IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **54** (2009) 4455–4476 doi:10.1088/0031-9155/54/14/007

A Markov decision process approach to temporal modulation of dose fractions in radiation therapy planning

M Kim[1,2], A Ghate[3] and M H Phillips[2]

[1] Department of Applied Mathematics, University of Washington, Seattle, WA, USA
[2] Department of Radiation Oncology, University of Washington, Seattle, WA, USA
[3] Department of Industrial and Systems Engineering, University of Washington, Seattle, WA, USA

E-mail: mk688@u.washington.edu

Received 9 February 2009, in final form 15 May 2009
Published 26 June 2009
Online at stacks.iop.org/PMB/54/4455

Abstract

The current state of the art in cancer treatment by radiation optimizes beam intensity spatially such that tumors receive high dose radiation whereas damage to nearby healthy tissues is minimized. It is common practice to deliver the radiation over several weeks, where the daily dose is a small *constant* fraction of the total planned. Such a 'fractionation schedule' is based on traditional models of radiobiological response where normal tissue cells possess the ability to repair sublethal damage done by radiation. This capability is significantly less prominent in tumors. Recent advances in quantitative functional imaging and biological markers are providing new opportunities to measure patient response to radiation over the treatment course. This opens the door for *designing fractionation schedules that take into account the patient's cumulative response to radiation up to a particular treatment day in determining the fraction on that day. We propose a novel approach that, for the first time, mathematically explores the benefits of such fractionation schemes*. This is achieved by building a stylistic Markov decision process (MDP) model, which incorporates some key features of the problem through intuitive choices of state and action spaces, as well as transition probability and reward functions. The structure of optimal policies for this MDP model is explored through several simple numerical examples.

1. Introduction

Radiation therapy (RT) is a toxic therapy that calls for strategies to minimize normal tissue damage while maximizing tumor cell kill. Several methods of treatment have been developed

 © 2009 Institute of Physics and Engineering in Medicine

over the last century to try to achieve this goal. Two of the most important can be described, at a high level, as *dose localization* and *fractionation*. Dose localization, as the name suggests, refers to delivering a highly concentrated dose to the target (tumor) with a sharp drop in dose beyond the targeted area, whereas fractionation relates to how the total planned dose to be delivered is divided into fractional doses. Thus, dose localization is a spatial optimization approach whereas fractionation is temporal.

In the last decade, significant improvements in dose localization have been made especially with the development of intensity modulated radiation therapy (IMRT) (Ahunbay *et al* 2007, Caglar and Allen 2007, Hou *et al* 2004, Langer *et al* 2002, Lu *et al* 2008, Yu *et al* 2006). IMRT enables treatment planners to optimize beam intensity within a single radiation field resulting in a more conformal dose to the tumor while sparing nearby organs-at-risk (OAR). In addition to IMRT, the development of image guided radiotherapy (IGRT) has improved the accuracy with which such conformal dose can be delivered to the tumor.

While dose localization is driven by the physical character of the dose such as its spatial distribution near the targeted area, fractionation is motivated by radiobiological effects of the dose on the patient. In particular, for a fixed amount of dose, tumor and healthy tissue response to radiation vary depending on the fractionation scheme used to deliver the dose. Although physical characteristics of the dose are well characterized and therefore widely used to estimate the effect of radiation on patients, they do not accurately capture biological effects such as repopulation, redistribution and reoxygenation (Hall and Giaccia 2005). Unfortunately, a complete biological modeling of these effects is very difficult owing to differences between types of tumors and types of normal tissues, differences between individual patients and variation in response among the cells within tumors or normal tissues. These hurdles have traditionally forced planners to use a simple fractionation scheme of constant daily dose, which does not account for variations in patient response during the treatment course. Typically, a small, constant daily dose is delivered over the course of several weeks. This fractionation schedule is based on traditional radiobiological models of radiation response in which normal tissue cells are capable of repairing some of the radiation damage between fractions, whereas tumor cells have much less repair ability. A variation on this strategy, hyperfractionation, magnifies this differential by dividing the total dose into even smaller daily fractions. Hypofractionation, where a larger than conventional dose per fraction is given, has received much attention recently because it shortens the overall radiation treatment course. By giving a large dose per fraction and therefore shortening the overall treatment time, damage to early responding tumors is maximized while relying on the dose localization of IMRT to preserve normal tissues (Hall and Giaccia 2005). However, in both cases a constant dose per fraction is currently used in practice. A mathematical investigation of non-constant, but non-adaptive fractionation schemes were conducted in Swan (1981), where a deterministic model of organ cell kill was assumed.

Interestingly, recent advances in quantitative functional imaging and biological markers are providing new opportunities to measure patient's response to radiation during and after RT (Corry *et al* 2008, Heron *et al* 2008, Lawrence *et al* 2008, South *et al* 2008, Tatum *et al* 2006). Access to such response information creates the possibility of improving the fractionation schedule in order to enhance the 'therapeutic ratio', that is, the difference between normal tissue effects and tumor cell kill. There is a clear possibility of designing patient-specific fractionation schedules based on the cumulative response up to that point in time without needing to address the detailed biological events leading up to that response. By incorporating such response information in our models, it is the intent of this work to mathematically explore the advantages to be gained, if any, by using non-standard, adaptive fractionation schedules that do not assume constant daily doses. To the best of our knowledge, this paper is the only

work that investigates this important aspect of RT from a decision-making perspective. As the first foray into this field, our focus is not on incorporating all complexities of the radiation treatment planning process as would be necessary for a clinical implementation, but rather on building simplistic yet insightful decision models that attempt to capture some salient features of the radiobiological effect of fractionation schemes.

More specifically, we employ finite-horizon Markov decision processes (henceforth MDPs), a well-known mathematical modeling technique from Operations Research (Bertsekas 1995, Puterman 1994), to compute 'optimal' fractionation schemes. MDPs are suitable for modeling decision problems where the system under consideration is assumed to evolve according to a Markov process affected by the decisions made by a system planner. In particular, the planning horizon is divided into discrete intervals of time and the planner observes the 'state' of the system before making a decision at the beginning of each interval. The system then evolves into the next state according to a transition probability that in general may depend on the current state, the decision chosen in that state and the next state. Crucially, this state evolution depends on events in the past only through the current state of the system and therefore detailed knowledge of events leading up to this state is irrelevant for current and future decisions. The system may accumulate an expected reward in this time interval as a result of the decision chosen in the current state. This procedure repeats every interval until the end of the planning horizon where a 'terminal reward' that depends on the final system state may be obtained. The goal is to compute a policy, that is, a decision in every system state possible at the beginning of each interval, that maximizes the total expected reward obtained over the course of the planning horizon. Note that such a policy is computed *a priori* but implemented over time after observing the actual state of the system at each decision epoch.

MDPs and related modeling techniques have recently been applied with much success to make optimal decisions in medicine. In Alagoz *et al* (2004), MDPs were used to calculate optimal liver transplant policies for end-stage liver disease patients, whereas in Shechter *et al* (2008), to obtain an optimal time to start HIV treatment. More specifically, such techniques have been applied to compute adaptive fractionation schemes that compensate for noise in dose delivery in earlier fractions due to patient movements and setup errors (Sir 2007, de la Zerda *et al* 2007, Ferris and Voelker 2004, Rehbinder *et al* 2004, Wu *et al* 2002).

We believe that MDPs provide a reasonable modeling approach for the fractionation problem considered in this paper. In particular, the patient being treated is the system under consideration and the planning horizon corresponds to the course of the radiation treatment. This planning horizon is divided into time intervals, where each interval corresponds to the time between two successive visits by the patient to receive radiation treatment. We assume that the 'conditions' of the tumors and OAR are observed by the planner before deciding the level of dose to be delivered in each fraction. For example, this may include information about the size of apoptosis or clonogens of high density or the level of hypoxia, as observed by emerging imaging techniques such as quantitative functional imaging (Buck *et al* 2003, Komar *et al* 2008, Laprie *et al* 2008, Minn *et al* 2008, Weber *et al* 2008). Thus, these conditions correspond to the state of our system and decisions correspond to the dose level in a fraction. Note however that it may be convenient in practice to perform such imaging and associated observations only once in a while rather than on every visit leading to a slightly different MDP model where each time interval of the planning horizon corresponds to the time between two successive observations. The tumors and OAR respond to this dose and may change their state by the patient's next visit. This state evolution is modeled using stylistic transition probability functions that make sense from a clinical perspective. Since the objective in RT is to damage the tumors while sparing the OAR by the end of the treatment course, our model does not

include any intermediate rewards, but rather only incorporates a terminal reward representing the patient's utility function as designed by the treatment planner.

This paper is organized as follows. The main components of our MDP model are developed in section 2. Our numerical examples are presented in section 3. Conclusions and future research directions are discussed in section 4.

2. Model formulation

We consider a patient with $n_1 \geqslant 1$ organs at risk, $n_2 \geqslant 1$ number of target volumes and a planning horizon of T treatments (or fractions) numbered $t = 1, 2, \ldots, T$. As an example, typical values for these quantities may include $n_1 = 5$ OAR, $n_2 = 1$ targets and $T = 30$ fractions, corresponding to six weeks of daily treatment delivered five days a week. In this paper, the phrases 'tth treatment', 'tth fraction' and 'tth period' will be used interchangeably. Our MDP model includes four components: system states, treatment planner's decisions, state transition probabilities and terminal reward function. These quantities are now defined precisely.

2.1. System states

The total number of possible states for the ith OAR is assumed to be $m_i \geqslant 1$, whereas $l_j \geqslant 1$ represents the total number of possible states for the jth tumor. Note that an OAR or a tumor may in reality be in an uncountably infinite number of possible conditions. However, we assume that the biological features used to characterize the states can be binned appropriately and hence represented 'finitely'. As an intuitive example, the states of the ith OAR may be modeled simply as 'high' and 'low' corresponding to the amount of apoptosis, resulting in $m_i = 2$. Similarly, the state of the jth tumor may also be modeled as 'large', 'medium' and 'small' indicating the hypoxic volume, leading to $l_j = 3$. For ease in writing mathematical equations and computer programs, OAR and tumor states are numbered sequentially starting with one thus leading to the sets S_i^o and S_j^τ of possible states for the ith OAR and jth tumor, respectively, defined as follows:

$$\text{For OAR } i = 1, 2, \ldots, n_1, \qquad S_i^o = \{1, 2, \ldots, m_i\}$$

$$\text{For tumor } j = 1, 2, \ldots, n_2, \qquad S_j^\tau = \{1, 2, \ldots, l_j\}.$$

Notation s_i^o is used to represent a generic element of the set S_i^o and similarly for s_j^τ. As a result, the state of our MDP model is $\left(s_1^o, s_2^o, \ldots, s_{n_1}^o, s_1^\tau, s_2^\tau, \ldots, s_{n_2}^\tau\right) \in \left(S_1^o \times S_2^o \times \cdots \times S_{n_1}^o \times S_1^\tau \times S_2^\tau \times \cdots \times S_{n_2}^\tau\right)$. We use set S to denote the Cartesian product $\left(S_1^o \times S_2^o \times \cdots \times S_{n_1}^o \times S_1^\tau \times S_2^\tau \times \cdots \times S_{n_2}^\tau\right)$. Hence S denotes the state space of our MDP model and a generic state in S will be denoted s. For brevity, we often use the vectors $s^o \equiv \left(s_1^o, s_2^o, \ldots, s_{n_1}^o\right)$ and $s^\tau \equiv \left(s_1^\tau, s_2^\tau, \ldots, s_{n_2}^\tau\right)$ to represent the states of all OAR and all tumors, respectively. Similarly, $S^o \equiv \left(S_1^o \times S_2^o \times \cdots \times S_{n_1}^o\right)$ and $S^\tau \equiv \left(S_1^\tau \times S_2^\tau \times \cdots \times S_{n_2}^\tau\right)$ are used to represent the sets of all such vectors.

Remark 2.1. The above states are ordered such that small state values of an OAR are associated with better patient conditions. On the other hand, larger state values of a tumor correspond to better patient conditions. Consequently, larger OAR states and smaller tumor states correspond to worsened patient conditions. Moreover, in the following we interpret all vector inequalities such as $s^o \geqslant \sigma^o$ for $s^o \in S^o$ and $\sigma^o \in S^o$ componentwise, meaning that $s^o \geqslant \sigma^o$ if and only if every component of s^o is at least as big as the corresponding component of σ^o.

2.2. Treatment planner's actions

In our model, the treatment planner chooses the dose in a fraction after observing the state of the OAR and the tumors. We assume that a non-zero dose is chosen in every fraction and therefore the set of possible actions, i.e., the 'action-space' in MDP parlance, is denoted $A = \{a_1, \ldots, a_\eta\} \subset \mathbb{R}^\eta_{++}$. As a simple example, $A = \{1, 2, 5\}$ would indicate that the treatment planner chooses from the three dose levels of 1 Gy, 2 Gy and 5 Gy in each fraction. To put such actions in perspective, hyperfractionation schemes often use a constant daily dose of about 1–1.5 Gy, whereas hypofractionation schedules use 3–30 Gy depending on the tumor type. More conventional fractionation schemes on the other hand deliver 1.8–2 Gy daily. Finally, an action from set A will be denoted by the letter a.

2.3. State transition probabilities

When the treatment planner chooses dose level $a \in A$ for the tth fraction after observing the state to be $s \in S$, the state before delivering the $(t+1)$st fraction is assumed to be $s' \in S$ with probability $P_t(s'|s, a)$. This transition probability can potentially be estimated from statistical data that show the correlation between dose levels and damage to organs for a wide range of population.

2.4. Terminal reward function

Our model uses $f(s)$ to denote a real-valued terminal reward function that reflects the patient's utility for the condition of his/her OAR and tumors as given by state $s \in S$ at the end of the Tth treatment. The model is designed to work with any utility function and many of our numerical examples in section 3 employ utility functions with shapes that have been found to be realistic and representative.

Our goal is to compute an optimal fractionation policy, i.e., a decision rule that assigns a dose level to every state $s \in S$ so as to maximize the expected patient utility. This can be achieved by solving the following standard recursive equations well-known in the MDP literature (Puterman 1994) for all $s \in S$ and $t = 1, 2, \ldots, T$:

$$V_t(s) = \max_{a \in A} \sum_{s'} P_t(s'|s, a) V_{t+1}(s'), \tag{1}$$

where $V_t(s)$ is the optimal expected patient utility resulting from the remaining treatment if the state at the beginning of the tth treatment is $s \in S$, with the boundary condition $V_{T+1}(s) = f(s)$. In other words, an action a^* that achieves the above maximum, i.e., satisfies the equation

$$V_t(s) = \sum_{s'} P_t(s'|s, a^*) V_{t+1}(s')$$

is optimal in state s at the beginning of the tth fraction. Results obtained from this approach are presented next.

3. Results

In this section, we present numerical examples that provide important insight into the behavior of optimal fractionation policies. The recursive equations in (1) can be solved easily using backward induction starting at the boundary condition whenever the state space S and action space A are not too large. Our numerical examples satisfy the following overriding assumptions derived from radiobiological intuition and clinical practice. It should be noted that these

assumptions are not necessary for successfully solving recursive equations (1), and are only included here for concreteness in constructing our examples.

Assumption 3.1. *The patient utility corresponding to OAR states that represent better health conditions for the patient is at least as large as the patient utility corresponding to OAR states that represent worse health conditions when all other states are fixed; similarly for the tumor states. Mathematically, since OAR states are ordered such that smaller OAR states represent better patient conditions, $f(\sigma^o, s^\tau) \geqslant f(s^o, s^\tau)$ for a fixed $s^\tau \in S^\tau$ whenever $\sigma^o \leqslant s^o$. Moreover, since tumor states are ordered such that larger tumor state values represent better patient conditions, $f(s^o, \sigma^\tau) \geqslant f(s^o, s^\tau)$ for a fixed $s^o \in S^o$ whenever $\sigma^\tau \geqslant s^\tau$.*

Assumption 3.2. *Given the states of different OAR and tumors before the t*th *fraction and a dose level in that fraction, the states of these OAR and tumors after that fraction are (conditionally) independent of one another. That is,*

$$\begin{aligned} &P_t\big(\sigma_1^o, \sigma_2^o, \ldots, \sigma_i^o, \ldots, \sigma_{n_1}^o, \sigma_1^\tau, \ldots, \sigma_{n_2}^\tau | s_1^o, s_2^o, \ldots, s_i^o, \ldots, s_{n_1}^o, s_1^\tau, \ldots, s_{n_2}^\tau, a\big) \\ &\quad = P_t\big(\sigma_1^o | s_1^o, a\big) \times P_t\big(\sigma_2^o | s_2^o, a\big) \times \cdots \times P_t\big(\sigma_{n_1}^o | s_{n_1}^o, a\big) \\ &\qquad \times P_t\big(\sigma_1^\tau | s_1^\tau, a\big) \times P_t\big(\sigma_2^\tau | s_2^\tau, a\big) \times \cdots \times P_t\big(\sigma_{n_2}^\tau | s_{n_2}^\tau, a\big). \end{aligned}$$

Assumption 3.2 reflects our intuition that given the current condition of an organ and the dose level, the future condition of that organ does not depend on other organs and hence does not appear too restrictive.

Assumption 3.3. *Between two successive treatments, an OAR state can either stay the same or get worse, that is, increase (see remark 2.1) by one. Note this implies that for OAR $i = 1, 2, \ldots, n_1$, $P_t(m_i | m_i, a) = 1$ for all $a \in A$. Similarly, between two successive treatments, a tumor state can either stay the same or get better, that is, increase (see remark 2.1) by one. Note this implies that for tumors $j = 1, 2, \ldots, n_2$, $P_t(l_j | l_j, a) = 1$ for all $a \in A$.*

Assumption 3.3 reflects the clinical and radiobiological observations that daily changes are small relative to the changes possible during the entire course of treatment. The assumption that the tumors do not get worse and the OAR do not get better is perhaps more restrictive but intuitive at least as a first level approximation derived from the fundamental tradeoff in RT especially given that a strictly positive dose is delivered in every fraction as stated in our action space in section 2. For any OAR $i = 1, 2, \ldots, n_1$, $s_i^o = m_i$ will be called 'OAR upper boundary state'. Similarly, for any tumor $j = 1, 2, \ldots, n_2$, $s_j^\tau = l_j$ will be called 'tumor upper boundary state'.

Assumption 3.4. *The probability for a (non-upper boundary) OAR state to change (as opposed to remaining the same) is higher with higher dose delivered in a fraction. Similarly, the probability for a (non-upper boundary) tumor state to change (as opposed to remaining the same) is higher with higher dose delivered in a fraction. In light of assumption 3.3 above, this means that for OAR $i = 1, 2, \ldots, n_1$,*

$$P_t\big(1 + s_i^o | s_i^o, a'\big) \geqslant P_t\big(1 + s_i^o | s_i^o, a''\big) \qquad \text{if} \quad a' \geqslant a'' \qquad \text{for} \quad s_i^o = 1, 2, \ldots, m_i - 1.$$

Similarly, for tumors $j = 1, 2, \ldots, n_2$,

$$P_t\big(1 + s_j^\tau | s_j^\tau, a'\big) \geqslant P_t\big(1 + s_j^\tau | s_j^\tau, a''\big) \qquad \text{if} \quad a' \geqslant a'' \qquad \text{for} \quad s_j^\tau = 1, 2, \ldots, l_j - 1.$$

Assumption 3.5. *We implicitly assume the well-known linear–quadratic model for dose–response relationship (Fowler 1989, Hall and Giaccia 2005) where the biological effect in fraction t, denoted by E_t, is given by $E_t = \alpha d_t + \beta d_t^2$ with appropriate parameters α and β*

for each organ and d_t is the dose in the tth fraction. Moreover, the ratio α/β is assumed to be large for tumors and small for OAR as is common in radiobiology (Hall and Giaccia 2005).

Note that the ratio α/β has the following interpretation: it is the unique value of d_t for which the contribution of the linear term αd_t to E_t equals that of the quadratic term βd_t^2. Moreover, for $d_t < \alpha/\beta$, the contribution of the linear term dominates whereas for $d_t > \alpha/\beta$, the contribution of the quadratic term dominates.

Assumption 3.6. *For dose level $a \in A$ in a fraction t, the transition probabilities for non-upper boundary states depend only on the changes in these states rather than the actual values of these states. That is, for any OAR $i = 1, 2, \ldots, n_1$, $P_t\big(z_i^o \big| s_i^o, a\big) = P_t\big(\chi_i^o \big| \sigma_i^o, a\big)$ whenever $z_i^o - s_i^o = \chi_i^o - \sigma_i^o$. Similarly, for any tumor $j = 1, 2, \ldots, n_2$, $P_t\big(z_j^\tau \big| s_j^\tau, a\big) = P_t\big(\chi_j^\tau \big| \sigma_j^\tau, a\big)$ whenever $z_j^\tau - s_j^\tau = \chi_j^\tau - \sigma_j^\tau$.*

Due to a large α/β ratio, and hence dominance of the linear term in E_i for most plausible values of dose d_i, it is reasonable to assume no dependence on the actual values but rather on the changes in tumor states for the transition probability. A similar assumption is more restrictive for OAR due to their generally smaller α/β values. For simplicity however, this assumption is employed even for OAR states. Non-stationary transition probability is used later in section 3.1.4 to account for the dependence on OAR state values to a certain extent.

3.1. Single OAR and single tumor

Our first example considers one OAR and one tumor, i.e., $n_1 = 1$ and $n_2 = 1$. Thus, our state space is given by $S = \big(S_1^o \times S_1^\tau\big)$. For simplicity, we consider the action space $A = \{H, L\}$, where H corresponds to a 'high' dose value (say 4 Gy) and L to a 'low' dose value (say 2 Gy). Moreover, we use additively separable terminal reward functions $f\big(s_1^o, s_1^\tau\big) = f_1^o\big(s_1^o\big) + f_1^\tau\big(s_1^\tau\big)$. This implies the intuitive notion that the patient's utility from the OAR being in a certain condition does not depend on the condition of the tumor. In this example, the terminal reward $f_1^o\big(s_1^o\big)$ associated with the OAR is parameterized by a parameter $c_1 > 0$, whereas $f_1^\tau\big(s_1^\tau\big)$ is parameterized by $c_2 > 0$. The ratio c_1/c_2 reflects the 'relative importance' of saving the OAR with respect to killing the tumor through RT. Specifically, when $c_1 > c_2$, the terminal reward (and hence the decision process) favors the OAR more than the tumor and vice versa. We need values of the following eight transition probabilities for non-upper boundary states s_1^o and s_1^τ: $P_t\big(\big(s_1^o, s_1^\tau\big)\big|\big(s_1^o, s_1^\tau\big), H/L\big)$, $P_t\big(\big(1{+}s_1^o, s_1^\tau\big)\big|\big(s_1^o, s_1^\tau\big), H/L\big)$, $P_t\big(\big(s_1^o, 1{+}s_1^\tau\big)\big|\big(s_1^o, s_1^\tau\big), H/L\big)$, and $P_t\big(\big(1 + s_1^o, 1 + s_1^\tau\big)\big|\big(s_1^o, s_1^\tau\big), H/L\big)$. Based on assumption 3.6, these eight probabilities are denoted by $P_t((0,0)|, H/L)$, $P_t((1,0)|, H/L)$, $P_t((0,1)|, H/L)$ and $P_t((1,1)|, H/L)$ respectively for brevity. Thus, in view of assumption 3.2, we obtain the following expressions:

$$P_t((0,0)|, H/L) = P_t^o(0|H/L) \times P_t^\tau(0|H/L) \tag{2a}$$

$$P_t((1,0)|, H/L) = (1 - P_t^o(0|H/L)) \times P_t^\tau(0|H/L) \tag{2b}$$

$$P_t((0,1)|, H/L) = P_t^o(0|H/L) \times (1 - P_t^\tau(0|H/L)) \tag{2c}$$

$$P_t((1,1)|, H/L) = \big(1 - P_t^o(0|H/L)\big) \times \big(1 - P_t^\tau(0|H/L)\big), \tag{2d}$$

where, as an example, $P_t^o(0|H)$ denotes the probability that a (non-upper boundary) OAR state remains the same when the dose level chosen is H. The other probabilities are interpreted similarly with superscript τ denoting the tumor. Note that the above four probability expressions thus have four probability values that we need to fix in order to get all other

transition probabilities. The following fixed values were used for all t in the next three subsections:

$$P_t^o(0|H/L) = 0.5/0.6 \tag{3a}$$

$$P_t^\tau(0|H/L) = 0.4/0.5. \tag{3b}$$

These numbers were chosen so that $P_t^o(0|H) = 0.5 > 0.4 = P_t^\tau(0|H)$ and $P_t^o(0|L) = 0.6 > 0.5 = P_t^\tau(0|L)$ to ensure that for a given dose level, the probability that the OAR condition deteriorates is smaller than the probability that the tumor condition improves. Moreover, the numbers ensure that $P_t^o(1|H) = 1 - P_t^o(0|H) = 1 - 0.5 = 0.5 > 0.4 = (1 - 0.6) = 1 - P_t^o(0|L) = P_t^o(1|L)$, and similarly, $P_t^\tau(1|H) > P_t^\tau(1|L)$, hence assumption 3.4 is satisfied. The resulting transition probabilities obtained from equations (2*a*)–(2*d*) are

$$P_t((0,0)|, H/L) = 0.2/0.3$$
$$P_t((1,0)|, H/L) = 0.2/0.2$$
$$P_t((0,1)|, H/L) = 0.3/0.3$$
$$P_t((1,1)|, H/L) = 0.3/0.2.$$

Our numerical results are presented for all possible initial OAR and tumor states although it is common in the MDP literature to assume a specific initial system state, for example in our context, the best possible OAR state and the worst possible tumor state. Section 3.1.1 employs convex quadratic terminal reward functions whereas section 3.1.2 uses concave quadratic terminal reward functions for both the OAR and the tumor. The effect of the ratio c_1/c_2 is investigated in section 3.1.3, whereas the effect of non-stationary transition probabilities is analyzed in section 3.1.4, both using concave quadratic terminal reward functions. All these examples used $m_1 = l_1 = 7, T = 4$, that is, seven OAR states, seven tumor states and four fractions. However our numerical experiments with other values for these parameters revealed fractionation policies with similar character. Finally, optimal policy tables obtained for our examples are available in the appendix.

3.1.1. Convex quadratic terminal rewards. The example in this section used convex terminal reward functions $f_1^o(s_1^o) = c_1(s_1^o - k)^2$ and $f_1^\tau(s_1^\tau) = c_2(s_1^\tau)^2$ resulting in $f(s_1^o, s_1^\tau) = c_1(s_1^o - k)^2 + c_2(s_1^\tau)^2$, where $k = m_1 + 1$ is a parameter that ensures all OAR terminal rewards are positive by shifting the focus of the associated parabola. Observe that the terminal reward function satisfies the monotonicity assumption 3.1. The resulting shapes of the two reward functions are shown in figure 1(a) whereas the optimal policy obtained after solving Bellman's equations (1) by backward induction is shown in table 1.

We now discuss the structure of the optimal policy as observed from table 1. We first consider the terminal fraction $t = 4$ and states (s_1^o, s_1^τ) where $s_1^o < 7$ and $s_1^\tau < 7$. In these states, a *monotone* policy is optimal. In particular, if H is optimal in state (s_1^o, s_1^τ), then H is also optimal in states $(\sigma_1^o, \sigma_1^\tau) \geqslant (s_1^o, s_1^\tau)$. In words, increasing the state does not decrease optimal dose. More specifically, for a fixed tumor state, dose level L is optimal in smaller (better) OAR states, and for a fixed OAR state, dose level L is optimal in smaller (worse) tumor states. This optimal policy structure results from convexity of our reward functions. Note that the optimal policy structure is different for the upper boundary states, i.e. $s_1^o = 7$ or $s_1^\tau = 7$ from the rest of the states. Since tumor state $s_1^\tau = 7$ corresponds to the best tumor condition, which cannot change, the focus of radiation treatment is solely to minimize damage to the OAR and hence dose L is optimal in $s_1^\tau = 7$. Similarly, dose H is optimal in OAR state $s_1^o = 7$ to maximize tumor control as the OAR is in the worst possible state that cannot get any worse.

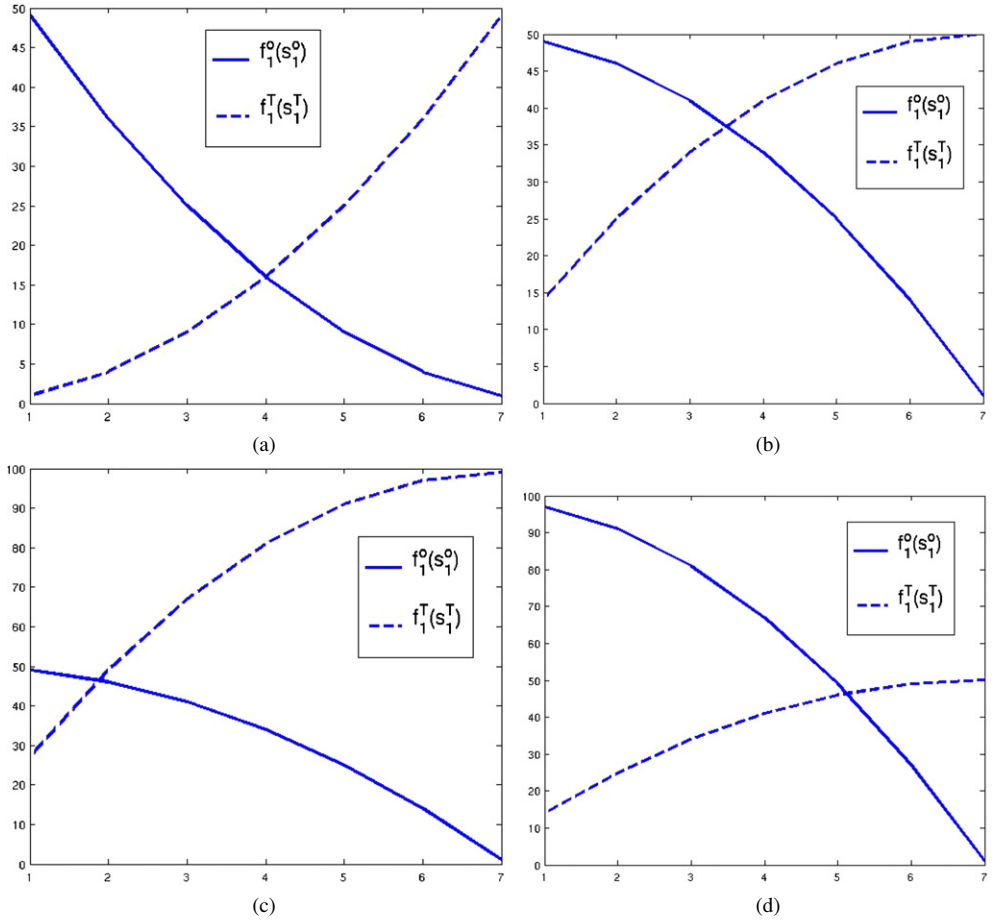

Figure 1. Different reward functions used in our examples. (a) Convex terminal reward functions for an OAR and a tumor. The plots show $f_1^o(s_1^o) = (s_1^o - 8)^2$ and $f_1^\tau(s_1^\tau) = (s_1^\tau)^2$ obtained by setting $k = 8, c_1 = 1, c_2 = 1$. (b) Concave terminal reward functions for an OAR and a tumor. The plots show $f_1^o(s_1^o) = 50 - (s_1^o)^2$ and $f_1^\tau(s_1^\tau) = 1 + 14s_1^\tau - (s_1^\tau)^2$ obtained by setting $k = 7, c_1 = 1, c_2 = 1$. (c) Concave terminal reward functions with the tumor weighted more than the OAR by setting $c_1 = 1, c_2 = 2$. The plots show $f_1^o(s_1^o) = 50 - (s_1^o)^2$ and $f_1^\tau(s_1^\tau) = 1 + 28s_1^\tau - 2(s_1^\tau)^2$. (d) Concave terminal reward functions with the OAR weighted more than the tumor by setting $c_1 = 2, c_2 = 1$. The plots show $f_1^o(s_1^o) = 99 - 2(s_1^o)^2$ and $f_1^\tau(s_1^\tau) = 1 + 14s_1^\tau - (s_1^\tau)^2$.

The optimal policy in upper boundary states at fractions $t = 1, 2, 3$ has this character as well. The monotone policy described above is also optimal in $t = 3$ for $s_1^o < 6, s_1^\tau < 6$; $t = 2$ for $s_1^o < 5, s_1^\tau < 5$; $t = 1$ for $s_1^o < 4, s_1^\tau < 4$.

To summarize, in the tth fraction, the structure of the optimal policy is different in states where $1 \leqslant s_1^o < 7 - (T - t)$ and $1 \leqslant s_1^\tau < 7 - (T - t)$ as compared to states where $7 - (T - t) \leqslant s_1^o \leqslant 7$ and $7 - (T - t) \leqslant s_1^\tau \leqslant 7$. This difference between optimal policy structure in states closer to the upper boundary and states away from the upper boundary will be termed 'boundary effect' in the following. Note that due to assumption 3.3, the boundary effect is carried over to at most $T + 1 - t$ states from the upper boundary in fraction t. Finally

note that the monotone policy observed in table 1 is counterintuitive and uninteresting from a practical viewpoint because it delivers a high level of dose when the tumor is in a better condition and OAR in a worse condition. Therefore, in the remainder of this paper, we focus on concave terminal reward functions that yield optimal policies that stand the test of clinical intuition.

3.1.2. Concave quadratic terminal rewards. Now we investigate the case where the terminal reward functions for both the OAR and the tumor are concave. A concave terminal reward function for an OAR implies that the same amount of deterioration in the OAR condition is penalized more when the OAR is already in a worse condition as compared to when it is in a better condition. Similarly, a concave terminal reward function for a tumor implies that the same amount of improvement in the tumor condition is appreciated less when the tumor is in a better condition as compared to when it is in a worse condition. We used functions $f_1^o(s_1^o) = 1 + c_1k^2 - c_1(s_1^o)^2$ and $f_1^\tau(s_1^\tau) = 1 + 2c_2ks_1^\tau - c_2(s_1^\tau)^2$ with parameters c_1 and c_2 whose ratio c_1/c_2 reflects the relative importance of the OAR to the tumor as in the case of convex quadratic terminal reward functions. Parameter k is employed to obtain positive rewards for all OAR and tumor states. The numerical results below were obtained by setting $k = 7, c_1 = 1, c_2 = 1$ resulting in reward functions shown in figure 1(b). The optimal policy obtained after solving Bellman's equations (1) is depicted in table 2.

Table 2 shows that in all fractions, dose L is optimal in tumor boundary states, i.e., $s_1^\tau = 7$ to spare the OAR as the tumor is already in the best possible condition that cannot change. Similarly, in all fractions, dose H is optimal in OAR boundary states, i.e., $s_1^o = 7$ to maximize tumor control as the OAR is in the worst possible state that cannot change. Note that this behavior is similar to the convex case. A monotone policy is optimal in the terminal fraction $t = 4$ in non-boundary states $s_1^o < 7$ and $s_1^\tau < 7$. However, note that the 'direction' of this monotone policy is opposite to that in the convex case above. In particular, if L is optimal in state $s \equiv (s_1^o, s_1^\tau)$, then L is also optimal in $(\sigma_1^o, \sigma_1^\tau) \equiv \sigma \geqslant s$. In other words, increasing the state does not increase optimal dose. More specifically, for a fixed tumor state, dose level L is optimal in larger (worse) OAR states, whereas H is optimal in smaller (better) OAR states. On the other hand, for a fixed OAR state, dose level L is optimal in larger (better) tumor states, whereas H is optimal in smaller (worse) tumor states. A similar monotone policy is optimal in $t = 3$ for $s_1^o < 6, s_1^\tau < 6$; $t = 2$ for $s_1^o < 5, s_1^\tau < 5$; $t = 1$ for $s_1^o < 4, s_1^\tau < 4$. Thus, as in the case with convex reward functions, the monotone optimal policy is observed only for $s_1^o < m_1 - (T - t)$ and $s_1^\tau < l_1 - (T - t)$ in fraction t due to the boundary effect. Note that the direction of this monotone policy is intuitive from a clinical perspective. This example will be referred to as the 'base case' in the remainder of this paper.

3.1.3. Relative importance of the OAR and the tumor. We investigate sensitivity of the optimal policy to the ratio c_1/c_2 in the concave OAR and tumor terminal reward functions discussed in section 3.1.2. In particular, instead of choosing $c_1 = 1$ and $c_2 = 1$ as in section 3.1.2, we first use $c_1 = 1, c_2 = 2$ resulting in reward functions shown in figure 1(c) and the optimal policy as in table 3, and then $c_1 = 2, c_2 = 1$ leading to reward functions in figure 1(d) and table 4. The fundamental structure of the optimal policy in both these cases is similar to the base case $c_1 = 1, c_2 = 1$. However, when the tumor is weighted more, i.e., $2 = c_2 > c_1 = 1$, H is optimal in several more states in table 3 as compared to table 2 for increased tumor control. On the other hand, when the OAR is weighted more, i.e.,

$2 = c_1 > c_2 = 1$, L is several more states in table 4 as compared to table 2 to reduce normal tissue complications.

3.1.4. Non-stationary transition probabilities. Since the OAR is assumed to have a small α/β ratio (assumption 3.5), the damage for a given dose is larger as the total dose delivered increases. One possible way to model this is to use OAR state transition probabilities that change over time, i.e., are non-stationary. In particular, for the same level of dose delivered in a fraction, we set the probability that a non-boundary OAR state worsens by one (as opposed to remaining the same) to be higher for later fractions during the treatment course. A simple scenario with this characteristic was modeled with two different transition probabilities for the OAR—one for the first two periods and another for the last two periods as follows:

$$P_t^o(0|H/L) = 0.5/0.6 \qquad \text{for} \quad t = 1, 2 \tag{4a}$$

$$P_t^o(0|H/L) = 0.2/0.3 \qquad \text{for} \quad t = 3, 4. \tag{4b}$$

All the other data were exactly as in the base case in section 3.1.2. The optimal policy is shown in table 5.

Comparing table 5 with table 2, it can be seen that it is now optimal to prescribe L in some states earlier in the treatment course where dose H was originally optimal in the base case. These results agree with the common practice that the low (constant) daily dose is prescribed due to the OAR with low α/β. As in the base case, H is optimal in $s_1^o = 7$ and L is optimal in $s_1^\tau = 7$ for all t. Similar monotone optimal policy is also observed with non-stationary transition probabilities in $s_1^o < 7 - (T - t)$ and $s_1^\tau < 7 - (T - t)$ in tth period.

3.2. Multiple OAR and single tumor

We now consider two OAR (denoted OAR1 and OAR2) and one tumor. In our standard notation, this means $n_1 = 2$ and $n_2 = 1$ and the state is given by $\left(s_1^o, s_2^o, s_1^\tau\right)$. As before, we use seven states for each OAR and the tumor and four fractions meaning $m_1 = 7, m_2 = 7, l_1 = 7$ and $T = 4$. The terminal reward function is again assumed to be additively separable yielding $f\left(s_1^o, s_2^o, s_1^\tau\right) = f_1^o\left(s_1^o\right) + f_2^o\left(s_2^o\right) + f_1^\tau\left(s_1^\tau\right)$. We use concave quadratic functions similar to the base case in section 3.1.2. Specifically, $f_1^o\left(s_1^o\right) = 1 + c_1k^2 - c_1\left(s_1^o\right)^2$, $f_2^o\left(s_2^o\right) = 1 + c_1k^2 - c_1\left(s_2^o\right)^2$, and $f_1^\tau\left(s_1^\tau\right) = 1 + 2c_2ks_1^\tau - c_2\left(s_1^\tau\right)^2$ with $c_1 = 1, c_2 = 1$ and $k = 7$. The state transition probabilities for OAR1 and the tumor are the same as in the base case in section 3.1.2. However, OAR2 is assumed to be more radiosensitive than OAR1. That is, for the same given dose, the probability that OAR2 state worsens by one is higher than OAR1 state worsens by one. This was ensured by setting $P_t^{o_2}(0|H/L) = 0.2/0.3$ in the notation introduced at the beginning of section 3.1 and superscript o_2 stands for OAR2.

Since the state space is three dimensional, it is difficult and too lengthy to completely illustrate the optimal policy obtained after solving Bellman's equations (1). Therefore we use the following approach—fix the state of either OAR1, OAR2 or the tumor and depict the structure of the optimal policy on a state-grid that represents conditions of the other two organs (as in earlier sections in the paper). For brevity, rather than fixing the state of OAR1, OAR2 or the tumor at all seven possible values one by one, we only use three values. In particular, tables 6–8 show the optimal policy when OAR1 state is fixed at $s_1^o = 1, 4, 7$ respectively, tables 9–11 show the optimal policy when OAR2 state is fixed at $s_2^o = 1, 4, 7$

respectively, and tables 12–14 show the optimal policy when the tumor state is fixed at $s_1^\tau = 1, 4, 7$ respectively.

These results illustrate a number of important points. Given that OAR1 is in its lowest state (OAR1 in its best condition), H is optimal when OAR2 is in a low state (OAR2 in good condition) and the tumor is in a low state (poor tumor control). As the OAR1 state becomes higher (OAR1 in poorer condition), L is optimal in most states in an effort to spare both OAR1 and OAR2. However, this tendency reverses when OAR1 is in its highest state (worst OAR1 state). The optimal policy now is to give H for more states to tradeoff between OAR2 and tumor since OAR1 is already in the highest state and can suffer no more damage. Similar results are shown for a fixed OAR2 state. Since OAR2 is more radiosensitive than OAR1, optimal policy is more aggressive, i.e., more occurrences of H when OAR2 is in its highest state (worst OAR2 condition) compared to when OAR1 is in its highest state. Similarly, L is optimal in more states when OAR1 is in the highest state than when OAR2 is in the highest state. For the fixed tumor state, L is clearly optimal in all states to spare OAR1 and OAR2 when the tumor is in the highest state (highest tumor cell death). When the tumor is in the lowest state (lowest tumor cell death), H is the optimal action when either OAR1 or OAR2 is in the highest state or both OAR1 and OAR2 are in the lowest state in an effort to improve the tumor control. When the tumor is in a middle state, L tends to be optimal in most states to spare both OAR1 and OAR2. For a fixed state of one organ, the optimal policy for the other two organs is always monotone in $s_i^o < 7 - (T - t)$ for $i = 1, 2$ and $s_1^\tau < 7 - (T - t)$ in the tth fraction as in section 3.1. The boundary effect is observed explicitly when OAR1 is fixed to its highest state as seen in table 8 since OAR1 cannot get any worse so it does not play a role in determining the optimal action. Consequently, the optimal policy in the OAR2–tumor state-grid is similar to the base case. Similar results are shown in table 11 when OAR2 is fixed to its highest state. When the tumor is fixed to its highest state, it is easy to guess that L would be optimal in the OAR1–OAR2 state-grid to spare the OAR which agrees with the numerical results shown in table 14.

4. Discussion, conclusions and future research

Traditional fractionation schedules in radiation therapy deliver a constant dose in each fraction without regard to patients' actual radiobiological response to radiation up to that specific fraction. One reason for the prevalence of such non-adaptive schedules is that detailed modeling of repopulation, redistribution and reoxygenation is very difficult owing to the number and variety of factors and the complex interactions that influence these biological processes. The only existing adaptive fractionation schemes focus on a different unrelated problem where the goal is to compensate for dose delivery errors owing to patient movements and setup in previous fractions (Sir 2007, de la Zerda *et al* 2007, Ferris and Voelker 2004, Rehbinder *et al* 2004, Wu *et al* 2002). Consequently, in contrast to our work, the 'system state' in such adaptive models corresponds to the total dose delivered in all previous fractions, and does not include the radiobiological condition of tumors or OAR.

Interestingly, new developments in quantitative functional imaging and biomolecular testing are increasingly providing opportunities to assess patients' cumulative radiobiological response to treatment up to a particular time without the need to model complex biological events that led to that condition. Our goal was to investigate, through mathematical decision models, whether access to this response information through the treatment course can be fruitfully exploited to design adaptive fractionation schemes. Moreover, the intent was to observe whether and how the resulting adaptive schemes differ from more traditional constant-dose fractionation schedules. We believe that owing to the probabilistic nature of radiation

response over time, Markov decision processes are an appropriate decision modeling approach to achieve this goal.

In this paper, we have taken initial steps in this direction by building a simplified MDP model that captures some key features of the probabilistic evolution of biological states of OAR and tumors as affected by the sequence of doses delivered in past fractions. The other components of this model include state transition probabilities that define this evolution and a terminal reward function that essentially assigns a 'utility value' to the OAR and tumor states at the end of treatment. We investigated the effect of the shape of the terminal reward function, varying the relative importance of sparing an OAR relative to killing tumor cells, varying the OAR transition probabilities across fractions and increasing the number of OAR on adaptive fractionation policies. The strength of our model lies in its ability to bring out several interesting radiobiological tradeoffs and calculate fractionation policies that make sense clinically, despite its simplicity and computational tractability. Moreover, the fractionation policies calculated in our examples have a specific monotone structure often of major mathematical interest in the general MDP literature. Below we delineate several future research directions that build upon the foundation laid in this paper.

The MDP model was designed to work with any utility function and our numerical examples incorporated some intuitive choices for these functions. A commonly used utility function in medicine is called quality adjusted life years (QALY) (Sassi 2006). In the calculation of QALY, weights ranging from 0 (death) to 1 (perfect health) are assigned to different health conditions. QALY is the product of years of life and the utility of the quality of that life. Such an endpoint is a natural one in health decision analysis and can reasonably be part of the terminal reward function of the MDP. However, assigning QALY or any other utility values to every combination of possible tumor and OAR states in our model would require extensive research including complex surveys and thorough analysis by clinicians, which is not the focus of this paper. Data published in the medical literature (Bell *et al* 2001, Stewart *et al* 2005, Tsuchiya and Dolan 2005) may provide a starting point for such an investigation in the future.

Our examples included up to two OAR and one tumor and only two dose levels along with simplifying assumptions on state transition probabilities to demonstrate important structural features of adaptive fractionation policies. In practice, up to 10–15 organs and several dose levels may need to be considered during treatment planning. This, in combination with more complex and accurate state transition probabilities, will make exact solution of Bellman's equations by backward induction computationally challenging and time consuming. Moreover, the problem of designing treatment plans that *simultaneously* strive for adaptive dose localization and adaptive fractionation by essentially optimizing radiation beamlet intensities spatially as well as temporally would be far more computationally challenging than optimization problems currently solved in state-of-the-art IMRT algorithms. These challenges will call for recently developed computational techniques for MDPs called approximate dynamic programming (ADP) (Powell 2007), which is the subject of our continuing research in this area. ADP methods have recently proven fruitful in computing adaptive fractionation schemes that compensate for dose delivery errors in previous fractions (Sir 2007, Ferris and Voelker 2004).

Our examples did not allow for a dose of zero Gy because this action choice would conflict with the overriding assumption that the states of OAR and tumor either remain the same or increase by one. We expect however that the inclusion of zero dose will not present any significant computational difficulty in the future and hence can be achieved in a straightforward way. This action choice would then be analogous to a treatment pause so that the normal tissue can recover.

An interesting variant of the present finite-horizon model would be its 'optimal stopping' extension where the treatment planners, after observing the condition of different organs, decide whether treatment should be continued or stopped, and in the case where the decision is to continue, what the dose level in the current fraction should be.

The MDP modeling ideas presented in this paper are also applicable to other related problems such as choosing an appropriate form of treatment, for example, chemotherapy, radiation therapy or surgery.

Acknowledgments

This work was supported in part by NIH grant 1-RO1-CA112505 and the University of Washington.

Appendix

Optimal policy tables for numerical examples presented in section 3 are provided here.

Table 1. Optimal policy for the four period example discussed in section 3.1.1 for the convex terminal rewards shown in figure 1(a). The policy for periods $t = 1, 2, 3, 4$ is shown in four separate tables. Each table depicts a 'state-grid' where tumor states s_1^τ from 1 to 7 are listed horizontally whereas OAR states s_1^o from 1 to 7 are shown vertically. The optimal action H or L is written in each cell in this state grid, whereas a cell with a · indicates a tie between H and L.

$t = 1$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	H	H	H	L	
5	H	H	H	H	H	H	L	
4	L	H	H	H	H	L	L	
3	L	L	H	H	H	L	L	
2	L	L	L	H	H	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	H	H	H	L	
5	H	H	H	H	H	H	L	
4	L	H	H	H	H	H	L	
3	L	L	H	H	H	L	L	
2	L	L	L	H	H	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	H	H	H	L	
5	L	H	H	H	H	H	L	
4	L	L	H	H	H	H	L	
3	L	L	L	H	H	H	L	
2	L	L	L	L	H	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	H	H	H	L	
5	L	H	H	H	H	H	L	
4	L	L	L	H	H	H	L	
3	L	L	L	H	H	H	L	
2	L	L	L	L	L	H	L	
1	L	L	L	L	L	H	L	
	1	2	3	4	5	6	7	s_1^τ

Table 2. Optimal policy for the four period example discussed in section 3.1.2 for the concave terminal rewards shown in figure 1(b). These concave terminal rewards use $c_1 = 1, c_2 = 1$ and hence place equal importance on the OAR and the tumor.

$t = 1$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	H	L	L	L	L	L	L	
3	H	H	L	L	L	L	L	
2	H	H	H	L	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	H	L	L	L	L	L	L	
3	H	H	L	L	L	L	L	
2	H	H	H	L	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	H	·	
6	L	L	L	L	L	L	L	
5	H	L	L	L	L	L	L	
4	H	H	L	L	L	L	L	
3	H	H	H	L	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 3. Optimal policy for the four period example discussed in section 3.1.3 for the concave terminal rewards shown in figure 1(c). These rewards use $c_1 = 1$ and $c_2 = 2$ hence placing more importance on the tumor.

$t = 1$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	H	L	L	L	
5	H	H	H	L	L	L	L	
4	H	H	H	L	L	L	L	
3	H	H	H	L	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	H	L	L	L	
5	H	H	H	L	L	L	L	
4	H	H	H	L	L	L	L	
3	H	H	H	H	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	L	L	L	L	
5	H	H	H	L	L	L	L	
4	H	H	H	L	L	L	L	
3	H	H	H	H	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	H	·	
6	H	H	H	L	L	L	L	
5	H	H	H	L	L	L	L	
4	H	H	H	H	L	L	L	
3	H	H	H	H	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 4. Optimal policy for the four period example discussed in section 3.1.3 for the concave terminal rewards shown in figure 1(d). These rewards use $c_1 = 2$ and $c_2 = 1$ hence placing more importance on the OAR.

$t = 1$

s_1^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	H	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	H	H	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	H	H	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	H	L	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 5. Optimal policy for the four period example discussed in section 3.1.4 for concave terminal reward functions shown in figure 1(b). Non-stationary state transition probabilities given in equations (4*a*) and (4*b*) are used for the OAR.

$t = 1$

s_1^o								
7	H	H	H	H	H	H	H	
6	H	H	H	H	L	L	L	
5	H	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	H	H	H	H	
6	H	H	H	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	H	H	H	
6	H	H	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	H	L	L	L	L	L	L	
3	H	H	L	L	L	L	L	
2	H	H	H	L	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	H	L	L	L	L	L	L	
4	H	H	L	L	L	L	L	
3	H	H	H	L	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 6. Optimal policy for the two OAR, one tumor example in section 3.2 when OAR1 state is fixed at $s_1^o = 1$ (best condition).

$t = 1$

s_2^o								
7	H	H	H	L	L	L	L	
6	H	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_2^o								
7	H	H	H	H	L	L	L	
6	H	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	H	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_2^o								
7	H	H	H	H	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	H	L	L	L	L	L	L	
1	H	H	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_2^o								
7	H	H	H	H	H	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 7. Optimal policy for the two OAR, one tumor example in section 3.2 when OAR1 state is fixed at $s_1^o = 4$ (mediocre condition).

$t = 1$

s_2^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_2^o								
7	H	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_2^o								
7	H	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_2^o								
7	H	H	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 8. Optimal policy for the two OAR, one tumor example in section 3.2 when OAR1 state is fixed at $s_1^o = 7$ (worst condition).

$t = 1$

s_2^o								
7	H	H	H	H	H	H	H	
6	H	H	H	H	L	L	L	
5	H	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_2^o								
7	H	H	H	H	H	H	H	
6	H	H	H	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_2^o								
7	H	H	H	H	H	H	H	
6	H	H	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	H	L	L	L	L	L	L	
3	H	H	L	L	L	L	L	
2	H	H	H	L	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_2^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	H	L	L	L	L	L	L	
4	H	H	L	L	L	L	L	
3	H	H	H	L	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 9. Optimal policy for the two OAR, one tumor example in section 3.2 when OAR2 state is fixed at $s_2^o = 1$ (best condition).

$t = 1$

s_1^o								
7	H	H	H	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	H	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	H	L	L	L	L	L	L	
1	H	H	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 10. Optimal policy for the two OAR, one tumor example in section 3.2 when OAR2 state is fixed at $s_2^o = 4$ (mediocre condition).

$t = 1$

s_1^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 11. Optimal policy for the two OAR, one tumor example in section 3.2 when OAR2 state is fixed at $s_2^o = 7$ (worst condition).

$t = 1$

s_1^o								
7	H	H	H	H	H	H	H	
6	H	H	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	H	L	L	L	L	L	L	
2	H	H	L	L	L	L	L	
1	H	H	H	L	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 2$

s_1^o								
7	H	H	H	H	H	H	H	
6	H	H	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	H	L	L	L	L	L	L	
3	H	H	L	L	L	L	L	
2	H	H	H	L	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 3$

s_1^o								
7	H	H	H	H	H	H	H	
6	H	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	H	L	L	L	L	L	L	
3	H	H	L	L	L	L	L	
2	H	H	H	L	L	L	L	
1	H	H	H	H	L	L	L	
	1	2	3	4	5	6	7	s_1^τ

$t = 4$

s_1^o								
7	H	H	H	H	H	H	H	
6	L	L	L	L	L	L	L	
5	H	L	L	L	L	L	L	
4	H	H	L	L	L	L	L	
3	H	H	H	L	L	L	L	
2	H	H	H	H	L	L	L	
1	H	H	H	H	H	L	L	
	1	2	3	4	5	6	7	s_1^τ

Table 12. Optimal policy for the two OAR, one tumor example in section 3.2 when the tumor state is fixed at $s_1^\tau = 1$ (worst condition).

$t = 1$

s_1^o								
7	H	H	H	L	H	H	H	
6	L	L	L	L	L	H	H	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	H	
2	L	L	L	L	L	L	H	
1	L	L	L	L	L	H	H	
	1	2	3	4	5	6	7	s_2^o

$t = 2$

s_1^o								
7	H	H	H	L	L	H	H	
6	L	L	L	L	L	L	H	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	H	
2	L	L	L	L	L	L	H	
1	H	L	L	L	L	H	H	
	1	2	3	4	5	6	7	s_2^o

$t = 3$

s_1^o								s_1^o
7	H	H	H	H	L	H	H	
6	L	L	L	L	L	L	H	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	H	
3	L	L	L	L	L	L	H	
2	H	L	L	L	L	L	H	
1	H	H	L	L	L	L	H	
	1	2	3	4	5	6	7	s_2^o

$t = 4$

s_1^o								
7	H	H	H	H	H	L	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	H	
4	L	L	L	L	L	L	H	
3	H	L	L	L	L	L	H	
2	H	H	L	L	L	L	H	
1	H	H	H	L	L	L	H	
	1	2	3	4	5	6	7	s_2^o

Table 13. Optimal policy for the two OAR, one tumor example in section 3.2 when the tumor state is fixed at $s_1^\tau = 4$ (mediocre condition).

$t = 1$

s_1^o								
7	L	L	L	L	L	H	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_2^o

$t = 2$

s_1^o								
7	L	L	L	L	L	L	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	H	
	1	2	3	4	5	6	7	s_2^o

$t = 3$

s_1^o								
7	H	L	L	L	L	L	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	H	
	1	2	3	4	5	6	7	s_2^o

$t = 4$

s_1^o								
7	H	H	L	L	L	L	H	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	H	
1	L	L	L	L	L	L	H	
	1	2	3	4	5	6	7	s_2^o

Table 14. Optimal policy for the two OAR, one tumor example in section 3.2 when the tumor state is fixed at $s_1^T = 7$ (best condition).

$t = 1$								
s_1^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_2^o

$t = 2$								
s_1^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_2^o

$t = 3$								
s_1^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_2^o

$t = 4$								
s_1^o								
7	L	L	L	L	L	L	L	
6	L	L	L	L	L	L	L	
5	L	L	L	L	L	L	L	
4	L	L	L	L	L	L	L	
3	L	L	L	L	L	L	L	
2	L	L	L	L	L	L	L	
1	L	L	L	L	L	L	L	
	1	2	3	4	5	6	7	s_2^o

References

Ahunbay E E, Chen G P, Thatcher S, Jursinic P A, White J, Albano K and Li X A 2007 Direct aperture optimization-based intensity-modulated radiotherapy for whole breast irradiation *Int. J. Radiat. Oncol. Biol. Phys.* **67** 1248–58

Alagoz O, Maillart L M, Schaefer A J and Roberts M S 2004 The optimal timing of living-donor liver transplantation *Manag. Sci.* **50** 1420–30

Bell C, Chapman R, Stone P, Sandberg E and Newmann P 2001 A comprehensive catalog of preference scores from published cost-utility analysis *Med. Decis. Making* **21** 288–94

Bertsekas D P 1995 *Dynamic Programming and Optimal Control* (Nashua, NH: Athena Scientific)

Buck A, Halter G, Schirrmeister H, Kotzerke J, Wurziger I, Glatting G, Mattfeldt T, Neumaier B, Reske S N and Hetzel M 2003 Imaging proliferation in lung tumors with PET: F18-FLT versus F18-FDG *J. Nucl. Med.* **44** 1426–31

Caglar H B and Allen A M 2007 Intensity-modulated radiotherapy for head and neck cancer *Clin. Adv. Hematol. Oncol.* **5** 425–31

Corry J, Rischin D, Hicks R J and Peters L J 2008 The role of PET-CT in the management of patients with advanced cancer of the head and neck *Curr. Oncol. Rep.* **10** 149–55

de la Zerda A, Armbruster B and Xing L 2007 Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework *Phys. Med. Biol.* **52** 4137–53

Ferris M C and Voelker M M 2004 Fractionation in radiation treatment planning *Math. Program.* **101** 387–413

Fowler J F 1989 The linear-quadratic formula and progress in fractionated radiotherapy *Br. J. Radiol.* **62** 679–94

Hall E J and Giaccia A J 2005 *Radiobiology for the Radiologist* (Baltimore, MA: Williams and Wilkins)

Heron D E, Andrade R S, Beriwal S and Smith R P 2008 PET-CT in radiation oncology: the impact on diagnosis, treatment planning, and assessment of treatment response *Am. J. Clin. Oncol.* **31** 352–62

Hou Q, Zhang C, Wu Z and Chen Y 2004 A method to improve spatial resolution and smoothness of intensity profiles in IMRT treatment planning *Med. Phys.* **31** 1339–47

Komar G, Seppanen M, Eskola O, Lindholm P, Gronroos T J, Forsback S, Sipila H, Evans S M, Solin O and Minn H 2008 18F-EF5: a new PET tracer for imaging hypoxia in head and neck cancer *J. Nucl. Med.* **49** 1944–51

Langer M, Lee E K, Deasy J O, Rardin R L and Deye J A 2002 Operations research applied to radiotherapy: an NCI-NSF sponsored workshop, February 7–9 *Int. J. Radiat. Oncol. Biol. Phys.* **57** 762–8

Laprie A, Catalaa I, Cassol E, McKnight T R, Berchery D, Marre D, Bachaud J M, Berry B and Moyal E C J 2008 Proton magnetic resonance spectroscopic imaging in newly diagnosed glioblastoma; predictive value for the site of postradiotherapy relapse in a prospective longitudinal study *Int. J. Radiat. Oncol. Biol. Phys.* **70** 773–81

Lawrence Y R, Werner-Wasik M and Dicker A P 2008 Biologically conformal treatment: biomarkers and functional imaging in radiation oncology *Future Oncol.* **4** 689–704

Lu R, Radke R J, Yang J, Happersett L, Yorke E and Jackson A 2008 Reduced-order constrained optimization in IMRT planning *Phys. Med. Biol.* **53** 6749–66

Minn H, Gronroos T J, Komar G, Eskola O, Lehtia K, Tuomela J, Seppanen M and Solin O 2008 Imaging of tumor hypoxia to predict treatment sensitivity *Curr. Pharm. Des.* **14** 2932–42

Powell W 2007 *Approximate Dynamic Programming* (New Jersey: Wiley)

Puterman M 1994 *Markov Decision Processes* (New Jersey: Wiley)

Rehbinder H, Forsgren C and Lof J 2004 Adaptive radiation therapy for compensation of errors in patient setup and treatment delivery *Med. Phys.* **31** 3363–71

Sassi F 2006 Calculating QALYs, comparing QALY and DALY calculations *Health Policy Plan.* **21** 402–8

Shechter S M, Bailey M D, Schaefer A J and Roberts M S 2008 The optimal time to initiate HIV therapy under ordered health states *Operations Res.* **56** 20–33

Sir M Y 2007 Optimization of radiotherapy considering uncertainties caused by daily setup procedures and organ motion *PhD Thesis* University of Michigan, Ann Arbor, MI

South C P, Partridge M and Evans P M 2008 A theoretical framework for prescribing radiotherapy dose distributions using patient-specific biological information *Med. Phys.* **35** 4599–611

Stewart S, Lenert L, Bhatnagar V and Kaplan R 2005 Utilities for prostate cancer health states in men aged 50 and older *Med. Care* **43** 347–55

Swan G W 1981 *Optimization of Human Cancer Radiotherapy* (Berlin: Springer)

Tatum J L *et al* 2006 Hypoxia: importance in tumor biology, noninvasive measurement by imaging, and value of its measurement in the management of cancer therapy *Int. J. Radiat. Oncol. Biol. Phys.* **82** 699–757

Tsuchiya A and Dolan P 2005 The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature *Med. Decis. Making* **25** 460–7

Weber M A, Giesel F L and Stieltjes B 2008 MRI for identification of progression in brain tumors: from morphology to function *Expert Rev. Neurotherapeutics* **8** 1507–25

Wu C, Jera R, Olivera G and Mackie T 2002 Re-optimization in adaptive radiotherapy *Phys. Med. Biol.* **47** 3181–95

Yu C, Shepard D, Earl M, Cao D, Luan S, Wang C and Chen D Z 2006 New developments in intensity modulated radiation therapy *Technol. Cancer Res. Treat.* **5** 451–64