European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Stratified breast cancer follow-up using a continuous state partially observable Markov decision process

Maarten Otten[a,b,*], Judith Timmer[b], Annemieke Witteveen[c]

[a] Center for Healthcare Operations Improvement and Research, University of Twente, P.O. Box 217, Enschede 7500 AE, The Netherlands
[b] Department of Stochastic Operations Research, University of Twente, P.O. Box 217, Enschede 7500 AE, The Netherlands
[c] Department of Health Technology and Services Research, University of Twente, P.O. Box 217, Enschede 7500 AE, The Netherlands

ARTICLE INFO

Article history:
Received 23 February 2018
Accepted 27 August 2019
Available online xxx

Keywords:
Decision processes
Medical decision making
Partially observable Markov decision process

ABSTRACT

Frequency and duration of follow-up for breast cancer patients is still under discussion. Currently, in the Netherlands follow-up consists of annual mammography for the first five years after treatment and does not depend on the personal risk of developing a locoregional recurrence or a second primary tumor. The aim of this study is to gain insight in how to allocate resources for optimal and personalized follow-up. We formulate a discrete-time Partially Observable Markov Decision Process (POMDP) over a finite horizon with both discrete and continuous states, in which the size of the tumor is modeled as a continuous state. Transition probabilities are obtained from data of the Netherlands Cancer Registry. We show that the optimal value function of the POMDP is piecewise linear and convex and provide an alternative representation for it. Under the assumption that the tumor growth follows an exponential distribution and that the model parameters can be described by exponential functions, the optimal value function can be obtained from the parameters of the underlying probability distributions only. Finally, we present results for a stratification of the patients based on their age to show how this model can be applied in practice.

1. Introduction

Patients that are treated successfully for breast cancer are followed clinically for a period of time, after their treatment, in order to detect possible reappearances of the tumor (Lu et al., 2009). A reappearance of the initial tumor on the same site is called a locoregional recurrence (LRR), whereas a tumor independent of the initial tumor is called a second primary (SP) tumor. Due to a high risk of distant metastases in case of an LRR, it is desirable to detect it in an early stage (Moossdorff et al., 2014). In the Netherlands, follow-up typically consists of annual mammograms for a period of five years (IKNL, 2017b). However, only a small part of the LRRs are detected by mammograms (Geurts et al., 2012). The majority is detected by the patient between check-ups. Furthermore, due to an increasing survival and incidence rate, the number of patients in follow-up increases and becomes more of a burden to healthcare. The follow-up also burdens the patients themselves, as the visits lead to anxiety and stress (Beaver et al., 2009; Pennery & Mallet, 2000). Although it is known that personal characteristics of the patient, such as the patient's age, size of the initial tumor and type of treatment of the initial tumor, are highly correlated with the probability of an LRR, there is no differentiation based on these factors in the current policy (Witteveen et al., 2015). Moreover, since 2012 the national guideline for cancer advises that the follow-up should be adjusted for several risk factors of a patient in order to personalize the follow-up as good as possible.

These observations together give rise to the question whether it is possible to improve the current follow-up policy and to use the available resources optimally by on the one hand avoiding overtreatment and on the other hand detecting recurrences as early as possible. With follow-up based on risk, clinicians can focus resources on patients with higher risk, while avoiding unnecessary and potentially harmful follow-up visits for women with very low risks. Our aim is to construct a sequential decision model in which we can decide whether it is optimal for the patient to test or to wait at every decision epoch. The mathematical framework we use to model this problem is a Partially Observable Markov Decision Process (POMDP), which is a generalization of a Markov Decision Process (MDP). In an MDP, decisions are made based on the current state, whereas in a POMDP, decisions are made based on partial information about the current state. For the problem that we consider, decisions are made based on mammograms and self-detections, which only provide partial information about the actual health state of a patient. Therefore, a POMDP is a very well suited framework for our problem (Steimle & Denton, 2017).

* Corresponding author at: Department of Stochastic Operations Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
E-mail address: j.w.m.otten@utwente.nl (M. Otten).

https://doi.org/10.1016/j.ejor.2019.08.046
0377-2217/

In previous research we modeled the problem described by a discrete-state POMDP (Otten et al., 2017). In that model an LRR is modeled as a two state Markov chain. In the first state the LRR is in an early stage and the prognosis is fairly good. In the second state the LRR is in a late stage and the prognosis is rather bad. We found this model usable for the problem but we also found that the outcome is quite sensitive to the transition probability between an early and a late LRR. These findings encouraged us to model the problem by a continuous-state POMDP, in which the health-state of the patient is modeled by a continuous model, to improve accuracy.

In Ayer, Alagoz, and Stout (2012) a discrete-state POMDP is developed to model a slightly different decision process: preventive screening for breast cancer. In Ayvaci, Alagoz, and Burnside (2012) the same problem is considered but from a budgetary instead of a patient perspective and the partially observability of the process is omitted. In Zhang, Denton, Balasubramanian, Shah, and Inman (2012) the patient and budgetary perspective are compared for a similar case, a POMDP approach for PSA screening. However, all of the models are based on an underlying discrete-state Markov chain and therefore simplifying the model for the health-state of a patient considerably. To the best of our knowledge there is no literature available that applies a continuous-state model to a medical decision making process. In Porta, Spaan, and Vlassis (2005) and Porta, Vlassis, Spaan, and Poupart (2006) a continuous-state model for robot planning is developed and some important analytical results are proved. In Duff (2002) some other useful results for continuous-state POMDPs are provided. These results also lay the basis for the solution method for the POMDP. However, necessary adjustments need to be made. In the first place, POMDPs based on medical decision making slightly differ from the standard framework of POMDPs in the way that the reward depends not only on the current state and action but also on the observation. Secondly, our model needs both a discrete component as well as a continuous component. The patient is either healthy or not, this is a discrete component. On the other hand, the growth of a tumor is modeled by a continuous model. The interaction between the discrete and continuous states is not present in the model by Duff (2002) and needs to be added.

Our contribution to this research is threefold. Firstly, we provide a more realistic model for the described problem. Instead of modeling the health state of the patient as a finite set of states, we model it as a continuum of states. Secondly, we prove some important results in order to derive a algorithm for finding the optimal testing schedule. Thirdly, we derive a simple algorithm for the optimal policy under some restrictions on the growth model of the tumor.

The remainder of this paper is organized as follows. In Section 2 we state some preliminary information on standard POMDPs and present the continuous-state POMDP model for our specific problem. In Section 3 we derive an expression for the optimal value function. We also provide an alternative representation of the optimal value function. This result will be used to derive a solution method. Under some restrictions on the dynamics of the POMDP, we then derive a simpler form of the optimal value function. In Section 4 we present the general algorithm for solving the POMDP and an algorithm for a special case. In Section 5 we illustrate how this model can be applied in practice. Finally, we summarize the results and conclude in Section 6.

2. Model

In this section we state the model for the problem described. To incorporate specific aspects of our problem we need to make some adjustments to the regular framework of POMDPs. For clarity we first describe a standard POMDP and based on this we present the model for our problem.

2.1. Preliminaries: POMDPs

A POMDP is generalization of a Markov Decision process (Aström, 1965). It describes a decision maker's interaction with a stochastic system of which the current state is not directly observable. The model is described by the following elements

- S, the set of system states.
- A, the set of actions.
- O, the set of observations.
- The observation model denoted by $K_t^a(o|s)$, i.e. the probability that observation o was made given that the state was s and action a was taken at time t.
- The underlying Markov Chain that models the transitions of the system's state, denoted by $P_t^{(a,o)}(s'|s)$, the probability that the next state is s' given that the previous state was s, action a was taken and observation o was made at time t.
- The reward function $r_t(s, a, o)$, which is the reward when the state is s, action a was taken and observation o made at time t.
- The belief $b(s)$. Because the decision maker cannot observe the system's state directly, the knowledge about the system is represented by the belief state. This is a probability distribution over the state space based on the internal dynamics of the system, the actions taken and the observations made. When the current state is s, action a is taken and observation o is made, the updated belief τ is computed with a Bayesian update (Smallwood & Sondik, 1973):

$$\tau[b, a, o](s') = \frac{\sum_s b(s) K_t^a(o|s) P_t^{(a,o)}(s'|s)}{\sum_s b(s) K_t^a(o|s)}.$$

The combination of an action and an observation induces an immediate reward, depending on the current state, and a future reward, depending on the next state. The value function describes the relation between the immediate reward, future reward and the belief state:

$$V_t^a(b) = \sum_s b(s) \sum_o K_t^a(o|s) \Big[r_t(s, a, o) + V_{t+1}(\tau[b, a, o]) \Big].$$

A policy is a function that maps a belief to an action. An optimal policy is one that maximizes the value function. This is described by the optimal value function, which gives for each belief the maximum value that the decision maker can obtain, $V_t^*(b) = \max_a \{V_t^a(b)\}$. The optimal policy can be defined by $\omega^*(b) = \arg\max_a V_t^*(b)$.

At first it may seem that solving the optimal value function is intractable, because it is defined on a continuous belief space. But, fortunately, it is proven that the optimal value function is piece-wise linear and convex (PWLC) (Smallwood & Sondik, 1973; Sondik, 1971). As a result the optimal value function can be written as

$$V_t^*(b) = \max_k \left\{ \sum_s b(s) \alpha_t^k(s) \right\},$$

for a certain finite set $\{\alpha^k\}$ of so-termed α-vectors. These α-vectors can be calculated in a recursive way. This provides a straightforward way to obtain the optimal policy.

2.2. Model formulation

In the problem described we model the growth of tumor as a continuous process. Because the patient can die during the process and the process terminates whenever the patient goes into treatment, we need to modify the standard POMDP framework in order

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

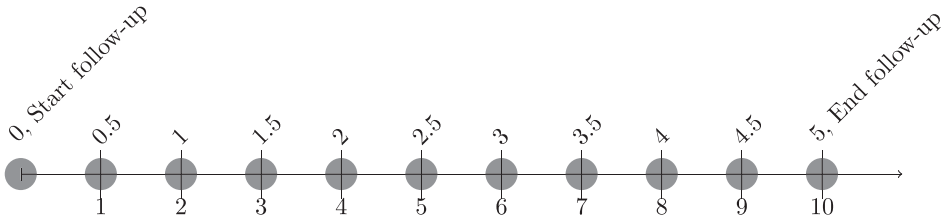

Fig. 1. The follow-up time line: above the line is the time (years) since initial treatment and below the line is the number of the corresponding decision epoch.

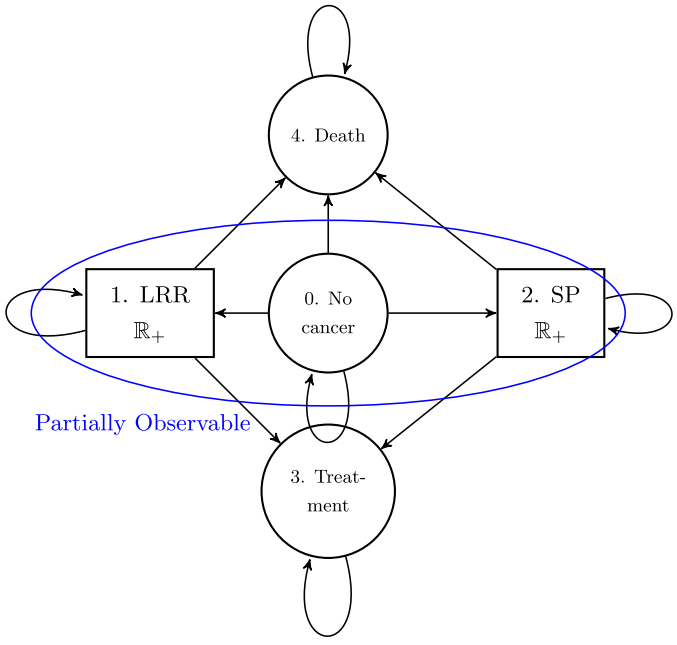

Fig. 2. State diagram of the underlying Markov process.

to model our problem correctly. The problem is therefore modeled by a discrete-time continuous state POMDP over a finite horizon, in which a decision maker aims to maximize the total expected number of quality-adjusted life years (QALYs).

Every 6 months a decision is made whether the patient should have a mammogram or not. Decisions are based on the patient's personal risk of an LRR or SP and on prior test results. In case of a positive mammogram or a self-detection an additional test, i.e. a biopsy, is performed. We assume that the additional test is perfect, so that if this test is positive it is certain that the patient has cancer. In this case we assume that the patient starts treatment immediately and leaves the decision process by transitioning to one of the absorbing states. If the additional test is negative, the process will proceed to the next decision epoch. The process also proceeds after a negative mammogram or after a decision to wait until the next decision epoch and no self-detection was made. For our notation we follow Ayer et al. (2012) and Otten et al. (2017). The complete model and the notation used are as follows:

Decision epochs Decisions are made twice a year and the decision process starts 6 months after initial treatment finished, $t = 1 \dots T$. The time between two subsequent decisions is denoted by $\sigma = 0.5$ year. The decision process terminates after 10 decision epochs (see Fig. 1).

Core state space The core state space is denoted by $S = \{0, S^{LRR}$ and S^{SP}, 3, 4$\}$, where $S^{LRR} = S^{SP} = \mathbb{R}_+$. It consists of three discrete states {0, 3, 4}, 0 stands for no (detectable) cancer, 3 for treatment of the patient and 4 stands for the death of the patient. S^{LRR}, S^{SP} are two continuous states (or better: a continuum of states) in the core state space that represent a measure, e.g. the size of the tumor, for the state of an LRR and an SP, respectively. To see how these different states are connected, see Fig. 2. As early detection of distant metastasis (DM) is not part of the follow-up and after both detection of DM and death the follow-up would end, both states, i.e., death and DM, are combined in the model. The true health state of the patient at time t is denoted by s_t. We model an LRR and an SP as continuous variables to incorporate the difference in expected remaining QALYs between earlier and later detection as good as possible. Note that the decision maker can directly observe whether a patient is in the state 'Treatment' or 'Death' but not whether a patient is in one of the other states. We therefore call the states $\{0, S^{LRR}, S^{SP}\}$ partially observable and denote this subset of the core state space as S^{PO}.

Information space The space consisting of possible probability distributions over the state space S is denoted by $\Pi(S)$. An element $\pi \in \Pi(S)$ is called an information state.

Belief space The space of all probability distributions over the partially observable states S^{PO} is denoted by $B(S^{PO})$. For clarity we define a belief vector $b = [b(0)\ b(S^{LRR})\ b(S^{SP})]$, which denotes the belief that a patient is in state 0, S^{LRR} or S^{SP} and belief functions $b_{LRR}(s)$, $b_{SP}(s)$ which denote the belief that a patient's true health state is $s \in \mathbb{R}_+$ given the patient is in the continuous state LRR or SP, respectively.

Actions The set of possible actions at time t is A_t. An element of the set is denoted by $a_t \in A_t = \{W, M\}$, where W stands for wait and M for mammogram. The action set is only defined for $s \in S^{PO}$ because the decision process terminates in the other states.

Observation space The set of possible observations, when action a is selected, is denoted by Θ_a. If $a_t = M$, the possible observations are a positive mammogram (M^+) or a negative mammogram (M^-). If $a_t = W$, the patient will perform a self-test. This can either result in a self-detection (SD^+) or not a self-detection (SD^-). We have $\Theta_M = \{M^+, M^-\}$ and $\Theta_W = \{SD^+, SD^-\}$. When the action corresponding with the observation is clear from the context we will denote both SD^- and M^- with $-$ and SD^+ or M^+ with $+$.

Observation probabilities The probability of making at time t observation o when decision a was taken while in state s, is denoted by $K_t^a(o|s)$. These probabilities are completely determined by the specificity of a mammogram, the fraction of healthy patients having a negative mammogram and the sensitivity of a mammogram, the fraction of patients with cancer having a positive mammogram. For example, $K_t^M(M^-|s =$ 'No cancer') is the probability of having a negative mammogram when the true health state of the patient is 'No cancer', this is the specificity of a mammogram. We denote the specificity of a mammogram by $spec_t(M)$ and of self-detection by $spec_t(SD)$. Similarly, the sensitivity of a mammogram is denoted by $sens_t(s, M)$ and of self-detection $sens_t(s, SD)$. Note that, unlike specificity, the sensitivity of a test depends on the true health state of the patient. The observation probabilities are:

$$
\begin{aligned}
&K_t^M(M^-|s=0) = spec_t(M)\\
&K_t^M(M^+|s=0) = 1 - spec_t(M)\\
&K_t^W(SD^-|s=0) = spec_t(SD)\\
&K_t^W(SD^+|s=0) = 1 - spec_t(SD)\\
&K_t^M(M^+|s) = sens_t(s, M) && s \in \{S^{SP}, S^{LRR}\}\\
&K_t^M(M^-|s) = 1 - sens_t(s, M) && s \in \{S^{SP}, S^{LRR}\}\\
&K_t^W(SD^+|s) = sens_t(s, SD) && s \in \{S^{SP}, S^{LRR}\}\\
&K_t^W(SD^-|s) = 1 - sens_t(s, SD) && s \in \{S^{SP}, S^{LRR}\}
\end{aligned}
$$

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

Core state transitions The distribution function of the transition at time t, when the current state is s, action a was taken and observation o made, is denoted by $P_t^{(a,o)}(s'|s)$. Because the state space contains both discrete and continuous states, these probability distributions can be discrete, continuous or a mixture of both. Since transitions within the partially observable state space are only possible from the discrete state 0 to the cancer states and not vice versa, it is only in this state that a mixture of a discrete and a continuous probability distribution occurs. In state 0 the transitions are as follows: with probability p_t^C, $C = LRR, SP$, the patient gets cancer and transitions to the corresponding continuous state and with probability $1 - p_t^{LRR} - p_t^{SP}$ the patient stays in state 0. When transitioning to the continuous state the outcome is a continuous random variable. This is also the case for transitions within the continuous states. So the growth of the tumor in state 0 is 0 with probability $1 - p_t^{LRR} - p_t^{SP}$ and X with probability p_t^C, where X is a continuous random variable with probability density function $f_t^C(x|0)$. The growth in state $s \in S^{LRR}$, S^{SP} is X', where X' is a continuous random variable with probability density function $f_t^C(x|s)$, $C = LRR, SP$.

Updated belief space The belief at time $t+1$, when the belief about patient's true health state at time t was b, action a was taken and observation o was made, is denoted by $\tau[b, a, o]$. With slight abuse of notation (we denote $b(0)K_t^a(o|0)$ as $\int_S b(s)K_t^a(o|s)ds$ for $S = 0$) we can denote the updated belief state as:

$$\tau[b,a,o](s') = \begin{cases} \frac{\sum_{s\in S^{PO}} \int_S b(s)K_t^a(o|s)P_t^{(a,o)}(s'|s)ds}{\sum_{s\in S^{PO}} \int_S b(s)K_t^a(o|s)ds} & if\ o = M^-, SD^-, \\ P_t^{(a,o)}(s'|0) & if, o = M^+, SD^+. \end{cases} \quad (1)$$

Rewards The expected number of QALYs between two decision epochs, when the true health state of the patient is s, action a is taken and observation o is made, is denoted by $r_t(s, a, o)$. To factor in the probability that a patient dies between two decisions we use the half-cycle correction method (Sonnenberg & Back, 1993). In this method, it is assumed that if the patient dies between two decision epochs half of the cycle length σ is accrued to the expected number of QALYs. From this, QALYs are subtracted for the disutility of a possible mammogram or biopsy. If the patient is in one of the cancer states ($s \in S^{LRR} \cup S^{SP}$) and observes a positive mammogram or makes a self-detection, then she is rewarded a lump-sum reward of $R_t(s)$. This is the life expectancy of the patient given that her true health state is s minus the disutility associated with a biopsy and a possible mammogram. So, no QALYs are rewarded over the next decision epoch when a true positive mammogram or self-detection is observed, i.e. $r_t(s, M, M^+) = r_t(s, W, SD^+) = 0$. The reward in the treatment and death states are zero.

Let the expected reward between times t and $t+1$, if the true health state is s and the action chosen is a, be denoted by $r_t(s, a) = \sum_{o\in\Theta_a} K_t^a(o|s)r_t(s, a, o)$.

Let $r_T(s)$ denote the total expected remaining QALYs at time T when the patient's true health state is s at time T.

Let $p_d(s)$ denote the probability that a patient dies between two decision epochs when the true health state is s, and let dis_M, dis_B be the disutility experienced when undergoing a mammogram and a biopsy, respectively. The rewards for $t = 1, \ldots, T-1$ are:

$$\begin{aligned} r_t(s, W, SD^-) &= p_t^d(s)\cdot 0.5\sigma + (1 - p_t^d(s))\cdot\sigma && s \in S^{PO} \\ r_t(0, W, SD^+) &= p_t^d(s)\cdot 0.5\sigma + (1 - p_t^d(s))\cdot\sigma - dis_B \\ r_t(s, M, M^-) &= p_t^d(s)\cdot 0.5\sigma + (1 - p_t^d(s))\cdot\sigma - dis_M && s \in S^{PO} \\ r_t(0, M, M^+) &= p_t^d(s)\cdot 0.5\sigma + (1 - p_t^d(s))\cdot\sigma - dis_M - dis_B \\ r(s,\cdot,\cdot) &= 0 && \text{otherwise.} \end{aligned} \quad (2)$$

3. Optimal value function

In this section we derive an expression for the optimal value function of the POMDP described in the previous section. Furthermore, we provide an alternative representation of the optimal value function, which we use to construct an algorithm to determine the optimal actions.

The optimal value function is denoted by $V_t^*(\pi)$, the maximum expected number of QALYs a patient can obtain when the information state is $\pi \in \Pi(S)$ at time t. Whenever we consider the belief state, we will denote it by $V_t^*(b)$. Our goal is to derive an expression for the optimal value function in every belief state. We will do this by deriving the optimality equations, which recursively connect the optimal value function at different decision epochs. The decision process terminates whenever the patient moves to one of the absorbing states. $V_t^*(\pi)$ can be expressed as:

$$V_t^*(\pi) = \begin{cases} R_t(3) & \pi(3) = 1, \\ R_t(4) & \pi(4) = 1, \\ V_t^*(b) & \exists s \in S^{PO} s.t. \pi(s) > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

We denote the maximum total expected QALYs a patient can obtain, when at time t in belief state b and choosing action a, by $V_t^a(b)$:

$$V_t^*(b) = \max_a \{V_t^a(b)\} \qquad t = 1 \ldots T-1, \text{ with}$$

$$\begin{aligned} V_t^a(b) = {} & b(0)K_t^a(-|0)\Big[r_t(0,a,-) + (1 - p_t^{LRR} - p_t^{SP})V_{t+1}^*(\tau[b,a,-]) \\ & + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)V_{t+1}^*(\tau[b,a,-])ds\Big] \\ & + \sum_{C\in\{LRR,SP\}} \Bigg(\int_{S^C} b_C(s)K_t^a(-|s)\Big[r_t(s,a,-) \\ & + \int_{S^C} f_t^C(s'|s)V_{t+1}^*(\tau[b,a,-])ds'\Big]ds\Bigg) \\ & + b(0)K_t^a(+|0)\Big[r_t(0,a,+) + (1 - p_t^{LRR} - p_t^{SP})V_{t+1}^*(\tau[b,a,+]) \\ & + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)V_{t+1}^*(\tau[b,a,+])ds\Big] \\ & + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds \\ V_T^a(b) = {} & b(0)r_T(0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)r_T(s)ds. \end{aligned} \quad (4)$$

We can simplify the optimality equations by moving the parts that do not depend on s outside the integral and by noting that $\int_S f_t(x'|x)dx' = 1$.

$$\begin{aligned} V_t^*(b) = \max_a \Bigg\{ & b(0)K_t^a(-|0)\Big[r_t(0,a,-) + V_{t+1}^*(\tau[b,a,-])\Big] \\ & + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)\Big[r_t(s,a,-) + V_{t+1}^*(\tau[b,a,-])\Big]ds \\ & + b(0)K_t^a(+|0)\Big[r_t(0,a,+) + V_{t+1}^*(\tau[b,a,+])\Big] \\ & + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds\Bigg\} \\ & t = 1 \ldots T-1, \end{aligned}$$

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

$$V_T^*(b) = b(0)r_T(0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s) r_T(s) ds. \tag{5}$$

The optimal value function at time $t = T$ can be interpreted as the weighted average of the immediate reward given a certain belief about the patient's true health state.

3.1. Alternative representation of the optimal value function

The key idea of value iteration, one of the most widely used methods for solving Markov decision processes, is to relate the optimal value function V^* at time t to V^* at time $t+1$ (Puterman, 1994). Because the belief state is in fact a probability space over the core state space, the optimal value function is defined on an infinite dimensional vector space $B(S^{PO})$. This prevents us from iterating over all possible belief states to determine the optimal value function directly. However, the optimal value function is piecewise linear and convex (PWLC) and can, therefore be represented as the maximum over a set of finite dimensional vectors. This result is formalized in the following theorem.

Theorem 3.1. *The optimal value function $V_t^*(b)$ is piece-wise linear and convex, and can thus be written as*

$$V_t^*(b) = \max_k \left\{ b(0)\alpha_0^{k,t}(0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)\alpha_C^{k,t}(s) ds \right\}, \tag{6}$$

for some set of functions $\alpha_C^{k,t}(s)$, $C \in \{0, LRR, SP\}$, $k = 1, 2, \ldots$. The term α-function is used to refer to such a function.

The proof goes by induction and is very similar to that of the discrete case, proven by Smallwood and Sondik (1973), and to that of the continuous case, proven by Porta, Spaan, and Vlassis (2004), and is therefore omitted.

We can now write the optimal value function in terms of the α-functions.

Proposition 3.1. *The following representation of the optimal value function is equivalent to the optimal value function given in (4).*

$$\begin{aligned}
V_t^*(b) = \max_a \Bigg\{ & b(0)K_t^a(-|0)\Big[r_t(0,a,-) + (1 - p_t^{LRR} \\
& - p_t^{SP})\alpha_0^{i(b,a,-),t+1}(0) \\
& + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{i(b,a,-),t+1}(s) ds\Big] \\
& + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s) K_t^a(-|s)\Big[r_t(s,a,-) \\
& + \int_{S^C} f_t^C(s'|s)\alpha_C^{i(b,a,-),t+1}(s') ds'\Big] ds \\
& + b(0)K_t^a(+|0)\Big[r_t(0,a,+) \\
& + \max_k \Big((1 - p_t^{LRR} - p_t^{SP})\alpha_{t+1}^k(0) \\
& + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t+1}(s) ds\Big)\Big] \\
& + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s) K_t^a(+|s) R_t(s) ds \Bigg\},
\end{aligned} \tag{7}$$

where

$$\begin{aligned}
i(b,a,o) = \arg\max_k \Bigg\{ & b(0)K_t^a(-|0)(1 - p_t^{LRR} - p_t^{SP})\alpha_0^{k,t+1}(0) \\
& + \sum_{C\in\{LRR,SP\}} \int_{S^C} \Big[b(0)K_t^a(-|0) f_t^C(s'|0) \\
& + \int_{S^C} b_C(s) K_t^a(-|s) f_t^C(s'|s) ds\Big]\alpha_C^{k,t+1}(s') ds' \Bigg\}.
\end{aligned} \tag{8}$$

Proof First, we derive an equivalent representation of $V_{t+1}^*(\tau[b,a,o])$ in terms of the α-functions. Substituting the expression for $\tau[b,a,o]$ from (1) into (6) gives:

$$V_{t+1}^*(\tau[b,a,o]) = \tag{9}$$

$$= \begin{cases}
\max_k \left\{ \frac{b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) ds}\alpha_0^{k,t+1}(0) \right. \\
\left. + \sum_{C\in\{LRR,SP\}} \int_{S^C} \frac{b(0)K_t^a(-|0) f_t^C(s'|0) + \int_{S^C} b_C(s)K_t^a(-|s) f_t^C(s'|s) ds}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) ds}\alpha_C^{k,t+1}(s') ds' \right\} \\
if\ o = - \\
\max_k \left\{ (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{k,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s) ds \right\} \\
if\ o = +.
\end{cases} \tag{10}$$

The denominators do not depend on s' and k, hence they can be moved outside the integral over s' and the maximum over k. Also, by changing the order of integration and substituting $i(b,a,o)$ from (8), we obtain the following:

$$V_{t+1}^*(\tau[b,a,o]) = \tag{11}$$

$$= \begin{cases}
\frac{1}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b(S)K_t^a(-|s) ds} \\
\quad \times \max_k \left\{ b(0)K_t^a(-|0)(1 - p_t^{LRR} - p_t^{SP})\alpha_0^{k,t+1}(0) \right. \\
\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} \Big[b(0)K_t^a(-|0) f_t^C(s'|0) \\
\quad \left. + \int_{S^C} b_C(s)K_t^a(-|s) f_t^C(s'|s) ds\Big]\alpha_C^{k,t+1}(s') ds' \right\} & if o = -, \\
\max_k \left\{ (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{k,t+1}(0) \right. \\
\quad \left. + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s) ds \right\} & if o = +.
\end{cases} \tag{12}$$

$$= \begin{cases}
\frac{b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})\alpha_0^{i(b,a,o),t+1}(0)}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) ds} \\
\quad + \frac{b(0)K_t^a(-|0) \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,o),t+1}(s') ds'}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) ds} \\
\quad + \frac{\sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} b_C(s)K_t^a(-|s) \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,o),t+1}(s') ds' ds}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b(S)K_t^a(-|s) ds} & if o = -, \\
\max_k \left\{ (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{k,t+1}(0) \right. \\
\quad \left. + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s) ds \right\} & if o = +.
\end{cases} \tag{13}$$

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

Rewriting the expression for the optimal value function (5) gives:

$$\begin{aligned}
V_t^*(b) &= \max_a \Big\{ b(0)K_t^a(-|0)\Big[r_t(0,a,-) + V_{t+1}^*(\tau[b,a,-])\Big] \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)\Big[r_t(s,a,-) \\
&\quad + V_{t+1}^*(\tau[b,a,-])\Big]ds \\
&\quad + b(0)K_t^a(+|0)\Big[r_t(0,a,+) + V_{t+1}^*(\tau[b,a,+])\Big] \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds\Big\} \\
&= \max_a \Big\{ b(0)K_t^a(-|0)r_t(0,a,-) \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)r_t(s,a,-)ds \\
&\quad + \Big[b(0)K_t^a(-|0) \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds\Big]V_{t+1}^*(\tau[b,a,-]) \\
&\quad + b(0)K_t^a(+|0)r_t(0,a,+) \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds \\
&\quad + b(0)K_t^a(+|0)V_{t+1}^*(\tau[b,a,+])\Big\}. \qquad (14)
\end{aligned}$$

Finally, by substituting the expression derived for $V_{t+1}^*(\tau[b,a,o])$ (13) in the rewritten expression for $V_t^*(b)$ (14) we have:

$$\begin{aligned}
V_t^*(b) &= \max_a \Big\{ b(0)K_t^a(-|0)r_t(0,a,-) \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)r_t(s,a,-)ds \\
&\quad + \Big[b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds\Big] \\
&\quad \times \Bigg[\frac{b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})\alpha_0^{i(b,a,o),t+1}(0)}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds} \\
&\quad + \frac{b(0)K_t^a(-|0)\sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,o),t+1}(s')ds'}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds} \\
&\quad + \frac{\sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) \int_{S^C} f_t^C(s'|s)\alpha_C^{i(b,a,o),t+1}(s')ds'ds}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds} \Bigg] \\
&\quad + b(0)K_t^a(+|0)r_t(0,a,+) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds \\
&\quad + b(0)K_t^a(+|0)\max_k \Big\{ (1-p_t^{LRR}-p_t^{SP})\alpha_0^{k,t+1}(0) \\
&\quad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds\Big\}\Big\} \qquad (15)
\end{aligned}$$

$$\begin{aligned}
&= \max_a \Big\{ b(0)K_t^a(-|0)r_t(0,a,-) \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)r_t(s,a,-)ds \\
&\quad + \Big[b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})\alpha_0^{i(b,a,o),t+1}(0) \\
&\quad + b(0)K_t^a(-|0)\sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,o),t+1}(s')ds' \\
&\quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) \int_{S^C} f_t^C(s'|s)\alpha_C^{i(b,a,o),t+1}(s')ds'ds\Big] \\
&\quad + b(0)K_t^a(+|0)r_t(0,a,+) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds \\
&\quad + b(0)K_t^a(+|0)\max_k \Big\{ (1-p_t^{LRR}-p_t^{SP})\alpha_0^{k,t+1}(0) \\
&\quad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds\Big\}\Big\}. \qquad (16)
\end{aligned}$$

By rearranging the terms and by factorization of the last expression we obtain the desired result. □

By combining Theorem 3.1 and Proposition 3.1 an explicit expression of the α-functions can be derived. The algorithm that will be used utilizes this representation for solving the POMDP.

Corollary 3.1. *Let $\alpha_C^{l^*(b),t}$ denote the optimizing α-function for belief state b. Then the α-functions can be expressed as:*

$$\begin{aligned}
\alpha_0^{l^*(b),t}(0) &= K_t^a(-|0)\Big[r_t(0,a,-) + (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i(b,a,-),t+1}(0) \\
&\quad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,-),t+1}(s')ds'\Big] \\
&\quad + K_t^a(+|0)\Big[r_t(0,a,+) + \max_k \Big\{ (1-p_t^{LRR}-p_t^{SP})\alpha_0^{k,t+1}(0) \\
&\quad \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds\Big\}\Big] \\
\alpha_C^{l^*(b),t}(s) &= K_t^a(-|s)\Big[r_t(s,a,-) + \int_{S^C} f_t^C(s'|s)\alpha_C^{i(b,a,-),t+1}(s')ds'\Big] \\
&\quad + K_t^a(+|s)R_t(s), \qquad (17)
\end{aligned}$$

where

$$l^*(b) = \arg\max_k \Big\{ b(0)\alpha_0^k(0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b(s)\alpha_C^k(s)ds\Big\}. \qquad (18)$$

The function $\alpha_t^{l^*(b)}$ denotes the maximum expected number of QALYs a patient can attain, following the optimal policy.

3.2. Special case: exponentially distributed transitions

As can be seen in the results of the previous section, the expressions for the α-functions are rather complicated. In general, there is no guarantee that we can calculate the optimal value function explicitly without using numerical approximation methods. However, under some reasonable conditions on the transitions, observations and rewards we can prove that the α-functions, and thereby the optimal value function, can be obtained explicitly. This result is presented in the following proposition and corollary.

Proposition 3.2. *If the transitions are exponentially distributed and the rewards and observation probabilities are described by exponential functions, then*

$$\alpha_C^{i,t}(s) = \sum_{k=1}^{5} \beta_C^{k,t} e^{-\gamma_C^{k,t}s} \qquad C \in \{LRR, SP\}, \qquad (19)$$

for all i and $t = 0 \ldots T-1$ and certain parameters β and γ.

Proof If the transitions are exponentially distributed and the rewards and observation probabilities are described by exponential

functions, they can be written as:

$$\begin{aligned} f_t^C(x|s) &= \lambda e^{-\lambda^1(x-s)} && x > s \\ K_t^a(+|s) &= 1 - \kappa_t e^{-\kappa_t^1 s} \\ K_t^a(-|s) &= 1 - K_t^a(+|s) \\ &= \kappa_t e^{-\kappa_t^1 s} \\ R_t(s) &= \rho_t e^{-\rho_t^1 s} \\ p_t^d(s) &= 1 - p_t^d e^{-\nu^1 s}. \end{aligned}$$

Substituting the expression for $p_t^d(s)$ into the expression for the rewards (2) gives

$$\begin{aligned} r(s,a,o) &= p_t^d(s)0.5\sigma + (1 - p_t^d(s))\sigma - \mu_o^a \\ &= \nu_t e^{-\nu_t^1 s} - \mu_o^a. \end{aligned}$$

For $t = T$ we have

$$\begin{aligned} \alpha_T^i(s) &= R_T(s) \\ &= \rho_t e^{-\rho_t^1 s}, \end{aligned}$$

which is of the desired form. Now suppose that $\alpha_C^{i,t+1}(s) = \sum_{k=1}^5 \beta_C^{k,t+1} e^{-\gamma_C^{k,t+1} s}$ for $C \in \{LRR, SP\}$ and a certain $t+1$, then we have by Corollary 3.1

$$\begin{aligned} \alpha_C^{i,t}(s) &= K_t^a(-|s)\left[r_t(s,a,-) + \int_{S^C} f_t^C(s'|s)\alpha_C^{i,t+1}(s')ds'\right] \\ &\quad + K_t^a(+|s)R_t(s) \\ &= \kappa_t e^{-\kappa_t^1 s}\left[\nu_t e^{-\nu_t^1 s} - \mu_o^a + \int_0^\infty \lambda e^{-\lambda^1(x-s)} \right. \\ &\quad \left. \times \sum_{k=1}^5 \beta_C^{k,t+1} e^{-\gamma_C^{k,t+1} x} dx\right] + \left(1 - \kappa_t e^{-\kappa_t^1 s}\right)\rho_t e^{-\rho_t^1 s} \\ &= \kappa_t \nu_t e^{-(\kappa_t^1 + \nu_t^1)s} - \mu_o^a \kappa_t e^{-\kappa_t^1 s} + \rho_t e^{-\rho_t^1 s} - \kappa_t \rho_t e^{-(\kappa_t^1 + \rho_t^1)s} \\ &\quad + \left[\sum_{k=1}^5 \frac{\beta_C^{k,t+1}}{\lambda^1 + \gamma_C^{k,t+1}}\right]\kappa_t \lambda e^{-(\kappa_t^1 - \lambda^1)s} \end{aligned} \tag{20}$$

which is also of the desired form. By induction we conclude that the proposition holds. □

Remark The proposition only holds if the parameters for the transition probability density functions (λ) are constants, so they do not depend on s or depend on s trough an exponential relation. Furthermore, instead of proving the proposition for the optimal α-function $\alpha^{l^*(b),t}$ we prove it for an arbitrary α-function. The reason for this is that this simplifies the proof somewhat and that when we solve the problem, we first generate all α-functions before determining the optimal one (see Section 4).

With this closed form for the α-functions in the continuous states we can readily derive an expression for the values of the α-functions in the discrete state $S = \{0\}$.

Corollary 3.2. *If the transitions are exponentially distributed and the rewards and observation probabilities are described by exponential functions, then*

$$\alpha_0^{i,t}(0) = \beta_0^t \alpha_0^{i,t+1}(0) + \gamma_0^t \tag{21}$$

for all i and $t = 0 \ldots T-1$ and certain parameters β and γ.

Proof. By Corollary 3.1 $\alpha_0^{l^*(b),t}(0)$ is given by

$$\begin{aligned} \alpha_0^{l^*(b),t}(0) &= K_t^a(-|0)\left[r_t(0,a,-) + (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{i(b,a,-),t+1}(0) \right. \\ &\quad \left. + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,-),t+1}(s')ds'\right] \\ &\quad + K_t^a(+|0)\left[r_t(0,a,+) \right. \\ &\quad \left. + \max_k \left\{(1 - p_t^{LRR} - p_t^{SP})\alpha_0^{k,t+1}(0) \right. \right. \\ &\quad \left. \left. \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds\right\}\right]. \end{aligned}$$

Once again, since we do not need an explicit expression for the optimal α-function $\alpha_0^{l^*(b),t}(0)$ but instead for an arbitrary α-function, we can leave out the maximum over k and the index $i(b, a, o)$. This gives a simpler expression for $\alpha_0^{i,t}(0)$:

$$\begin{aligned} \alpha_0^{i,t}(0) &= \sum_o K_t^a(o|0)\left[r_t(0,a,o) + (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{i,t+1}(0) \right. \\ &\quad \left. + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i,t+1}(s')ds'\right] \\ &= (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{i,t+1}(0) \\ &\quad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i,t+1}(s')ds' \\ &\quad \sum_o K_t^a(o|0)r_t(0,a,o) \\ &= (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{i,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \lambda \sum_{k=1}^5 \frac{\beta_C^{k,t+1}}{\lambda^1 + \gamma_C^{k,t+1}} \\ &\quad + \nu_t - \kappa_t \mu_-^a - (1 - \kappa_t)\mu_+^a \\ &= \beta_0^t \alpha_0^{i,t+1}(0) + \gamma_0^t. \end{aligned}$$

Here, the second equation follows from the fact that $K_t^a(+|s) + K_t^a(-|s) = 1$. □

4. Algorithm

In this section we use the results from the previous section to construct an algorithm that generates the α-functions iteratively. Furthermore, we provide an algorithm that can obtain the α-functions efficiently for the special case mentioned in Section 3.2.

The general algorithm is based on the fact that the optimal value function V^* is PWLC. The algorithm was first stated by Smallwood and Sondik (1973) and later Monahan (1982) and Lovejoy (1991) simplified and adjusted it. All these algorithms where developed for discrete-state POMDPs. Because we modeled our problem as a continuous-state POMDP some modifications are needed but the main principles of the work cited remain valid for our case. The algorithm generates all possible α-functions using Eq. (17), deletes the non-optimal α-functions and uses the remaining α-functions and the expression of $V_t^*(b)$ in Theorem 3.1 to construct the optimal value function. The complete algorithm is stated below.

Algorithm α-functions algorithm.

1. **Initialize.** $\alpha_C^{1,T}(s) = r_T(s)$, for all $C \in \{0, LRR, SP\}$ $s \in S^C$, $\mathcal{A}_T = \{\alpha_C^1\}$ and $t = T - 1$
2. **Generate.** Generate $\mathcal{A}_t = \{\alpha_C^{1,t}, \alpha_C^{2,t} \ldots\}_{C\in\{0,LRR,SP\}}$ (by (22), see below) and mark all α-functions.
3. **Eagle's reduction.**
   (a) Select a marked α-function $\alpha_C^{i,t}$. If none exists go to step 4. Otherwise,
   (b) unmark the selected α-function and if there exists an $\alpha_C^{j,t}$ such that $\alpha_C^{i,t}(s) \le \alpha_C^{j,t}(s)$ for all $s \in S^C$ delete the selected α-function. Go to step 3(a).
4. **Time update.** If $t > 1$, then $t = t - 1$ and go to step 2, otherwise stop.

We now describe step 2 from the algorithm in more detail. Let

$$\mathcal{A}_{t+1} = \{\alpha_C^{1,t+1}, \alpha_C^{2,t+1}, \ldots\}_{C=\{0,LRR,SP\}}$$

denote the set of α-functions at time $t+1$. Now instead of determining the optimal α-function $\alpha^{I^*(b)_t}$ by Eq. (17) we generate the α-function for every combination of an action and an $\alpha_C^{i,t+1}$, let this be denoted by $\alpha_C^{(a,i),t}$. So we have

$$\mathcal{A}_t = \left\{\alpha_C^{(W,i),t}, \alpha_C^{(M,i),t}\right\}_{C\in\{0,LRR,SP\}}^{i=1\dots|A_{t+1}|},$$

with

$$\begin{aligned}\alpha_0^{(a,i),t}(0) &= \sum_o K_t^a(o|0)\Big[r_t(0,a,o) + (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i,t+1}(0) \\ &\quad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i,t+1}(s')ds'\Big] \\ \alpha_C^{(a,i),t}(s) &= K_t^a(-|s)\Big[r_t(s,a,-) + \int_{S^C} f_t^C(s'|s)\alpha_C^{i,t+1}(s')ds'\Big] \\ &\quad + K_t^a(+|s)R_t(s), \qquad C\in\{LRR,SP\}. \end{aligned} \tag{22}$$

When all the α-functions are generated for every decision epoch and the (completely) dominated ones are deleted, the optimal value function follows directly from the representation in Theorem 3.1. Furthermore, since every α-function has an action associated with it (22), the optimal action is easy to determine.

4.1. Exponential transitions

We now use the special structure of the α-functions in the exponential case, recalling Proposition 3.2 and Corollary 3.2, to determine the parameters that describe the α-functions.

For clarity we restate the expressions for the transition probability density functions and the expressions for the rewards, observation probabilities and probability of death, for which we now explicitly mention where they depend on:

$$\begin{aligned} f_t^C(x|s) &= \lambda^C e^{-\lambda^{C,1}(x-s)}, && x>s,\\ K_t^a(+|s) &= 1-\kappa_t^C e^{-\kappa_t^{C,1}s}\\ K_t^a(-|s) &= 1-K_t^a(+|s)\\ &= \kappa_t^C e^{-\kappa_t^{C,1}s}\\ R_t(s) &= \rho_t^C e^{-\rho_t^{C,1}s}\\ p_t^d(s) &= 1-p_t^{C,d}e^{-\nu^{C,1}s}\\ r(s,a,o) &= \nu_t^C e^{-\nu_t^{C,1}s} - \mu_o^{C,a}. \end{aligned}$$

The algorithm to determine the parameters of the optimal value function in the exponential case is stated in the following pseudocode:

Algorithm α-functions algorithm in the exponential case.

1. **Initialize.** $\alpha_C^{1,T}(s) = r_T(s) = \rho_T e^{-\rho_T^1 s}$, define $\beta_0^T(1) = \rho_T^0$, $\gamma_0^T(1) = \rho_T^{0,1}$, $\beta_C^{1,T}(1) = \rho_T^C$, $\gamma_C^{1,T}(1) = \rho_T^{C,1}$, for $C\in\{LRR,\ SP\}$, $\mathcal{A}_T = \{\alpha^1\}$, $i=1$ and $t=T-1$.
2. **Generate.** Generate $\mathcal{A}_t = \{\alpha_C^{1,t}, \alpha_C^{2,t}\dots\}_{C\in\{0,LRR,SP\}}$ by generating the β and γ parameters (20) and mark all α-functions.
   for $i=1$ to $|\mathcal{A}_{t+1}|$

   for $a = W, M$

$$\begin{aligned}\beta_0^t(a,i) &= (1-p_t^{LRR}-p_t^{SP})\\ \gamma_0^t(a,i) &= \sum_{C\in\{LRR,SP\}} p_t^C\lambda^C \sum_{k=1}^{5} \frac{\beta_C^{k,t+1}}{\lambda^{C,1}+\gamma_C^{k,t+1}}\\ &\quad + \nu_t^C - \kappa_t^C\mu_-^{C,a} - (1-\kappa_t^C)\mu_+^{C,a}\end{aligned}$$

$$\beta_C^{k,t}(a,i) = \begin{cases} \kappa_t^C\nu & k=1\\ -\mu_o^{C,a}\kappa_t^C & k=2\\ \rho_t^C & k=3\\ -\kappa_t^C\rho_t^C & k=4\\ \left[\sum_{k=1}^5 \frac{\beta_C^{k,t+1}}{\lambda^{C,1}+\gamma_C^{k,t+1}}\right]\kappa_t^C & k=5 \end{cases}$$

$$\gamma_C^{k,t}(a,i) = \begin{cases} \kappa_t^{C,1}+\nu^{C,1} & k=1\\ \kappa_t^{C,1} & k=2\\ \rho_t^{C,1} & k=3\\ \kappa_t^{C,1}+\rho_t^{C,1} & k=4\\ \kappa_t^{C,1}-\lambda^{C,1} & k=5 \end{cases}$$

   end

   end
3. **Eagle's reduction.**
   (a) Select a marked α-function $\alpha_C^{i,t}$. If none exists go to step 4. Otherwise,
   (b) unmark the selected α-function and if there exists an $\alpha_C^{j,t}$ such that $\alpha_C^{i,t}(s) \le \alpha_C^{j,t}(s)$ for all $s\in S^C$ delete the selected α-function. Go to step 3(a).
4. **Time update.** If $t>1$, then $t=t-1$ and go to step 2, otherwise stop.

5. Case study

To illustrate how the model can be applied in practice, we present the optimal follow-up plan for a stratification of the patients based on their age. We limit ourselves to the case in which the transitions within the continuous states (i.e. the growth model for the tumors) are exponentially distributed and where the observation probabilities, probability of death and the rewards are described by exponential relations (see Section 3.2). We first describe the parameters that are needed for the model and then the results.

5.1. Parameters

As stated before, our aim is to determine the optimal follow-up scheme for a patient based on the personal risk of recurrence. For this we need, in addition to the derived model, parameters based on the characteristics of the patient. In this section we elaborate on the parameters we need and their sources (see Table 1).

Based on the age we can estimate the probability that a patient will die between decision epochs. We obtain these probabilities from CBS (2017). If the age of patients in a certain group differs we use the probability of death for the average age, e.g. when the age in a group is between 40 and 50 we use the probability of death of a 45 year old woman.

The state transition probabilities between the discrete and the continuous states, i.e. the probability that a patient gets a second primary tumor or a LRR between two decision epochs, are obtained from the Netherlands Cancer Registry (NCR) (based on data from women first diagnosed with early breast cancer between 2003 and 2006 in all Dutch hospitals ($n = 37,230$)) (IKNL, 2017a; Witteveen et al., 2015). The estimates for the transitions within the continuous states (i.e. the grow rates of a second primary tumor and of a LRR) are also obtained from (IKNL, 2017a).

Estimations of the disutility associated with mammography vary between 0.5 and 1.5 days (Mandelblatt et al., 2002). We estimate it at 1 day. For biopsies the estimations of disutility vary

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

Table 1
The model parameters.

Parameter	Symbol	Source
Probability of death	ν	CBS (2017)
State transitions in S^{PO}	P	IKNL (2017a) and Witteveen et al. (2015)
Growth rate	λ	IKNL (2017a)
Disutility of a mammogram	$\mu_o^{C,M}$	Mandelblatt et al. (2002)
Disutility of a biopsy	$\mu_+^{C,\cdot}$	Velanovich (1995)
Specificity and sensitivity of mammography	κ^C	Kolb et al. (2002)
Specificity and sensitivity of self-detection	κ^C	ibid.
Survival rates	ρ^1	IKNL (2017a)
Life expectancy	ρ	CBS (2017)

Table 2
The optimal number of mammograms for various values of the parameters λ and ρ^1 compared to the optimal number of mammograms for women age 50–59.

λ	0.970	0.980	0.985	0.990	1.000
Optimal number of mammograms	−5	+0	5	+0	+1
ρ^1	0.25	0.3	0.35	0.40	0.45
Optimal number of mammograms	+1	+1	5	−2	−5
κ	0.93	0.95	0.97	0.99	1.00
Optimal number of mammograms	−2	−2	5	+0	+1

between 2 and 4 weeks (Velanovich, 1995), in our model we estimate it at 3 weeks. We assume that the mentioned disutilities do not depend on the age of the patient. We use Kolb, Lichy, and Newhouse (2002) to obtain the specificity and sensitivity of both mammography and self-detection.

The rewards that a patient receives upon leaving the decision process, either by detection of cancer or at the end of the follow-up phase are based on the healthy life expectancy. We use the healthy life expectancy at the beginning and at the end of the follow-up to derive a linear expression for the life expectancy of a healthy patient at each decision epoch. The expected remaining life years for patients in the different cancer states, i.e. the lump-sum and end rewards, are modeled to be exponentially decreasing with the growth of the tumor. These exponential relations are based on the 10-year survival rates for the different groups, which are also obtained from IKNL (2017a).

5.2. Results

Since the optimal policy will vary for different subgroups of patients, we present the results for four basic categories. These categories serve as an illustration and since age is a well risk factor we choose this for our stratification. The reader should bear in mind that the model can be applied to much more specified categories of patients.

The patients in the first category are up to 50 years old, in the second category 50–59 years old, in the third category 60–69 years old and in the fourth category 70 years old and above.

Since the probability of getting cancer is small (≈ 0.01) and the specificity of both mammography and self-detection is high (≈ 0.99), the majority (approximately 85%) of patients will never have a positive mammogram or a self-detection. We therefore present the optimal policy for a patient that never has a positive mammogram or a self-detection. The optimal policies for these patients, for each of the four categories, are given in Fig. 3. The bar charts represent the probability of cancer in every interval. This probability is divided in the probability of a LRR (in blue) and of a SP (in red). Note that the probability of a SP, between the seventh and eight decision epoch, for the above 70 category is not zero but very small, such that it is invisible in the bar chart. On top of the probabilities the optimal action at each decision epoch is given.

We see that it is optimal to intensify the screening when the probability of an LRR peaks and just after that. Also, as the age of the patient increases, the optimal number of mammograms decreases. This is because the probability of a recurrence is lower for older patients and the remaining life expectancy is lower, so there is less to gain by early detection. Given the proportion of patients being 26.24%, 28.45%, 22.60% and 22.70%, respectively, for the age categories and they currently receive five annual mammograms, the new policy would result in 17% less mammograms. Using different stratifications could also result in different numbers. But very large differences are not expected. Also, the ethical questions such as offering a minimum of follow-up could change the actual policy.

5.3. Sensitivity analysis

For several of the input parameters there is no specific data available so that we have to use rather rough estimates. We therefore conduct a basic sensitivity analysis in order to provide insight in how sensitive the optimal policy is to small changes in these input parameters. We use the number of mammograms as a metric to compare the optimal policies for various values of the input parameters. We do this for the input parameters λ, the growth rate of the tumor, ρ^1, the rate at which the life expectancy decreases for a patient with cancer and κ, the specificity of a mammogram.

We compare the optimal number of mammograms for various values of the input parameters λ and ρ^1 with the case of patients aged 50–59 as described in the previous section. The results are given in Table 2. The values of the parameters are selected such that they are within the bounds provided in literature, see Section 5.1. We see that the optimal number of mammograms is not very sensitive to the tumor growth parameter λ for $0.980 \leq \lambda \leq 1.000$ but that there is a sharp decrease for $\lambda = 0.970$. The optimal number of mammograms is more sensitive to variations in the input parameters ρ^1 and κ. The main conclusion we can draw from this sensitivity analysis is that the model is quite sensitive to some of the input parameters for which only rough estimates are available. This means that before using this decision

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

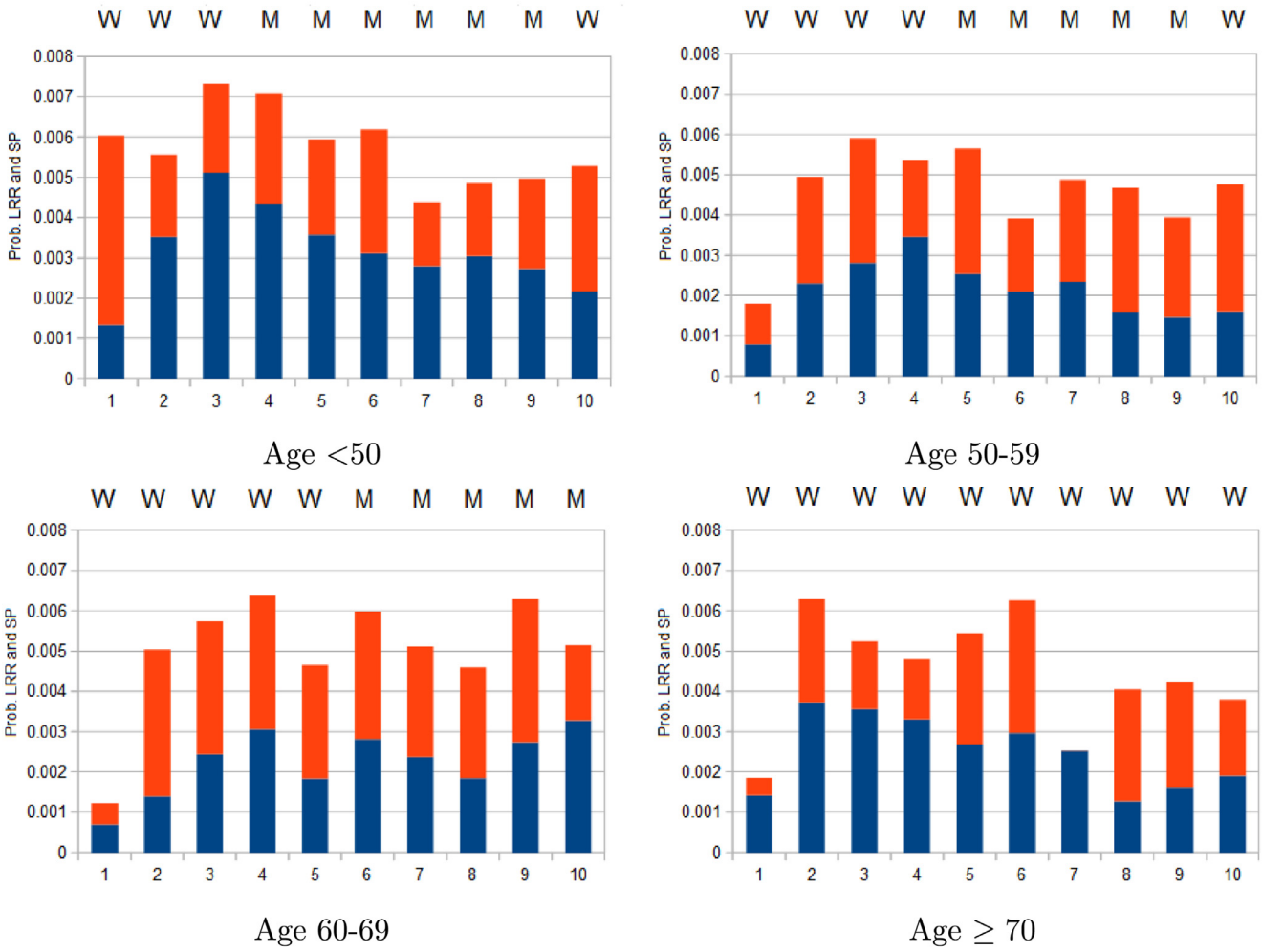

Fig. 3. Probability of a LRR (blue) and a SP (red) and the optimal policy for different age categories. W stands for wait, M stands for mammogram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

model in practice further research has to be done into obtaining better estimates for the model parameters.

6. Conclusions and discussion

Currently, follow-up for breast cancer patients consists of annual mammography for five years. Even though previous research shows that the probability of recurrence is highly correlated with the personal characteristics of a patient, follow-up is not differentiated. Follow-up tailored to the individual case is suggested by the national guidelines but without implementation in practice. In earlier research this sort of problems were modeled by discrete-state POMDPs (Ayer et al., 2012; Otten et al., 2017). Because of limitations, discussed earlier, we model the problem as a continuous-state POMDP. For this POMDP we derive an expression for the optimal value function. For this optimal value function we prove an alternative representation described by the α-functions. From this alternative description an iterative scheme can be deduced to obtain the optimal value function for every belief state at every decision epoch. In general, the solution algorithm for the optimal value function can only be carried out with numerical methods. We prove that under some restrictions on the dynamics of the underlying Markov chain, we can calculate the optimal value function exactly. In particular, we assume the transition probabilities to be exponentially distributed and that the rewards are described by an exponential function. Similar results may be derived for various specific transition probability distributions, depending on the context of the problem.

As an illustration of how this model can be used in practice, we determine the optimal policy for groups of patients. Because the age of the patients is known to be of large influence on the risk of a recurrence we make a stratification of the patients based on their age. The outcome suggests that it is optimal to test the patient more often just after the peak of risk of a recurrence and to reduce the number of tests when the age increases. For the oldest group of patients it seems optimal to not test at all.

Compared to the discrete model (Otten et al., 2017) there are some differences and some similarities. As with the discrete model the results suggest that it is optimal to reduce the number of mammograms as the age of the patient increases. Both models also suggest that the testing should be intensified just after the peak in the probability of a recurrence. The optimal number of mammograms differs, especially for the eldest group of patients. Because the input parameters for the discrete and the continuous model are based on the same data, the differences in the optimal policy between the two models is mainly due to the different manner in which the growth of the tumors is modeled, namely in a discrete and a continuous manner.

In our model, the time between decision epochs is fixed. A possibility for future research would be to model the problem as a continuous time POMDP where decision can take place at any time, or with variable time lengths between decision epochs.

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046

A large limitation of our study is that the estimates for some of the model parameters are quite inexact and that the outcomes are rather sensitive for these parameters; this is in particular the case for the life expectancy model. Therefore, further study is needed before the model can be used in practice.

References

Aström, K. (1965). Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications, 10*(1), 174–205.

Ayer, T., Alagoz, O., & Stout, N. K. (2012). A POMDP approach to personalize mammography screening decisions. *Operations Research, 60*(5), 1019–1034. https://doi.org/10.1287/opre.1110.1019.

Ayvaci, M. U. S., Alagoz, O., & Burnside, E. S. (2012). The effect of budgetary restrictions on breast cancer diagnostic decisions. *MSOM, 14*(4), 600–617. doi:10.1287/msom.1110.0371.

Beaver, K., Tysver-Robinson, D., Campbell, M., Twomey, M., Williamson, S., Hindley, A., et al. (2009). Comparing hospital and telephone follow-up after treatment for breast cancer: Randomised equivalence trial. *BMJ, 338*, a3147. doi:10.1136/bmj.a3147.

CBS (2017). Statline. [accessed 7-March-2017] http://statline.cbs.nl/Statweb/.

Duff, M. (2002). *Optimal learning: Computational procedures for Bayes-adaptive Markov decision processes*. University of Massassachusetts. Ph.D. thesis.

Geurts, S. M. E., de Vegt, F., Siesling, S., Flobbe, K., Aben, K. K. H., van der Heiden-van der Loo, M., et al. (2012). Pattern of followup care and early relapse detection in breast cancer patients. *Breast Cancer Research and Treatment, 136*, 859–868. doi:10.1007/s10549-012-2297-9.

IKNL (2017a). Nederlandse kankerregistratie. [accessed 7-March-2017] http://www.cijfersoverkanker.nl/.

IKNL (2017b). Richtlijnen oncologische zorg. [accessed 7-March-2017] http://www.oncoline.nl/.

Kolb, T. M., Lichy, J., & Newhouse, J. H. (2002). Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology, 225*, 165–175. doi:10.1148/radiol.2251011667.

Lovejoy, W. S. (1991). A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research, 28*(1), 47–65. https://doi.org/10.1007/BF02055574.

Lu, W. L., Jansen, L., Post, W. J., Bonnema, J., van de Velde, J. C., & Bock, G. H. D. (2009). Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: A meta-analysis. *Breast Cancer Research and Treatment, 114*, 403–412. https://doi.org/10.1007/s10549-008-0023-4.

Mandelblatt, J. S., Wheat, M. E., Monane, M., Moshief, R. D., Hollenberg, J. P., & Tang, J. (2002). Breast cancer screening for elderly women with and without comorbid conditions: A decision analysis model. *Annals of Internal Medicine, 116*(9), 722–730. https://doi.org/10.7326/0003-4819-116-9-722.

Monahan, G. E. (1982). A survey of partially observable Markov decision processes: Theory, models and algorithms. *Management Science, 28*(1), 1–16. https://doi.org/10.1287/mnsc.28.1.1.

Moossdorff, M., van Roozendaal, L. M., Strobbe, L. J. A., Aebi, S., Cameron, D. A., Dixon, J. M., et al. (2014). Maastricht delphi consensus on event definitions for classification of recurrence in breast cancer research. *Journal of the National Cancer Institute, 106*(12), 1–7. https://doi.org/10.1093/jnci/dju288.

Otten, J. W. M., Witteveen, A., Vliegen, I. M. H., Siesling, S., Timmer, J. B., & IJzerman, M. J. (2017). *Stratified breast cancer follow-up using a partially observable MDP*. In R. J. Boucherie, & N. M. van Dijk (Eds.) (pp. 223–244). Springer International Publishing.

Pennery, E., & Mallet, J. (2000). A preliminary study of patients' perceptions of routine follow-up after treatment for breast cancer. *The European Journal of Oncology Nursing, 4*(3), 138–145. https://doi.org/10.1054/ejon.2000.0092.

Porta, J. M., Spaan, M. T. J., & Vlassis, N. (2004). Value iteration for continuous-state POMDPs. *Technical report*. Informatics Institute, University of Amsterdam.

Porta, J. M., Spaan, M. T. J., & Vlassis, N. (2005). Robot planning in partially observable continuous domains. In *Proceedings of the 2005 robotics: Science and systems* (pp. 217–224). MIT Press.

Porta, J. M., Vlassis, N., Spaan, M. T., & Poupart, P. (2006). Point-based value iteration for continuous POMDPs. *The Journal of Machine Learning Research, 7*, 2329–2367.

Puterman, M. L. (1994). *Markov decision processes: discrete stochastic dynamic programming* (1st). New York, NY, USA: John Wiley & Sons, Inc..

Smallwood, R. D., & Sondik, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research, 21*(5), 1071–1088. https://doi.org/10.1287/opre.21.5.1071.

Sondik, E. J. (1971). *The optimal control of partially observable Markov processes*. Stanford University. Ph.D. thesis.

Sonnenberg, F. A., & Back, J. R. (1993). Markov models in medical decision making, a practical guide. *Medical Decision Making, 13*(4), 322–338. https://doi.org/10.1177/0272989X9301300409.

Steimle, L. N., & Denton, B. T. (2017). *Markov decision processes for screening and treatment of chronic diseases*. In R. J. Boucherie, & N. M. van Dijk (Eds.) (pp. 189–222). Springer International Publishing.

Velanovich, V. (1995). Immediate biopsy versus observation for abnormal findings on mammograms: An analysis of potential outcomes and costs. *The American Journal of Surgery, 170*(4), 327–332. https://doi.org/10.1016/S0002-9610(99)80298-0.

Witteveen, A., Vliegen, I. M. H., Sonke, G. S., Klaase, J. M., IJzerman, M. J., & Siesling, S. (2015). Personalisation of breast cancer follow-up: A time-dependent prognostic nomogram for the estimation of annual risk of locoregional recurrence in early breast cancer patients. *Breast Cancer Research and Treatment, 152*, 627–636. https://doi.org/10.1007/s10549-015-3490-4.

Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., & Inman, B. A. (2012). Optimization of PSA screening policies: A comparison of the patient and societal perspectives. *Medical Decision Making, 32*(1), 337–349. https://doi.org/10.1177/0272989X11416513.

Please cite this article as: M. Otten, J. Timmer and A. Witteveen, Stratified breast cancer follow-up using a continuous state partially observable Markov decision process, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2019.08.046