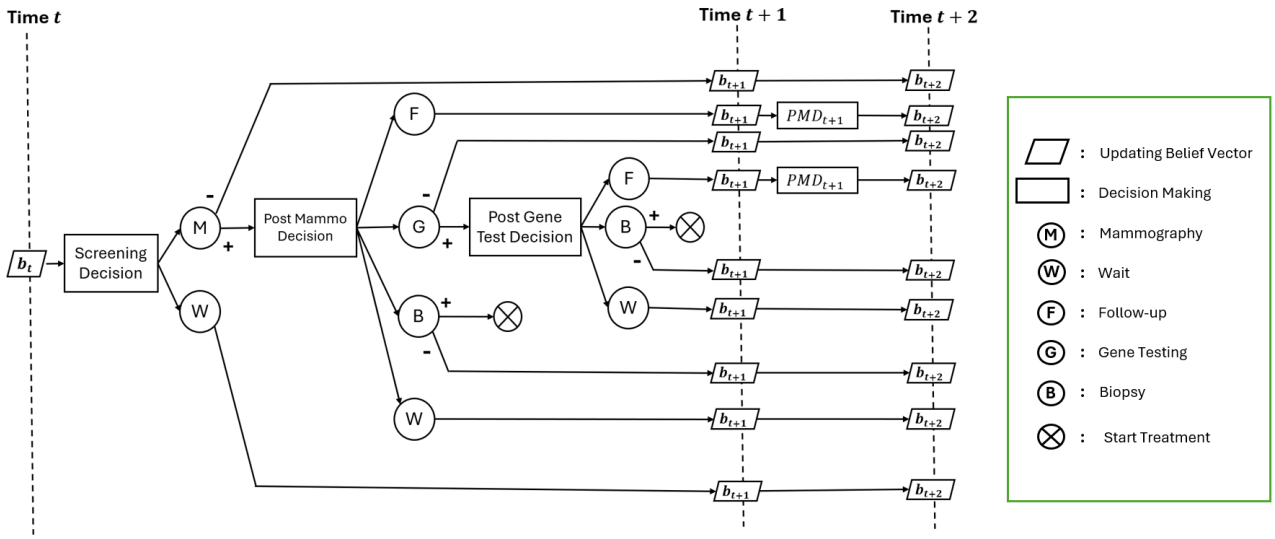
4. Model Formulation

We formulate the combined screening and diagnostic problem as a discrete-time, finite-horizon partially observable Markov decision process (POMDP) model, where a single decision-maker determines the timing of screenings and the method of diagnosis for a woman before breast cancer is confirmed. The decision-maker's objective is to maximize the patient's total expected weighted reward, assuming an unconstrained budget. The weights assigned to the rewards are exogenously determined to account for varying preferences in the trade-off between benefits and risks. We define two types of decision epochs: *screening epoch*s, which occur every six months, and *diagnostic epoch*s, which take place one day after performing a mammogram or diagnostic test. In addition to the decision to do nothing or wait until the next screening epoch (**W**), the considered screening and diagnostic tests include mammography (**M**), follow-up diagnostic mammography (**F**), gene testing (**G**), and biopsy (**B**).

4.1. Decision Process Description

At the beginning of each *screening epoch*, we update our belief about the patient's cancer risk, represented as a probability distribution over the patient's true health state. Next, we decide whether to screen the patient or not (**M, W?**). If we choose not to screen (**W**), the patient progresses to the epoch ($t+2$), meaning no action is taken for 12 months. If we decide to screen (**M**), we must then determine the post-mammography action (**F, G, B, W?**). If follow-up screening (**F**) is chosen, the patient moves to the next epoch ($t+1$) for a repeat mammogram. If a biopsy (**B**) is performed and the result is positive, the patient enters treatment, and the process stops; if negative, the patient moves to the epoch ($t+2$). Choosing the wait option (**W**) also sends the patient to epoch ($t+2$). If gene testing (**G**) is chosen, we must decide on a post-gene testing action (**F, B, W?**). The consequences of **F**, **B**, and **W** at this stage are similar to the previous decision point. There are only two absorbing states that terminate the process: death and a positive biopsy result. Figure 2 represents the detailed process diagram of our POMDP model.

Figure 2 Process Diagram of our POMDP Model



4.2. POMDP Formulation

In this section, first, we will explain our novel approach in defining the nested decisions or actions of our POMDP model, then we will explain the notation and components of our POMDP model followed by the optimality equations.

Our approach in defining the nested decisions: To integrate screening and diagnostic decisions, we define three nested layers of decision-making processes within each six-month epoch, as shown in Figure 2. In each epoch, we may engage in one layer of decision-making (if we choose not to screen), two layers (a screening decision followed by a post-mammography decision, if we decide to screen but not proceed with gene testing), or three layers (a screening decision, a post-mammography decision, and a post-gene testing decision, if we decide to screen and then proceed with gene testing).

Time steps: We first define fixed six-month decision epochs, referred to as *screening epochs*. However, the time step index k is also incremented whenever a mammogram, gene test, or biopsy is performed. In other words, time step increments whenever screening or diagnostic decision-making processes occur.

- $k \in \mathcal{K} = \{0, 1, 2, \ldots, N\}$ where $N < \infty$.

States: We define the states as the patient's breast cancer risk, rounded to the nearest integer between 0 and 100. In addition to this, we include death (D) and post-biopsy (PB) as the two absorbing states. Let $\mathcal{S}$ represents the state space, and s_k the patient's state at time step k, then:

- $s_k \in \mathcal{S} = \{0, 1, 2, \ldots, 100, D, PB\}$

Actions: We define a feasible action set for each decision-making layer. Let $\mathcal{A}$ be the main action set, and a^{SC}, a^{PM}, and a^{PG} represent the feasible action sets for the three decision-making layers: screening, post-mammography, and post-gene testing, respectively. Then, we define:

- $a^{SC} = \{W^{SC},\ M^{SC}\}$
- $a^{PM} = \{W^{PM},\ F^{PM}, G^{PM}, B^{PM}\}$
- $a^{PG} = \{W^{PG},\ F^{PG}, B^{PG}\}$
- $\mathcal{A} = a^{SC} \cup a^{PM} \cup a^{PG}$

Observations: We define observations as the outcomes of screening and diagnostic tests. Each testing method has its own possible observation set. For mammography and gene testing, we assume that a cancer risk, a number between 0 and 100, can be assigned to the patient based on the test results, while biopsy results are classified as either positive or negative. Let O represent the overall observation set, and o^M, o^G, and o^Brepresent the observation sets for mammography, gene testing, and biopsy, respectively. Therefore, we define:

- $o^M = o^G = \{0, 1, 2, \ldots, 100\}$
- $o^B = \{B+, B-\}$
- $O = o^M \cup o^G \cup o^B$

Core Transition Probabilities: Let $\mathcal{P}$ be the core transition function. This function determines the true state of the system in the next epoch, i.e. $\mathcal{P} = P_k^{(a,o)}(s'|s)$, that is the probability of going to state s' at time $k+1$ when the current state is s, action a is taken, and observation o is seen in time k. In our setting, o matters since we have some deterministic test results (for biopsy) and subsequent transitions.

Observation Probabilities: Let $\mathcal{Q} = P_k^a(o|s)$ be the observation probabilities function, i.e. the probability of observing outcome o of test a at time k when the true state is s. This function basically determines our predictions of test results that are based on the tests' specificity and sensitivity. We consider indices k for the accuracy of test to define them as age-dependent. For instance,

- $P_k^M\left(o^M = 0 \middle| s = 0\right) = spec_k(M)$
- $P_k^M\left(o^M = cancerous \middle| s = 0\right) = 1 - spec_k(M)$

where $spec_k(M)$ is specificity of mammogram for the patient in time k.

Reward Functions: We define our total reward function based on six items and their associated costs: (1) the number of mammogram false-positives (f), representing the anxiety cost; (2) the number of mammogram false-negatives (d), representing the cost of a one-period delay in cancer detection; (3) the number of benign biopsy outcomes (u), representing the cost of unnecessary

painful follow-up examinations; (4) the number of performed mammograms (m), representing the immediate cost (disutility) of each mammogram; (5) the number of performed biopsies (b), representing the immediate cost (disutility) of each biopsy; and (6) the number of performed gene tests (g), representing the immediate cost (disutility) of each gene test. Let $r_k(s,a,o)$ be the immediate reward in time k when the patient's true state is s, we choose action a, and observation o is seen, and let the $r_k(s,a)$ be the total reward for being is state s and choosing action a, then we will have:

- $r_k(s,a,o) = \begin{cases} f & \text{when true state is 0 but the test result is not 0 (false-positive).} \\ d & \text{when true state cancerous but we do nothing (delay in detection).} \\ u & \text{when that true state is 0 but we perform biopsy (benign result).} \\ m & \text{every time we perform mammography (disutility of mammography).} \\ b & \text{every time we perform biopsy (disutility of biopsy).} \\ g & \text{every time we perform gene testing (disutility of gene testing).} \end{cases}$
- $r_k(s,a) = \sum_{o \in O} P_k^a(o|s) r(s,a,o)$

The following are some examples of reward function values in specific scenarios:

- $r_k\left(0, M, o^M = \{cancerous\}\right) = f + m$: occurs when the true state is cancer-free, we decide to screen, and the mammogram result shows high cancer risk (false-positive mammogram).
- $r_k(0, B, B-) = u + b$: occurs when the true state is cancer-free, we perform a biopsy, and the result is benign.
- $r_k\left(s \in \{cancerous\}, W^{SC}, .\right) = d$: occurs when the true state is cancerous, but after screening we decide to wait until next epoch, regardless of the mammography result.
- $r_k\left(s \in \{cancerous\}, M, o^M = 0\right) = d + m$: occurs when the true state is cancerous, we decide to screen, and the mammogram result shows nothing (false-negative mammogram).
- $r_k(s \in \{cancerous\}, B, B-) = d + b$: occurs when the true state is cancerous, we perform biopsy, and the biopsy result is benign (false-negative biopsy).

Belief vector: Let b_k be our belief in time k about the true state of the patient, which is basically a probability distribution over all possible states of the patient.

Updated belief vector: Let b'_k be our expectation about the next true state of the system when in time k we chose action a and saw observation o, which is calculated using the Bayes rule based on the current belief vector and observation in time k as follows.

- $b'_k = \frac{\sum_S b(s) P_k^a(o|s) P_k^{a,o}(s'|s)}{\sum_S b(s) P_k^a(o|s)}$

4.3. Optimality Equations

We now present the optimality equations for the simplified primary case discussed in §3.3, where instead of three layers, we have only two layers of decision-making; screening (W, M?), and diagnostic (Wm,B?). Let $V_t^*(b)$ and $V_t^a(b)$ be the maximum total expected reward in time k, when our current belief vector is b, and when we choose action a, respectively.

$V_k^*(b) = \max\{V_k^W(b), V_k^M(b)\}$, where

$V_k^W(b) = \sum_{s\in S} b(s)\ [r_k(s,W,o) + \sum_{s'\in S} P_k^{(W,o)}(s'|s) V_{k+1}^*(b')]$, and

$V_k^M(b) = \sum_{s\in S} b(s) \sum_{o\in O_M} P_k^M(o|s)\ r_k(s,M,o) + \max\{V_{k+1}^B(b), V_{k+1}^{Wm}(b)\}$, where

$V_{k+1}^B(b) = \sum_{s\in S} b(s) \sum_{o\in O_B} P_k^B(o|s)\ [r_{k+1}(s,B,o) + \sum_{s'\in S} P_{k+1}^{(B,o)}(s'|s) V_{k+2}^*(b')], and$

$V_{k+1}^{Wm}(b) = \sum_{s\in S} b(s)\ [r_{k+1}(s,Wm,o) + \sum_{s'\in S} P_{k+1}^{(Wm,o)}(s'|s) V_{k+2}^*(b')]$

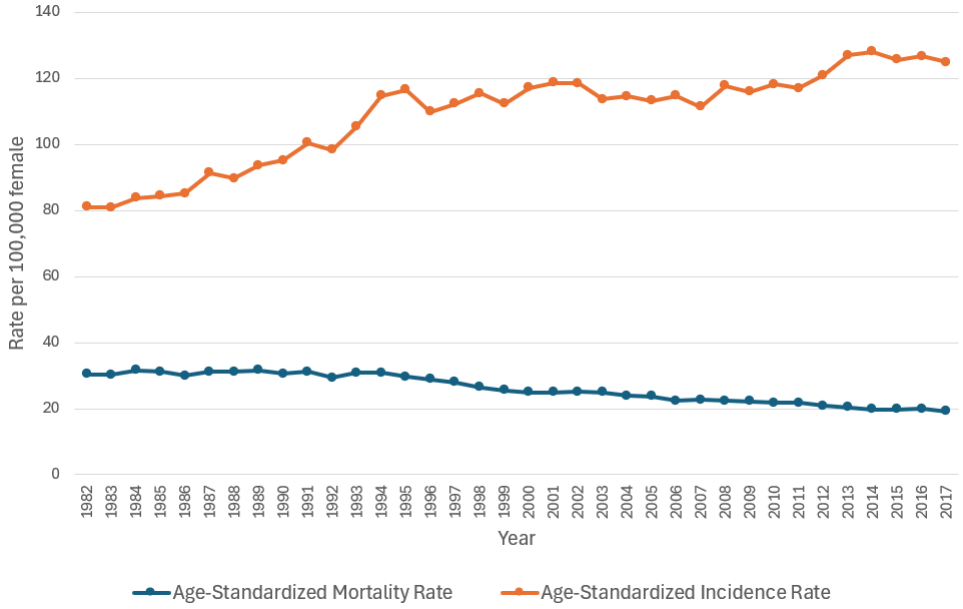
5. Results and Discussion

6. Conclusion

Notes

Appendix A: Illustration of Breast Cancer Incidence and Mortality Rates in Australia

Figure 3 **Age-standardized incidence and mortality rate for breast cancer (female), Australia, 1982 to 2017.**



Note. Data from Australian Institute of Health and Welfare (AIHW 2024).

Appendix B: BI-RADS Table

Assessment	**Likelihood of Cancer (%)**	**Diagnostic Recommendation**
Category 0: Incomplete	N/A	Recall
Category 1: Negative	Essentially 0	Routine screening
Category 2: Benign	Essentially 0	Routine screening
Category 3: Probably Benign	≥ 0 but ≤ 2	Short-interval (6-month) follow-up
Category 4: Suspicious	> 2 but < 95	Biopsy
Category 5: Highly Suggestive of Malignancy	≥ 95	Biopsy
Category 6: Known Biopsy-Proven Malignancy	N/A	Surgical extraction

Table 1 BI-RADS Categories, Likelihood of Cancer, and Diagnostic Recommendations

Appendix C: Breast Cancer Screening Guidelines

Program	Age Group	Guideline
Alberta Breast Cancer Screening Program (ABCSP) October 2023	40-44	Referral needed
	45-74	Every 2 years
	75+	Referral needed
Canadian Task Force on Preventive Health Care (CTFPHC) May 2024	40-49	Not recommended
	50-74	Every 2 or 3 years
	75+	Referral needed
BreastScreen Australia Program 2025	40-49	Have option for every 2 years
	50-74	Every 2 years
	75+	Can continue every 2 years if they are healthy
American Cancer Society December 2023	40-44	Have option for every year
	45-54	Every year
	55+	Every 2 years as long as healthy
United States Preventive Services Task Force (USPSTF) April 2024	40-74	Every 2 years
	75+	Insufficient evidence to assess benefits to harms
American College of Physicians (ACP) 2019	40-49	Referral needed
	50-74	Every 2 years
	75+	Discontinue

Table 2 Breast Cancer Screening Guidelines

References

AIHW (2024) Australian Institute of Health and Welfare Report: Cancer data in Australia. *Web Report* Available at: https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/about (Accessed: November, 2025).

Alagoz O, Ayer T, Erenay F (2011) Operations research models for cancer screening. Cochran J, Cox L, Keskinocak J, Kharoufeh J, Smith J, eds., *Wiley Encyclopedia of Operations Research and Management Science* (ohn Wiley Sons).

Alagoz O, Chhatwal J, Burnside ES (2013) Optimal policies for reducing unnecessary follow-up mammography exams in breast cancer diagnosis. *Decision Analysis* 10(3):200–224.

Alberdi R, Llanes A, Ortega R (2011) Effect of radiologist experience on the risk of false-positive results in breast cancer screening programs. *Eur Radiol* 21:2083–2090.

Ayer T, Alagoz O, Stout NK, Burnside ES (2015) Heterogeneity in women's adherence and its role in optimal breast cancer screening policies. *Management Science* 65(5):1339–1362.

Ayer T, Oguzhan A, Natasha K S (2012) Or forum—a pomdp approach to personalize mammography screening decisions. *Operations Research* 60(5):1019–1034.

Ayvaci M, Alagoz O, Ahsen M, Burnside E (2018) Preference-sensitive management of post-mammography decisions in breast cancer diagnosis. *Production and Operations Management* 27(12):2313–2338.

Ayvaci MU, Alagoz O, Burnside ES (2012) The effect of budgetary restrictions on breast cancer diagnostic decisions. *Manufacturing Service Operations Management* 14(4):600–617.

Brewer NT, T S, SE L (2007) Systematic review: The long term effects of false-positive mammograms. *Ann. Internal Medicine* 146(7):502–510.

CBC (2024) Advocates hope alberta will follow new guidelines, lower breast cancer screening age to 40. *CBC* Available at: https://www.cbc.ca/lite/story/1.7206704 (Accessed: September 13th, 2024).

CCS (2023) Cancer statistics 2023. *Canadian Cancer Society, Canadian Cancer Statistics Advisory Committee, Statistics Canada, and the Public Health Agency of Canada* Available at: https://cancer.ca/en/research/cancer-statistics (Accessed: September 13th, 2024).

Chhatwal J, Alagoz O, Burnside ES (2010) Optimal breast biopsy decision-making based on mammographic features and demographic factors. *Operations Research* 58(6):1577–1591.

Corsetti V, Houssami N, Ferrari A, Ghirardi M, Bellarosa S, Angelini O, Bani C, Sardo G Pand Remida, Galligioni E, Ciatto S (2008) Breast screening with ultrasound in women with mammography-negative dense breasts: evidence on incremental cancer detection and false positives, and associated cost. *European journal of cancer* 44(4):539–544.

CTFPHC (2024) The draft recommendations for breast cancer screening (may 30, 2024). *Canadian Task Force on Preventive Health Care* Available at: https://canadiantaskforce.ca/guidelines/published-guidelines/breast-cancer-update-2024/ (Accessed: September 13th, 2024).

Department of Health and Aged Care (2025) BreastScreen Australia program. *Australian Government* Available at: https://www.health.gov.au/our-work/breastscreen-australia-program (Accessed: November, 2025).

Duffy S (2020) Effect of mammographic screening from age 40 years on breast cancer mortality (uk age trial): Final results of a randomised, controlled trial. *The Lancet Oncology* 21(9):1165–1172.

Eby PR, Ghate S, Hooley R (2022) The benefits of early detection: Evidence from modern international mammography service screening programs. *Journal of Breast Imaging* 4(4):346–356.

Ellison LF, Saint-Jacques N (2023) Five-year cancer survival by stage at diagnosis in canada. *Health reports* 34(1):3–15.

Gail M, Rimer B (1998) Risk-based recommendations for mammographic screening for women in their forties. *Clinical Oncology* 16(9):3105–3114.

Helvie M, Hendrick R, Banerjee M (2014) Reduction in late-stage breast cancer incidence in the mammography era: implications for overdiagnosis of invasive cancer. *Cancer* 120(14):2649–2656.

Henderson J, Webber E, Weyrich M, Miller M, Melnikow J (2024) Screening for breast cancer: Evidence report and systematic review for the us preventive services task force. *JAMA* 331(22):1931–1946.

Ivy JS (2009) Can we do better? optimization models for breast cancer screening. Romeijn HE, Pardalos PM, eds., *Handbook of Optimization in Medicine*, chapter 2, 25–52 (Springer).

Jayasinghe UW, Taylor R, Boyages J (2005) Is age at diagnosis an independent prognostic factor for survival following breast cancer? *ANZ J. Surgery* 75(9):762–767.

Kerlikowske K, Carney PA, Geller B, Mandelson M, Taplin S, Malvin K, Ernster V (2000) Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Ann. Internal Medicine* 133(11):855–863.

Koleva-Kolarova R, Zhan Z, Greuter M, Feenstra T, De Bock G (2015) Simulation models in population breast cancer screening: A systematic review. *Breast* 24(4):354–363.

Liu C, Sun M, Arefan D, Margarita Z, Sumkin J, Wu S (2024) Deep learning of mammogram images to reduce unnecessary breast biopsies: a preliminary study. *Breast Cancer Res* 26(82).

Løberg M, Lousdal M, Bretthauer M (2015) Benefits and harms of mammography screening. *Breast Cancer Res* 63(17).

Magny S, Shikhman R, Keppke A (2023) Breast imaging reporting and data system. *StatPearls [internet]* Available at: https://www.ncbi.nlm.nih.gov/books/NBK459169/ (Accessed: September 13th, 2024).

Marmot M, Altman D, Cameron D, Dewar J, Thompson S, Wilcox M (2013) The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 108(11):2205–2240.

Michaelson J, Halpern E, Kopans D (1999) Abreast cancer: Computer simulation method for estimating optimal intervals for screening. *Radiology* 212:551–560.

NYTimes (2024) In reversal, expert panel recommends breast cancer screening at 40. *New York Times* Available at: https://www.nytimes.com/2024/04/30/health/breast-cancer-mammography-ultrasound.html?searchResultPosition=4 (Accessed: September 13th, 2024).

Peer P, van Dijck J, Hendricks J, Holland R, Verbeek A (1993) Age-dependent growth of primary breast cancer. *Cancer* 71:3547–3551.

Pocobelli G, Weiss N (2015) Breast cancer mortality in relation to receipt of screening mammography: a case–control study in saskatchewan, canada. *Cancer Cause Control* 26:231–237.

Prevention (2016) Why getting a mammogram may cause more trouble than it's worth. *Prevention.com* Available at: https://www.prevention.com/health/a20466779/the-trouble-with-getting-a-mammogram-at-40 (Accessed: September 13th, 2024).

Tabár L (2018) Swedish two-county trial: Impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 286(3):808–822.

Thapa S, Bonde M, Leppin A (2025) Implementing interventions to increase genetic testing for breast cancer among high-risk populations: A systematic review of implementation strategies, outcomes, and gaps. *Genet Med* 27(11):101573.

Tunç S, Alagoz O, Burnside ES (2022) A new perspective on breast cancer diagnostic guidelines to reduce overdiagnosis. *Production and Operations Management* 31:2361–2378.

USPSTF (2024) Screening for breast cancer: Us preventive services task force recommendation statement 2024. *JAMA* 332(22):1918–1930.