UNIVERSITY OF TWENTE.

Master Thesis

Applied Mathematics

Stratified Breast Cancer Follow-Up Using a Continuous-state POMDP.

Author:
J.W.M. Otten

Supervisors:
Dr. J.B. Timmer
A. Witteveen, Msc.

March 27, 2017

Abstract

Frequency and duration of follow-up for patients with breast cancer is still under discussion. Current 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 (LRR) or second primary tumor. Aim of this study is to formulate a continuous-state POMDP, in which at every epoch a decision is made whether or not to test, based on the personal risk factors of the patient. We show that the optimal value function of the POMDP is piece-wise linear and convex (PWLC). This result provides an alternative expression of the optimal value function, which leads to a solution algorithm for the POMDP. Under some conditions the optimal value function can be obtained by a simple solution algorithm. We present results for this case to illustrate how the model can be applied in practice.

1 Introduction

After curative treatment for breast cancer, patients are followed clinically for a period of time to detect locoregional recurrences (LRRs) in an early phase [1]. A LRR is defined as the reappearance of breast cancer on the same site as the primary tumor [2]. Because a LRR has a high risk of distant metastases it is important that it is detected in an early stage [1]. Currently, in the Netherlands, patients have regular follow-up for at least five years after their treatment [3]. However only a minority of the LRRs discovered are detected at a scheduled check-up, more often the patient detects it in between check-ups [4]. Furthermore, due to an increasing incidence and survival rate, the number of patients currently in the follow-up phase increases and becomes more of a burden to health care.

Even though it is known that certain factors, such as tumor size, the patient's age and treatment of the primary tumor are highly correlated with the risk of a LRR, there is no differentiation in the follow-up policy for different categories of patients [5]. Since 2012 the national guideline of the Netherlands proposes that the follow-up should be tailored to the individual situation of the patient and that personal risk factors should be taken into account but it does not give any specific recommendations on how to effectuate this.

These observations together give rise to the question whether it is possible to improve the current follow-up policy, both from patient and health care perspective. Our aim is to develop a sequential decision process in which a decision maker, the patient and/or a physician, chooses at every decision epoch whether or not to have a check-up. We model this problem using a partially observable Markov decision process (POMDP), which is a generalization of a Markov decision process and allows us to model a sequential decision making process in which the information about the true state of the system is incomplete. Because the true health state of a patient, i.e. whether a patient is disease-free or not, is only partially observable, a POMDP is ideally suited to this problem [6].

In previous research we modeled the problem described by a discrete-state POMDP [7]. In this model a 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 an late stage and the prognosis is rather bad. We found this model usable for the problem however we also found that the outcome is quite sensitive to the transition probability between the early and the late state. 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.

Ayer et al. [8] have developed a POMDP model to a similar problem, a mammography decision model for preventive screening for breast cancer. Ayvaci et al. [9] give an analysis of the same problem as Ayer et al. but under budgetary constraints, however they model it as a normal Markov decision process and thus simplifying the problem considerably. Zhang et al. [10] made a comparison of the patient and societal perspectives for a similar case, PSA screening policies, via a POMDP approach. However, all of the models are based on an underlying discrete-state Markov chain and therefore simplifying the health-state of the 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. Porta et al. [11] [12] developed a continuous-state model for robot planning and

proved some important analytical results. Duff [13] also provides some useful results for continuous-state POMDPs. These results also lay the basis for the solution method for the POMDP. This research provides very useful models and results for our problem 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 [8]. 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 leads to some complications that need to be addressed.

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 by a finite set of states, we model it as a continuum of states. Secondly, we proof some important results in order to derive a solution algorithm for finding the optimal testing schedule. Thirdly we derive a simple solution algorithm for the optimal policy under some restriction on the growth model of the tumor.

The remainder of this report is organized as follows. In §2 we state some preliminary information on standard POMDPs and present the continuous-state POMDP model for our specific problem. in §3 we derive the optimality equations. We also provide an alternative representation of the optimality equations. 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 optimality equations. In §4 we present the general algorithm for solving the POMDP and an algorithm for a special case. In §5 we will present an illustration how this model can be applied in practice. Finally we summarize the results and conclude in §6.

2 Model

In this section we describe the model for the given problem. The problem described is modeled by a Partially Observable Markov Decision Process (POMDP). To incorporate some specific aspects of our problem we need to make some adjustments to the regular framework of POMDPs. For clarity we will first describe a standard POMDP and based on this we will present the model for our problem.

2.1 Preliminaries: POMDPs

A POMDP is generalization of a Markov Decision process [14]. It models 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 system states.
- A, the set of actions.
- O, the set of observations.
- An observation model described by $K^a(o|s)$, the probability that observation o was done given that the state was s and action a was taken.
- An underlying Markov Chain that models the transitions of the system's state. This is described by $P^{(a,o)}(s'|s)$ which is the probability that the next

state is s' given that the previous state was s and action a was taken and observation o was done.

• A reward function $r(s, a, o)$, which is the reward when the state is s and action a was taken and observation o done.

Because the decision maker can not observe the system's state directly, the knowledge about the system is represented by the so-called belief state. The belief state is a probability distribution over the state space based on the internal dynamics of the system, the actions taken and the observations done. When the current state is s and action a is taken and observation o is done, the new belief, τ is computed with a Bayesian update [15]

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

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

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

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 as $\pi^*(b) = \arg\max_a V_t^*(b)$.

At first it may seem that solving the optimal value function is intractable, because it is defined on a continuous belief space. However it can be proven that the optimal value function is piece-wise linear and convex (PWLC)[15][16]. Because the optimal value function is PWLC it 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-called α-vectors. These α-vectors can be calculated in a recursive way. This provides a straightforward way to obtain the optimal policy.

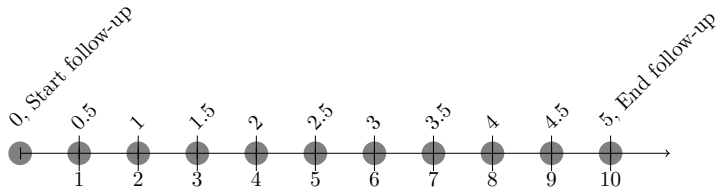
2.2 Model Formulation

In the problem described we want to model the growth of tumor as a continuous process also, because the patient can die during the process and the process terminates whenever the patient goes into treatment, we need to modify the POMDP framework described in order to model our problem correctly. The problem described 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).

Twice a year, a decision is made whether a patient should have a mammogram or should wait for another 6 months. The decision made is based on

the patient's current risk of cancer, which, among others, depends on several personal risk factors and prior test results. When the decision is made that the patient should have a mammogram and the result is positive or if a self-detection is made, it is followed by a biopsy. We assume that the biopsy after a positive mammogram or self-detection is perfect. When this perfect test is also positive (i.e. the patient has cancer) we assume that she starts treatment immediately and leaves the decision process by moving to the treatment state, otherwise the decision process proceeds to the next decision epoch. Also when the mammogram is negative or the decision is to wait for another 6 months and no self-detection is made, the decision process proceeds to the next decision epoch where the same decision has to be made. For our notation we follow Ayer et al.[8] and Otten et al. [7]. Throughout this report we refer to a person in our model as 'she', because breast cancer is very rare for men, and as 'the patient', irrespective of her true health state. The complete model and the notation used is as follows:

• Decision epochs, $t = 1 \cdots T$, $T = 10$. We assume that decisions are made twice every year and that the decision process starts 6 months after treatment of the primary tumour finished, so $t = 5$ denotes 2.5 years after primary treatment (see timeline below). Let σ denote the time between two succesive decision epochs, $\sigma = 0.5$ year. The decision horizon is at $t = 10$ because the annual check-ups are stopped after 5 years[3] (depending on the age of the patient the check-ups after this are annual, biennial or stopped).



• Core state space, $S = \{0, S^{LRR}, S^{SP}, 3, 4\}$, where $S^{LRR} = S^{SP} = \mathbb{R}_+$. The core state space consists of three discrete states $\{0, 3, 4\}$, where 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 and represents a measure, e.g. the size of the tumor, for the state of a LRR and a SP, respectively. To see how these different states are connected see figure 1.s_t is the true health state of the patient at time t. We model a LRR and a SP as continuous variables to incorporate the difference in expected remaining QALYs between earlier and later detection as good as possible because early detection of a tumor yields a better prognosis[1]. 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, $\Pi(S)$, the space of all probability distributions over the core state space S. A function $\pi \in \Pi(S)$ is called an information state.

• Belief space, $B(S^{PO})$, the space of all probability distributions over the partially observable states, S^{PO}. A function $b \in B(S^{PO})$ is in fact the same function as $\pi \in \Pi(S)$, only defined on a subset of partially observable states. This reduction of the information function makes sense because the probability

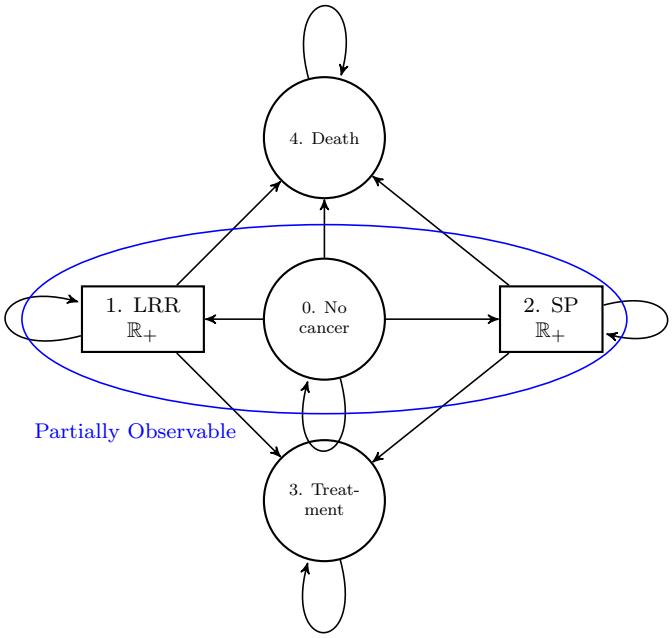


Figure 1: State diagram of the underlying Markov process.

that the true state $s = 3, 4$ is either 0 or 1.
For clarity we define a belief vector $b = [b(0)\ b(S^{LRR})\ b(S^S)]$, which denotes the belief that a patient is in state $0, 1$ or 2 and belief functions $b_{LRR}(s), b_{SP}(s)$ which denote the belief that a patients true health state is $s \in \mathbb{R}_+$ given the patient is in the continuous state LRR or SP respectively.

• Actions, A_t, the set of actions at time t. $a_t \in A_t = \{W, M\}$, where W stands for wait and M for mammography. The action set is only defined for $s \in S^{PO}$ since the process terminates whenever the patient dies or goes to the treatment state.

• Observation space, Θ_a, the set of possible observations when action a is selected. If $a_t = M$, the possible observations are a positive mammogram (M^+) or a negative mammogram (M^-). If $a_t = W$, the patient can either make a self-detection (SD^+) or no self-detection (SD^-). So 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^-, M^- with $-$ and SD^+, M^+ with $+$.

• Observation probabilities. $K_t^a(o|s)$ is the probability of making, at time t, observation o when decision a was taken while in state 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 mammography

by $spec_t(M)$ and of self-detection by $spec_t(SD)$. Similarly, the sensitivity of mammography 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. From these observations we can obtain the observation probabilities as follows:

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

• Core state transitions. $P_t^{(a,o)}(\cdot|s)$ is the distribution function of the transition at time t when the current state is s and action a was taken and observation o observed. Because the state space contains both discrete and continuous states, these probability distributions can be discrete, continuous or a mixture of both. However, 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 continuous probability distribution occurs. In state 0 the transitions are as follows: With a certain probability, say 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 actual 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 $f^C(x|0)$. The growth in state $s \in S^{LRR}, S^{SP}$ is X, where X is a continuous random variable with probability density function $f^C(x|s)$, $C = LRR, SP$.

• Updated belief space. $\tau[b, a, o]$ defines the belief (i.e. the probability distribution over the partially observable states) at time $t + 1$ when the belief about patient's true health state at time t was b and action a was taken and observation o was made. In particular, $\tau[b, a, o](s) = P_{t+1}(s|b, a, o)$ for $s = 0$ and$\tau[b, a, o](s) = f_{t+1}(s|b, a, o)$ for $s \neq 0$. 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:

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

• Rewards. $r_t(s, a, o)$ is 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 was made. To factor in the probability that a patient dies between two

decisions we use the half-cycle correction method [17]. The idea of this correction method is that if the patient dies between two decision epochs it is assumed that half of the cycle length σ is accrued to the expected number of QALYs. From this QALYs are subtracted for the disutility of a mammogram and a biopsy, when a patient should have one of these. Note that 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. 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 rewards in the treatment and death state are zero.

The expected reward between time t and $t+1$ when the true health state is s and the action chosen is a is 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 dis_M, dis_B the disutility experienced when undergoing a mammogram and a biopsy, respectively. The rewards for $t = 1, \cdots, T-1$ are:

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

3 Optimality Equations

We want to derive optimality equations for the number of QALYs a patient can obtain in order to determine the optimal policy of a patient. Let $V_t^*(\pi)$ denote this quantity when her information state is $\pi \in \Pi(S)$ at time t. Likewise let $V_t^*(b)$ denote the same quantity when the patient's belief state is $b \in B(S^{PO})$. Because the process terminates when in one of the treatment states or the death state, $V_t^*(\pi)$ can be expressed as:

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

Let $V_t^a(b)$ denote the maximum total expected QALYs a patient can obtain when at time t in belief state b and action a is chosen. Then $V_t^*(b)$ can be written as:

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

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

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

$$
\begin{aligned}
V_t^*(b) = \max_a \Bigg\{ & b(0)K_t^a(-|0)\Big[r_t(0,a,-) + V_{t+1}^*(\tau[b,a,-])\Big] \\
& + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)\Big[r_t(s,a,-) + V_{t+1}^*(\tau[b,a,-])\Big] ds \\
& + b(0)K_t^a(+|0)\Big[r_t(0,a,+) + V_{t+1}^*(\tau[b,a,+])\Big] \\
& + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds\Bigg\} \\
& \qquad\qquad\qquad\qquad\qquad\qquad t = 1\cdots T-1 \\
V_T^*(b) = b(0)r_T(0) & + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)r_T(s)ds \qquad (5)
\end{aligned}
$$

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. At time $t = 1\ldots T-1$ it is the probability that a patient is in a state, times the probability that a certain observation occurs, times the immediate and future rewards associated with this line of events.

3.1 Alternative Representation of the Optimality Equations

The key idea of value iteration, which is one of the most widely used methods for solving any type of Markov decision process, is relating the optimal value function V^* at time t to V^* at time $t+1$ [18]. As derived in the previous section the optimal value function of this particular problem is defined on the continuous space $B(S^{PO})$. So for solving the optimal value function at time t one would need the optimal value function at time $t+1$ on a continuous space and therefore a infinite dimensional vector would be needed to store these value functions. Fortunately it can be proven that for a POMDP the optimal value function is piecewise linear and convex (PWLC) and can therefore be represented as the maximum over a finite number of finite-dimensional vectors. This result is stated 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*

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

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

The theorem can be proven by induction. The proof goes in a similar way as the proof in the discrete case as proven by Smallwood et al.[15] and as the proof in the continuous case as proven by Porta et al. [19].

The optimality equations can now be written in terms of the α-functions.

Proposition 3.1 *The following representation of the optimality equations is equivalent to the optimality equations given in (4).*

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

Where

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

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:

$$V_{t+1}^*(\tau[b,a,o]) = \begin{cases} \max_k \Bigg\{ \frac{b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds}\alpha_0^{k,t+1}(0) \\ \quad + \sum_{C\in\{LRR,SP\}} \int_{S^C} \frac{b(0)K_t^a(-|0)f_t^C(s'|0) + \int_{S^C} b_C(s)K_t^a(-|s)f_t^C(s'|s)ds}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds}\alpha_C^{k,t+1}(s')ds' \Bigg\} & \text{if } o = - \\ \max_k \Big\{(1-p_t^{LRR}-p_t^{SP})\alpha_0^{k,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds\Big\} & \text{if } o = + \end{cases} \tag{9}$$

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

$$V^*_{t+1}(\tau[b,a,o]) = \begin{cases} \frac{1}{b(0)K^a_t(-|0)+\sum\limits_{C\in\{LRR,SP\}}\int\limits_{S^C} b(S)K^a_t(-|s)ds} \\ \quad \times \max_k \Big\{ b(0)K^a_t(-|0)(1-p^{LRR}_t-p^{SP}_t)\alpha^{k,t+1}_0(0) \\ \quad + \sum\limits_{C\in\{LRR,SP\}}\int\limits_{S^C} \Big[b(0)K^a_t(-|0)f^C_t(s'|0) \\ \quad + \int_{S^C} b_C(s)K^a_t(-|s)f^C_t(s'|s)ds \Big] \alpha^{k,t+1}_C(s')ds' \Big\} & \text{if } o=- \\ \max_k \Big\{ (1-p^{LRR}_t-p^{SP}_t)\alpha^{k,t+1}_0(0) \\ \quad + \sum\limits_{C\in\{LRR,SP\}} p^C_t \int\limits_{S^C} f^C_t(s|0)\alpha^{k,t}_C(s)ds \Big\} & \text{if } o=+ \end{cases} \tag{10}$$

$$V^*_{t+1}(\tau[b,a,o]) = \begin{cases} \frac{b(0)K^a_t(-|0)(1-p^{LRR}_t-p^{SP}_t)\alpha^{i(b,a,o),t+1}_0(0)}{b(0)K^a_t(-|0)+\sum\limits_{C\in\{LRR,SP\}}\int\limits_{S^C} b_C(s)K^a_t(-|s)ds} \\ \quad + \frac{b(0)K^a_t(-|0)\sum\limits_{C\in\{LRR,SP\}} p^C_t \int\limits_{S^C} f^C_t(s'|0)\alpha^{i(b,a,o),t+1}_C(s')ds'}{b(0)K^a_t(-|0)+\sum\limits_{C\in\{LRR,SP\}}\int\limits_{S^C} b_C(s)K^a_t(-|s)ds} \\ \quad + \frac{\sum\limits_{C\in\{LRR,SP\}} p^C_t \int\limits_{S^C} b_C(s)K^a_t(-|s)\int_{S^C} f^C_t(s'|0)\alpha^{i(b,a,o),t+1}_C(s')ds'ds}{b(0)K^a_t(-|0)+\sum\limits_{C\in\{LRR,SP\}}\int\limits_{S^C} b(S)K^a_t(-|s)ds} & \text{if } o=- \\ \max_k \Big\{ (1-p^{LRR}_t-p^{SP}_t)\alpha^{k,t+1}_0(0) \\ \quad + \sum\limits_{C\in\{LRR,SP\}} p^C_t \int\limits_{S^C} f^C_t(s|0)\alpha^{k,t}_C(s)ds \Big\} & \text{if } o=+ \end{cases} \tag{11}$$

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

$$
\begin{aligned}
V_t^*(b) = \max_a \Bigg\{ & b(0) K_t^a(-|0) \Big[r_t(0,a,-) + V_{t+1}^*(\tau[b,a,-]) \Big] \\
& + \sum_{C \in \{LRR, SP\}} \int_{S^C} b_C(s) K_t^a(-|s) \Big[r_t(s,a,-) + V_{t+1}^*(\tau[b,a,-]) \Big] ds \\
& + b(0) K_t^a(+|0) \Big[r_t(0,a,+) + V_{t+1}^*(\tau[b,a,+]) \Big] \\
& + \sum_{C \in \{LRR, SP\}} \int_{S^C} b_C(s) K_t^a(+|s) R_t(s) ds \Bigg\} \\
= \max_a \Bigg\{ & b(0) K_t^a(-|0) r_t(0,a,-) + \sum_{C \in \{LRR, SP\}} \int_{S^C} b_C(s) K_t^a(-|s) r_t(s,a,-) ds \\
& + \Big[b(0) K_t^a(-|0) + \sum_{C \in \{LRR, SP\}} \int_{S^C} b_C(s) K_t^a(-|s) ds \Big] V_{t+1}^*(\tau[b,a,-]) \\
& + b(0) K_t^a(+|0) r_t(0,a,+) + \sum_{C \in \{LRR, SP\}} \int_{S^C} b_C(s) K_t^a(+|s) R_t(s) ds \\
& + b(0) K_t^a(+|0) V_{t+1}^*(\tau[b,a,+]) \Bigg\} \qquad (12)
\end{aligned}
$$

Finally, by substituting the expression derived for $V_{t+1}^*(\tau[b,a,o])$ (11) into the rewritten expression for $V_t^*(b)$ (12) we have:

$$
\begin{aligned}
V_t^*(b) = \max_a \Bigg\{ & b(0)K_t^a(-|0)r_t(0,a,-) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)r_t(s,a,-)ds \\
& + \left[b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds \right] \qquad (13) \\
& \times \Bigg[\frac{b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})\alpha_0^{i(b,a,o),t+1}(0)}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds} \\
& + \frac{b(0)K_t^a(-|0) \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,o),t+1}(s')ds'}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds} \\
& + \frac{\sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) \int_{S^C} f_t^C(s'|s)\alpha_C^{i(b,a,o),t+1}(s')ds'ds}{b(0)K_t^a(-|0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)ds} \Bigg] \\
& + b(0)K_t^a(+|0)r_t(0,a,+) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds \\
& + b(0)K_t^a(+|0) \max_k \Bigg\{ (1-p_t^{LRR}-p_t^{SP})\alpha_0^{k,t+1}(0) \\
& \qquad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds \Bigg\} \Bigg\} \\
= \max_a \Bigg\{ & b(0)K_t^a(-|0)r_t(0,a,-) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s)r_t(s,a,-)ds \\
& + \Bigg[b(0)K_t^a(-|0)(1-p_t^{LRR}-p_t^{SP})\alpha_0^{i(b,a,o),t+1}(0) \\
& + b(0)K_t^a(-|0) \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i(b,a,o),t+1}(s')ds' \\
& + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(-|s) \int_{S^C} f_t^C(s'|s)\alpha_C^{i(b,a,o),t+1}(s')ds'ds \Bigg] \\
& + b(0)K_t^a(+|0)r_t(0,a,+) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b_C(s)K_t^a(+|s)R_t(s)ds \\
& + b(0)K_t^a(+|0) \max_k \Bigg\{ (1-p_t^{LRR}-p_t^{SP})\alpha_0^{k,t+1}(0) \\
& \qquad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s|0)\alpha_C^{k,t}(s)ds \Bigg\} \Bigg\} \qquad (14)
\end{aligned}
$$

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 repre-

sentation for solving the POMDP.

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

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

Where

$$l^*(b) = \arg\max_k \left\{ b(0)\alpha_0^k(0) + \sum_{C\in\{LRR,SP\}} \int_{S^C} b(s)\alpha_C^k(s)ds \right\} \qquad (16)$$

$\alpha_t^{l^*(b)}$ can be interpreted as the maximum expected number of QALYs a patient can attain when she follows 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, we have 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*

$$\alpha_C^{i,t}(s) = \sum_{k=1}^{5} \beta_C^{k,t} e^{-\gamma_C^{k,t} s} \qquad C \in \{LRR, SP\} \qquad (17)$$

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, then 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

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

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 clearly 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

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

Which is also of the desired form. So 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. Also note that 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)). So for solving the problem we do not need an explicit expression for the α-function $\alpha^{l^*(b),t}$.

With this closed form for the α-functions over the continuous states we can easily 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*

$$\alpha_0^{i,t}(0) = \beta_0^{k,t}\alpha_0^{i,t+1}(0) + \gamma_0^{k,t} \tag{18}$$

$$\tag{19}$$

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

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

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

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 $\alpha_0^{i,t}(0)$

$$\begin{aligned}
\alpha_0^{i,t}(0) &= \sum_o K_t^a(o|0)\Big[r_t(0,a,o) + (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i,t+1}(0) \\
&\qquad + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i,t+1}(s')ds'\Big] \\
&= (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i,t+1}(s')ds' \\
&\quad \sum_o K_t^a(o|0)r_t(0,a,o) \\
&= (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \int_0^\infty \lambda e^{-\lambda^1 x}\sum_{k=1}^5 \beta_C^{k,t+1} e^{-\gamma_C^{k,t+1}x}dx \\
&\quad + \nu_t - \kappa_t\mu_-^a - (1-\kappa_t)\mu_+^a \\
&= (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \lambda\sum_{k=1}^5 \beta_C^{k,t+1}\int_0^\infty e^{-(\lambda^1+\gamma_C^{k,t+1})x}dx \\
&\quad + \nu_t - \kappa_t\mu_-^a - (1-\kappa_t)\mu_+^a \\
&= (1-p_t^{LRR}-p_t^{SP})\alpha_0^{i,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \lambda\sum_{k=1}^5 \frac{\beta_C^{k,t+1}}{\lambda^1+\gamma_C^{k,t+1}} \\
&\quad + \nu_t - \kappa_t\mu_-^a - (1-\kappa_t)\mu_+^a \\
&= \beta_0^{k,t}\alpha_0^{i,t+1}(0) + \gamma_0^{k,t}
\end{aligned}$$

□

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

4 Algorithm

The general algorithm we use is based on the fact that the optimal value function V^* is PWLC. The idea was first described by Smallwood and Sondik [15] and later Monahan[20] and Lovejoy [21] somewhat simplified the algorithm. All these ideas where developed for discrete-state POMDPs. Because we modeled our problem as a continuous-state POMDP some modifications are needed, however the main principles of the work cited remain valid for our case. The basic outline of the algorithm is that first all possible α-functions are generated using equation(15) then non-optimal α-functions are deleted and finally the optimal value function is constructed using the remaining α-functions and the expression of $V_t^*(b)$ in theorem 3.1, see the pseudo-code below.

Algorithm. α-functions algorithm

---

1. **Initialize.** $\alpha_C^{1,T}(s) = r_T(s)$, for all $C \in \{0, LRR, SP\}$ $s \in S^C$, $A_T = \{\alpha_C^1\}$ and $t = T-1$
2. **Generate.** Generate $A_t = \{\alpha_C^{1,t}, \alpha_C^{2,t}\cdots\}_{C\in\{0,LRR,SP\}}$ (using (20)) 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 $\alpha_C^{i,t}(s) \leq \alpha_C^{i,t}(s)$ 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.

Generating the α-functions. Let $A_{t+1} = \{\alpha_C^{1,t+1}, \alpha_C^{2,t+1}, \cdots\}_{C=\{0,LRR,SP\}}$ denote the set of α-functions at time $t+1$. Now instead of determining the optimal α-function $\alpha^{l^*(b)_t}$ by equation (15) 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

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

with

$$\alpha_0^{(a,i),t}(0) = \sum_o K_t^a(o|0)\Bigg[r_t(0,a,o) + (1 - p_t^{LRR} - p_t^{SP})\alpha_0^{i,t+1}(0) + \sum_{C\in\{LRR,SP\}} p_t^C \int_{S^C} f_t^C(s'|0)\alpha_C^{i,t+1}(s')ds'\Bigg]$$

$$\alpha_C^{(a,i),t}(s) = K_t^a(-|s)\left[r_t(s,a,-) + \int_{S^C} f_t^C(s'|s)\alpha_C^{i,t+1}(s')ds'\right] + K_t^a(+|s)R_t(s) \qquad C \in \{LRR, SP\} \tag{20}$$

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, namely:

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

and because every α-function has an action associated with it (20), the optimal action is easy to determine.

4.1 Exponential Transitions

In theorem 3.2 we have shown that in the special case that the transitions are exponentially distributed and the rewards and observation probabilities are described by exponential functions, then the α-functions can be obtained without explicitly calculating the integrals in the expression for the α-functions (20).Instead, the α-functions then are described by the parameters β and γ and can

be written as

$$\alpha_0^{(a,i),t}(0) = \beta_0^t(a,i)\alpha_0^{i,t+1}(0) + \gamma_0^t(a,i)$$
$$\alpha_C^{(a,i),t}(s) = \sum_{k=1}^{5} \beta_C^{k,t}(a,i) e^{-\gamma_C^{k,t}(a,i)s} \qquad C \in \{LRR, SP\}$$

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:

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

In the proofs of proposition 3.1 and corollary 3.1 we have derived explicit forms for the various parameters that describe the α-functions in the different states and at each decision epoch. In the pseudo-code below an algorithm is described to obtain the parameters sequentially in the special case.

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 $C \in \{LRR, SP\}$
   $A_T = \{\alpha^1\}$, $i = 1$ and $t = T - 1$.

2. **Generate.**
   for $j = 1$ to $||A_{t+1}||$

   for a=W,M

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

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

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

end

end

3. **Time update.** If $t > 1$, then $t = t - 1$ and go to step 2, otherwise stop.

---

5 Results

As an illustration how the model can be applied in practice, we will present results for a stratification of the patients based on their age. We will also 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). However for each stratification of the patients and for any growth model, as long as the parameters are available, the model can be applied. In this section we will first describe the parameters that are needed for the model and then the results.

Parameter	Source
Probability of death	CBS [22]
State transitions in S^{PO}	NCR [23], Witteveen et al. 2015[5]
Disutility of a mammogram	Mandelblatt et al. 1992 [24]
Disutility of a biopsy	Velanovich 1995 [25]
Specificity and sensitivity of mammography	Kolb et al. 2002 [26]
Specificity and sensitivity of self-detection	ibid.
Survival rates	NCR [23]
Life expectancy	CBS [22]

Table 1: Sources of model parameters.

5.1 Parameters

In this section we present the input parameters and their sources. Table 1 provides a list with the sources of the model parameters. For each set of patient characteristics the parameters will differ.

The probability that a healthy patient dies between two decision epochs depends on the age of the patient and is obtained from Statistics Netherlands (Centraal Bureau voor de Statistiek) [22]. Whenever the age of patients in a certain group differs we will 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 locoregional recurrence between two decision epochs, are obtained from the Netherlands Cancer Registry (NCR) [23][5]. The estimates for the transitions within the continuous states (i.e. the grow rates of a second primary tumor and of a locoregional recurrence) are also obtained from the NCR [23]. The estimations of the disutility of a mammogram vary between 0.5 and 1.5 days [24], so we use an estimate of one day. The disutility of a biopsy is estimated between two and four weeks [25], in our model we take the average of three weeks. We assume that these disutilities are the same for all ages. The specificity and sensitivity of both mammography and self-detection are obtained from Kolb et al. [26].

The lump-sum rewards and the end rewards are based on the life expectancy of a healthy patient. The expected remaining life years of an average patient at the start of the follow-up and at the end are used to construct a linear function of the life expectancy of a healthy patient at time $t = 1 \cdots T$. The expected remaining life years for patients in the different cancer states, i.e. the lump-sum and the 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 the NCR [23].

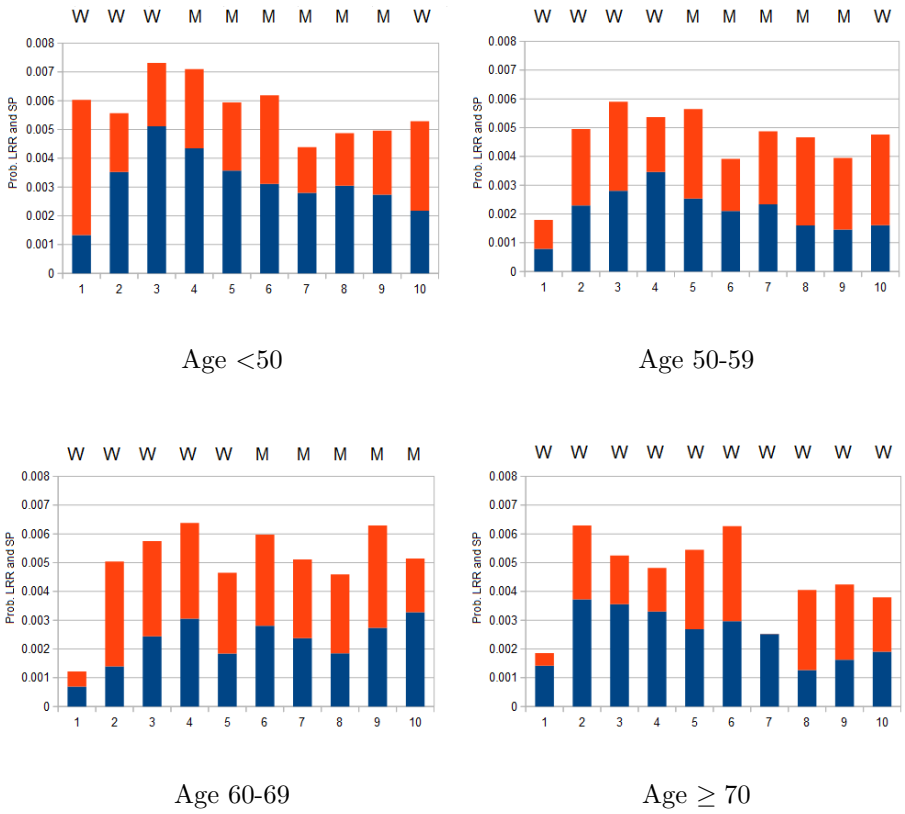
For several of the input parameters the precise values are not available. The core state transitions, for instance, are based on the current policy of annual mammography. The probabilities will therefore be slightly shifted in time e.g. if in reality a patient is most likely to get a LRR after 14 months it will not be detected for at least 10 months when the next mammogram is taken, so the transition probabilities will suggest a later time at which the patient is most likely to get a LRR. Also it is very hard to give a precise estimation of the growth model for both a LRR and a SP.

5.2 Results

Since the optimal policy will vary for different categories of patients, we will present the results for four basic categories. These categories serve as an illustration and since age is known to be of great influence on the risk of a LRR we choose this factor as an illustration. 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 upto 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 figure 2. 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). Above the probabilities the optimal action at each decision epoch is given.



Age <50 Age 50-59

Age 60-69 Age $\geq$ 70

Figure 2: Probability of a LRR (blue) and a SP (red) and the optimal policy for different age categories. W means it is optimal to wait, M means it is optimal to make a mammogram.

As one can see it is optimal to intensify the screening when the probability of a LRR peaks and just after that. Also that 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 their is less to gain by early detection.

6 Conclusions & Discussion

Currently breast cancer patients in the Netherlands have annual check-ups after treatment. Even though it is known that many factors, such as age, characteristics of the primary tumor and of treatment are of great influence on the risk of a LRR, the follow-up is the same for all patients. Individual mammography follow-up decisions based on personal risk characteristics are proposed by the national guideline of the Netherlands but without results in practice. In earlier research this sort of problems are modeled by discrete-state POMDPs [8] [7]. 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 proof an alternative representation described by so called α-functions. From this alternative description an iterative scheme can be deduced in order to obtain the optimal value function for every belief state and at every decision epoch. In general, the solution algorithm for the optimal value function can only be calculated with numerical methods. However, we proof that under some restrictions on the dynamics of the underlying Markov chain we can calculate the optimal value function without approximating. The restriction that we make on the dynamics of the underlying Markov chain are that we assume that the transition probabilities are exponentially distributed and that the rewards are described by an exponential relationship. 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 may be used in practice we calculate 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 op patients it seems optimal to not test at all.

Compared with the discrete model [7] 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.

In our study we choose to have a constant fixed time between two decision epochs. It would of course be more preferable to have the possibility of testing at time in the follow-up phase. However, from literature it is known that there are no exact solution methods for continuous-time POMDPs. The solution methods for continuous-time POMDPs all use discretization and therefore the POMDP reduces to the model we present in this study. Because we model over a finite time horizon the time between two decision epochs can be reduced to any sensible length without making the solution algorithm intractable.

A possible further refinement of this model would be to investigate variable time steps. This would exploit the benefits of a continuous time model without getting an intractable model.

The biggest 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 growth model of

the LRR and SP tumors. Therefore, without further study to obtain better estimates, the model cannot be used to give recommendations for exact testing policies.

References

[1] W. L. Lu, L. Jansen, W. J. Post, J. Bonnema, J. C. van de Velde, and G. H. D. Bock, "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*, vol. 114, pp. 403–412, 2009, http://dx.doi.org/10.1007/s10549-008-0023-4.

[2] M. Moossdorff, L. M. van Roozendaal, L. J. A. Strobbe, S. Aebi, D. A. Cameron, J. M. Dixon, A. E. Giuliano, B. G. Haffty, B. E. Hickey, C. A. Hudis, V. S. Klimberg, B. Koczwara, T. Kühn, M. E. Lippman, A. Lucci, M. Piccart, B. D. Smith, V. C. G. Tjan-Heijnen, C. J. H. van de Velde, K. J. V. Zee, J. B. Vermorken, G. Viale, A. C. Voogd, I. L. Wapnir, J. R. White, and M. L. Smidt, "Maastricht delphi consensus on event definitions for classification of recurrence in breast cancer research," *Journal of the National Cancer Institute*, vol. 106, no. 12, pp. 1–7, 2014, http://dx.doi.org//10.1093/jnci/dju288.

[3] IKNL, "Richtlijnen oncologische zorg," 2017, [accessed 7-March-2017]. [Online]. Available: http://www.oncoline.nl/

[4] S. M. E. Geurts, F. de Vegt, S. Siesling, K. Flobbe, K. K. H. Aben, M. van der Heiden-van der Loo, A. L. M. Verbeek, J. A. A. M. van Dijck, and V. C. G. Tjan-Heijnen, "Pattern of followup care and early relapse detection in breast cancer patients," *Breast Cancer Res Treat*, vol. 136, pp. 859–868, 2012, http://dx.doi.org/10.1007/s10549-012-2297-9.

[5] A. Witteveen, I. M. H. Vliegen, G. S. Sonke, J. M. Klaase, M. J. IJzerman, and S. Siesling, "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*, vol. 152, pp. 627–636, 2015, http://dx.doi.org/10.1007/s10549-015-3490-4.

[6] L. N. Steimle and B. T. Denton, *Markov Decision Processes for Screening and Treatment of Chronic Diseases*. Springer International Publishing, 2017, pp. 189–222. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-47766-4_6

[7] J. W. M. Otten, A. Witteveen, I. M. H. Vliegen, S. Siesling, J. B. Timmer, and M. J. IJzerman, *Stratified Breast Cancer Follow-Up Using a Partially Observable MDP*. Springer International Publishing, 2017, pp. 223–244. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-47766-4_7

[8] T. Ayer, O. Alagoz, and N. K. Stout, "A pomdp approach to personalize mammography screening decisions," *Operations Research*, vol. 60, no. 5, pp. 1019–1034, 2012, http://dx.doi.org/10.1287/opre.1110.1019.

[9] M. U. S. Ayvaci, O. Alagoz, and E. S. Burnside, "The effect of budgetary restrictions on breast cancer diagnostic decisions," *MSOM*, vol. 14, no. 4, pp. 600–617, 2012, http://dx.doi.org/10.1287/msom.1110.0371.

[10] J. Zhang, B. T. Denton, H. Balasubramanian, N. D. Shah, and B. A. Inman, "Optimization of psa screening policies: a comparison of the patient and societal perspectives," *Medical Decision Making*, vol. 32, no. 1, pp. 337–349, 2012, http://dx.doi.org/10.1177/0272989X11416513.

[11] J. M. Porta, M. T. J. Spaan, and N. Vlassis, "Robot planning in partially observable continuous domains," in *Robotics: Science and Systems*. MIT Press, 2005, pp. 217–224.

[12] J. M. Porta, N. Vlassis, M. T. Spaan, and P. Poupart, "Point-based value iteration for continuous pomdps," *J. Mach. Learn. Res.*, vol. 7, pp. 2329–2367, Dec. 2006.

[13] M. Duff, "Optimal learning: Computational procedures for bayes-adaptive markov decision processes," Ph.D. dissertation, University of Massassachusetts, 2002.

[14] K. Åström, "Optimal control of markov processes with incomplete state information," *Journal of Mathematical Analysis and Applications*, vol. 10, no. 1, pp. 174 – 205, 1965. [Online]. Available: http://dx.doi.org/10.1016/0022-247X(65)90154-X

[15] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable markov processes over a finite horizon," *Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973, http://dx.doi.org/10.1287/opre.21.5.1071.

[16] E. J. Sondik, "The optimal control of partially observable markov processes," Ph.D. dissertation, Stanford University, 1971.

[17] F. A. Sonnenberg and J. R. Back, "Markov models in medical decision making, a practical guide," *Medical Decision Making*, vol. 13, no. 4, pp. 322–338, 1993, http://dx.doi.org/10.1177/0272989X9301300409.

[18] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1994.

[19] J. M. Porta, M. T. J. Spaan, and N. Vlassis, "Value iteration for continuous-state POMDPs," Informatics Institute, University of Amsterdam, Tech. Rep. IAS-UVA-04-04, Dec. 2004.

[20] G. E. Monahan, "A survay of partially observable markov decision processes: Theory, models and algorithms," *Management Science*, vol. 28, no. 1, pp. 1–16, 1982, http://dx.doi.org/10.1287/mnsc.28.1.1.

[21] W. S. Lovejoy, "A survey of algorithmic methods for partially observed markov decision processes," *Annals of Operations Research*, vol. 28, no. 1, pp. 47–65, 1991, http://dx.doi.org/10.1007/BF02055574.

[22] CBS, "Statline," 2017, [accessed 7-March-2017]. [Online]. Available: http://statline.cbs.nl/Statweb/

[23] IKNL, "Nederlandse kankerregistratie," 2017, [accessed 7-March-2017]. [Online]. Available: http://www.cijfersoverkanker.nl/

[24] J. S. Mandelblatt, M. E. Wheat, M. Monane, R. D. Moshief, J. P. Hollenberg, and J. Tang, "Breast cancer screening for elderly women with and without comorbid conditions: A decision analysis model," *Annals of internal medicine*, vol. 116, no. 9, pp. 722–730, 2002, http://dx.doi.org/10.7326/0003-4819-116-9-722.

[25] V. Velanovich, "Immediate biopsy versus observation for abnormal findings on mammograms: An analysis of potential outcomes and costs," *The American Journal of Surgery*, vol. 170, no. 4, pp. 327–332, 1995, http://dx.doi.org/10.1016/S0002-9610(99)80298-0.

[26] T. M. Kolb, J. Lichy, and J. H. Newhouse, "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*, vol. 225, pp. 165–175, 2002, http://dx.doi.org/10.1148/radiol.2251011667.