Manufacturing & Service Operations Management

The Effect of Budgetary Restrictions on Breast Cancer Diagnostic Decisions

Mehmet U. S. Ayvaci, Oguzhan Alagoz, Elizabeth S. Burnside,

Please scroll down for article—it is on subsequent pages

The Effect of Budgetary Restrictions
on Breast Cancer Diagnostic Decisions

Mehmet U. S. Ayvaci, Oguzhan Alagoz

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

Elizabeth S. Burnside

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

We develop a finite-horizon discrete-time constrained Markov decision process (MDP) to model diagnostic decisions after mammography where we maximize the total expected quality-adjusted life years (QALYs) of a patient under resource constraints. We use clinical data to estimate the parameters of the MDP model and solve it as a mixed-integer program. By repeating optimization for a sequence of budget levels, we calculate incremental cost-effectiveness ratios attributable to consecutive levels of funding and compare actual clinical practice with optimal decisions. We prove that the optimal value function is concave in the allocated budget. Comparing to actual clinical practice, using optimal thresholds for decision making may result in approximately 22% cost savings without sacrificing QALYs. Our analysis indicates short-term follow-ups are the immediate target for elimination when budget becomes a concern. Policy change is more drastic in the older age group with the increasing budget, yet the gains in total expected QALYs related to larger budgets are predominantly seen in younger women along with modest gains for older women.

Key words: Markov decision processes; linear programming; mixed-integer programming; constrained MDPs; breast cancer; diagnostic decisions; cost-effectiveness; medical decision making; mammography; service operations

History: Received: March 28, 2011; accepted: November 6, 2011. Published online in *Articles in Advance* April 13, 2012.

1. Introduction

With more than 1.1 million diagnoses and 410,000 deaths worldwide every year, breast cancer is the most frequently diagnosed cancer (23% of female cancers) and the leading cause of cancer mortality among women (14% of female cancer deaths) (Parkin et al. 2005). One in eight American women are expected to develop breast cancer in their lifetime.

Although breast cancer can be fatal at late stages, it has a relatively favorable prognosis when compared to other types of cancer. Survival rate is estimated to average 73% in developed countries and is reported to be approximately 90% in the United States (Jemal et al. 2007). Mortality rates have been decreasing since the 1990s largely because of medical advances, which allow for early diagnosis and better treatment (Berry et al. 2005). When detected before the development of symptoms, the disease can be cured. If diagnosed at an early stage (stage I), the five-year survival rate is 100%, whereas, in the more advanced metastatic stage, the five-year survival rate drops to 21% (Hayat et al. 2007).

Various organizations including the American Cancer Society (ACS) recommend mammography as the primary means of detecting breast cancer. ACS also recommends that women should undergo clinical breast examination triennially between ages 20 and 40 and annually after age 40. For average-risk women, the principal method for detecting preclinical breast cancer is screening mammography. This is recommended on an annual basis after age 40 and continues as long as the patient is a candidate for basic breast cancer treatment (Smith et al. 2010) and it has been shown to significantly reduce breast cancer mortality (Nelson et al. 2009). Additional benefits, including monetary savings due to early detection and reassurance in the absence of breast cancer, make screening mammography attractive from the societal perspective (Greif 2010). For high-risk women, ultrasound and magnetic resonance imaging have recently been considered a complement to mammography screening. However, there is limited evidence for these technologies to justify their use in low-risk populations (Smith et al. 2003). Although the use of digital imaging techniques and computer-aided detection as well as tomosynthesis (3-D mammography) and positron emission tomography (PET) have some potential for advancement in detection and diagnosis of breast cancer, mammography remains the primary modality of choice.

Following screening mammography, radiologists assess the risk/probability of cancer associated with the findings on a subjective scale and decide on the appropriate management as they weigh the trade-off between early detection of cancer and false positives. Based on the patient's risk of cancer, radiologists typically choose one of the following three options: (1) routine follow-up mammography in a year; (2) short-term follow-up mammography (in six months); or (3) immediate diagnostic actions including biopsy (i.e., pathological examination of the breast tissue obtained by needle biopsy or surgical excision).

Making diagnostic decisions is not a trivial task as the radiologist attempts to interpret the extensive amount of information collected from the mammography. Inevitably, the quality and interpretation of mammography varies with radiologist's skills and experience (Brown et al. 1995). Several studies report substantial variability among radiologists' interpretation of mammography images and diagnostic decisions (Elmore et al. 1994, Beam et al. 1996). One study investigated disparity among final assessment of mammography for the patients with cancer and reported 75% agreement between readings by the same radiologist over a period and only 65% agreement between readings by different radiologists (Kerlikowske et al. 1998). Later, Berg et al. (2000) reviewed diagnostic management decisions and found that only 55% were in agreement. A more recent study

on a large cohort of radiologists that interpreted more than one million mammography examinations found the rate of false positive assessments ranges from 1.7% to 24.7% and emphasized the significance of interpretation variability and the need for improvements in performance (Elmore et al. 2009).

The American College of Radiology (ACR) has developed a format called Breast Imaging Reporting and Data System (BI-RADS) lexicon (ACR 1998) to standardize mammography interpretation and diagnostic decisions. BI-RADS lexicon includes 43 hierarchical features describing lesions that appear on mammograms. Radiologists record their findings using this lexicon, assign a final assessment category, and recommend a diagnostic action (as presented in Figure 1). Although standardization of mammography reporting using BI-RADS lexicon greatly enhances the decision-making process, it is not sufficient for reducing interpretation variability.

As a solution, decision-aid systems called computer-aided diagnosis (CADx) models can be used (Ayer et al. 2010a). In particular, CADx models use the 43 BI-RADS features collected from a mammogram as the input and produce a numerical estimate of breast cancer risk, i.e., the probability of cancer, which helps radiologists better parse the extensive amount of information. Various CADx models such as artificial neural networks (Baker et al. 1995, Ayer et al. 2010b) logistic regression (Chhatwal et al. 2009), Bayesian networks

Figure 1   Radiologists' Diagnostic Decisions Using BI-RADS Assessment Categories
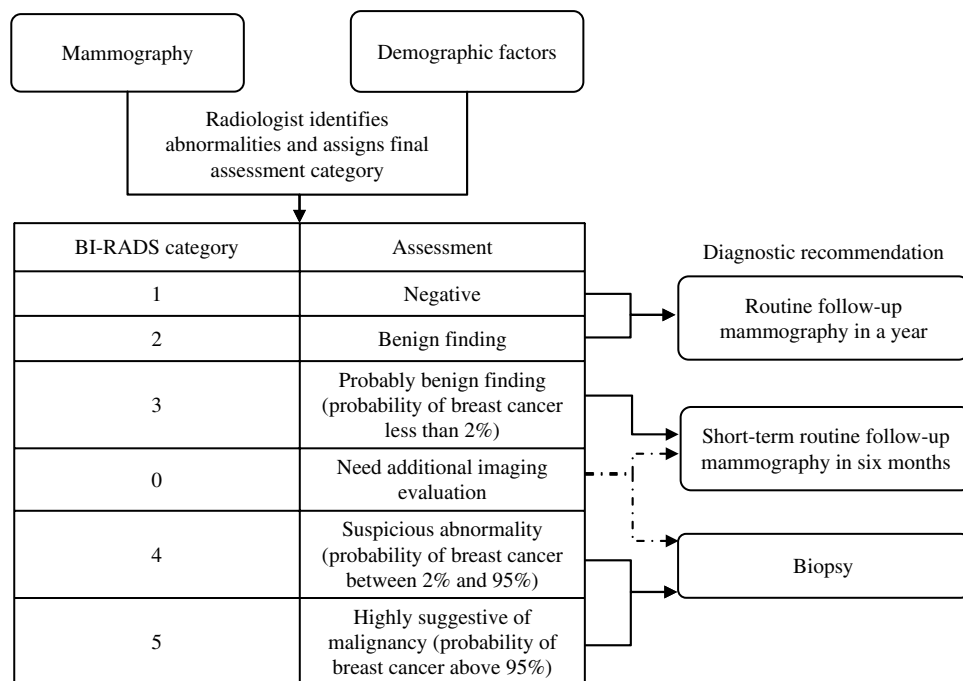
Notes. The dashed lines stemming from BI-RADS category 0 reflect the inconsistency in the corresponding management decisions (Geller et al. 2002). BI-RADS category 0 should be an intermediary assessment prompting additional imaging such as diagnostic mammography and may result in any of the other BI-RADS categories.

(BN) (Burnside et al. 2009), and many others suggest improvements in discrimination of breast cancer. Consequently, a more important decision has to be made by the radiologist for which CADx models do not offer a solution: managing the disease through diagnostic actions while taking into consideration the evolution of cancer risk and other patient-specific factors. This task is not easy as further evidenced by the following statistics on false positives and biopsies with benign outcomes.

Census data and biopsy utilization rates among women aged between 40 and 75 translates to 400,000 biopsies per year in the United States (U.S. Census Bureau 2000, Ghosh et al. 2005). On average, only 32.6% of the biopsies reveal cancer. The proportion of women with abnormal screening examinations and who were eventually diagnosed with cancer is 4.8% (Rosenberg et al. 2006). The percentage of abnormal exams and the number of surgical biopsies that do not reveal cancer are twice as frequent in the United States than in the United Kingdom despite similar cancer detection rates (Smith-Bindman et al. 2003). Breast lesions, which may not be diagnosed and/or treated over the lifetime of a woman, referred to as overdiagnosis, comprise 25% of the mammographically detected breast cancers (Welch and Black 2010). False positives and overdiagnosis may cause additional morbidity and anxiety without influencing mortality (Ernster et al. 2000). In relation to our study, the associated expenditures add a significant amount on the total cost of screening mammography as described below.

1.1. Role of Cost in Diagnostic Decisions

Studying costs as an outcome in the healthcare domain is difficult because it equates the value of life with a monetary value. Still, the rising costs and the mounting economic pressure make it important to understand the trade-offs related to various levels of funding so that optimal decisions can be made. For an organized screening procedure such as mammography, several alternatives are suggested for cost containment. Reducing the unit costs and discontinuing the inappropriate medical services are immediate ways to cut healthcare expenditures (Breen and Brown 1994). For mammography screening, once the starting age, the frequency of screening, and the high-risk groups to be screened are identified, the opportunity to alleviate the cost burden of mammography lies in making better diagnostic decisions. Moreover, it is equally important to determine the impact of budget changes on diagnostic decisions and consequently the health outcomes. Identifying ways to make better diagnostic decisions at reduced cost and

assessing the consequences in terms of life expectancy are important for both the policy makers and healthcare providers.

Aggregate cost of mammography screening in the United States is estimated to be $3 billion to $5 billion dollars (Burnside et al. 2001). There are many factors that contribute to this estimate; about 20% are due to diagnostic procedures following an abnormal mammogram (Poplack et al. 2005). After the cost of a mammogram, the total cost is most sensitive to recall rate (percentage of screening mammography needing further workup), cost of a recall, and cancer to biopsy ratio (percentage of biopsies revealing cancer) (Burnside et al. 2001). Failures in diagnostic decisions in the form of false positives and overdiagnoses are not rare as discussed above and therefore are important determinants of the total cost of mammography screening. One study reports that for every $100 dollars spent on screening, an additional cost of $33 dollars is incurred because of false positives (Elmore et al. 1998).

When we consider breast cancer diagnosis from a population perspective, it is evident from the aforementioned statistics that both the economic and quality of life implications should be weighed. That is, the government programs or insurance companies or any other third party who are partially or fully paying for the healthcare services would prefer to avoid unwarranted costs while maximizing benefits to the patients undergoing mammography screening. This problem poses a trade-off commonly observed in medical decisions: When human health is the issue, care providers are primarily concerned with the patient's health. In the world of limited resources though, policy makers seek the best strategies to ensure patient health while allocating resources in the best possible way.

Most of the existing optimization studies that consider cost as part of the modeling framework either optimize cost by ignoring health outcomes (or treatment effectiveness) or convert all costs into a measure of health outcome (or vice versa) using a value for willingness-to-pay. In other words, these studies optimize a single objective. Both approaches are often difficult to justify from an ethical standpoint when the objective is to optimize the patient's well-being and, both have been subject to controversy (Neumann and Weinstein 2010). The U.S. Panel on Cost-Effectiveness in Health and Medicine recommended quality-adjusted life years (QALYs) as the sole benchmark for health outcomes (Weinstein et al. 1996). More recent legislation called the Patient Protection and Affordable Care Act (ACA) prohibited the development or use of value for willingness-to-pay thresholds (Neumann and Weinstein 2010). In the following, we explore the sensitive trade-off between

public interest and patient's well being complying with ACA enactment.

1.2. Problem Definition

We present a novel framework for evaluating the cost-effectiveness of diagnostic procedures in the context of breast cancer diagnostic decisions after mammography. As part of this framework, we maximize the total expected QALYs of a patient under resource constraints. Specifically, we develop a finite-horizon discrete-time constrained Markov decision process (MDP) to model the problem i.e., given mammography features, what is the optimal course of action: routine screening, short-term follow-up or biopsy under budgetary restrictions?

We use clinical data to estimate the parameters of the MDP model and solve it as a mixed-integer program (MIP) subject to budget restrictions. By repeating the optimization of the value function for various budget levels in a feasible range, we conduct incremental cost-effectiveness analysis and compare actual clinical practice with optimal decisions obtained from our model. The steps in our study are detailed in §1.3.

Our contribution in this paper is five-fold. First, we provide a mathematical modeling scheme for a medical decision-making problem under budget constraints. Our work is unique in jointly examining health outcomes and economical consequences of diagnostic decisions after mammography, while solely optimizing the health outcomes. To the best of our knowledge, this scheme is the first in treatment optimization literature to consider lifetime expected costs while maximizing a patient's total expected QALYs via a constrained finite-horizon MDP model. In addition, MIP has not been used in formulating finite-horizon constrained MDPs in the past which makes our modeling technique unique.

Second, our framework has a potential for use in other healthcare settings. Because medical treatment decisions are often sequential and uncertain, MDP models are proposed as an appropriate tool for modeling these decisions (Schaefer et al. 2004). There are many successful applications of MDPs in the clinical domain ranging from dialysis therapy management (Lee et al. 2008) to organ transplantation (Sandıkçı et al. 2008) where life years or QALYs are maximized. These applications of sequential decision making under uncertainty could be further extended to consider funding of such services, and hence similar cost-effectiveness analyses can be conducted. For example, in the case of organ transplantation, an important issue for resource-constrained countries is to decide on the best timing of the transplant to maximize the total expected QALYs of the patient where life support could be costly and resources are limited. Our framework would optimally make these

sequential decisions for various budget levels. In addition, our modeling framework could be useful in certain healthcare settings involving real cash-limited budgets. A good example is controlling the spread of sexually transmitted diseases (STDs) including HIV, gonorrhea, genital chlamydia infections, syphilis, and chancroid in developing countries. A number of primary and secondary prevention strategies exist such as prevention through advertising condom use, prevention of mother-to-infant transmission, prevention of transmission by blood transmission, and treatment of curable STDs. Allocation of resources to these strategies and optimal control would be based on the costs and effects of the relevant options (Over and Piot 1996). Therefore, our framework could prove useful in such a setting where the optimal strategy should vary according to disease epidemiology and access to resources.

In fact, the applicability of our framework could go beyond the monetary terms. In medical decision-making problems, decision makers often face constraints due to patient-specific factors, existing guidelines, or other resource restrictions. For instance, restrictions on dose and frequency of treatments, availability of clinical services, patient-specific treatment restrictions, effectiveness requirements in probabilistic terms, and other medico-legal or policy requirements can be considered constraints to such expected life maximization problems. Our framework can serve as a venue for reaching clinically sensible decisions under various forms of constraints.

Third, the computational approach for incremental cost-effectiveness by iteratively solving for optimal diagnostic decisions for various budget levels is a novel approach in cost-effectiveness literature. Conventional cost-effectiveness studies are usually static and are implemented on fixed performance measures such as averages. We develop a methodology where policy makers could measure the effectiveness while accounting for optimal courses of action and changing risk of disease. This flexibility enables decision making at two levels: finding optimal actions given the budget restrictions or finding the best resource allocation given the patient subgroups.

Fourth, it is possible to identify cost savings without compromising patients' well-being by using our methodology. This has implications for both developing and developed countries. When different types of cancer are compared in terms of new cases (incidence), deaths (mortality), and number of people living with cancer within five years of diagnosis (prevalence), the cancer profile may or may not differ for developed and developing countries. For cancer types where poor outcomes are more likely (i.e., lung, liver), the differences are diminished. For cancer types where benefits

from diagnosis and treatment are likely to materialize, however, the differences between incidence, mortality, and prevalence for developed and developing nations are more apparent. Breast cancer, from the perspective of optimal diagnostic decisions under budgetary restrictions, falls under the latter group where the results would apply to both developed and developing countries, albeit in different ways. In developed countries, flexible budgets will let optimal diagnosis be the emphasis whereas in developing countries, the emphasis will shift toward cost concerns while doing the best possible for the patient.

Finally, our study is the first in breast cancer literature to investigate optimal breast cancer diagnostic decisions under resource constraints. Previous operations research literature examined questions pertaining to breast cancer either from a screening or diagnostics perspective. In the case of breast cancer screening, the questions of when to begin and cease screening, and how often to perform screening are investigated (Ayer et al. 2012, Maillart et al. 2008). In the case of diagnostic decisions after mammography, however, Chhatwal et al. 2010 and Alagoz et al. (2010) are the only studies that address optimal biopsy decisions based on mammographic features and demographic factors. Both studies consider the breast cancer diagnostic decision-making problem from the patient's perspective yet ignore the costs of diagnosis in making this decision.

1.3. Modeling Framework

Our modeling framework consists of four steps in addition to a preliminary step as depicted in Figure 2. (1) Initially, we develop a finite-horizon discrete-time MDP to model the radiologist's diagnostic decision problem. That is, given the patient descriptors and mammographic findings, what is the optimal course of action? (2) We formulate this finite-horizon MDP model using a linear program (LP). (3) Then, we add the budget constraint. It is well known that the LP formulation of an MDP with extra constraints does not generate an optimal deterministic policy (Puterman 1994). Instead, such an LP model leads to an optimal randomized policy (i.e., choosing an action with a certain probability at each state). Because this randomized policy is not clinically intuitive, we introduce binary decision variables and solve the problem as an MIP to obtain clinically sensible optimal policies. (4) Finally, we conduct a cost-effectiveness analysis using a feasible range of budget restrictions. We use clinical data to estimate parameters as described in §4, which comprises Step 0 of our modeling framework. An important component of our work is related to the use of LP and MIP for solving our MDP model.

There are two main approaches for solving constrained MDPs. The first is the Lagrangian approach

Figure 2   Modeling Framework
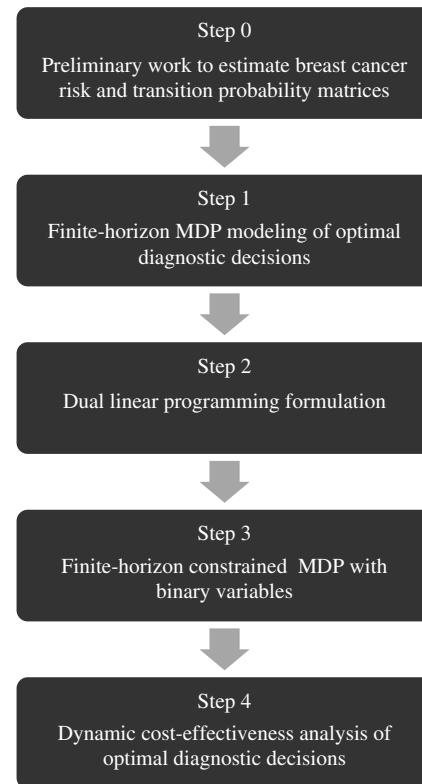
as introduced by Beutler and Ross (1985, 1986). However, this approach lacks explicit computational tools (Altman 1999). The second is the LP approach that was first introduced by Manne (1960) as a solution method for infinite-horizon Markov decision models. Although dynamic programming is conceptually simpler when solving finite-horizon problems, complexities such as the existence of constraints make the LP a more feasible option. Derman and Klein (1965) adapted LP formulation of stochastic shortest route problems to finite-horizon problems showing that LP is a viable option for solving finite-horizon constrained MDPs. Kallenberg (1981) proposed an approach based on occupation measures, which is a computationally efficient way of solving constrained MDPs with discounted or total cost criteria (Altman 1999). We adapt Kallenberg's modeling scheme as outlined in Kallenberg (1994). However, we follow a new approach and include binary decision variables to obtain deterministic optimal policies. This is required because of the idiosyncrasies of decision making in the health domain.

The remainder of this paper is organized as follows. In §2, we formulate the breast cancer diagnostic decisions as a finite-horizon MDP. In §3, we give the primal and dual LP formulations of the MDP model. We extend this model to include constraints, which could be applied in different settings. Using state-action frequency interpretation at each decision epoch,

we incorporate expected cost as a constraint into the problem formulation and present the complete formulation as MIP to obtain deterministic policies. Section 4 describes the estimation of model parameters followed by numerical experiments in §5. We finalize with concluding remarks in §6.

2. MDP Formulation

We use a discrete-time finite-horizon MDP with non-stationary rewards and transition probabilities to formulate the breast cancer diagnostic decision problem. We refer to the radiologist as the decision maker who undertakes diagnostic decisions with the patient. We assume that both the radiologist and the patient are risk neutral, and that the patient strictly adheres to the radiologist's recommendations.

The decision process in our model proceeds as follows: When a patient undergoes mammography screening, the radiologist estimates the probability of cancer for this patient based on risk factors and the mammographic findings using her subjective assessment or a computer-aided diagnostic model as described in §4. Then, based on this probability of cancer estimate, the radiologist makes one of the following three recommendations: (1) she can send the patient for biopsy immediately to confirm whether the patient has cancer; (2) she can recommend the patient to undergo another mammography within the next six months; (3) she can ask the patient to come back one year later for routine mammography. The patient accumulates rewards represented by the total life expectancy minus disutilities of diagnostic decisions associated with each action. Without loss of generality, we consider only the disutilities of diagnostic actions and ignore the treatment-related or age-related utilities to consider the trade-off between diagnosis-related utilities and life expectancy in isolation. The decision process ends either when the patient develops cancer or if she dies. We present a depiction of the decision process in Figure 3. Note that, although in this section we provide the MDP formulation when routine screening considers annual mammography, our modeling framework can easily be extended to consider other screening frequencies (such as biennial or triennial) by slightly changing the transition probabilities, reward function, and optimality equations. The formal definition of the MDP model is given below: Let $k \in K = \{0, 1, 2, \ldots, N\}$ represent the decision epochs where $N < \infty$.

States: Let $x_k \in \mathcal{X}$ denote the state of the patient at time k, where $\mathcal{X}$ is the state space. We define $\mathcal{X} = \{0, 1, 2, \ldots, 100, D, PC\} = \{s_k, D, PC\}$, where $s_k \in \{0, 1, \ldots, 100\}$ is the risk score at time k when the patient undergoes mammography, and D and PC represent the death state and the postcancer state following biopsy, respectively. Note that s_k represents the
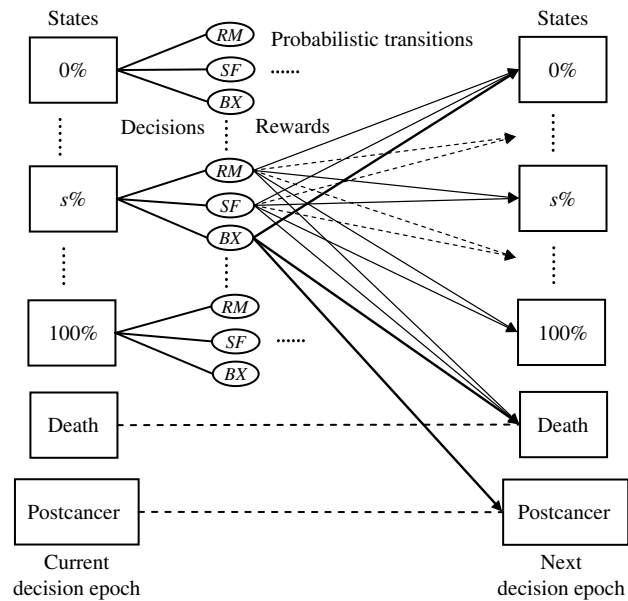
Note. The rectangles represent the states, ovals represent the decisions, solid lines represent choices, and arrows represent the probabilistic transitions.

discretized probability of breast cancer before diagnosis rounded to closest integer between 0 and 100. One can estimate the probability of breast cancer as described in §4.

Actions: The radiologist chooses action $a_{x_k}(k) \in A_{x_k}(k) \subset A = \{$routine mammography ($RM$), short-term follow-up (SF), biopsy (BX)$\}$ when the state is x_k at time k. Table 1 shows the action space corresponding to state-time pairs of the MDP model.

Transition Probabilities: If the selected action is BX and the biopsy has a malignant outcome, the patient starts treatment and quits the process. If the selected action is SF or RM, the patient undergoes another mammography after six months (i.e., one decision epoch) or after one year (i.e., two decision epochs), respectively. Death and postcancer states are absorbing and no decision is associated with these states.

Let $p_k^{SF}(x_{k+1} \mid x_k)$ denote the probability that the patient will be in state $x_{k+1} \in \mathcal{X}$ at time $k+1$ and let $p_k^{SF}(D \mid x_k)$ be the probability of death before the decision epoch $k+1$, given that the patient's state is x_k and the action is SF at time k. Then, $p_k^{SF}(D \mid x_k) = \pi(x_k) \cdot \{p_k^{ct}(D)\} + \{1 - \pi(x_k)\} \cdot \{p_k^{cf}(D)\}$ for $x_k \in \mathcal{X} \setminus \{D, PC\}$,

Table 1 **State and Action Space of the MDP Model**

State (x_k)	Action space
$s_k, k \in \{K \setminus \{N, N-1\}\}$	RM, SF, BX
$s_k, k = N-1$	SF, BX
$s_k, k = N$	—
D, PC	—

where $\pi(x_k) = x_k/100$ represents the probability of cancer at time k when the current state is x_k, $p_k^{cf}(D)$ represents the probability of death during decision epoch k when the patient is cancer free, and $p_k^{ct}(D)$ represents the probability of death during decision epoch t when the patient has breast cancer, but has not started receiving treatment (i.e., when the patient is unaware of the disease). We define $p_k^{SF}(PC \mid x_k) = 0$ for all $x_k \in \mathscr{X} \setminus \{D, PC\}$.

When the patient is in state $x_k \in \mathscr{X} \setminus \{D, PC\}$ and RM is chosen at time k, the probability that she will be in state $x_{k+2} \in \mathscr{X}$ at time $k + 2$ is given by $p_k^{RM}(x_{k+2} \mid x_k)$ which is defined as follows:

$$p_k^{RM}(x_{k+2} \mid x_k) = \sum_{x_{k+1} \in \mathscr{X}} p_k^{SF}(x_{k+1} \mid x_k) p_{k+1}^{SF}(x_{k+2} \mid x_{k+1}).$$

Let $p_k^{RM}(D \mid x_k)$ be the probability of death before decision epoch $k+2$, given that the patient's state is x_k and the action is RM at time k. Then,

$$p_k^{RM}(D \mid x_k) = p_k^{SF}(D \mid x_k) + [1 - p_k^{SF}(D \mid x_k)]$$
$$\cdot \sum_{x' \in \mathscr{X} \setminus \{D\}} p_k^{SF}(x' \mid x_k) p_{k+1}^{SF}(D \mid x').$$

We define $p_k^{RM}(x_{k+1} \mid x_k) = 0$ for all x_{k+1} and $p_k^{RM}(PC \mid x_k) = 0$ for all $x_k \in \mathscr{X} \setminus \{D, PC\}$.

When the patient is in state $x_k \in s_k$ and BX is chosen at time k, the state transitions occur according to $p_k^{SF}(x_{k+1} \mid 0)$ with probability $1 - \pi(x_k)$, the likelihood of a benign outcome. When the biopsy is positive, given the patient was in state x_k at time k, the patient moves to the absorbing state PC with probability $\pi(x_k)$, the likelihood of malignant outcome.

Reward Functions: Let $\tilde{q}_k(x_k, SF)$ represent the intermediate expected reward accrued between time k and $k + 1$ when the patient's state is x_k and action SF is selected at time k. We assume half decision interval length for the expected life in the case of death and full decision interval length in the case of no death. This is referred to as *half-cycle correction* and implies that the death could occur, on average, halfway through the cycle. Then, $\tilde{q}_k(x_k, SF)$ is calculated as the weighted average of expected life by weighing life years until the subsequent decision epoch according to the probability of death during the respective time frame.

The corresponding intermediate QALYs accrued, $q_k(x_k, SF)$, is then defined as follows:

$$q_k(x_k, SF) = \tilde{q}_k(x_k, SF) - \bar{u}^{SF},$$

where $\bar{u}^{SF}$ represents the disutility associated with SF.

Similarly, let $\tilde{q}_k(x_k, RM)$ represent the intermediate expected reward accrued between time k and $k + 1$ when the patient's state is x_k and RM action is selected at time k.

The corresponding intermediate QALYs accrued, $q_k(x_k, RM)$, is then defined as follows:

$$q_k(x_k, RM) = \tilde{q}_k(x_k, RM) - \bar{u}^{RM},$$

where $\bar{u}^{RM}$ represents the disutility associated with RM.

The rewards associated with action BX depend on the outcome of the biopsy. Namely, if the biopsy's outcome is negative (benign), then the patient moves to state 0. The intermediate expected reward when the biopsy reveals negative, $q_k(x_k, BX, 0)$, is calculated as follows:

$q_k(x_k, BX, 0)$
$$= 0.5 \cdot P(\text{alive in the current decision}$$
$$\text{epoch given benign outcome})$$

$$+ 0.25 \cdot P(\text{death in the current decision}$$
$$\text{epoch given benign outcome})$$

$$q_k(x_k, BX, 0) = 0.5 \cdot \left\{ \sum_{x'=0}^{S} p_k^{SF}(x' \mid 0) \right\} + 0.25 \cdot \{ p_k^{SF}(D \mid 0) \}.$$

If the biopsy's outcome is positive (malignant), the patient is absorbed into the state PC and the associated intermediate expected reward is represented by $q_k(x_k, BX, PC)$. Because the postcancer state is absorbing, the total expected postbiopsy reward represents the expected remaining life under treatment for a given age, which is defined as a lump sum. The total expected postcancer life when the patient's state is x_k at time k, $\tilde{q}_k(x_k, BX)$, is then defined as follows:

$$\tilde{q}_k(x_k, BX) = (1 - \pi(x_k)) \cdot q_k(x_k, BX, 0)$$
$$+ \pi(x_k) \cdot q_k(x_k, BX, PC).$$

The corresponding intermediate expected QALYs accrued, $q_k(x_k, BX)$, is then defined as follows:

$$q_k(x_k, BX) = \tilde{q}_k(x_k, BX) - \bar{u}_k^{BX},$$

where $\bar{u}_k^{BX}$ represents age-dependent disutility of biopsy. If the patient is in the postcancer state or death state, the patient does not accrue any additional rewards. Terminal reward at period $k = N$ is represented by q_N. We depict the mechanics of the rewards process in Figure 4.

Optimality Equations: We use $\gamma \in [0, 1]$ as the discounting factor for the rewards over six months. Let Q^π be the total expected discounted QALYs that can be attained under policy π, i.e.,

$$Q^\pi := E_\alpha^\pi \left[\sum_{k=0}^{N-1} \gamma^k q_k(x_k, a_{x_k}(k)) + \gamma^N q_N \right],$$

where α is the initial distribution for x_0 such that $\alpha(j)$ satisfies $\sum_{j \in s_0} \alpha(j) = 1$, $\alpha(D) = \alpha(PC) = 0$, and $\alpha(j) \geq 0$

Figure 4   Rewards Process Associated With the Diagnostic Decisions in Our MDP Formulation
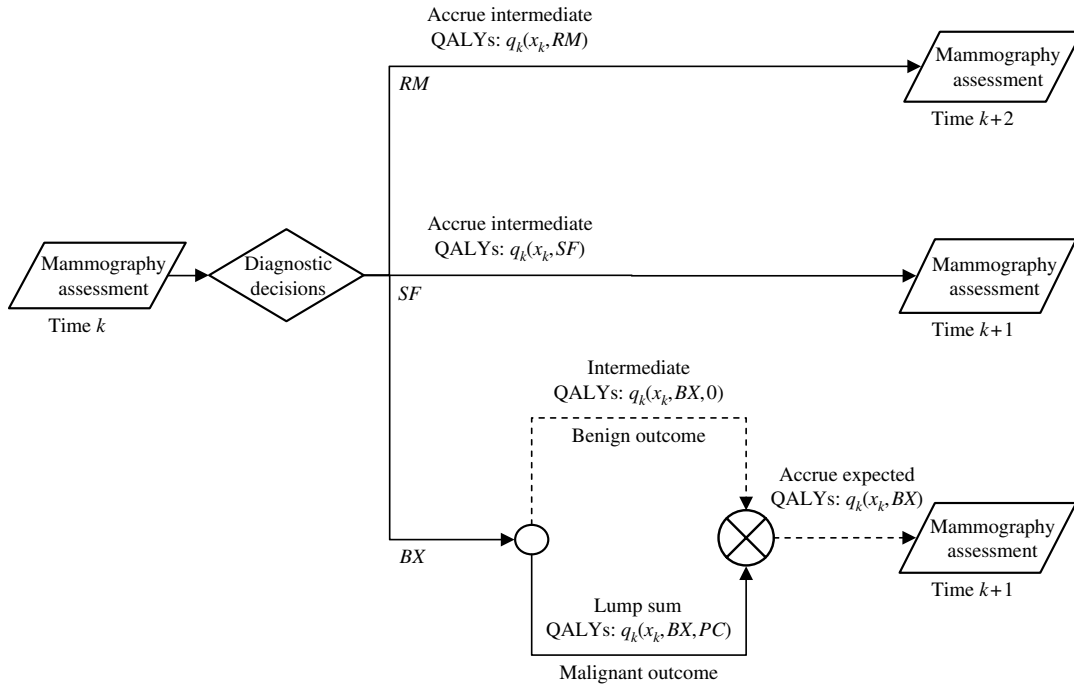
Notes. The dashed line represents the case that the mammography screening continues if the biopsy reveals a benign outcome. All of the rewards consider age-based mortality due to cancer and other causes. For routine mammography and short-term follow-up decisions, respective intermediate rewards are immediately accrued. For biopsy decision, however, accrued rewards are calculated as the expected value of intermediate QALYs corresponding to benign and malignant outcomes.

for $j \in \mathscr{X}$. Let $Q_k(x_k)$ be the maximum total expected reward that the patient can attain when her current state is x_k at time k. Given the discount rate γ, one can find the optimal policy corresponding to the expected total reward of a finite-horizon MDP with finite state space and finite action space by dynamically solving the Bellman optimality equations (Puterman 1994) given by

$$Q_k(s_k)$$
$$= \max\left\{ q_k(s_k, BX) + \gamma(1 - \pi(s_k)) \sum_{x' \in \mathscr{X}} p_k^{SF}(x' \mid 0) Q_{k+1}(x'), \right.$$
$$q_k(s_k, SF) + \gamma \sum_{x' \in \mathscr{X}} p_k^{SF}(x' \mid s_k) Q_{k+1}(x'), q_k(s_k, RM)$$
$$\left. + \gamma^2 \sum_{x' \in \mathscr{X}} p_k^{RM}(x' \mid s_k) Q_{k+2}(x') \right\},$$
$$k = 0, 1, \ldots, N-2; \quad (1)$$

$$Q_k(s_k) = \max\left\{ q_k(s_k, BX) + \gamma(1 - \pi(s_k)) \right.$$
$$\cdot \sum_{x' \in \mathscr{X}} p_k^{SF}(x' \mid 0) Q_{k+1}(x'), q_k(s_k, SF)$$
$$\left. + \gamma \sum_{x' \in \mathscr{X}} p_k^{SF}(x' \mid s_k) Q_{k+1}(x') \right\},$$
$$k = N-1; \quad (2)$$

$$Q_k(D) = Q_k(PC) = 0 \quad k = 0, 1, \ldots, N-1. \quad (3)$$

For $k = N$, we add a boundary condition as follows:

$$Q_N(D) = 0 \quad \text{and} \quad Q_N(x_N) = q_N,$$
$$\text{for all } x_N \in \mathscr{X} \setminus \{D\}. \quad (4)$$

The policy maximizing (1)–(4) is the optimal policy.

Now, assume that there is an additional cost constraint that limits the expected expenditures under policy π. Let $c_k(x_k, a_{x_k}(k))$ represent the total cost of decision $a_{x_k}(k)$ at time k and state x_k, and C^π represent the total expected discounted costs over N decision epochs, where

$$C^\pi := E_\alpha^\pi\left[\sum_{k=0}^{N-1} \lambda^k c_k(x_k, a_{x_k}(k)) \right]$$

is at most $C(\delta)$ and λ is the discount factor for cost. The constrained MDP model can be defined as follows:

$$Q^{\pi^*} := \sup_\pi \{Q^\pi \mid C^\pi \leq C(\delta)\}, \quad (5)$$

where Q^{π^*} is the optimal total expected QALYs. Once the constraint is introduced, Bellman equations are no longer useful for solving the formulation presented in Equation (5). We use LP and MIP formulations to solve the MDP problem as described in the next section.

3. MIP Formulation

Finite-Horizon Primal Linear Program (FPLP). If resource constraints are ignored, we can solve the radiologist's diagnostic decision problem using the following LP formulation:

$$Q^* = \min \sum_{j \in s_0} \alpha(j) Q_0(j)$$

$$\text{s.t.} \quad Q_k(x_k) - \gamma \sum_{j \in \mathscr{X}} p_k^{SF}(j \mid x_k) Q_{k+1}(j) \geq q_k(x_k, SF),$$
$$\forall x_k \in s_k, \ \forall k \in \{K \backslash N\}$$

$$Q_k(x_k) - \gamma(1 - \pi(x_k)) \sum_{j \in \mathscr{X}} p_k^{SF}(j \mid 0) Q_{k+1}(j)$$
$$- \gamma \pi(x_k) Q_{k+1}(PC) \geq q_k(x_k, BX),$$
$$\forall x_k \in s_k, \ \forall k \in \{K \backslash N\}$$

$$Q_k(x_k) - \gamma^2 \sum_{j \in \mathscr{X}} p_k^{RM}(j \mid x_k) Q_{k+2}(j) \geq q_k(x_k, RM),$$
$$\forall x_k \in s_k, \ \forall k \in \{K \backslash \{N, N-1\}\}$$

$$Q_k(j) - \gamma Q_{k+1}(j) \geq 0 \quad \forall j \in \{D, PC\},$$
$$\forall k \in K \backslash \{N\}$$

$$Q_N(x_N) \geq q_N, \quad \forall x_N \in s_N$$

$$Q_N(D), Q_N(PC) \geq 0,$$

$$Q_k(x_k) \text{ is free} \quad \forall x_k \in \mathscr{X}, \ \forall k \in K,$$

where Q^* is the optimal objective value of the decision model and $\alpha(j)$ satisfies $\sum_{j \in s_0} \alpha(j) = 1$, $\alpha(D) = \alpha(PC) = 0$, and $\alpha(j) \geq 0$ for $j \in \mathscr{X}$.

Again, ignoring budgetary constraints, we can solve the radiologist's diagnostic decision problem using the dual of the FPLP, called *Finite-Horizon Dual Linear Program* (FDLP), which has the following representation:

$$Q^* = \max \sum_{x_k \in s_k} \sum_{a_{x_k}(k) \in A_{x_k}(k)} \sum_{k=0}^{N-1} q_k(x_k, a_{x_k}(k)) u_k(x_k, a_{x_k}(k))$$
$$+ \sum_{x_N \in s_N} q_N u_N(x_N),$$

$$\text{s.t.} \sum_{a_j(0) \in A_j} u_0(j, a_j(0)) = \alpha(j), \quad \forall j \in s_0; \qquad (6)$$

$$u_0(j) = \alpha(j), \quad \forall j \in \{D, PC\}; \qquad (7)$$

$$\sum_{a_j(1) \in A_j} u_1(j, a_j(1))$$
$$- \gamma \sum_{x_0 \in s_0} \big[p_0^{SF}(j \mid x_0) u_0(x_0, SF) + (1 - \pi(x_0))$$
$$\cdot p_0^{SF}(j \mid 0) u_0(x_0, BX) \big] = 0, \quad \forall j \in s_1; \qquad (8)$$

$$u_1(D) - \gamma \bigg[u_0(D) + \sum_{x_0 \in s_0} \big[p_0^{SF}(D \mid x_0) u_0(x_0, SF)$$
$$+ (1 - \pi(x_0)) p_0^{SF}(D \mid 0) u_0(x_0, BX) \big] \bigg] = 0; \quad (9)$$

$$u_1(PC) - \gamma \bigg[u_0(PC) + \sum_{x_0 \in s_0} [\pi(x_0) u_0(x_0, BX)] \bigg] = 0; \qquad (10)$$

$$\sum_{a_j(k) \in A_j} u_k(j, a_j(k))$$
$$- \gamma \sum_{x_{k-1} \in s_{k-1}} \big[p_{k-1}^{SF}(j \mid x_{k-1}) u_{k-1}(x_{k-1}, SF)$$
$$+ (1 - \pi(x_{k-1})) p_{k-1}^{SF}(j \mid 0) u_{k-1}(x_{k-1}, BX) \big]$$
$$- \gamma^2 \sum_{x_{k-2} \in s_{k-2}} p_{k-2}^{RM}(j \mid x_{k-2}) u_{k-2}(x_{k-2}, RM) = 0,$$
$$\forall j \in s_k, \ \forall k \in \{K \backslash \{0, 1, N\}\}; \quad (11)$$

$$u_N(j) - \gamma \sum_{x_{N-1} \in s_{N-1}} \big[p_{N-1}^{SF}(j \mid x_{N-1}) u_{N-1}(x_{N-1}, SF)$$
$$+ (1 - \pi(x_{N-1})) p_{N-1}^{SF}(j \mid 0) u_{N-1}(x_{N-1}, BX) \big]$$
$$- \gamma^2 \sum_{x_{N-2} \in s_{N-2}} p_{N-2}^{RM}(j \mid x_{N-2}) u_{N-2}(x_{N-2}, RM) = 0,$$
$$\forall j \in s_k; \quad (12)$$

$$u_k(D) - \gamma [u_{k-1}(D)$$
$$+ \sum_{x_{k-1} \in s_{k-1}} \big[p_{k-1}^{SF}(D \mid x_{k-1}) u_{k-1}(x_{k-1}, SF)$$
$$+ (1 - \pi(x_{k-1})) p_{k-1}^{SF}(D \mid 0) u_{k-1}(x_{k-1}, BX) \big]]$$
$$- \gamma^2 \sum_{x_{k-2} \in s_{k-2}} p_{k-2}^{RM}(D \mid x_{k-2}) u_{k-2}(x_{k-2}, RM) = 0,$$
$$\forall k \in \{K \backslash \{0, 1\}\}; \quad (13)$$

$$u_k(PC) - \gamma \bigg[u_{k-1}(PC)$$
$$+ \sum_{x_{k-1} \in s_{k-1}} \pi(x_{k-1}) u_{k-1}(x_{k-1}, BX) \bigg] = 0,$$
$$\forall k \in \{K \backslash \{0, 1\}\}; \quad (14)$$

$$u_k(s_k, a_{s_k}(k)) \geq 0 \quad \forall s_k,$$
$$\forall k \in K \backslash \{N\}, \ \forall a_{s_k}(k) \in A_{s_k}(k); \quad (15)$$

$$u_k(x_k) \geq 0 \quad \forall x_k \in \{D, PC\}, \ \forall k \in K \backslash \{N\}; \quad (16)$$

$$u_N(x_N) \geq 0 \quad \forall x_N \in \mathscr{X}, \qquad (17)$$

where $\alpha(j)$ satisfies $\sum_{j \in s_0} \alpha(j) = 1$, $\alpha(D) = \alpha(PC) = 0$, and $\alpha(j) \geq 0$ for $j \in \mathscr{X}$.

Note that FDLP will produce nonnegative optimal $u_k(x_k, a_{x_k}(k))$ values such that for each x_k and $a_{x_k}(k)$ pair, there will be at most one $u_k(x_k, a_{x_k}(k))$ strictly positive among possible $a_{x_k}(k) \in A_{x_k}(k)$, which is a deterministic policy. One can interpret $u_k(x_k, a_{x_k}(k))$ as the probability/frequency of taking decision $a_{x_k}(k)$ when the system is in state x_k at time k (Derman and Klein 1965, Kallenberg 1981). More specifically, $u_k(x_k, a_{x_k}(k))$ is defined as the long-term average time, occupation measure, that the patient will be in state x_k

at time k and $a_{x_k}(k)$ is the optimal diagnostic decision for the radiologist. Using this definition, the equations in FDLP have an intuitive interpretation. The objective value is simply the total expected discounted rewards. The constraints ensure that each occupation measure for state x_k at time k is probabilistically connected to the occupation measures of the possible states at time $k+1$ or $k+2$.

Occupation measure allows one to add the constraints that use expectations for different criteria such as expected cost being less than a certain amount. Therefore, we introduce the following constraint to limit the expenditures associated with diagnostic decisions:

$$\sum_{x_k \in s_k} \sum_{a_{x_k}(k) \in A_{x_k}(k)} \sum_{k=0}^{N-1} \lambda^k c_k(x_k, a_{x_k}(k)) u_k(x_k, a_{x_k}(k)) \leq C(\delta),$$

(18)

where $C(\delta) = \delta$ is the total expected budget allocation for breast cancer diagnostic procedures during the lifetime of a woman. The equivalent LP formulation of the constrained MDP model of Equation (5) is given by Equations (6)–(18). Let $F(\delta)$ represent the feasible set of **u** induced by these equations as a function of δ, where $\mathbf{u} = [[u_k(s_k, a_{s_k}(k)), u_k(D), u_k(PC), u_N(x_N)]_{a_{s_k}(k) \in A_{s_k}(k), k \in K}]$ is the row vector of decision variables and $\mathbf{u}^T$ is its transpose with the same entries. More explicitly,

$$F(\delta) = \{\mathbf{u} \mid \mathbf{u} \text{ satisfies } (6)\text{–}(18)\}.$$

For any δ that produces a nonempty $F(\delta)$, we define

$$Q(\delta) = \max_{\mathbf{u} \in F(\delta)} \sum_{x_k \in s_k} \sum_{a_{x_k}(k) \in A_{x_k}(k)} \sum_{k \in K} q_k(x_k, a_{x_k}(k)) u_k(x_k, a_{x_k}(k))$$

$$+ \sum_{x_N \in \mathscr{X} \setminus D} q_N u_N(x_N),$$

which gives the optimal total expected QALYs as a function of δ.

Theorem 1 shows how the optimal value function changes with the allocated budget. As the available budget increases, the marginal gains in terms of the total expected QALYs diminish, therefore the optimal value function is concave in the allocated budget. Intuitively, this implies that as the available resources increase, the health outcomes improve, yet the magnitude of improvements in health outcomes diminish (see Figure 6(a) in §5).

THEOREM 1. *The optimal total QALYs, $Q(\delta)$, is a concave function of δ on the feasible set.*

Note that with the introduction of additional constraints, the model will no longer generate a deterministic optimal policy. Instead, an optimal policy will involve a randomized decision rule for a certain state x_k. In a clinical context, specifically breast cancer diagnostic decisions, randomized policies are not acceptable, because leaving a clinical decision to chance is not justifiable. To obtain a deterministic solution, we introduce binary decision variables and add the following set of constraints to the model:

$$u_k(x_k, a_{x_k}(k)) \leq M.b_k(x_k, a_{x_k}(k))$$

$$\forall x_k \in s_k, \ \forall k \in K, \ \forall a_{x_k}(k) \in A_{x_k}(k), \ (19)$$

$$\sum_{a_{x_k}(k) \in A_{x_k}(k)} b_k(x_k, a_{x_k}(k)) = 1 \quad \forall x_k \in s_k, \ \forall k \in K, \qquad (20)$$

$$b_k(x_k, a_{x_k}(k)) \in \{0, 1\}, \qquad (21)$$

where $b_k(x_k, a_{x_k}(k))$ are the binary logical variables to enforce deterministic policies. Equations (6)–(21) form an MIP called *Constrained Finite-Horizon Dual Mixed-Integer Program* (CFDMIP) modeling the diagnostic decisions following mammography under budgetary constraints. Let $G(\delta)$ represent the feasible set of **u** induced by these equations as a function of δ. More explicitly,

$$G(\delta) = \{\mathbf{u} \mid \mathbf{u} \text{ satisfies Equations } (6)\text{–}(21)\}.$$

For any δ that produces a nonempty $G(\delta)$, we define

$$QB(\delta) = \max_{\mathbf{u} \in G(\delta)} \sum_{x_k \in s_k} \sum_{a_{x_k}(k) \in A_{x_k}(k)} \sum_{k \in K} q_k(x_k, a_{x_k}(k)) u_k(x_k, a_{x_k}(k))$$

$$+ \sum_{x_N \in \mathscr{X} \setminus D} q_N u_N(x_N),$$

which gives the optimal total expected QALYs as a function of δ. Lemma 1 illustrates the functional form of $QB(\delta)$ as compared to its linear counterpart $Q(\delta)$.

LEMMA 1. *$QB(\delta)$ is a right continuous step function bounded by $Q(\delta)$.*

Lemma 2 shows that the optimal value function of the CFDMIP exhibits diminishing gains similar to Theorem 1. The intuitive explanation in the case of concavity holds such that as the resources are increased diminishing health outcomes are observed.

LEMMA 2. *The optimal total QALYs, $QB(\delta)$, of the CFDMIP model is subadditive in δ that are binding.*

All proofs are provided in the online appendix (available at http://dx.doi.org/10.1287/msom.1110.0371).

4. Estimation of Model Parameters

We solve the CFDMIP model parameters using clinical data. Our data set consists of 65,892 consecutive mammographic findings from 18,269 patients. Data were collected between April 5, 1999, and February 9, 2004,

from the Medical College of Wisconsin, Milwaukee (MCW). After excluding the duplicate, incomplete, and unmatching cases, we are left with 62,219 cases reported in the BI-RADS format. The outcomes are verified either through biopsy or through state cancer registries that track and record cancer cases across the United States. The state cancer registry allows disclosure of cancers missed during mammography examinations. To the best of our knowledge, our database is the largest in the nation with registry matching.

4.1. Estimating the Probability of Breast Cancer and Transition Probabilities

An important input for our MDP model is the probability of breast cancer, i.e., risk scores. Although we could use the assessment of radiologists as a proxy for perceived risk of breast cancer based on the mammographic findings, this is a difficult task and is prone to error as discussed in §1

We use the BN of Chhatwal et al. (2010) for probabilistic evaluation of a given mammogram instead of the likelihood of breast cancer as defined by BI-RADS assessment categories. BNs are mathematical modeling tools used for predicting the probability of an outcome based on observed variables. They are composed of two main parts: a dependence structure given by an acyclic graph and probabilistic relations specified by conditional probability tables. The model predicts the probability of breast cancer given the mammographic findings and the patient's demographic factors. We discretize the probability estimate by rounding to the nearest integer, therefore achieving a more granular risk scoring than the BI-RADS assessments.

We use the BN model to estimate the risk scores for each mammography report in MCW data, which is then used to estimate probability transition matrices of our MDP model. Because MCW data are composed of consecutive mammography reports collected over almost five years, many individual patients had multiple mammography reports. The risk scores associated with the mammography reports that belonged to the same patient and their evolution over time helped us to estimate a discrete probability distribution. Occasionally, patient visits did not occur on a regular basis. These irregularities were corrected by linear interpolation through time. We normalize the transition probabilities for each action at every age by using the age-specific mortality for patients with breast cancer (Jemal et al. 2007) and for cancer-free patients (life tables from the National Center for Health Statistics (NCHS) and Center for Disease Control and Prevention (CDC) Arias 2006).

4.2. Estimating the Rewards

We calculate the expected intermediate reward for each diagnostic decision using the half-cycle corrected

expected life until the next decision epoch following the action. We weight life years with respect to the breast cancer risk and age-dependent death probabilities. When biopsy reveals cancer, we calculate the lump-sum reward as the expected remaining life using the probability of death calculations as suggested in Arias (2006).

We account for quality of life reductions due to diagnostic decisions by subtracting disutilities associated with diagnostic decisions from expected lifetime. We assume moderate disutilities of 0.25 days for both routine mammography, $\bar{u}^{RM} = 0.25$, and for follow-up, $\bar{u}^{SF} = 0.25$ (Mandelblatt et al. 1992). We assume the disutility associated with biopsy at age 40 as two weeks (Gram et al. 1990). Because biopsy is associated with increased risk of complications in older ages, we assume that the associated disutility will increase as women get older. Therefore, we impose a linear function on the biopsy disutility over ages where the function attains twice as much the disutility value at age 100 when compared to the disutility at age 40.

4.3. Estimating the Cost Parameters

For cost parameters, we use expected costs of diagnostic decisions as suggested by Poplack et al. (2005), who base their estimation on Medicare fee schedules. Without loss of generality, we ignore the costs for routine screening mammography because routine screening is actually the do-nothing decision for cost-constrained biopsy decisions. We apply a 3% inflation rate to bring follow-up imaging and biopsy costs to 2010 dollars resulting in $362 and $1,258, respectively.

5. Numerical Study

In this section, we describe the implementation of the constrained MDP model of §2 by solving it using the CFDMIP model of §3; parameters as summarized in §4. We use CPLEX as the MIP solver through GAMS Version 23.3 when solving the proposed models. The computation time required for solving the CFDMIP model to the optimality ranged from around half an hour to a day excluding the time needed to construct the input. In our numerical experiments, we have two primary objectives: to demonstrate how optimal diagnostic decisions change under budget constraints and to conduct a cost-effectiveness analysis of diagnostic procedures in the context of a sequential decision-making problem. In addition, we compare the performance of our model to that of the MCW data and investigate the age-based differences in number of deaths, number of cancers detected, and life gains at various budget levels.

5.1. Optimal Diagnostic Decisions for Breast Cancer Under Budget Constraints

According to BI-RADS lexicon, findings with less than 2% probability of breast cancer, but somewhat higher
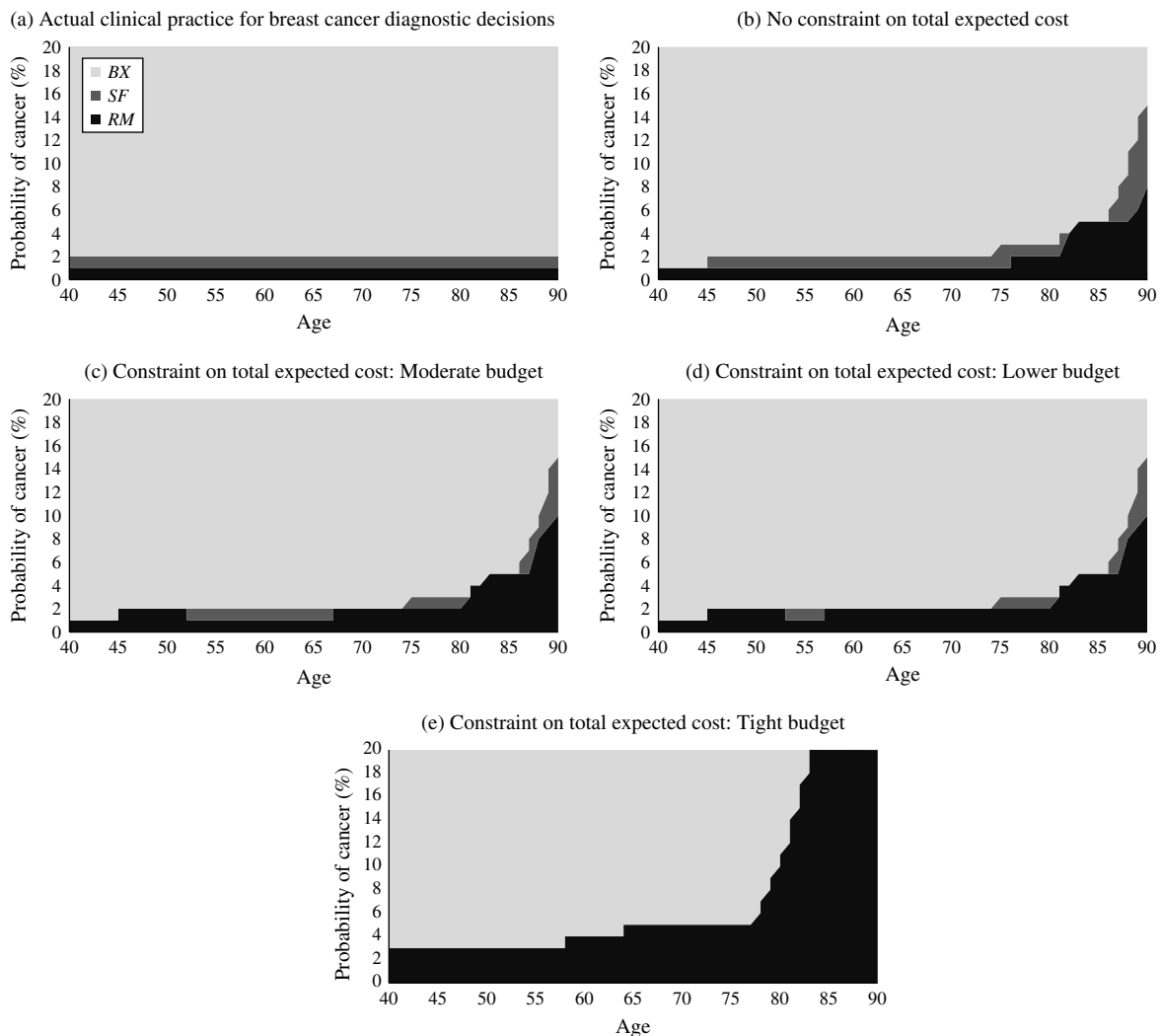
than the overall cancer detection rate (0.6%), are recommended follow-up; findings with more than 2% probability of cancer are recommended biopsy. Retrospective studies report an average of 1% probability of breast cancer as the baseline of mammographic recommendations for short-interval follow-up (Yasmeen et al. 2003, Vizcaíno et al. 2001). Therefore, we use 1% and 2% probability of breast cancer as the decision thresholds for follow-up and biopsy decisions to reflect clinical routine, respectively. Figure 5(a) depicts the current clinical practice tailored to these thresholds.

Figure 5(b) depicts the optimal diagnostic decisions when follow-up and routine mammography disutilities are 0.25 days under no budgetary constraints. The figure shows that for each age there exists a particular probability of cancer, namely, a threshold, over which the optimal decision is short-term follow-up, and a secondary threshold over which the optimal decision is biopsy. For example, for women between

45 and 73 years old the optimal diagnostic decision is biopsy, when the risk of breast cancer is above 2%. For the same age group, the optimal diagnostic decision is follow-up, when the risk is between 1% and 2%. Finally, the decision is to do nothing and wait for the next routine mammography when the risk is below 1%. For older women, these thresholds shift upward.

For women younger than 45, we observe that the optimal decision is to biopsy for probability of cancer over 1%; the follow-up decision is never used when there is no budget constraint. This suggests that, for women in this age range, the optimal policy is more aggressive in the sense that it suggests an invasive procedure, biopsy, rather than watchful waiting via short-term follow-up. This coincides with the medical intuition that in younger women, breast cancer is more aggressive (Anders et al. 2008) and therefore more aggressive diagnostic actions (i.e., biopsy) are justified. Moreover, revealing the existence or nonexistence of

Figure 5   Optimal Breast Cancer Diagnostic Decisions as a Function of Age and Probability of Breast Cancer at Different Budget Levels

(a) Actual clinical practice for breast cancer diagnostic decisions

(b) No constraint on total expected cost

(c) Constraint on total expected cost: Moderate budget

(d) Constraint on total expected cost: Lower budget

(e) Constraint on total expected cost: Tight budget

cancer via biopsy rather than close watch by follow-up decision could be sensible because of the longer life expectancy of this age group.

In general, radiologists are expected to evaluate mammograms independent of the age/other demographic factors and recommend diagnostic actions, namely, assess the BI-RADS category, based on the mammogram. Ideally, however, age could be considered in predicting the likelihood of breast cancer and should play a role when assessing the BI-RADS category. This is because the incidence of breast cancer changes by age (Hayat et al. 2007) and the trade-off between the benefits and harms of diagnosing and treating breast cancer depend on age (Jayasinghe et al. 2005). Optimal diagnostic decisions as depicted in Figure 5(b) partially reflect how experienced radiologists practice in general and prescribe, while Figure 5(a) represents the existing guidelines.

Our findings on age-based diagnostic recommendations warrant further discussion. There is evidence of systematic differences between diagnostic recommendations at different age groups (Taplin et al. 2002) with no reference to the outcomes following these recommendations. Although we are not aware of direct evidence that age-based recommendations produce better health outcomes, there are several facets of the issue that could indirectly shed light on our findings. First, there is no consensus in the medical community on the age at which we should start/end mammography screening and how often we should screen (Mandelblatt et al. 2009, Warner 2011). The controversy still exists after several decades since the first use of mammography for screening breast cancer (Marshall 2010). The decision to screen and its frequency are inherently related to diagnostic decisions following mammography and the debate suggests the variation in opinion. Furthermore, clinical practice shows evidence on age-based use of screening services (Bynum et al. 2005). Second, Jayasinghe et al. (2005) presented evidence on age-based differences in breast cancer prognosis. The findings indirectly imply that because the aggression of cancer differs by age, the aggressiveness with which we biopsy for a given cancer risk should also differ by age. Last, medical literature suggests that both the incidence of breast cancer (Hayat et al. 2007) and the sensitivity of mammography (Kerlikowske et al. 2000) vary by age. Therefore, the threshold to biopsy would indirectly be influenced by age because of the incidence rate and the sensitivity of mammography.

We compare the expected cost and the effects, i.e., the total expected QALYs resulting from following the BI-RADS recommendations to those of the optimal diagnostic decisions for a 40-year-old cancer-free woman. Expected lifetime cost of diagnostic decisions is $1,536.84 when the BI-RADS recommendations

are followed, whereas it is $1,247.21 when the optimal diagnostic decisions are followed. Corresponding QALYs resulting from following the BI-RADs recommendations and the optimal diagnostic decisions are 41.15317 and 41.15497, respectively. Following optimal diagnostic decisions leads to approximately $160,905 savings per life years. Therefore, if the optimal diagnostic decisions based on age and breast cancer risk are followed, cost savings without sacrificing QALYs can be attained. A possible explanation for such savings lies in the age-based differences as can be seen in Figures 5(a) and 5(b). Using the optimal thresholds for decision making can result in approximately 22% cost savings. Taking more aggressive diagnostic actions at younger ages and less aggressive actions at older ages, as suggested by the optimal diagnostic decisions, can decrease false positives, improve patient outcomes, and decrease cost. Next, we analyze the effect of budget constraint on optimal diagnostic decisions.

Figure 5 depicts optimal policies as a function of the budget. Figures 5(c) and 5(d) suggest that the initial reductions in the budget would lessen the use of follow-up recommendations and replace it with routine mammography. There is some support/evidence from the medical community reinforcing our findings on follow-up recommendations in the context of resource limitations (Rubin 1999, Yasmeen et al. 2003). When the cost constraint is more pressing as in Figure 5(e), follow-up recommendations would be entirely foregone and the biopsy thresholds would shift up. For budget reductions at higher magnitude, the cost constraint affects all age groups, yet we observe a more dramatic change in biopsy thresholds at older ages.

An interesting observation in Figure 5 is the changing nature of the optimal policies at different budget levels. In particular, we observe that the unconstrained problem produces mostly monotone policies and as the constraint is imposed, this monotone structure no longer holds. There could be several reasons for this phenomenon. As we explained earlier in the manuscript, introduction of constraints into the LP formulation of the MDP disturbs the problem structure to produce randomized policies. Therefore, one explanation for lack of monotonicity is the disturbed problem structure. It could also be due to varying incidence rates in different age groups. Breast cancer is mostly seen in middle age woman (Edwards et al. 2002) and the disconnect between breast cancer incidence rates and mortality ratios could explain the lack of monotonicity as the constraint takes effect.

Using MCW data, we compare the performance of our model to the radiologists for various budget levels in Table 2, where the numbers in parentheses refer to missed cancers for routine mammography and short-term follow-up actions, and negative result for biopsy

Table 2   Performance of the CFDMIP Model on the MCW Data for Various Budget Levels

Assessment by	Budget level	RM		SF		BX	
Radiologists	—	29,105	(11)	12,534	(75)	676	(272)
CFDMIP	High	39,659	(25)	861	(7)	1,795	(326)
CFDMIP	Moderate	40,000	(27)	520	(5)	1,795	(326)
CFDMIP	Low	40,354	(30)	175	(2)	1,795	(326)
CFDMIP	Lower	40,681	(34)	0	(0)	1,634	(324)
CFDMIP	Tight	41,191	(51)	0	(0)	1,124	(307)
CFDMIP	Very tight	41,691	(129)	0	(0)	624	(229)

action. The radiologists recommended 29,105 routine mammographies and 12,534 follow-ups as compared to 39,659 routine mammographies and 861 follow-up recommendations by the CFDMIP model. As a result, the radiologists missed a total of 86 cancers compared to the model, which missed a total of 32 cancers. The increased performance of the CFDMIP model is achieved at a reduced expected cost. The rest of the results in Table 2 suggest that further cost reductions are possible at comparable performance levels.

5.2.   A Dynamic Incremental Cost-Effectiveness Analysis

The impact of budget reduction is also reflected in the optimal objective value, namely, total expected QALYs. We ran the CFDMIP model of §3 for different budget levels, i.e., for various $C(\delta)$s. Our CFDMIP model with no budget constraint returned the expected cost of $1,247.21 for optimal diagnostic decisions over the lifetime of a 40-year-old woman. We selected the $180–$1,380 range with $60-wide difference for consecutive budget levels as the value set for the budget experiments, i.e., $C(\delta_i) = 180 + i \cdot 60$ where $i \in \{0, \dots, 20\}$. We define ith incremental cost-effectiveness ratio, ICER$_i$, for a given $C(\delta_i)$ as ICER$_i = 60/(Q(\delta_i) - Q(\delta_{i-1}))$. It is well known that conventional cost-effectiveness analysis compares relative costs and effects that belong to different courses of programs, however we use optimal actions under consecutive budget levels as the mode of comparison.

Such analysis can help us understand the effect of budget on patient outcomes.

The solid line in Figure 6(a) depicts the trend in patient outcomes as a function of budget levels. The figure suggests decreasing marginal returns in total expected QALYs as the available budget increases. The dashed lines in Figure 6(a) represent the incremental cost-effectiveness analysis for the respective $C(\delta_i)$s using our formulation. Over the budget level of $C($720)$, additional spending is hardly justifiable because of an exponential increase in spending per QALYs saved. The arbitrary choice of $60 difference between consecutive budget levels was for illustrative purposes only. We observe similar trends when the difference between consecutive budget levels varies.

Sensitivity of optimal value function to the disutility of follow-up is depicted in Figure 6(b). The trend is analogous to the sensitivity of optimal value function to budget, yet in the reverse. We observe diminishing losses in total expected QALYs as the disutility of follow-up increases. This is because, as the disutility of follow-up increases, follow-ups are recommended less often and after a certain disutility level, follow-ups are no longer recommended.
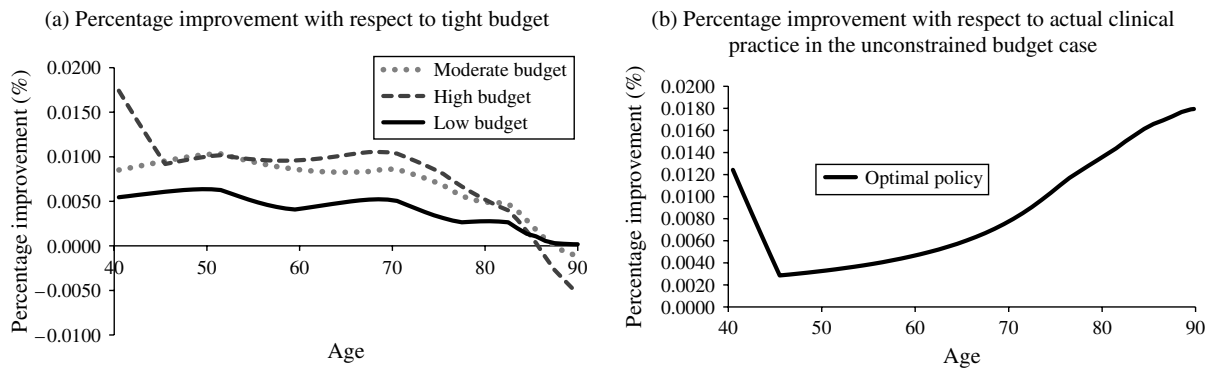
5.3.   Age Effects

We also investigate which age groups benefit most when the available budget increases and optimal recommendations are followed as compared to the current guidelines. First, we investigate age-based improvements by analyzing the percentage improvement in total expected QALYs from tight budget (TB) to low budget (LB), moderate budget (MB), and high budget (HB) cases. We denote the expected remaining QALYs of a risk-free woman at time k given that she was risk-free at age 40 and optimal policy π^* is followed under the budget constraint B as $Q_{k,rf}^{\pi^*(40,B)}$. Therefore, we define the percentage improvement in expected remaining QALYs at time k when switching from TB to LB as $(Q_{k,rf}^{\pi^*(40,LB)} - Q_{k,rf}^{\pi^*(40,TB)})/Q_{k,rf}^{\pi^*(40,TB)}$, from TB to MB as $(Q_{k,rf}^{\pi^*(40,MB)} - Q_{k,rf}^{\pi^*(40,TB)})/Q_{k,rf}^{\pi^*(40,TB)}$,

Figure 6   Sensitivity of Optimal Total Expected QALYs
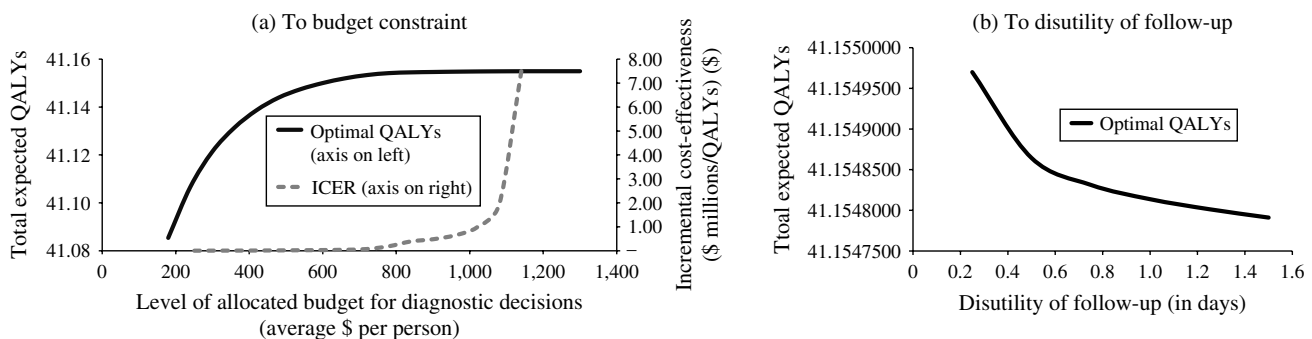
(a) To budget constraint

(b) To disutility of follow-up

Figure 7 **Percentage Improvement of Total Expected QALYs as a Function of Age for a Risk-Free Woman**

(a) Percentage improvement with respect to tight budget

(b) Percentage improvement with respect to actual clinical practice in the unconstrained budget case

and from TB to HB as $(Q_{k,rf}^{\pi^*(40,\,HB)} - Q_{k,rf}^{\pi^*(40,\,TB)})/Q_{k,rf}^{\pi^*(40,\,TB)}$. The corresponding trends are depicted in 7(a). We find that, as the available budget increases, younger women benefit more. We observe lower improvements in the older age group, interestingly, with some deterioration in the oldest age group. This may appear counterintuitive at first, however, it could be due to potential harms of screening mammography when the life expectancy and the risk of breast cancer is very low.

Second, we compare optimal policy under no budget constraint to the actual clinical practice. The comparison reveals improvements over actual clinical practice at all ages with relatively more emphasis on the very young or those over age 65 (Figure 7(b)). The counterintuitive result of higher improvement in younger women with a nadir at around age 45 can be explained by the potential gains in QALYs because of aggressive diagnostic actions in this age group. The slowly increasing trend beyond age 45 can be explained by the increasing prevalence of breast cancer by age.

We further investigate age-based differences with respect to the number of deaths and number of cancers detected. In Table 3, we compare the expected number of women who will die or be diagnosed with breast cancer under varying budget levels and actual clinical practice. We estimate the number of deaths and number of cancers detected based on the estimates of CFDMIP model for 100,000 women. For each

age group, as expected, increasing budget averts more deaths. Moreover, our model under the high budget case averts more deaths than actual practice in women younger than 70, whereas actual practice averts more deaths than our model in women over 70. Overall, our model averts more deaths than the actual practice.

There is no clear trend over different budget levels in terms of cancer detection. When low budget and high budget results are compared, more cancers are detected in women younger than 70, and fewer cancers are detected in women over 70 under the high budget scenario. We observe that more cancers are detected in actual practice than those detected under high budget, which may appear to be counterintuitive for women younger than 70. That is, actual practice detects more cancers yet averts fewer deaths for women younger than 70. However, this may be due to a highly contentious phenomenon called overdiagnosis (Warren and Eleti 2006) as described in §1, resulting in wasted resources and increased risk of overtreatment.

6. Conclusion

In this study, we consider the problem of optimizing diagnostic decisions after mammography from the patient's perspective while also considering the societal impact. In particular, we develop a finite-horizon constrained MDP model that optimizes the diagnostic decisions of a radiologist by maximizing the total

Table 3 **Average Performance of the CFDMIP Model for 100,000 Risk-Free Women at Age 40 for Various Budget Levels**

	Age									
	No. of deaths					No. of cancers detected				
	40–50	50–60	60–70	70–80	80–90	40–50	50–60	60–70	70–80	80–90
Low budget	2,240	4,711	10,485	21,123	33,462	270	368	342	292	164
Moderate budget	2,221	4,685	10,433	21,061	33,436	285	400	344	267	161
High budget	2,219	4,678	10,426	21,017	33,426	309	385	360	280	152
Actual	2,229	4,681	10,428	21,005	33,345	321	391	360	297	182

QALYs of a patient under budget constraints. In line with the ACA enactment, we adopt a novel approach and consider cost as a constraint while maximizing total expected QALYs. This approach allows us to conduct a cost-effectiveness analysis by comparing the ratio of cost resulting from following optimal diagnostic policies to the effect in terms of total expected QALYs. By repeating this experiment for a sequence of reasonable budget levels, we formed incremental cost-effectiveness ratios attributable to consecutive levels of funding.

The patient outcomes as measured by total expected QALYs improve as the level of funding increases, yet with diminishing returns in the allocated budget. Comparing proposed optimal policies to actual clinical practice, we find that the use of optimal thresholds for decision making could result in approximately 22% cost savings without sacrificing QALYs. Our analysis indicates short-term follow-ups are the immediate target for elimination when budget becomes a concern. Policy change is more drastic in the older age group with the increasing budget, yet the gains in total expected QALYs related to larger budgets are predominantly seen in younger women along with modest gains for older women. Moreover, our findings make a potential contribution to the recent contoversy following the U.S. Preventive Services Task Force's new policy recommendation on screening (Marshall 2010). The new recommendations suggest that women in their 40s should not be screened because of the high false positive rates and associated economic and psychological costs in this age group. However, we demonstrate that most benefits in terms of QALYs can be achieved in this age group.

Although our results are promising, there are several issues for either the radiologist or the patient that need to be addressed before implementation. First, there may be practical and ethical difficulties if radiologists are to implement our findings. Despite this, budget observant optimal recommendations would be ethical and practical when the trade-off is transparent to the radiologist. That is, when the proposed changes to the policy under the influence of societal motives are medically rational, implementation could be more feasible. Second, whether a patient tolerates budget-observant policies depends on personal preferences. Many patients may perceive less aggressive diagnostic decisions as less care, and therefore may prefer not to choose them. Conversely, there are patients who prefer less aggressive diagnostic approaches because of high disutility for invasive follow-up procedures such as biopsy that may lead to severe complications such as bleeding. In any case, the societal perspective, which is used to answer most of the health policy questions, requires consideration of healthcare issues at the population level while accounting for individual preferences. In fact, this is how existing guidelines, including

those recommending a 2% threshold for biopsy and 1% threshold for follow-up for all ages, are developed. Medically rational policies defined at the population level would take into account the factors that are relevant and would also apply to individuals. Finally, the aforementioned points can be extended to include the very real medico-legal risks that radiologists face. If policy recommendations are set at a population level prioritizing the interest of society as a whole, the medico-legal environment could support radiologists in applying those recommendations.

The advantages of the modeling framework we introduce manifest themselves mostly for policy makers in a clinically sound way. First, explicit representation of patient dynamics complies with personalized decision making, which is strongly advocated for and considered to have a pivotal role in the future of healthcare (Cortese 2007). A recent editorial in the *Annals of Internal Medicine* calls for tailoring recommendations based on individual risk to help improve breast cancer prevention (Kerlikowske 2009). Our research is a step toward personalized diagnostic recommendations while maximizing benefits and minimizing harm. Second, the cost-effectiveness approach we introduce suggests a flexible policy selection scheme allowing investigation of multiple criteria. Finally, our methodology allows incorporation of various types of policy restrictions. We consider only the expected cost constraints while maximizing patient outcomes. However, one can model various policy restrictions as constraints to assess the clinical and monetary implications of those policies.

Our research could serve as a guide for optimal diagnostic decisions in developed countries or in developing countries where resources are limited. By using our modeling scheme, policymakers could assess QALYs or the cost implications of various policies, or make resource allocation decisions to improve care. The ability to include constraints in such an optimization model could present opportunities for investigating different aspects of policy making other than cost.

Electronic Companion
An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/msom.1110.0371.

References

Alagoz, O., J. Chhatwal, E. S. Burnside. 2010. Optimizing the level of follow-up mammography recommendations with comparison to current clinical practice. Working paper, University of Wisconsin–Madison, Madison.

Altman, E. 1999. *Constrained Markov Decision Processes*. Chapman & Hall/CRC Press. Boca Raton, FL.

American College of Radiology. 1998. *Breast Imaging Reporting and Data System (BI-RADS)*. American College of Radiology, Reston, VA.

Anders, C. K., D. S. Hsu, G. Broadwater, C. R. Acharya, J. A. Foekens, Y. Zhang, Y. Wang, et al. 2008. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J. Clinical Oncology* **26**(20) 3324–3330.

Arias, E. 2006. United States life tables, 2003. *Natl. Vital Statist. Rep.* **54**(14) 1–40.

Ayer, T., O. Alagoz, N. K. Stout. 2012. A POMDP approach to personalize mammography screening decisions. *Oper. Res.*, ePub ahead of print June 12, http://dx.doi.org/10.1287/opre.1110.1019.

Ayer, T., M. U. S. Ayvaci, Z. X. Liu, O. Alagoz, E. S. Burnside. 2010a. Computer-aided diagnostic models in breast cancer screening. *Imaging in Medicine* **2**(3) 313–323.

Ayer, T., O. Alagoz, J. Chhatwal, J. W. Shavlik, C. E. Kahn Jr, E. S. Burnside. 2010b. Breast cancer risk estimation with artificial neural networks revisited. *Cancer* **116**(14) 3310–3321.

Baker, J. A., P. J. Kornguth, J. Y. Lo, M. E. Williford, C. E. Floyd. 1995. Breast cancer: Prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* **196**(3) 817–822.

Beam, C. A., P. M. Layde, D. C. Sullivan. 1996. Variability in the interpretation of screening mammograms by U.S. radiologists: Findings from a national sample. *Arch. Internal Medicine* **156**(2) 209–213.

Berg, W. A., C. Campassi, P. Langenberg, M. J. Sexton. 2000. Breast imaging reporting and data system: Inter- and intraobserver variability in feature analysis and final assessment. *Amer. J. Roentgenology* **174**(6) 1769–1777.

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

Beutler, F. J., K. W. Ross. 1985. Optimal policies for controlled Markov chains with a constraint. *J. Math. Anal. Appl.* **112**(1) 236–252.

Beutler, F. J., K. W. Ross. 1986. Time-average optimal constrained semi-markov decision processes. *Adv. Appl. Probab.* **18**(2) 341–359.

Breen, N., M. L. Brown. 1994. The price of mammography in the United States: Data from the national survey of mammography facilities. *Milbank Quart.* **72**(3) 431–450.

Brown, M. L., F. Houn, E. A. Sickles, L. G. Kessler. 1995. Screening mammography in community practice: Positive predictive value of abnormal findings and yield of follow-up diagnostic procedures. *Amer. J. Roentgenology* **165**(6) 1373–1377.

Burnside, E., J. Belkora, L. Esserman. 2001. The impact of alternative practices on the cost and quality of mammographic screening in the United States. *Clinical Breast Cancer* **2**(2) 145–152.

Burnside, E. S., J. Davis, J. Chhatwal, O. Alagoz, M. J. Lindstrom, B. M. Geller, B. Littenberg, K. A. Shaffer, C. E. Kahn Jr., C. D. Page. 2009. Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. *Radiology* **251**(3) 663–672.

Bynum, J. P. W., J. B. Braunstein, P. Sharkey, K. Haddad, A. W. Wu. 2005. The influence of health status, age, and race on screening mammography in elderly women. *Arch. Internal Medicine* **165**(18) 2083–2088.

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

Chhatwal, J., O. Alagoz, M. J. Lindstrom, C. E. Kahn, K. A. Shaffer, E. S. Burnside. 2009. A logistic regression model to aid breast cancer diagnosis based on the national mammography database format. *Amer. J. Roentgenology* **192**(4) 1117–1127.

Cortese, D. A. 2007. A vision of individualized medicine in the context of global health. *Clinical Pharmacology and Therapeutics* **82**(5) 491–493.

Derman, C., M. Klein. 1965. Some remarks on finite horizon Markovian decision models. *Oper. Res.* **13**(2) 272–278.

Edwards, B. K., H. L. Howe, L. A. Ries, M. J. Thun, H. M. Rosenberg, R. Yancik, P. A. Wingo, A. Jemal, E. G. Feigal. 2002. Annual report to the nation on the status of cancer, 1973–1999, featuring implications of age and aging on U.S. cancer burden. *Cancer* **94**(10) 2766–2792.

Elmore, J. G., C. K. Wells, C. H. Lee, D. H. Howard, A. R. Feinstein. 1994. Variability in radiologists' interpretations of mammograms. *New England J. Medicine* **331**(22) 1493–1499.

Elmore, J. G., M. B. Barton, V. M. Moceri, S. Polk, P. J. Arena, S. W. Fletcher. 1998. Ten-year risk of false positive screening mammograms and clinical breast examinations. *New England J. Medicine* **338**(16) 1089–1096.

Elmore, J. G., S. L. Jackson, L. Abraham, D. L. Miglioretti, P. A. Carney, B. M. Geller, B. C. Yankaskas, et al. 2009. Variability in interpretive performance at screening mammography and radiologists characteristics associated with accuracy. *Radiology* **253**(3) 641–651.

Ernster, V. L., J. Barclay, K. Kerlikowske, H. Wilkie, R. Ballard-Barbash. 2000. Mortality among women with ductal carcinoma in situ of the breast in the population-based surveillance, epidemiology, and end results program. *Arch. Internal Medicine* **160**(7) 953–958.

Geller, B. M., W. E. Barlow, R. Ballard-Barbash, V. L. Ernster, B. C. Yankaskas, E. A. Sickles, P. A. Carney, et al. 2002. Use of the American College of Radiology BI-RADS to report on the mammographic evaluation of women with signs and symptoms of breast disease. *Radiology* **222**(2) 536–542.

Ghosh, K., L. J. Melton III, V. J. Suman, C. S. Grant, S. Sterioff, K. R. Brandt, C. Branch, T. A. Sellers, L. C. Hartmann. 2005. Breast biopsy utilization: A population-based study. *Arch. Internal Medicine* **165**(14) 1593–1598.

Gram, I. T., E. Lund, S. E. Slenker. 1990. Quality of life following a false positive mammogram. *Br J Cancer* **62**(6) 1018–1022.

Greif, J. M. 2010. Mammographic screening for breast cancer: An invited review of the benefits and costs. *The Breast* **19**(4) 268–272.

Hayat, M. J., N. Howlader, M. E. Reichman, B. K. Edwards. 2007. Cancer statistics, trends, and multiple primary cancer analyses from the surveillance, epidemiology, and end results (SEER) program. *The Oncologist* **12**(1) 20–37.

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

Jemal, A., R. Siegel, E. Ward, T. Murray, J. Xu, M. J. Thun. 2007. Cancer statistics, 2007. *CA: A Cancer J. for Clinicians* **57**(1) 43–66.

Kallenberg, L. C. M. 1981. Unconstrained and constrained dynamic programming over a finite horizon. Technical Report 81-46, University of Leiden, The Netherlands.

Kallenberg, L. C. M. 1994. Survey of linear programming for standard and nonstandard Markovian control problems. Part i: Theory. *Math. Methods Oper. Res.* **40**(1) 1–42.

Kerlikowske, K. 2009. Evidence-based breast cancer prevention: The importance of individual risk. *Ann. Internal Medicine* **151**(10) 750–752.

Kerlikowske, K., D. Grady, J. Barclay, V. Ernster, S. D. Frankel, S. H. Ominsky, E. A. Sickles. 1998. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J. Natl. Cancer Inst.* **90**(23) 1801–1809.

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

Lee, C. P., G. M. Chertow, S. A. Zenios. 2008. Optimal initiation and management of dialysis therapy. *Oper. Res.* **56**(6) 1428–1449.

Maillart, L. M., J. S. Ivy, S. Ransom, K. Diehl. 2008. Assessing dynamic breast cancer screening policies. *Oper. Res.* **56**(6) 1411–1427.

Mandelblatt, J. S., M. E. Wheat, M. Monane, R. D. Moshief, J. P. Hollenberg, J. Tang. 1992. Breast cancer screening for elderly women with and without comorbid conditions. *Ann. Internal Medicine* **116**(9) 722–730.

Mandelblatt, J. S., K. A. Cronin, S. Bailey, D. A. Berry, H. J. de Koning, G. Draisma, H. Huang, et al. 2009. Effects of mammography screening under different screening schedules: Model estimates of potential benefits and harms. *Ann. Internal Medicine* **151**(10) 738–747.

Manne, A. S. 1960. Linear programming and sequential decisions. *Management Sci.* **6**(3) 259–267.

Marshall, E. 2010. Brawling over mammography. *Sci.* **327**(5968) 936–938.

Nelson, H. D., K. Tyne, A. Naik, C. Bougatsos, B. K. Chan, L. Humphrey. 2009. Screening for breast cancer: An update for the U.S. preventive services task force. *Ann. Internal Medicine* **151**(10) 727–737.

Neumann, P. J., M. C. Weinstein. 2010. Legislating against use of cost-effectiveness information. *New England J. Medicine* **363**(16) 1495–1497.

Over, M., P. Piot. 1996. Human immunodeficiency virus infection and other sexually transmitted diseases in developing countries: Public health importance and priorities for resource allocation. *J. Infectious Diseases* **174**(Supplement 2) 162–175.

Parkin, D. M., J. Ferlay, P. Pisani. 2005. Global cancer statistics, 2002. *CA: A Cancer J. for Clinicians* **55**(2) 74–108.

Poplack, S. P., P. A. Carney, J. E. Weiss, L. Titus-Ernstoff, M. E. Goodrich, A. N. A. Tosteson. 2005. Screening mammography: Costs and use of screening-related services. *Radiology* **234**(1) 79–85.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Rosenberg, R. D., B. C. Yankaskas, L. A. Abraham, E. A. Sickles, C. D. Lehman, B. M. Geller, P. A. Carney, et al. 2006. Performance benchmarks for screening mammography. *Radiology* **241**(1) 55–66.

Rubin, E. 1999. Six-month follow-up: An alternative view. *Radiology* **213**(1) 15–18.

Sandıkçı, B., L. M. Maillart, A. J. Schaefer, O. Alagoz, M. S. Roberts. 2008. Estimating the patients price of privacy in liver transplantation. *Oper. Res.* **56**(6) 1393–1410.

Schaefer, A. J., M. D. Bailey, S. M. Shechter, M. S. Roberts. 2004. Modeling medical treatment using Markov decision processes. M. Brandeau, F. Sainfort, W. Pierskalla, eds. *Operations Research and Health Care: A Handbook of Methods and Applications*. Kluwer Academic, Boston, 597–616.

Smith, R. A., V. Cokkinides, D. Brooks, D. Saslow, O. W. Brawley. 2010. Cancer screening in the United States, 2010: A review of current American cancer society guidelines and issues in cancer screening. *CA: A Cancer J. for Clinicians* **60**(2) 99–119.

Smith, R. A., D. Saslow, K. A. Sawyer, W. Burke, M. E. Costanza, W. P. Evans, R. S. Foster, E. Hendrick, H. J. Eyre, S. Sener. 2003. American cancer society guidelines for breast cancer screening: Update 2003. *CA: A Cancer J. for Clinicians* **53**(3) 141–169.

Smith-Bindman, R., P. W. Chu, D. L. Miglioretti, E. A. Sickles, R. Blanks, R. Ballard-Barbash, J. K. Bobo, et al. 2003. Comparison of screening mammography in the United States and the United Kingdom. *J. Amer. Medical Assoc.* **290**(16) 2129–2137.

Taplin, S. H., L. E. Ichikawa, K. Kerlikowske, V. L. Ernster, R. D. Rosenberg, B. C. Yankaskas, P. A. Carney, et al. 2002. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology* **222**(2) 529–535.

U.S. Census Bureau. 2000. Projections of the total resident population by 5-year age groups and sex with special age categories: Middle series, 1999 to 2100. Report, U.S. Census Bureau, Washington, DC. Accessed February 2, 2012, http://www.census.gov/population/www/projections/natsum-T3.html.

Vizcaíno, I., L. Gadea, L. Andreo, D. Salas, F. Ruiz-Perales, D. Cuevas, C. Herranz, F. Bueno, Screening Program Working Group. 2001. Short-term follow-up results in 795 nonpalpable probably benign lesions detected at screening mammography. *Radiology* **219**(2) 475–483.

Warner, E. 2011. Breast-cancer screening. *New England J. Medicine* **365**(11) 1025–1032.

Warren, R., A. Eleti. 2006. Overdiagnosis and overtreatment of breast cancer: Is overdiagnosis an issue for radiologists? *Breast Cancer Res.* **8**(2) 205.

Weinstein, M. C., J. E. Siegel, M. R. Gold, M. S. Kamlet, L. B. Russell. 1996. Recommendations of the panel on cost-effectiveness in health and medicine. *J. Amer. Medical Assoc.* **276**(15) 1253–1258.

Welch, H. G., W. C. Black. 2010. Overdiagnosis in cancer. *J. Natl. Cancer Inst.* **102**(9) 605–613.

Yasmeen, S., P. S. Romano, M. Pettinger, R. T. Chlebowski, J. A. Robbins, D. S. Lane, S. L. Hendrix. 2003. Frequency and predictive value of a mammographic recommendation for short-interval follow-up. *J. Natl. Cancer Inst.* **95**(6) 429–436.