A Markovian decision model of adaptive cancer treatment and quality of life*

Péter Bayer† Joel S. Brown‡
Johan Dubbeldam§ Mark Broom¶

May 2, 2022

Abstract

This paper develops and analyzes a Markov chain model for the treatment of cancer. Cancer therapy is modeled as the patient's Markov Decision Problem, with the objective of maximizing the patient's discounted expected quality of life years. Patients make decisions on the duration of therapy based on the progression of the disease as well as their own preferences. We obtain a powerful analytic decision tool through which patients may select their preferred treatment strategy. We illustrate the tradeoffs patients in a numerical example and calculate the value lost to a cohort in suboptimal strategies. In a second model patients may make choices to include drug holidays. By delaying therapy, the patient temporarily forgoes the gains of therapy in order to delay its side effects. We obtain an analytic tool that allows numerical approximations of the optimal times of delay.

1 Introduction

1.1 Motivation

Patients face challenging decisions regarding cancer treatments. This is especially so when cure is uncertain or nearly impossible, regardless of treatment. Such is the case for most patients with metastatic disease. Patients' decisions invariably balance quality of life with quantity of

*We thank Nathaniel Mon Père for useful discussions and assisting in conducting our simulations. We thank Kateřina Staňková and Jeffrey West for comments. Péter Bayer acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future program (Investissements d'Avenir, grant ANR-17-EURE-0010) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 789111 - ERC EvolvingEconomics).

†Toulouse School of Economics, 1 Esplanade de l'université 31080 Toulouse, France. E-mail: peter.bayer@tse-fr.eu.

‡Department of Integrated Mathematical Oncology, Moffitt Cancer Center, 12902 USF Magnolia Drive, Tampa, FL 33612, United States. E-mail: Joel.Brown@moffitt.org.

§Delft University of Technology, Mekelweg 5, 2628 CD Delft, The Netherlands. E-mail: J.L.A.Dubbeldam@tudelft.nl.

¶City, University of London, Northampton Square, London EC1V 0HB, United Kingdom. E-mail: Mark.Broom.1@city.ac.uk.

life. Therapies are invasive, costly, and often significantly reduce a patient's well- being both during and after therapy. Many chemotherapies bring hair loss, nausea, malaise and lethargy. Hormone therapies may leave the patient feeling uneasy, agitated, weak, and with diminished sex drive and performance. Radiation therapy can leave the patient acutely ill or with long-term health issues from damaged tissues such as urinary problems following radiation of the prostate, bladder or pelvic area. Partial or complete surgical removal of the colon, breasts, prostate, liver, brain and even amputations of limbs can leave permanent physical and psychological disabilities.

Survival time remains the prevailing measure of success in cancer therapy. Due to the unambiguity and availability of data it is the least controversial and most accessible metric. Mathematical models of cancer therapy often report on their proposed regimens' effects on (simulated) survival or progression time. Clinical trials of new drugs and methods of delivery are similarly evaluated on this basis. Yet, there is reason to believe that oncologists and patients do not make treatment decisions to maximize survival time. In particular, decisions to refuse therapy are often influenced by concerns over quality of life (Shumay et al., 2001) and cure probability (Frenkel, 2013) possibly at the expense of expected survival time. While the prevailing response to such decisions had been a call for oncologists to "better communicate" with their patients, whether the prescribed therapy indeed aligns with the patient's objectives is not so clear.[1]

Here, we provide a theoretical foundation to formally capture these dilemmas. We employ the mathematical tools of dynamic optimization, statistics and decision theory. With these, we build a model of cancer treatment by which these dilemmas can be explicitly modeled and analyzed. Our model will not capture all elements of such dilemmas. Our intention is to advance a modeling approach that introduces methods and concepts by which the discussion surrounding patient empowerment and individuality, quality versus quantity of life, and therapeutic strategies can be advanced.

There are two key aspects of our approach. First, it is akin to dynamic programming in that it seeks to optimize an objective function (the patient's utility) from the present state through the range of possible outcomes in the future emanating from a particular decision now (Gluzman et al., 2020). This thinking ahead by anticipating and steering has become a cornerstone of evolutionary therapies (Staňková et al., 2019). Second, we balance quality and quantity of life (see Billingham and Abrams (2002) for a review of modelling approaches). Simes (1985) provides an early application of decision theory to balancing this tradeoff in cancer therapy. Glasziou et al. (1998) use a gradient analysis by considering how a decision now influences the product of survivorship and quality of life. Empirically, this issue becomes important in encouraging patients to pursue curative options when they otherwise refuse treatment (Huijer and van Leeuwen, 2000; Dias et al., 2021); or, respecting a patient's right to balance quality and quantity of life issues (Chaikh et al., 2016; Terpos et al., 2021); or the inclusion of financial considerations when care is considered palliative (Patnaik et al., 1998).

1.2 Background

The tools and concepts of game theory and decision theory have proven extremely valuable in cancer research. The objective has been to utilize game theory's insights in understanding

[1]Suh et al. (2017) report on 149 of 617 of lung cancer patients who refused treatment. Gilbar (1991) found no difference in quality of life of patients who refused chemotherapy ($n = 19$), later ceased chemotherapy ($n = 51$), and those who completed chemotherapy ($n = 51$) (also see Frenkel (2013), and Delisle et al. (2020)).

the eco-evolutionary dynamics of cancer. The practical application of this research is first, to calibrate the parameters (doses, timing, duration) of existing therapy regimens (see e.g. adaptive therapy, (Gatenby et al., 2009)) and, second, to find new points of attack against the disease in search of new therapy regimens.

One development towards this first goal views cancer therapy as a game played between the disease and the treating physician (Orlando et al., 2012). A useful framework is to model the game as a leader-follower (Stackelberg) game with the treating physician as a strategic decision maker and cancer as a reactive player. Via natural selection, the cancer evolves resistance to the physician's past and current treatment strategies (Staňková et al., 2019). The approach identifies the patient benefits that the physician can realize by assuming the role of leader in the therapy-cancer game. The physician anticipates the cancer's possible evolutionary responses and uses this to the patient's advantage. In the absence of such an approach, we often observe physicians in the reactive role and following a prescribed or standard treatment strategy, changing only after observing disease progression as measured by tumor burdens rather than the evolutionary strategies of the cancer cells.

In the following modeling sections, we advance this thread of the literature by viewing the therapy-cancer game as a Markovian process. In Markovian models of cancer (Kay, 1986; Andersen et al., 1991), all relevant information regarding the prognosis of the patient is encoded in health states, usually including a healthy state, various states of disease progression, and a death state. The patient's transitions between these states follows a stochastic process. The transition probabilities between states may be calibrated from cohort data (Duffy et al., 1995) for simulations of likely disease progression. The resultant toolkit has applications in both medicine (Llorca and Delgado-Rodríguez, 2001) and health economics (Le Lay et al., 2007).

Traditionally, in Markovian models, the transition probabilities are assumed to depend only on the current state of the patient, not on previous disease history. This is both a simplifying and restrictive assumption. Too few health states may obscure state differences created by the patient's past history of therapy. Too many health states requires an overly large and unwieldy transition matrix that may fail to produce insights applicable to a large cohort of patients. Cooper et al. (2003, 2004) resolved this by introducing a small number of payoff-states (responsive, stable and progressive disease; and dead) while letting the transition probabilities change based on the length of the treatment, measured in the number of treatment cycles.

To this existing framework we add the element of choice by the patient.[2] Markov decision processes (MDPs) (Bellman, 1957) combine the tools of stochastic processes and decision theory. In this model the Markovian transition probabilities depend upon both the current state and the strategy of a payoff-maximizing decision maker. The patient receives payoffs, measured in quality adjusted life years (QALYs), from spending time in states, with more healthy states giving higher payoffs. The tension in these problems is introduced when the decision-maker faces a choice between strategies that lead to immediate payoff gains and strategies that lead to better future prospects but at the cost of foregoing immediate gains. These tradeoffs occur with cancer therapy. The patient taking therapy makes immediate financial and QALY sacrifices in hopes of a higher probability of cure and greater life expectancy. If the decision-maker's objectives can be represented by time-discounting future expected payoffs and the set of states is finite, then optimal policies will exist (Blackwell, 1962, 1965), and will be unique for generic calibrations of

[2]In the remainder of the paper we refer to the patient as the sole decision-maker without explicitly mentioning the treating oncologist, tumor board, or any other participants of the decision making process.

payoff values (Ortega-Gutiérrez et al., 2016).

In this paper we use MDPs to reduce the game to the decision problem of a single strategic decision maker, the patient. We treat the evolutionary processes of cancer as an exogenous and stochastic element, whose behavior, conditional on the selected treatment strategy. We introduce exponential discounting to model a preference for earlier QALYs over later ones. The treatment strategy should maximize discounted expected QALYs, we are able to derive optimal treatment strategies.

In Section 2, we consider the duration of treatment. The patient's payoff is the difference between their QALYs and the cost of the treatment. The patient decides whether to continue treatment in hopes of a higher cure probability or longer life expectancy, or to abandon treatment to forgo the cost of therapy. The adaptive dynamics of the cancer become a key factor. As the patient's disease progresses, cancer's responsiveness to therapy changes. Following Cooper et al. (2003) our model has an infinite series of health states, in addition to the absorbing 'cured' and 'dead' states. There are two types of non-absorbing states, characterized by two detection levels, detectable and undetectable disease burden. Both types of state have infinite copies, characterized by an integer i which corresponds to the state of the disease based on the patient's therapy history. While undetectable, we assume that therapy cannot be given; the disease will, in time become detectable. Without therapy, the detectable state i disease will, in time, lead to the death of the patient. With therapy, the detectable state i disease will transition to state cure, death, or state $i+1$ undetectable or state $i+1$ detectable disease state. We call the therapy received while the disease is in detectable state i a *round* of therapy. Thus, the first time the patient may receive care is in detectable state 0. The rates of transition are dependent on the value of i, the number of rounds the patient has received. The dilemma of the patient is to select the value i^* beyond which, no more therapy is taken.

This model permits evaluating treatment strategies of different duration. The model is highly efficient as the patient's payoff-maximizing treatment strategies may be derived analytically as a solution to a linear system of equations for all parameter settings. Furthermore, if the patient's likelihood of recovery declines with the progressive state of the detectable disease, an assumption that is motivated both by the onset of resistance to therapy as well as observed outcomes of cancer therapy, the globally optimal payoff-maximizing duration of therapy equals the myopically optimal one, the strategy the patient follows if they only plan one decision node ahead, i.e. take the next round of therapy if and only if taking it once more is better than ceasing immediately. This is an attractive property that avoids any time-inconsistencies of treatment. If the monotonicity conditions are met, then there will exist a unique progressive disease state beyond which the patient loses expected QALYs due to overtreatment if treatment is not stopped. Interestingly, while the average expected payoff loss of overtreatment across all patients may be marginal due to time-discounting and the cohort's attrition up to the time when overtreatment is reached, the realized payoff loss for patients who do reach that stage is substantial. We demonstrate this through a simulation.

Cancer therapy is often highly toxic for the patient. This lost quality of life is one of the main reasons for patients to refuse or abandon therapy. Our second model (Section 3) includes loss of QALYs where the payoff of the patient depends upon their current health state and the current level of toxicity. We assume that a patient's toxicity level increases with each round of therapy and declines as past rounds of drugs in the patient decay over time. Including toxicity makes the cost of therapy conditional on its effect. Surviving patients have to bear the QALY reduction

longer, and patients who are not cured may have to resort to taking on higher levels of toxicity and additional QALY reductions from additional rounds of therapy. This affords patients an additional option for managing the QALY-cost of therapy. They may choose to postpone rather than abandon therapy as a means of allowing their toxicity burden to depreciate. However, by doing so they also postpone any benefits of therapy to their recovery.

This model addresses what already happens in practice. For certain therapies, drug holidays are mandated as a means for reducing the risk of mortality from toxicity. Physicians may also temporarily cease drug use if the patient's health seems overly compromised, and patients themselves will temporarily refuse treatment as a consequence of feeling sick from the drug. By including the loss of patient QALYs due to cumulative toxicity, our model no longer conforms to classic MDPs. It is no longer analytically tractable. But, we do provide methods for a numerical approximation that allows the patient to optimize simultaneously the timing and duration of cancer therapy.

The paper proceeds as follows. In Section 2, we develop a Markov model of cancer therapy that includes therapy-dependent likelihoods of cure and mortality that can change with time. A key element of this model is the inclusion of patient-specific quality adjusted life years (QALYs). We then use of Markov decision processes (MDPs) to seek dynamic and optimal therapy scheduling that anticipates and allows for multiple points of decision making. In Section 3, we expand the model to include the effects of drug toxicity on QALYs and presents a finer tuning of the optimal duration and timing of therapy. In Section 4, we discuss the significance of the results, their relationship to other relevant work, and provide prospectus for future theoretical and empirical research. All mathematical proofs are provided in the appendix.

2 State-dependent payoffs

We assume that the patient has a solid tissue detectable tumor without specifying the exact kind of cancer. The progression of the disease is modeled as a Markov-process in continuous time. The states encode the patient's quality of life and prognosis-relevant data, while the transition rates describe their prognosis and depend upon the patient's chosen treatment strategy. The set of cancer's progressive states (henceforth, states) is $S = \{0, \{1^{(i)}, 2^{(i)}\}_{i=0}^{\infty}, 3\}$. The states are interpreted as follows:

- 0: Healthy, cancer free state.
- $1^{(i)}$: Undetectable cancer after i rounds of therapy.
- $2^{(i)}$: Detectable cancer after i rounds of therapy. The patient chooses whether to take another round of therapy.
- 3: Death of the patient.

Without therapy, the natural progression has state $1^{(i)}$ leading eventually to state $2^{(i)}$ which leads to state 3, an absorbing state. The healthy absorbing state 0 may only be reached by therapy. The patient or the treating physician cannot distinguish between the cured and the undetectable states, 0 and $1^{(i)}$, and hence we assume that the patient does not receive therapy until the next detectable state, $2^{(i)}$ is reached.

Upon entering state $2^{(i)}$ the patient alters the progression of the disease by accepting therapy. When receiving therapy in state $2^{(i)}$ the patient may transition to any one of the four states 0 (cure), $1^{(i+1)}$ (partial therapy success), $2^{(i+1)}$ (partial therapy failure), or 3 (death). On reaching the $i+1$th progressive states, the patient potentially faces different transition rates and probabilities. For instance, as resistance evolves, disease burden increases, or new metastases occur, the cure rate may decline and mortality may increase.

We define a *treatment strategy* by a function $x\colon \{2^{(i)}\}_{i=0}^{\infty} \to \{\textit{therapy}, \textit{no therapy}\}$. In words, for each detectable state $2^{(i)}$ the patient chooses whether or not to take therapy. If the choice is *no therapy*, the patient remains in state $2^{(i)}$ until the end of his or her life. If the choice is *therapy*, the patient remains in state $2^{(i)}$ until the ith treatment round is completed, i.e. until he or she transitions to one of 0, $1^{(i+1)}$, $2^{(i+1)}$, or 3. As a state $2^{(i+1)}$ can only be reached if the patient chooses to receive therapy in state $2^{(i)}$, we restrict attention to treatment strategies such that for every $i \geq 0$ with $x(2^{(i)}) = \textit{no therapy}$ we have $x(2^{(i+1)}) = \textit{no therapy}$. We can therefore describe a treatment strategy by the index of the detectable state $2^{(i)}$ at which the patient ceases treatment. We will thus denote strategies by x_i, indicating that the patient chooses therapy in every state $2^{(j)}$ for $j < i$. Strategy x_0 describes a patient forgoing therapy entirely, while x_∞ describes a patient who always opts for therapy in a detectable state. Let the patient's set of permissible strategies be denoted by $X = \{x_i\}_{i=0}^{\infty}$.

Time is continuous. We assume that the states encode all progression-relevant information to the disease. Hence the process, conditional on the treatment strategy, is Markovian. The transition rates by which the patient moves between the states are as follows:

1. $1^{(i)} \to 2^{(i)}$ at rate δ_i,
2. if $x(2^{(i)}) = \textit{no therapy}$, then $2^{(i)} \to 3$ at rate ω_i
3. if $x(2^{(i)}) = \textit{therapy}$, then
    a. $2^{(i)} \to 0$ at rate λ_i,
    b. $2^{(i)} \to 1^{(i+1)}$ at rate β_i,
    c. $2^{(i)} \to 2^{(i+1)}$ at rate γ_i,
    d. $2^{(i)} \to 3$ at rate μ_i.

Let the term $\alpha_i = \lambda_i + \beta_i + \gamma_i + \mu_i$ describe the overall rate of exit from state $2^{(i)}$. The model is summarized by Figure 1.[3]

Spending time in each health state provides payoffs to the patient measured in QALYs. In this section we assume that the payoff values of each state are independent of the chosen treatment strategy. This assumption is relaxed in the following section. We assume that the rate of accruing QALYs is highest when cured, lower when burdened by undetectable disease,

[3]The connection with the more well-known discrete-time Markov Decision Processes is the following: In expectation, a patient who does not take therapy at state $2^{(i)}$ spends time $1/\omega_i$ in $2^{(i)}$ before progressing to 3. The transition probability from $2^{(i)}$ to 3 is thus 1 without therapy. Similarly, a patient in $1^{(i)}$ transitions to $2^{(i)}$ with probability 1, spending at expected time of $1/\delta_i$ in $1^{(i)}$. A patient who takes therapy in $2^{(i)}$ spends an expected $1/\alpha_i$ time in this state before transitioning to one of 0, $1^{(i+1)}$, $2^{(i+1)}$, 3 with probabilities λ_i/α_i, β_i/α_i, γ_i/α_i and μ_i/α_i, respectively. The time spent in each state is exponentially distributed with parameter corresponding to the total transition rate out of the state: δ_i for state $1^{(i)}$, ω_i for state $2^{(i)}$ without therapy and α_i for state $2^{(i)}$ with therapy.

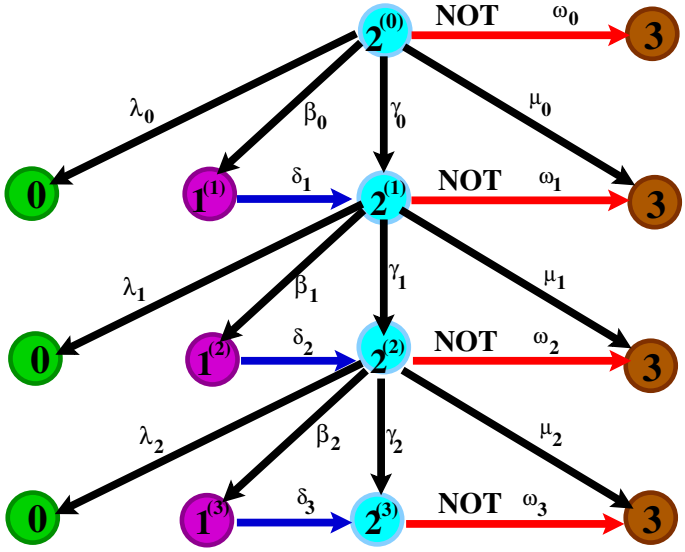

Figure 1: Schematic showing transition rates of the first 3 decision nodes of therapy. Each 0 node and each 3 node on the figure represent one absorbing state, while the figure shows multiple copies for better visibility. From an undetectable disease state, $1^{(i)}$, the patient will eventually progress to detectable state $2^{(i)}$. If the patient opts for therapy, he or she progresses to one of the four states 0, $1^{(i+1)}$, $2^{(i+1)}$, or 3. Otherwise, by choosing the **no t**herapy option, he or she eventually progresses to state 3.

lower still when having detectable disease, and 0 if dead. For $0 \leq v \leq u \leq 1$, the rate of QALY accrual is given by the function $q\colon S \to [0,1]$ given by

$$q(s) = \begin{cases} 1 & \text{if } s = 0 \\ u & \text{if } s \in \{1^{(i)}\}_{i=0}^{\infty} \\ v & \text{if } s \in \{2^{(i)}\}_{i=0}^{\infty} \\ 0 & \text{if } s = 3 \end{cases}$$

called the patient's *instantaneous payoff function*.

Upon selecting the treatment strategy x_i, the patient's progression through the states is a stochastic (Markovian) process. A realization of the patient's progression is called a play, described by a class of functions $s\colon [0,\infty) \times X \to S$. The value $s(t, x_i)$, denotes the patient's state at time $t \in [0,\infty)$ under treatment strategy x_i. Given strategy x_i, realization $s(\cdot, x_i)$ and $j \in \{1,\ldots,i\}$, let $t_j(s(\cdot, x_i))$ denote the time at which the patient arrives in $2^{(j)}$. Whenever it does not cause confusion we suppress the argument and write only t_j to denote the time of arrival in this state. We note that for realizations in which the patient dies before completing the planned treatment, i.e. transitions from a state $2^{(i')}$ to state 3 with $i' < i$, then we take $t_j = \infty$ or all $j \in \{i'+1,\ldots,i\}$. Next, let the time of administering treatment round j be denoted by $\tau_j(s(\cdot, x_i))$ (to take place in state $2^{(j-1)}$); as before, whenever it does not cause confusion, we write τ_j. Again, we take $\tau_j = \infty$ for all treatment rounds that the patient does not receive, either by the choice of treatment strategy, or by transitioning to state 3 sooner.

Taking therapy is costly. Each time the patient accepts therapy he or she incurs an instantaneous cost c. This may represent the monetary cost to pay for one round, lost income, or temporary discomfort caused by the therapy. Unlike the therapy's effects, which are delayed in the sense that the patient does not transition out of the detectable cancer state immediately upon choosing the round of therapy, the cost of a round of therapy is incurred the exact instant

the patient decides to take that round. This assumption is made for analytic tractability and is relaxed in Section 3.

We assume that the patient has a preference for earlier rewards, modeled via exponential discounting with a patient-specific discount factor $\rho > 0$.[4]

Given a strategy x_i and realization $s(\cdot, x_i)$, the patient's payoffs are given as

$$U(s(\cdot, x_i)) = \int_0^\infty \mathrm{e}^{-\rho t} q(s(t, x_i))\mathrm{d}t - \sum_{j=1}^{i} c\mathrm{e}^{-\rho\tau_j}. \tag{1}$$

We note that for any treatment strategy x_i with therapy ($i > 1$), we have $\tau_1 = t_0 = 0$. Due to $\rho > 0$, $U(s(\cdot, x_i))$ is finite for every realization of the stochastic process if a finite strategy x_i is chosen and for almost every realization if x_∞ is chosen.

For $j \leq i$ let

$$U^j(s(\cdot, x_i)) = \int_{t_j}^\infty \mathrm{e}^{-\rho(t-t_{j+1})} q(s(t, x_i))\mathrm{d}t - \sum_{j'=j+1}^{i} c\mathrm{e}^{-\rho(\tau_{j'}-t_j)}$$

denote the future payoffs of a patient who evaluates his or her prospect starting from state $2^{(j)}$, thus starting the game at time t_j which, if the patient accepts the next round of therapy, coincides with τ_{j+1}.

The patient chooses x_i to maximize his or her expected payoffs given by

$$V(x_i) = \mathbb{E}_{s(\cdot,x_i)} U(s(\cdot, x_i)). \tag{2}$$

As before, for $j \leq i$ we let

$$V^j(x_i) = \mathbb{E}_{s(\cdot,x_i)} U^j(s(\cdot, x_i))$$

denote the expected payoff of a patient who starts evaluating their prospects from state $2^{(j)}$.

From these formulations we derive main result of this section.

Proposition 2.1 (Recursive evaluation)**.** *For a fixed treatment strategy x_i with $i > 0$, the expected future payoffs in round $j < i$ is given as follows:*

$$V^j(x_i) = \frac{v}{\alpha_j + \rho} + \frac{\lambda_j}{\alpha_j + \rho} \cdot \frac{1}{\rho} + \frac{\beta_j}{\alpha_j + \rho}\left(\frac{u}{\delta_{j+1} + \rho} + \frac{\delta_{j+1}}{\delta_{j+1} + \rho} V^{j+1}(x_i)\right) + \frac{\gamma_j}{\alpha_j + \rho} V^{j+1}(x_i) - c, \tag{3}$$

$$V^i(x_i) = \frac{v}{\omega_i + \rho}, \text{ if } i \text{ is finite.} \tag{4}$$

Proposition 2.1 allows for the evaluation of the patient's payoffs in any state for any finite treatment through a linear recursive system. The right hand side of (3)'s five components are (1) the discounted expected payoff the patient collects in state $2^{(j)}$ before transitioning to any other state, (2) the discounted expected value of reaching state 0 (3) the discounted expected value

[4]Note that there is a mathematical equivalence between patients having explicit preferences for earlier rewards over later ones expressed by exponential discounting with rate ρ and between patients maximizing expected payoffs without time preferences but assuming a constant non-cancer-related mortality rate of ρ.

of transitioning to state $1^{(j+1)}$ followed by a transition into state $2^{(j+1)}$, (4) discounted expected value of a direct transition to state $2^{(j+1)}$, and (5) the instantaneous cost of the treatment. In (4), as there are no further rounds of therapy and the patient will progress to state 3. Thus the right hand side contains only the discounted expected value the patient collects in state $2^{(i)}$ before death. In the appendix we calculate each component and formally prove this result.

If for two treatment strategies, x_i, x_j, we have $V(x_i) \geq V(x_j)$ ($V(x_i) > V(x_j)$) we say that the patient *(strictly) prefers* i to j and denote it by $x_i \succsim x_j$ ($x_i \succ x_j$). We say that x_i is *optimal* if $x_i \succsim x_j$ for every j.

Proposition 2.1 allows for optimal treatment strategies to be derived efficiently even though, due to the time-heterogeneity of the transition rates, a closed form of (3) cannot be given. However, (3)-(4) can be transformed into a straightforward comparison between two "successive" strategies x_i and x_{i+1}. This allows for a myopic condition of stopping or continuing therapy that is shown in the next proposition.

Proposition 2.2 (Myopic stopping condition)**.** *For a finite i we have $x_i \succsim x_{i+1}$ if and only if*

$$v\frac{\alpha_i - \omega_i}{\omega_i + \rho} + c(\alpha_i + \rho) \geq u\frac{\beta_i}{\delta_{i+1} + \rho} + v\frac{1}{\omega_{i+1} + \rho}\left(\frac{\beta_i \delta_{i+1}}{\delta_{i+1} + \rho} + \gamma_i\right) + \frac{\lambda_i}{\rho}. \tag{5}$$

Under the myopic strategy, the patient compares stopping now in state $2^{(i)}$ with a strategy of taking therapy now and then stopping at the next detectable state $2^{(i+1)}$. The advantage of stopping treatment (left-hand-side of (5)) comes from the extra value from spending time in $2^{(i)}$ (v term, possibly negative if no therapy results in spending less time in expectation), plus the normalized saved cost of the treatment. The advantage of maintaining therapy for one more detectable state (right-hand-side of (5)) comes from the value of spending time in $1^{(i+1)}$ (u term), the value of spending time in $2^{(i+1)}$, either indirectly through $1^{(i+1)}$ or by a direct transition (v terms), and the value of possibly becoming cured.

Proposition 2.2 can be used to determine if, at any point, stopping therapy immediately is better than continuing for one more round with the intention of stopping therapy after that state. A sequence of such successive comparisons allows for a local optimization of the treatment strategy, but it may not result in a global optimum. The myopic optimization strategy will miss a better treatment strategy if, for instance, stopping treatment is better than continuing for one more round, but worse than continuing for two.

Under certain plausible monotonicity conditions, such local comparisons will produce the global optimum, e.g. if continuing for one more round is always better than stopping, then treatment should never be stopped. The last result of this section provides sufficient monotonicity conditions under which the optimal treatment strategy can be calculated through the myopic strategy.

Take the following homogeneity/monotonicity conditions:

- (H1): $u = v = 1$,
- (H2): $\delta_i = \delta$, $\omega_i = \omega$,
- (M1): $M(i) \leq M(i+1)$,
- (M2): $M(i) \geq M(i+1)$,

for all $i \in \mathbb{N}$, and

$$M(i) = \frac{\beta_i}{\alpha_i + \rho} \cdot \frac{\omega}{\delta + \rho} + \frac{\lambda_i}{\alpha_i + \rho} \cdot \frac{\omega}{\rho} + \frac{\omega - \mu_i}{\alpha_i + \rho}.$$

The value $M(i)$ is a measure of the advantage of taking therapy at state $2^{(i)}$; it is a weighted sum of the progression rates corresponding to at least partial therapy success (i.e. leading to states $1^{(i+1)}$ and 0) and the difference between the death rate without and with therapy.

The first condition pertains to the patient's preferences. Under (H1) the patient maximizes discounted life expectancy by spending as much time in states other than 3 as possible. Under (H2), the rate of progression from undetectable cancer to detectable, and the rate of death while living with untreated cancer, are constants and independent of i, the prior or current state of the disease. Under monotonicity condition (M1) the patient is improving under continuous therapy, transition probabilities become more favorable with each round. Under (M2) the reverse holds, the patient's prognosis worsens with each round.

Proposition 2.3 (Myopic optimization)**.** *Assume (H1) and (H2).*

1. *Under (M1) there exists an $i' \in \mathbb{N} \cup \{\infty\}$ such that for every $j < i \leq i'$ we have $x_i \prec x_j$ and for every $i > j \geq i'$ we have $x_i \succsim x_j$.*

2. *Under (M2) there exists an $i' \in \mathbb{N} \cup \{\infty\}$ such that for every $j < i \leq i'$ we have $x_i \succ x_j$ and for every $i > j \geq i'$ we have $x_i \precsim x_j$.*

The proof of Proposition 2.3 relies on the successive comparisons of Proposition 2.2. Under the first set of conditions, $V(x_i)$ is quasi-convex in i, while under the second it is quasi-concave. In either case we can provide the optimal treatment strategy, as shown by the next corollary.

Corollary 2.4 (Myopic optimization)**.** *Assume (H1) and (H2).*

1. *Under (M1), if $V(x_0) > V(x_\infty)$, then x_0 is the only optimal treatment strategy, if $V(x_0) < V(x_\infty)$, then x_∞ is an optimal treatment strategy, in case of equality both are optimal.*

2. *Under (M2), there exists $j' \geq i'$ such that $x_{i'}, x_{i'+1}, \ldots, x_{j'}$ are all optimal treatment strategies.*

To approximate $V(x_\infty)$, one may take the sequence $i = \{1, 2, \ldots\}$ and evaluate $V(x_i)$ through (3)-(4). As we have a positive discount rate, ($\rho > 0$), $V(x_\infty)$ will be the limit the sequence $V(x_i)_{i=1}^{\infty}$. In the second statement we find an optimal i' through a sequence of pairwise comparisons. As long as continuing therapy for one more round is weakly better than stopping immediately, the patient should continue. In this situation, a myopic strategy will identify the globally optimal treatment strategy, or a series of them.

Corollary 2.4 applies only if the homogeneity and monotonicity conditions (H1), (H2), and one of (M1) or (M2) hold. With (M1), the likelihood of cure or undetectable disease first decreases with each round, then increases after passing a threshold, meaning that either the treatment strategy x_0 or x_∞ is optimal. The latter case may arise when progressive disease states manifest as shrinkage of the overall tumor burden, elimination of the most life-threatening tumors or metastases, or therapy increases in efficacy.

With (M2), ceasing therapy at some point becomes optimal. Past the threshold, with each round the tumor burden, number of metastases, and resistance to therapy is increasing. This

means that the rate of adding QALYs declines with therapy in each progressive disease state i. For instance, if continued therapy only kills sensitive cells while leaving resistance cancer cells unharmed, then therapy, in time, results in diminishing returns. Under this condition, there exists an interior optimal treatment strategy beyond which further treatment is to the detriment of the patient.

Example 2.5 (Overtreatment)**.** To illustrate the model, we simulated the effects of overtreatment and calculated the loss of overall quality of life. To reduce the number of key factors we introduce a final homogeneity condition, (H3): $\beta_i = \beta$, $\gamma_i = \gamma$, $\mu_i = \mu$. Under (H3), only cure rate λ_i, depends on the patient's progressive disease state i. We let $\lambda_i = \lambda^{i+1}$ for some initial value λ. Table 1 gives the values for the time-homogeneous parameters of this simulation. The effects of varying λ and c are shown in Figure 2. As expected, the optimal duration of therapy increases with λ and decreases with c.

Parameter	ρ	δ	β	γ	μ	ω
Value	0.05	0.15	0.15	0.12	0.13	0.13

Table 1: For Example 2.5, we fixed ρ at 0.05 and randomized the transition parameters δ, β, γ, and μ between 0.1 and 0.2. We then set $\omega = \mu$, under which a decreasing $M(i)$ is guaranteed as long as λ_i is also decreasing with i which we have due to assuming $\lambda_i = \lambda^{i+1}$ for various levels of λ.

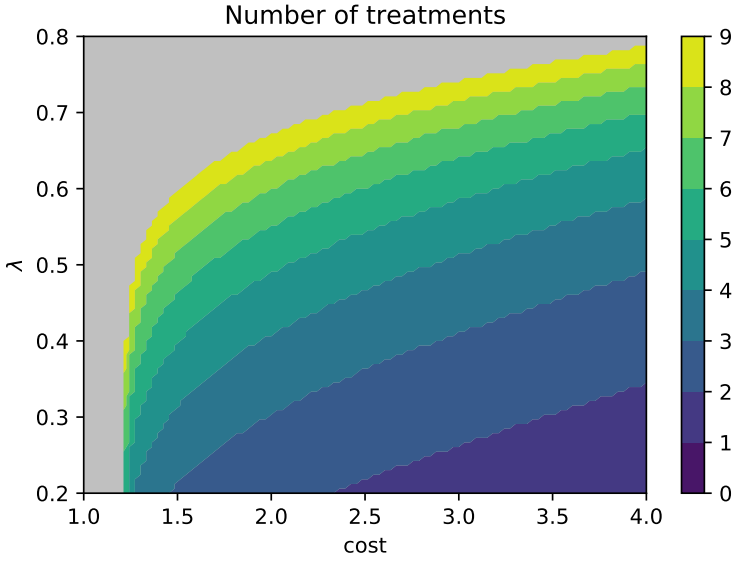

Figure 2: Optimal number of treatment rounds in the cost-based model for parameter values shown in Table 1. Gray areas show the regions in which more than 9 rounds of therapy is optimal. The progressive disease state at which the patient should cease therapy increases with the likelihood of cure and decreases with the cost to the patient in terms of money, lost income or ill-health from therapy.

For $\lambda = 0.4$ and $c = 3$ the parameters satisfy (M2) and the unique optimal strategy is x_2. Expected values of treatment strategies x_0 through x_7 are reported in Table 2 in relative terms to a healthy patient's payoff.[5]

[5]A healthy individual remains in state 0 and thus collects a payoff of 1 indefinitely. Taking into account time-discounting, this person has an absolute payoff of $1/\rho = 20$.

$V^j(x_i)$		j 0	1	2	3	4	5	6	7
	x_0	27.78%							
	x_1	49.95%	27.78%						
	x_2	52.24%	36.16%	27.78%					
x_i	x_3	52.17%	35.88%	27.04%	27.78%				
	x_4	51.91%	34.95%	24.59%	22.36%	27.78%			
	x_5	51.74%	34.31%	22.93%	18.69%	20.27%	27.78%		
	x_6	51.64%	33.96%	21.99%	16.62%	16.03%	19.39%	27.78%	
	x_7	51.59%	33.77%	21.49%	15.51%	13.77%	14.92%	19.04%	27.78%

Table 2: Expected values of treatment strategies x_0 to x_7 evaluated at different points of disease progression relative to a healthy individual's total payoffs with $\lambda = 0.4$ and $c = 3$. Taking 2 rounds is optimal, but further rounds diminish the present value (period 0) payoffs only marginally. Patients under continuous therapy who reach round 3 and beyond, if overtreated, have significantly lower prospects than patients who stop therapy.

As shown by Table 2, any treatment strategy that begins with therapy (x_i, $i > 0$) is better than no therapy at all (x_0)). The strategy of only having therapy in the first two progressive states (x_2) is the unique optimal strategy. However, as the cure rate, λ_i declines sharply in i, most patients who are not cured in the first two rounds lose the opportunity to do so in future rounds (Table 3).[6] For such patients, the cost of future therapy rounds is higher than the present value of the gains of postponing progression to state 3. If the standard of care is continuing therapy indefinitely, patients who survive beyond state $2^{(2)}$ are being overtreated and incur significant payoff losses. Patients in state $2^{(2)}$ lose 6.29% points under strategy x_7 when compared to the then-optimal x_2, patients in $2^{(3)}$ lose 12.27%, while patients who in $2^{(4)}$ lose the most at 14.01% of a healthy person's lifetime payoffs. Treatment strategies x_1 through x_7

Round	0	1	2	3	4	5
Cure rate	0.40	0.16	0.06	0.03	0.01	0.00
Cure probability	50.00%	28.57%	13.79%	6.02%	2.50%	1.01%
Death probability	16.25%	23.21%	28.02%	30.55%	31.69%	32.17%
Progression probability	33.75%	48.21%	58.19%	63.44%	65.82%	66.82%
Cohort size	100.00%	33.75%	16.27%	9.47%	6.01%	3.95%
Cured	0.00%	50.00%	59.64%	61.89%	62.46%	62.61%
Dead	0.00%	16.25%	24.08%	28.64%	31.54%	33.44%

Table 3: A simulated cohort's survival statistics under 'always treat' with $\lambda = 0.4$ up to state $2^{(5)}$.

[6]Note that this does not mean that subsequent rounds of therapy offer no benefits as patients under therapy have a longer life expectancy then those who are not even if $\lambda_i = 0$.

all provide very similar ex-ante evaluations despite the staggering payoff losses described above. This happened for two reasons: (1) the losses affect a minority of the population (only 9.47% of the cohort is neither cured nor dead after the third round, 6.01% after the fourth, 3.95% after the fifth), (2) the losses occur with a time delay starting at the time of reaching $2^{(2)}$, hence the differences are in the discounted future expected payoffs. Thus, the losses that occur due to overtreatment are obscured, delayed, and concentrated on a minority of patients making policy change to move away from the 'always treat' strategy in the standard of care very difficult.

It should also be noted that, while in our model and simulation, overtreatment is costly in lifetime payoff terms, the fraction of patients cured are larger the more rounds of therapy are taken. The strategy x_2 results in 59.64% of patients cured, while x_7 results in 62.66% (treatment strategies with more rounds of treatment offer only minuscule increases to the total cure rate). Furthermore, a payoff-maximizing patient who stops after reaching $2^{(2)}$ forgoes the cure percentage of 13.79% of the next round, showcasing that the objectives of oncologists and patients might differ and lead to highly different choices of treatment strategy; oncologists who maximize cure rate or expected patient survival time will choose 'always treat', while payoff-maximizing patients will abandon treatment relatively early.

Finally, we highlight the importance of the patient-specific decision parameters, particularly the discount rate, ρ. For large values of ρ, x_0 is optimal as the costs of even a single treatment are not recouped by the present values of higher expected QALYs. For $\rho = 0.096$, x_1 is optimal, then, as ρ decreases, treatment strategies with more and more treatment rounds become optimal with each treatment strategy being rationalizable with the right δ (Table 4). As ρ approaches zero, the optimal treatment strategy approaches the 'always treat' strategy.

Optimal strategy	Upper bound	Lower bound
x_0		0.098
x_1	0.097	0.072
x_2	0.071	0.048
x_3	0.047	0.031
x_4	0.030	0.018
x_5	0.017	0.011
x_6	0.010	0.006
x_7	0.005	

Table 4: Intervals of the patient's discount factor ρ with the corresponding optimal treatment strategy.

3 Toxicity-dependent payoffs

In our first model (Section 2), the cost of therapy was a constant and accrued only when therapy was being administered. This applies when the cost is monetary or under the simplifying assumption that the onset and cessation of any ill-health caused by the drug's toxicity switches instantly. Under these circumstances, the patient decides on which progressive disease state to cease therapy. Immediately upon progressing to the next detectable disease state the patient chooses whether or not to immediately undergo therapy for the duration of the detectable disease state. The disease transition rates changed only when the patient entered a new state.

Here we extend the model by separating the cost of therapy between the material cost and those directly affecting the patient's quality of life via therapy *toxicity*. We do this by considering the more realistic case where drug-induced malaise starts with therapy, and then declines with time upon ceasing therapy. In particular, when a round of therapy is unsuccessful in curing the disease, the lasting side-effects of the therapy can influence the decision to continue with therapy. The patient's level of therapy-induced toxicity negatively influences QALYs even when therapy has stopped.

Because of cumulative toxicity effects, a patient may decide on the timing of receiving therapy upon entering a new disease state. Drug holidays, for instance, provide a reprieve for the patient. The patient, upon entering a new disease state may delay the resumption of therapy. For our model, this means the patient spends some time experiencing the *no therapy* rate of progression before deciding to take the next round of treatment, after which the *therapy* transition rules apply. By including cumulative effects of drug toxicity, our next model captures the rational motivation behind taking drug holidays. We seek to find the optimal time for the patient to delay therapy upon entering the next detectable disease state, $2^{(i)}$.

We will add cumulative toxicity to the model by assuming that each round of therapy adds to toxicity, while its level decays exponentially over time. Let $i(t)$ denote the number of rounds of therapy taken up to time t. For $z_0, \hat{z} \geq 0$ and $\zeta > 0$ we define

$$z(z_0, t) = z_0 \mathrm{e}^{-\zeta t} + \sum_{i=1}^{i(t)} \hat{z} \mathrm{e}^{-\zeta(t-\tau_i)}, \tag{6}$$

The value $z(z_0, t)$ is called the patient's toxicity level, a negative payoff component to the patient's quality of life. Each round of therapy adds a fixed amount $\hat{z}$ to the patient's toxicity. Its starting level is denoted by z_0 and it decays exponentially with a constant rate ζ.

As an important component of the patient's well-being, the patient's choice on whether to continue therapy at the next detectable disease state, $2^{(i)}$, and when to start that therapy will be contingent on their current level of toxicity. Upon entering a state $2^{(i)}$, instead of a binary choice of take therapy or not, the patient chooses the to time delay therapy. By delaying for a time $\hat{t}$, during that time, the patient obeys the progression rule as if the *no therapy* choice was taken, i.e. moves to state 3 at rate ω_i. If the patient does not progress during this time, then they transition through the game tree in accordance with the *therapy* choice, i.e. moves to state 0, $1^{(i+1)}$, $2^{(i+1)}$, and 3 at rates λ_i, β_i, γ_i, and μ_i, respectively.

Formally, the patient's strategy is now described by a function $x\colon \{2^{(i)}\}_{i=0}^{\infty} \times [0, \infty) \to [0, \infty]$. For round i and toxicity level z the value $x(i, z)$ is the duration of the drug holiday in state $2^{(i)}$ before re-starting therapy. If this value is 0, then therapy begins immediately upon entering this disease state, if it is infinity, then the patient ceases taking therapy upon progressing to state $2^{(i)}$.

We now have two consistency conditions. First, as before, we restrict attention to strategies such that if for some i we have $x(i, z) = \infty$ for every z, then for every $j > i$ and every z' we have $x(j, z') = \infty$ as well, meaning that if the patient rejects therapy in state $2^{(i)}$, then the patients also rejects therapy in all future detectable diseases states. We call a treatment strategy *finite* if there exists i such that $x(i, z) = \infty$ for every z, i.e. the patient stops therapy after a finite number of rounds.

Second, we restrict attention to treatment strategies that are internally consistent within treatment rounds. Given treatment strategy x, if the patient arrives in a state $2^{(i)}$ with toxicity

level z, the patient will take the $i+1$th round of therapy with a delay of $x(i,z)$. If this is a finite, positive value, then the patient will wait and take the treatment round once his or her toxicity reaches a threshold value. During the wait the patient's toxicity will decrease and thus, pass through decision points with the same state $2^{(i)}$ and some lower toxicity levels $z' < z$. Our internal consistency restriction makes sure that the choices $x(i,z')$ conform to the original decision $x(i,z)$. In a way, we interpret the treatment strategies' waiting times as the patient's commitment to take the next round of therapy after that time has elapsed without changing his or her mind.

To formalize this idea, let i be given. We say that a treatment strategy x is internally consistent in treatment round i if for all $z > 0$ $x(i,z) = t < \infty$ implies

$$x(i,z') = t - \frac{1}{\zeta}\ln\left(\frac{z}{z'}\right),$$

for all $z' \in [z\mathrm{e}^{-\zeta t}, z]$. Intuitively, this amounts to assuming that the patient will indeed take the prescribed therapy after the waiting time has elapsed without making any decisions during the wait that would be inconsistent with this.

A treatment strategy x is internally consistent if it is internally consistent in all treatment rounds. In the remainder of this section we restrict attention to treatment strategies x that satisfy both consistency conditions. Note that if $x(i,z) = 0$ for some z, the condition is vacuous. This is because in this case the patient immediately takes the next round of therapy (hence paying the instantaneous cost and incurring the toxicity penalty right away) and thus could not change his/her mind during the waiting period. As a result, there is no restriction on the value of $x(i,z')$ for any $z' < z$. We further note that assuming internal consistency of the patient's treatment strategies is without loss of generality as the Markovian nature of the patient's problem means that all optimal treatment strategies will automatically satisfy it, but restricting the set of treatment strategies to internally consistent ones makes writing the patient's payoffs substantially easier.

The patient's *instantaneous payoff function when affected by toxicity*, $q\colon S \times [0,\infty) \to \mathbb{R}$, is given as

$$q(s,z) = \begin{cases} 1-z & \text{if } s \in \{0, \{1^{(i)}\}_{i=0}^{\infty}, \{2^{(i)}\}_{i=0}^{\infty}\} \\ 0 & \text{if } s = 3. \end{cases}$$

The patient collects a payoff of 1 in any health state other than 3, minus the amount of toxicity he or she currently has. We note that we allow for this value to be negative, indicating extreme discomfort for the patient, something that he or she may temporarily be willing to accept in the hopes of future recovery. In state 3, the patient collects a payoff of zero. We therefore replace the state-dependent quality-of-life-terms under therapy of our base model, u and v, with the toxicity-adjusted quality of life, $1-z$.

Given a treatment strategy x, state-realization $s(\cdot,x)$ and initial toxicity level z_0, the patient's *instantaneous payoff when affected by toxicity* is given by

$$U(s(\cdot,x),z_0) = \int_0^{\infty} \mathrm{e}^{-\rho t} q(s(t,x),z(z_0,t))\mathrm{d}t - \sum_{j=1}^{\infty} c\mathrm{e}^{-\rho\tau_j}, \tag{7}$$

where, as before τ_j denotes the time of administering the jth round of therapy. Due to $\rho > 0$, $U(s(\cdot,x),z(\cdot,x),0)$ is finite for every realization in every finite strategy and almost every realization for every strategy.

We define a patient's prospects starting in a general state $2^{(i)}$, at time t' conditional on the fact that their current toxicity level equals z_i before the $i+1$th round of therapy is taken as

$$U^i(s(\cdot,x),z_i,t') = \int_{t'}^{\infty} \mathrm{e}^{-\rho(t-t')} q(s(t,x),z(z_i,t))\mathrm{d}t - \sum_{j=i+1}^{\infty} c\mathrm{e}^{-\rho(\tau_j - t')}, \tag{8}$$

Given z_0, the patient chooses x to maximize their *discounted expected payoff*:

$$V(x,z_0) = \mathbb{E}_{s(\cdot,x)} U(s(\cdot,x),z_0).$$

A patient starting in state $2^{(i)}$ at time period t' with toxicity level z_i faces prospects given as

$$V^i(x,z_i) = \mathbb{E}_{s(\cdot,x)} U^i(s(\cdot,x),z_i,t'),$$

We note, that, due to the Markovian nature of the model, given the patient's chosen treatment strategy, starting state, and starting toxicity level, the expected payoff is independent of t'.

In the following proposition we establish how to evaluate a treatment strategy of a patient affected by toxicity.

Proposition 3.1 (Evaluation of treatment strategies under toxicity)**.** *At disease state $2^{(i)}$, for a treatment strategy x, with starting toxicity level z_i and where the patient waits time $\hat{t}$ before taking round $i+1$ (i.e. $x(i,z_i)=\hat{t}$), the patient's discounted expected payoff is calculated by the following recursive formula:*

$$\begin{aligned} V^i(x,z_i) = {} & \frac{1-\mathrm{e}^{-(\omega_i+\rho)\hat{t}}}{\omega_i+\rho} - \frac{z_i\left(1-\mathrm{e}^{-(\omega_i+\rho+\zeta)\hat{t}}\right)}{\omega_i+\rho+\zeta} + \mathrm{e}^{-(\omega_i+\rho)\hat{t}}\left(-c + \frac{1}{\alpha_i+\rho} - \frac{z_i\mathrm{e}^{-\zeta\hat{t}}+\hat{z}}{\alpha_i+\rho+\zeta}\right. \\ & +\lambda_i\left(\frac{1}{\rho(\alpha_i+\rho)} - \frac{z_i\mathrm{e}^{-\zeta\hat{t}}+\hat{z}}{(\rho+\zeta)(\alpha_i+\rho+\zeta)}\right) + \frac{\gamma_i}{\alpha_i}\int V^{i+1}(x, z_i\mathrm{e}^{-\zeta(\tau+\hat{t})}+\hat{z})\mathrm{e}^{-\rho y} f(y)\mathrm{d}y \\ & \left. + \frac{\beta_i}{\alpha_i}\left(\frac{\alpha_i}{(\alpha_i+\rho)(\delta_{i+1}+\rho)} - \frac{\alpha_i(z_i\mathrm{e}^{-\zeta\hat{t}}+\hat{z})}{(\alpha_i+\rho+\zeta)(\delta_{i+1}+\rho+\zeta)} + \int V^{i+1}(x, z_i\mathrm{e}^{-\zeta(y+\hat{t})}+\hat{z})\mathrm{e}^{-\rho y} g(y)\mathrm{d}y\right)\right). \end{aligned} \tag{9}$$

with probability measures

$$\begin{aligned} f(y) &= \alpha_i e^{-\alpha_i y}, \text{ for } y \geq 0, \\ g(y) &= \begin{cases} \frac{\delta_{i+1}\alpha_i}{\delta_{i+1}-\alpha_i}\left(e^{-\alpha_i y} - e^{-\delta_{i+1} y}\right) & \text{if } \alpha_i \neq \delta_{i+1} \\ \alpha_i^2 y e^{-\alpha_i y} & \text{if } \alpha_i = \delta_{i+1} \end{cases}, \text{ for } y \geq 0. \end{aligned}$$

Proposition 3.1 shows the relationship between the payoffs generated by treatment strategies in successive detectable disease states. The first component is the expected payoff the patient collects while waiting for the next round of therapy. The second component is the sum of three parts: the expected payoff of transitioning to state 0, the expected payoff of a direct transition to state $2^{(i+1)}$, and the expected payoff of a transition to state $2^{(i+1)}$ via state $1^{(i+1)}$.

It is clear that the cumulative toxicity model allows for significantly less analytic tractability than the instantaneous cost model of Section 2. This is most apparent when comparing the recursive formulae of Propositions 2.1 and 3.1. While the former shows a simple linear dependence

on successive disease states, the latter necessitates numerical methods of approximation. At the end of this section we examine a numerical example relying on such methods.

In special cases the toxicity model does provide analytically tractable results. Namely, a myopic calibration of the next round's delay, with the assumption that no further rounds will be taken. Thus, in the next lemma we evaluate finite treatment strategies close to the end of treatment. These provide optimal stopping conditions for myopic treatment strategies, and provide insights into a global optimization of treatment strategies. For $i \in \mathbb{N}$ let $X_i = \{x \colon x(i,z) = \infty \text{ for all } z\}$, i.e. treatment stops after i rounds. Due to the first consistency restriction these sets are nested, i.e. $X_i \subseteq X_{i+1}$ for every i.

Let

$$A_i(\rho) = \frac{1}{\alpha_i + \rho}\left(1 + \frac{\lambda_i}{\rho} + \frac{\gamma_i}{\omega_i + \rho} + \beta_i\left(\frac{1}{\delta_{i+1} + \rho} + \frac{\delta_{i+1}}{(\delta_{i+1} + \rho)(\omega_i + \rho)}\right)\right),$$

and

$$B_i(\rho) = \frac{1}{\omega_i + \rho}.$$

The following lemma gives an evaluation of three special strategies that form the cornerstones of myopic calibration of optimal delay.

Lemma 3.2 (Evaluating treatment strategies)**.** *1. For $x \in X_i$*

$$V^i(x, z_i) = B_i(\rho) - z_i B_i(\rho + \zeta). \tag{10}$$

2. For $x \in X_{i+1}$ with $x(i, z_i) = 0$

$$V^i(x, z_i) = A_i(\rho) - (z_i + \hat{z})A_i(\rho + \zeta) - c. \tag{11}$$

3. For $x \in X_{i+1}$ with $x(i, z_i) = \hat{t}$

$$\begin{aligned} V^i(x, z_i) = B_i(\rho)\left(1 - \mathrm{e}^{-(\omega_i+\rho)\hat{t}}\right) - z_i B_i(\rho + \zeta)\left(1 - \mathrm{e}^{-(\omega_i+\rho+\zeta)\hat{t}}\right) \\ + \mathrm{e}^{-(\omega_i+\rho)\hat{t}}\left(A_i(\rho) - (z_i \mathrm{e}^{-\zeta\hat{t}} + \hat{z})A_i(\rho + \zeta) - c\right). \end{aligned} \tag{12}$$

Lemma 3.2 shows straightforward evaluations of three treatment strategies for a patient currently in disease state $2^{(i)}$: (1) therapy is ceased immediately (i.e., after a total i previous rounds), (2) the final round of therapy (the $i+1$th) is applied immediately, and (3) the final round of therapy (the $i+1$th) is applied with a delay of $\hat{t}$ (i.e, the patient takes a drug holiday of duration $\hat{t}$). This result allows us to formulate a calibration of the optimal delay before the next round of therapy under the myopic assumption that no further rounds will be taken, as shown in the next proposition:

Proposition 3.3 (Myopic calibration of delay)**.** *Of the strategies with at most i rounds of therapy:*

1. If $B_i(\rho) - A_i(\rho) + \hat{z}A_i(\rho + \zeta) + c$ and $B_i(\rho + \zeta) - A_i(\rho + \zeta)$ are both negative, then the optimal time to administer the last round of therapy is to wait until the patient's toxicity level reaches a threshold $\overline{z}$ with

$$\overline{z} = \frac{B_i(\rho+\zeta)}{B_i(\rho)} \cdot \frac{B_i(\rho) - A_i(\rho) + \hat{z}A_i(\rho+\zeta) + c}{B_i(\rho+\zeta) - A_i(\rho+\zeta)},$$

or, if the patient's toxicity is below this level, then administer the last round of therapy immediately.

2. If $B_i(\rho) - A_i(\rho) + \hat{z}A_i(\rho+\zeta) + c > 0$ *and* $B_i(\rho+\zeta) - A_i(\rho+\zeta) < 0$*, then stopping at the* $i-1$*th round is better than continuing with the* i*th round.*

3. If $B_i(\rho) - A_i(\rho) + \hat{z}A_i(\rho+\zeta) + c < 0$ *and* $B_i(\rho+\zeta) - A_i(\rho+\zeta) > 0$*, then treatment should be administered immediately.*

4. If $B_i(\rho) - A_i(\rho) + \hat{z}A_i(\rho+\zeta) + c$ *and* $B_i(\rho+\zeta) - A_i(\rho+\zeta)$ *are both positive, then treatment should be administered immediately if the patient's toxicity is above the threshold* z' *and never if it is below it, with*

$$z' = \frac{B_i(\rho) - A_i(\rho) + \hat{z}A_i(\rho+\zeta) + c}{B_i(\rho+\zeta) - A_i(\rho+\zeta)}.$$

Proposition 3.3 plays a similar role as Section 2's Proposition 2.2. It identifies a myopically optimal stopping condition at a particular disease state without an intention of resuming therapy in subsequent disease states. Moreover, it determines the myopically optimal waiting time through analytic methods. Under condition (1) treatment is to be delayed until toxicity is sufficiently diminished, under (2) it is to be canceled no matter the patient's toxicity level, under (3) it is to be administered immediately no matter the patient's toxicity level, and finally, under (4) it is to be administered only for patients with high toxicity level. The final point shows a perverse case, resulting from the fact that patients with high negative instantaneous payoffs prefer to immediately receive the next round even though the transition parameters are such that doing so decreases the patient's life expectancy.

Example 3.4. In this example we demonstrate the value gained from calibrating the duration of the treatment holiday in a detectable disease state. As in Example 2.5, we let $\lambda_i = \lambda^{i+1}$ for an initial value λ. Consider the transition parameters shown in Table 5.

Parameter	ρ	δ	β	γ	μ	ω
Value	0.05	0.1	0.1	0.2	0.3	0.2

Table 5: The calibration of Example 3.4.

For a benchmark, we first consider the no toxicity case with $\lambda = 0.67$, meaning that we evaluate this example through Section 2's model. Then, as in Example 2.5, (M2) is satisfied. In Table 6, for each treatment strategy x_0 through x_8, we report the corresponding range of the treatment costs, c that lead produce it as the unique payoff-maximizing strategy.
Now consider the case of toxicity. To showcase its effect we set $c = 0$, i.e. the incentive of stopping treatment comes solely from the patient's decreased quality of life due to toxicity. We take $z_0 = 0$, $\hat{z} = 0.5$, and $\zeta = 0.03$. Under these parameters, the total payoff-reduction of one round of therapy from toxicity is $\hat{z}/(\rho+\zeta) = 6.25$, called its *present cost*. However, this cost is realized in full only by patients with a death rate of zero as patients who move to state 3 experience the quality-of-life reduction from toxicity for a shorter time. Patients in non-absorbing states face a constant death rate of $\mu = \omega = 0.2$ and hence face an *expected present*

Cost range	Optimal strategy	Payoff range (% of healthy)	Total cured (%)
0.84 – 1.13	x_8	64.13% – 62.28%	65.90%
1.14 – 1.55	x_7	62.22% – 59.61%	65.90%
1.56 – 2.13	x_6	59.55% – 55.93%	65.89%
2.14 – 2.92	x_5	55.87% – 50.92%	65.84%
2.93 – 3.94	x_4	50.85% – 44.47%	65.69%
3.95 – 5.20	x_3	44.40% – 36.58%	65.12%
5.21 – 6.65	x_2	36.52% – 27.87%	62.87%
6.66 – 8.22	x_1	27.81% – 20.01%	52.76%
8.23+	x_0	20.00%	0.00%

Table 6: Payoff-maximizing treatment strategies for various cost ranges, their corresponding ex-ante payoff ranges relative to a healthy individual, and total cure percentages.

cost of $\hat{z}/(\rho + \zeta + \omega) = 1.79$. As such, based on Table 6 we can expect at least 2 rounds of therapy, and at most 6.

Through Proposition 3.3, we can analytically derive a myopically optimal treatment plan, i.e. the optimal waiting times before each round under the assumption that there will be no further rounds of therapy attempted. As the benefits of therapy are declining with each round, the globally optimal strategy will be to delay therapy in future rounds more and more, thus the myopic assumption that therapy will cease after the current round under consideration will matter less and less. As such, myopic optimization will produce increasingly accurate estimates of the globally optimal treatment strategy as the patient takes more rounds. In Table 7 we report the threshold levels of toxicity in each round. With $z_0 = 0$ and $\hat{z} = 0.5$ the first two

Round	Cure rate	Cure probability	Threshold toxicity
1	0.67	0.53	0.87
2	0.45	0.43	0.76
3	0.30	0.33	0.65
4	0.20	0.25	0.48
5	0.14	0.18	0.22
6	0.09	0.13	negative

Table 7: Threshold toxicity levels below which the next round of treatment can be delivered under myopically optimal treatment strategies. Above this level, a payoff-maximizing myopic patient waits until toxicity drops to the threshold level before taking therapy.

rounds are delivered as soon as possible to the patient as the threshold of round 1 is 0.87, while that of round 2 is 0.76, and the maximum toxicity that the patient can have after round 1 is 0.5. From round 3 onward, however, the patient may be better off delaying, if their toxicity exceeds the threshold corresponding to round $i + 1$'s at the time of arrival to state $2^{(i)}$.

For a specific case consider a patient in state $2^{(2)}$, deciding on the delay of the third round. This patient has taken therapy in two detectable disease states and their toxicity level increased twice by $\hat{z} = 0.5$, however, in the intermittent times of waiting for the transitions (in states $2^{(0)}, 2^{(1)}$, possibly visiting $1^{(1)}$ or $1^{(2)}$ or both as well), the patient's toxicity level has declined.

In our example we set $z_2 = 0.73$. The patient is facing a cure rate of $\lambda_3 = 0.3$. By Table 7, this patient's payoff is maximized by waiting until the toxicity level reaches 0.65 to take the third round. The patient's present value, depending on their delay of taking the third round is shown in Figure 3.

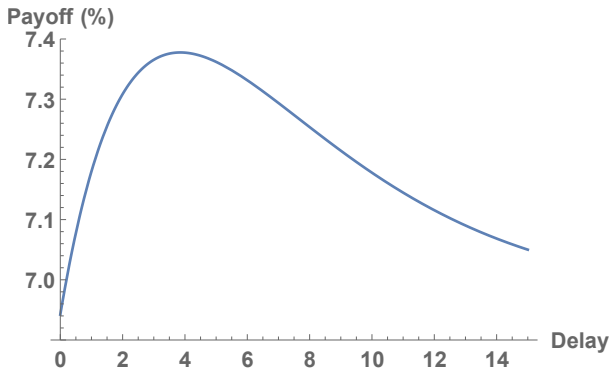

Figure 3: The patient's payoffs relative to a healthy individual's after completing two rounds as a function of round 3's delay with toxicity rate $z_2 = 0.73$ and facing a cure rate of $\lambda_2 = 0.3$. Expected payoffs are maximized at a delay of $\hat{t}_3 = \ln(z_2/\bar{z})/\zeta = 3.86$

We note that the patient's decision to delay the third round may seem surprising, considering that the probability of cure is still high (0.33), and that during the waiting time of 3.86 their probability of death is even higher ($e^{-3.86\omega} = 0.54$). It is clear that such a decision is not supported by practices that maximize probability of cure or survival time. The decision to delay is cast in a more favorable light by considering that receiving the toxicity hit of the third round immediately would yield a quality of life of -0.23. Even at the threshold toxicity of 0.65 the patient's quality of life turns temporarily negative. The delay lowers the patient's present cost of therapy enough for a payoff-maximizing patient to resume therapy.

Example 3.4 showcases both the possible benefits of delaying therapy (Figure 3) and a myopically optimal patient's behavior (Table 7). It also highlights the comparison between the models of Sections 2 and 3. The former prescribes the number of treatment rounds based on the flat one-time cost the patient incurs per round, while the latter prescribes the timing of these rounds. Note, however, that unlike in Section 2, where we were able to derive a condition that ensured that the myopically optimal behavior produces the globally optimal one (Proposition 2.3), there is no analogous result to guarantee that Table 7's results correspond to the globally optimal behavior in the toxicity model. In the next example, we evaluate the same calibration via a numerical approximation and show that its results are in agreement with the myopically optimal waiting times.

Example 3.5. Consider the same transition parameters as shown in Table 5. As in Example 3.4, we take $\lambda = 0.67$, $\hat{z} = 0.5$ and $\zeta = 0.03$ with $z_0 = 0$. Table 8 reports the expected optimal delays of a maximum of six treatment rounds through a numerical approximation (see the appendix for a summary of the methodology of the approximation).
The interpretation of the prescribed treatment strategy starting at round 0 (first row of Table 8) is as follows: Given the patient's toxicity level of $z_0 = 0$, in expectation, the patient is advised to wait time $\hat{t}_i$ before receiving the $i+1$th round of therapy. Note that the prescribed waiting times for distant treatment rounds are subject to change. At the onset, they are merely an expected

			Round	0	1	2	3	4	5
			Cure rate	0.67	0.45	0.30	0.20	0.14	0.09
i	z_i	Payoff	Cure perc.	52.76%	42.80%	33.39%	25.14%	18.37%	13.10%
0	0.00	42.70%		**0**	0	0	11.27	20576	∞
1	0.32	24.12%			**0**	0	13.76	17.67	∞
1	0.40	21.16%			**0**	0.84	13.76	118.88	∞
1	0.48	18.28%			**0**	3.68	13.91	13.77	∞
2	0.60	11.09%				**0**	15.44	176.77	∞
2	0.68	8.62%				**1.44**	16.96	22.31	∞
2	0.76	6.73%				**5.14**	16.96	22.19	∞
2	0.84	5.13%				**8.48**	16.96	21.88	∞
2	0.92	3.63%				**11.51**	16.96	24.12	∞

Table 8: Delay times and payoffs of approximate optimal strategies, $x^*(i, z_i)$ conditional on starting therapy in round i with toxicity level z_i. Bold numbers are actionable choices, all other delays are expected values subject to change. A patient progressing through the disease states re-optimizes in each detectable state and tailors their behavior based on the current level of toxicity.

time of optimal delay given the patient's *expected* progression, on which, based on backwards induction, the optimal time of delay of the first round, $\hat{t}_0 = 0$, can be calculated. Thus, only this first delay is actionable information. Should the patient reach the next decision node, their toxicity level may be quite different from the expected levels, hence, subsequent decisions need to be taken according to the *realized* toxicity levels.

To illustrate, we report three re-optimized treatment strategies given toxicity levels $z_1 = 0.32, 0.40$, and 0.48 after round 1 (rows 2 to 4 of Table 7). This large divergence in toxicities is based on the fact that patients who do not respond to the treatment (and thus progress to state $2^{(1)}$ directly) are expected to have larger toxicity levels than those who do (and thus reach $2^{(1)}$ indirectly through $1^{(1)}$), as the latter group's toxicity is allowed to decline for a longer time.[7] As shown in the table, these patients are all advised to take round 2 immediately, but their expected delays in future rounds, as well as their expected payoffs, diverge.

Those patients who progress further again need to re-optimize based on their realized levels of toxicity. We approximate optimal treatment strategies for patients who start after round 2 with toxicity levels $z_2 = 0.60, 0.68, 0.76, 0.84$, and 0.92. At this stage, the prescribed delays before taking round 3 are different, hence the different patients' payoff-maximizing behavior diverges. The approximate delay times of the next round line up with the myopically optimal ones (retrieved from Proposition 3.3) up to the 3rd decimal point, indicating that the approximate optimal solution and the myopically optimal one agree closely, provided that λ_i is decreasing.

[7]The expected time spent in state $1^{(i)}$ is $1/\delta = 10$ in this example, while toxicity level upon leaving state $1^{(i)}$ if it was at level z' upon entering it is $z'\delta/(\delta + \zeta)$, so an initial toxicity level of around 0.5 decreases to around 0.38.

4 Discussion

In this paper we develop a decision-making tool of cancer therapy. We model the development of the disease as a random, Markovian process, capturing the prognosis-relevant data with four types of health states (cure, undetectable tumor burden, detectable burden and death). This approach unifies the more classical Markovian models of cancer therapy (Cooper et al., 2003, 2004) with the novel game theoretic analysis of cancer (Orlando et al., 2012; Staňková et al., 2019), adding the element of patient choice to the former, and simplifying cancer's evolutionary dynamics to a stochastic, Markovian process in the latter. This framing of cancer strategies in response to therapy allows us to focus on the patient's choices. We then rely on classic results from Markov Decision Processes for the existence of a unique optimal treatment strategy.

The model's main disease-specific inputs are estimates of transition rates associated with and without therapy. These rates consider transitions from detectable disease to cure, death, undetectable disease, and the next detectable progressive disease state; and transitions from undetectable disease to a detectable disease state. Such data would need to be estimated from large patient cohorts that consider cure, complete responses, partial responses, stable disease states, disease progression and mortality from therapeutic regimens. These regimens could include the application of the same therapeutics regardless of disease state, or changes to the therapies in response to changed disease state. Sources of data can include clinical trials, patient outcome data compiled by governments, and published peer-reviewed papers on specific cancers using large patient cohorts (e.g. breast cancer, Urru et al. (2018); lung cancer, Sun et al. (2016); pancreatic cancer, He et al. (2020)). For instance, The National Cancer Database sponsored by the American College of Surgeons and the American Cancer Society can be used to analyze cancer patients, their treatments, and outcomes. With more than 34 million records, the database accrues more than 70 percent of newly diagnosed cancer cases within the United States (National Cancer Database (facs.org)). The Children's Oncology Group, supported by the National Cancer Institute, provides access to data on childhood and adolescent cancers from cancer centers across North America (>14,000 new patients per year), Australia, New Zealand, and Europe.

The patient-specific inputs include the patient-specific perceptions of the cost of therapy. These can include financial hardships (Ell et al., 2008; McNulty and Khera, 2015; Smith et al., 2021), emotional stress (Delgado-Guay et al., 2015; Traeger et al., 2009) and toxicity (Cleeland et al., 2012). For our model, challenges exist in terms of patients revealing or perceiving these costs. Surveys exist for evaluating these costs. Examples include for immune-checkpoint therapies (Hansen et al., 2020), for breast cancer patients undergoing diverse therapies (Mokhatri-Hesari and Montazeri, 2020; Bjelic-Radisic et al., 2020), and for thyroid cancer treated with lenvatinib (Giani et al., 2021). Additionally, the model requires patients to reveal or have a sense for how they discount time (see Vaughn et al. (2020) for the case of breast cancer patients).

With knowledge of transition rates of disease states and patient-specific parameters regarding time discounting and therapy costs, our first model provides a simple recursive formula to analytically evaluate the performance of various treatment strategies. This tool then allows the patient to choose their preferred therapy duration. Under some monotonicity and homogeneity assumptions, a myopic (looking just one disease state ahead) evaluation of the treatment strategies also produces the globally optimal outcome, further simplifying the decision-making progress. In a second model, where the patient's instantaneous payoffs were determined by their current toxicity levels from therapy, the evaluation of treatment strategies becomes more com-

plicated and requires numerical tools. Nevertheless, optimal duration of therapy and optimal timing of treatment rounds can be estimated. Myopically optimizing the next round's delay can be performed analytically, and can provide a good approximation to a globally optimal treatment strategy if the cure rate in future detectable disease states decreases.

We raise four discussion points on the modeling choices made in the paper. The first is the decision to include no more than four types of health states. One reason for this is to keep our models tractable. A second reason is that a practical application of a model with more health states requires more cohort data. Given the same amount of cohort data, calibrating a model with more than four health states comes with a loss of statistical power. In the case of large cohorts, collecting patient data of a given cancer type, this may not be a problem. However, in the case of cohorts stratified by age, sex, or by other variables, diluting the data in favor of including a larger number of health states may not be desirable. We further argue that more health states raises classification problems, while the four present in our paper is the lowest number that is needed. In cases where data are abundant and classification unproblematic, our model may be extended to include more state types in a straightforward manner.

Secondly, we reflect upon one of our model's main limitations, its Markovian nature. Particularly, upon the fact that the transition probabilities between states are time-independent. In reality this is not necessarily the case for the parameters of disease progression. In some cases, the likelihood of progressing increases with time spent in the disease state, and vice-versa (see Cleophas and Van Ouwerkerk (2007) for a discussion of the issue of using exponential models in clinical research). For our purposes, it is a necessary assumption for us to apply Markovian methods and models, and it represents the first order approximation of time dependent transition rates. Furthermore, these transition rates will be highly patient specific and depend on age, sex, time at which the disease was first detected, disease burden at detection, genetic predispositions, immune competency, etc. Hence, a given detectable cancer state for one patient may be quite different from another requiring appropriate adjustments to the transition rates. Yet, there may be pools of patients that provide cohorts from which to generally estimate these rates, at least for common cancers.

Thirdly, we raise the issue of personalized medicine. As we state above, the transition rates of our model are to be calibrated from cohort data. The ability to personalize our model depends on the availability cohort data corresponding to the patient's characteristics. For some cancers and for some strata this cannot be taken as given. In these cases, our models can still serve as useful benchmarks against which the patient and their physician may evaluate their options given the patient's own characteristics and responses. Even when the ability to personalize our model's transition rates is low, some of our model's variables such as the patient's instantaneous payoff parameters and discount rate can be calibrated to match the patient's preferences and characteristics. When personalization is high, the differences between these patient-specific traits may still mean that two patients belonging to the same demographic will favor different treatment strategies.

Fourthly, we address the relationship of the patient's toxicity level in our second model and the transition rates. In our model, these are mathematically independent in the sense that after a given number of rounds of therapy, progression rates are not affected by toxicity. In practice, toxicity caused by therapy is strongly related to the patient's prognosis. This mismatch is caused by the fact that our model combines "objective" parameters regarding disease prognosis with "subjective" ones that reflect the patients' preferences. Toxicity of therapy is related to

both. We therefore use the abstract term toxicity to reflect on the subjective aspect, measuring the patient's quality of life under therapy. Introducing explicit dependence between toxicity and transition would be problematic both for the tractability of the model and in mixing the objective concerns with subjective ones. For example, two patients may be very similar in their disease progression but may report varying levels of discomfort due to therapy, or vice versa, which may influence their choice of treatment. As the transition rates do depend on the number of rounds of therapy, our toxicity measure and the patient's prognosis are statistically not independent.

Finally, we reflect on our stated goal, to address the dilemmas arising from the difficulty in finding a suitable measure of success of cancer therapy. Our approach, maximizing the patient's discounted expected QALYs is rooted in a classic economic approach that treats individuals as rational utility maximizers. As such, we propose it as a good candidate to evaluate cancer therapy in a way that explicitly captures the patients' well-being. As an additional value, even if such an approach cannot be adopted in oncology formally, a model such as this can help identify and understand points of disagreement between cancer patients and their treating physicians in selecting a treatment strategy.

The model has several key utilities. First, there can be circumstances where a patient's optimal choice is to cease therapy even when cure may still be possible. This may pose ethical dilemmas for the physician. Generally, quantity versus quality of life tradeoffs come most into play when therapy is palliative and the disease state is assessed as incurable. In the model, a patient's willingness to cease therapy may be in part due to financial distress (Beeler et al., 2020). This creates health disparities between those with and without access to inexpensive health care, or employer supported sick leave. Second, the model can predict, on a patient to patient basis, the duration and timing of drug holidays. Current practice often has a pre-determined protocol for taking breaks in therapy regardless or patient preferences, or manages them haphazardly based on the patient's level of discomfort or abnormal bloodwork.

Our approach inherits the limitations and criticism of its two main components, QALYs and expected utility maximization. The former includes difficulty in measurement, interpersonal comparison, and equity concerns. Similarly, expected utility theory has its detractors, both in static settings (such as the well-known Allais, Ellsberg, and St. Petersburg paradoxes) and dynamic ones (such as time-inconsistent preferences). Addressing the former in the cancer context is part of a deeper discussion on the appropriateness of using QALYs. We argue that, while its shortcomings do not make it suitable to replace less controversial measures, such as survival time, considering QALYs in addition to survival time has significant added value. Addressing the latter in our setting requires a deeper mapping of the individual decision-making process. Methods of behavioral economics, psychology, and other decision sciences use model and tools based explicitly on expected utility theory. Thus, our model and its predictions, can serve as useful benchmarks for future research in the decision theory of cancer.

A Appendix

Proposition 2.1

We first show the second part of the statement, that is:

$$V^i(x_i) = \frac{v}{\omega_i + \rho},$$

for a finite i.

The patient collects a constant stream of instantaneous payoffs v while still in state $2^{(i)}$, and 0 after he or she transitions to state 3. Let ξ denote the time the patient spends in $2^{(i)}$. As $\xi \sim \text{Exp}(\omega_i)$, we have

$$V^i(x_i) = \mathbb{E}_\xi\left(\int_0^\xi v\mathrm{e}^{-\rho t}\mathrm{d}t\right) = \int_0^\infty\int_0^\xi v\mathrm{e}^{-\rho t}\mathrm{d}t\ \omega_i\mathrm{e}^{-\omega_i\xi}\mathrm{d}\xi = v\omega_i\int_0^\infty\left[-\frac{\mathrm{e}^{-\rho t}}{\rho}\right]_0^\xi \mathrm{e}^{-\omega_i\xi}\mathrm{d}\xi$$
$$= \frac{v\omega_i}{\rho}\int_0^\infty\left(1-\mathrm{e}^{-\rho\xi}\right)\mathrm{e}^{-\omega_i\xi}\mathrm{d}\xi = \frac{v\omega_i}{\rho}\left(\frac{1}{\omega_i}-\frac{1}{\omega_i+\rho}\right) = \frac{v}{\omega_i+\rho}.$$

To show the first part we calculate each of the following four components separately: (1) the discounted payoffs collected in state $2^{(j)}$ before transitioning; (2) those collected after transitioning to state 0; (3) those collected after transitioning to state $1^{(j+1)}$, followed by transitioning to state $2^{(j+1)}$; (4) those collected after a direct transition to $2^{(j+1)}$.

Calculating (1) amounts to evaluating

$$\mathbb{E}_\xi\left(\int_0^\xi v\mathrm{e}^{-\rho t}\mathrm{d}t\right) = \int_0^\infty\int_0^\xi v\mathrm{e}^{-\rho t}\mathrm{d}t\ \alpha_j\mathrm{e}^{-\alpha_j\xi}\mathrm{d}\xi = \frac{v}{\alpha_j+\rho},$$

with very similar steps as before, where now we have $\xi \sim \text{Exp}(\alpha_j)$.

To calculate (2) we need to evaluate

$$\mathbb{E}_\xi\left(\int_\xi^\infty \mathrm{e}^{-\rho t}\mathrm{d}t\right) = \int_0^\infty\int_\xi^\infty \mathrm{e}^{-\rho t}\mathrm{d}t\ \alpha_j\mathrm{e}^{-\alpha_j\xi}\mathrm{d}\xi = \alpha_j\int_0^\infty\left[-\frac{\mathrm{e}^{-\rho t}}{\rho}\right]_\xi^\infty \mathrm{e}^{-\alpha_j\xi}\mathrm{d}\xi$$
$$= \frac{\alpha_j}{\rho}\int_0^\infty \mathrm{e}^{-\rho\xi}\mathrm{e}^{-\alpha_j\xi}\mathrm{d}\xi = \frac{\alpha_j}{\rho}\frac{1}{\alpha_j+\rho}$$

as once more we have $\xi \sim \text{Exp}(\alpha_j)$. Multiplying by λ_j/α_j, the probability that state 0 is reached, we get

$$\frac{1}{\rho}\cdot\frac{\lambda_j}{\alpha_j+\rho}.$$

Component (3) has two parts: the payoffs collected while the patient is in state $1^{(j+1)}$, and the payoff he or she collects after transitioning to $2^{(j+1)}$. Taking $\xi \sim \text{Exp}(\alpha_j)$ and $\xi' \sim \text{Exp}(\delta_{j+1})$, the former amounts to

$$\mathbb{E}_{\xi,\xi'}\left(\int_\xi^{\xi+\xi'} u\mathrm{e}^{-\rho t}\mathrm{d}t\right) = \int_0^\infty\int_0^\infty\int_\xi^{\xi+\xi'} u\mathrm{e}^{-\rho t}\mathrm{d}t\ \alpha_j\mathrm{e}^{-\alpha_j\xi}\mathrm{d}\xi\ \delta_{j+1}\mathrm{e}^{-\delta_{j+1}\xi'}\mathrm{d}\xi' = u\alpha_j\delta_{j+1}\int_0^\infty\int_0^\infty$$
$$\left[-\frac{\mathrm{e}^{-\rho t}}{\rho}\right]_\xi^{\xi+\xi'} \mathrm{e}^{-\alpha_j\xi}\mathrm{d}\xi\ \mathrm{e}^{-\delta_{j+1}\xi'}\mathrm{d}\xi' = \frac{u\alpha_j\delta_{j+1}}{\rho}\int_0^\infty\int_0^\infty\left(\mathrm{e}^{-(\alpha_j+\rho)\xi}-\mathrm{e}^{-(\alpha_j+\rho)\xi}\mathrm{e}^{-\delta_{j+1}\xi'}\right)\mathrm{d}\xi\ \mathrm{e}^{-\delta_{j+1}\xi'}\mathrm{d}\xi'$$
$$= \frac{u\alpha_j\delta_{j+1}}{\rho}\cdot\frac{1}{\alpha_j+\rho}\int_0^\infty\left(\mathrm{e}^{-\delta_{j+1}\xi'}-\mathrm{e}^{-(\rho+\delta_{j+1})\xi'}\right)\mathrm{d}\xi' = \frac{u\alpha_j\delta_{j+1}}{\rho}\cdot\frac{1}{\alpha_j+\rho}\left(\frac{1}{\delta_{j+1}}-\frac{1}{\delta_{j+1}+\rho}\right)$$
$$= \frac{\alpha_j}{\alpha_j+\rho}\cdot\frac{u}{\delta_{j+1}+\rho}.$$

This, multiplied by the probability of reaching state $1^{(j+1)}$, β_j/α_j gives

$$\frac{\beta_j}{\alpha_j+\rho}\cdot\frac{u}{\delta_{j+1}+\rho}.$$

The second part, the payoff the patient receives after transitioning to $2^{(j+1)}$ amounts to receiving a payoff of $V^{j+1}(x_i)$ with time delay $\xi+\xi'$, that is, in expectation:

$$\frac{\alpha_j}{\alpha_j+\rho}\cdot\frac{\delta_{j+1}}{\delta_{j+1}+\rho}V^{j+1}(x_i).$$

Multiplying by the probability of reaching state $1^{(j+1)}$ (from which reaching state $2^{(j+1)}$ is certain), we get

$$\frac{\beta_j}{\alpha_j+\rho}\cdot\frac{\delta_{j+1}}{\delta_{j+1}+\rho}V^{j+1}(x_i).$$

The sum of the two parts gives the third component of (3) as desired.

In component (4), a direct transition to state $2^{(j+1)}$ provides a payoff of $V^{j+1}(x_i)$ with a delay of ξ with $\xi\sim\text{Exp}(\alpha_j)$, equaling

$$\frac{\alpha_j}{\alpha_j+\rho}V^{j+1}(x_i).$$

Multiplied by the probability of reaching $2^{(j+1)}$ directly, γ_j/α_j, we get

$$\frac{\gamma_j}{\alpha_j+\rho}V^{j+1}(x_i).$$

Finally, subtracting the cost of a round of therapy, c, incurred immediately, we get the right hand side of (3).

Proposition 2.2

As the two treatment strategies are identical in the first i periods, $V(x_i)\geq V(x_{i+1})$ if and only if $V^i(x_i)\geq V^i(x_{i+1})$. By Proposition 2.1 the left hand side amounts to $v/(\omega_i+\rho)$, while the right hand side is

$$V^i(x_{i+1})=\frac{v}{\alpha_i+\rho}+\frac{\lambda_i}{\alpha_i+\rho}\cdot\frac{1}{\rho}+\frac{\beta_i}{\alpha_i+\rho}\left(\frac{u}{\delta_{i+1}+\rho}+\frac{\delta_{i+1}}{\delta_{i+1}+\rho}V^{i+1}(x_{i+1})\right)+\frac{\gamma_i}{\alpha_i+\rho}V^{i+1}(x_{i+i})-c.$$

By plugging in $V^{i+1}(x_{i+1})=v/(\omega_{i+1}+\rho)$ we have that $V^i(x_i)\geq V^i(x_{i+1})$ if and only if

$$\frac{v}{\omega_i+\rho}\geq\frac{v}{\alpha_i+\rho}+\frac{\lambda_i}{\alpha_i+\rho}\cdot\frac{1}{\rho}+\frac{\beta_i}{\alpha_i+\rho}\left(\frac{u}{\delta_{i+1}+\rho}+\frac{\delta_{i+1}}{\delta_{i+1}+\rho}\frac{v}{\omega_{i+1}+\rho}\right)+\frac{\gamma_i}{\alpha_i+\rho}\cdot\frac{v}{\omega_{i+1}+\rho}-c.$$

Multiplying by $\alpha_i+\rho$ and rearranging produces the inequality stated by the proposition.

Proposition 2.3

Applying (H1) and (H2) to (5), by Proposition 2.2 we have $x_i \precsim x_{i+1}$ if and only if

$$\frac{\beta_i+\gamma_i+\lambda_i+\mu_i-\omega}{\omega+\rho}+c(\alpha_i+\rho)\le\frac{\beta_i}{\delta+\rho}+\frac{1}{\omega+\rho}\left(\frac{\beta_i\delta}{\delta+\rho}+\gamma_i\right)+\frac{\lambda_i}{\rho}.$$

Multiplying by $(\omega+\rho)/(\alpha_i+\rho)$ and rearranging gives

$$c\le\frac{1}{\omega+\rho}\left(\frac{\beta_i}{\alpha_i+\rho}\cdot\frac{\omega}{\delta+\rho}+\frac{\lambda_i}{\alpha_i+\rho}\cdot\frac{\omega}{\rho}+\frac{\omega-\mu_i}{\alpha_i+\rho}\right)=\frac{1}{\omega+\rho}M(i). \tag{13}$$

1. Let $i'\in\mathbb{N}$ be the smallest number such that $x_{i'}\precsim x_{i'+1}$. Then we have $c\le M(i')/(\omega+\rho)$. Under (M1) $M(i)$ is increasing in i, thus every successive treatment strategy with more than i' rounds is no worse than the one preceding it, hence for every $i>j\ge i'$ we have $x_j\precsim x_i$. By the choice of i', for every $j\le i'>0$ we have then $x_j\prec x_{j-1}$, implying that for every $i<j\le i'$ we have $x_j\prec x_i$.

2. Let $i'\in\mathbb{N}$ be the smallest number such that $x_{i'}\succsim x_{i'+1}$. Then we have $c\ge M(i')/(\omega+\rho)$. Under (M2) $M(i)$ is decreasing in i, thus every successive treatment strategy with more than i' rounds is no better than the one preceding it, hence for every $i>j\ge i'$ we have $x_j\succsim x_i$. By the choice of i', for every $j\le i'>0$ we have then $x_j\succ x_{j-1}$, implying that for every $i<j\le i'$ we have $x_j\succ x_i$.

Proposition 3.1

The value is the sum of five values: (1) the payoff received in state $2^{(i)}$ while waiting for the next round of therapy. We calculate the positive part of the payoff (i.e, without toxicity). Take $\xi\sim\mathrm{Exp}(\omega_i)$, then

$$\begin{aligned}\mathbb{E}_\xi\int_0^{\min\{\xi,\hat t\}}\mathrm{e}^{-\rho t}\mathrm{d}t&=\int_0^{\hat t}\omega_i\mathrm{e}^{-\omega_i\xi}\int_0^{\xi}\mathrm{e}^{-\rho t}\mathrm{d}t\mathrm{d}\xi+\int_{\hat t}^{\infty}\omega_i\mathrm{e}^{-\omega_i\xi}\int_0^{\hat t}\mathrm{e}^{-\rho t}\mathrm{d}t\mathrm{d}\xi\\&=\frac{1}{\rho}\left(1-\mathrm{e}^{-\omega_i\hat t}+\frac{\omega_i}{\omega_i+\rho}\left(\mathrm{e}^{-(\omega_i+\rho)\hat t}-1\right)+\mathrm{e}^{-\omega_i\hat t}-\mathrm{e}^{-(\omega_i+\rho)\hat t}\right)\\&=\frac{1-\mathrm{e}^{-(\omega_i+\rho)\hat t}}{\omega_i+\rho}.\end{aligned}$$

With very similar calculations we may get the negative (toxicity) part of this component:

$$\mathbb{E}_\xi\int_0^{\min\{\xi,\hat t\}}z_i\mathrm{e}^{-(\rho+\zeta)t}\mathrm{d}t=\frac{z_i\left(1-\mathrm{e}^{-(\omega_i+\rho+\zeta)\hat t}\right)}{\omega_i+\rho+\zeta}.$$

(2), the payoff received in state $2^{(i)}$ after taking therapy but before transitioning to any of the states 0, $1^{(i+1)}$, $2^{(i+1)}$, or 3 as a result. Again, just taking the positive component, with $\xi\sim\mathrm{Exp}(\alpha_i)$ this is

$$\mathbb{E}_\xi\int_{\hat t}^{\xi+\hat t}\mathrm{e}^{-\rho t}\mathrm{d}t=\mathrm{e}^{-\rho\hat t}\int_0^{\infty}\alpha_i\mathrm{e}^{-\alpha_i\xi}\int_0^{\xi}\mathrm{e}^{-\rho t}\mathrm{d}t\mathrm{d}\xi=\mathrm{e}^{-\rho\hat t}\frac{1}{\alpha_i+\rho}.$$

For the toxicity component that the patient started with, we get

$$\mathbb{E}_\xi \int_{\hat{t}}^{\xi+\hat{t}} z_i \mathrm{e}^{-(\rho+\zeta)t} \mathrm{d}t = \mathrm{e}^{-(\rho+\zeta)\hat{t}} \frac{z_i}{\alpha_i + \rho + \zeta}.$$

Adding the toxicity caused by therapy $\hat{z}$ at time time $\hat{t}$ we get

$$\mathbb{E}_\xi \int_{\hat{t}}^{\xi+\hat{t}} \hat{z} \mathrm{e}^{-\rho t} \mathrm{e}^{-\zeta(t-\hat{t})} \mathrm{d}t = \hat{z} \mathrm{e}^{-\rho\hat{t}} \mathbb{E}_\xi \int_0^{\xi} \mathrm{e}^{-(\rho+\zeta)t} \mathrm{d}t = \mathrm{e}^{-\rho\hat{t}} \frac{\hat{z}}{\alpha_i + \rho + \zeta}.$$

Adding these three and multiplying with the probability of the patient reaching the time to take therapy, $\mathrm{e}^{-\omega_i \hat{t}}$ we get

$$\mathrm{e}^{-(\omega_i+\rho)\hat{t}} \left(\frac{1}{\alpha_i + \rho} - \frac{z_i \mathrm{e}^{-\zeta\hat{t}} + \hat{z}}{\alpha_i + \rho + \zeta} \right).$$

(3), the payoff received upon a transition to state 0. Again, with $\xi \sim \mathrm{E}(\alpha_i)$ this is (positive and negative parts together):

$$\mathbb{E}_\xi \int_{\xi+\hat{t}}^{\infty} \mathrm{e}^{-\rho t} - z_i \mathrm{e}^{-(\rho+\zeta)t} - \hat{z} \mathrm{e}^{-\rho t - \zeta(t-\hat{t})} \mathrm{d}t = \alpha_i \mathrm{e}^{-\rho\hat{t}} \left(\frac{1}{\rho(\alpha_i + \rho)} - \frac{z_i \mathrm{e}^{-\zeta\hat{t}} + \hat{z}}{(\rho+\zeta)(\alpha_i + \rho + \zeta)} \right).$$

Multiplying with the probability reaching the time to administer round i, $\mathrm{e}^{-\omega_i\hat{t}}$, and by the probability of transitioning to state 0 given that the patient receives round i, λ_i/α_i, we get

$$\lambda_i \mathrm{e}^{-(\omega_i+\rho)\hat{t}} \left(\frac{1}{\rho(\alpha_i + \rho)} - \frac{z_i \mathrm{e}^{-\zeta\hat{t}} + \hat{z}}{(\rho+\zeta)(\alpha_i + \rho + \zeta)} \right).$$

(4), the payoff received upon a transition to state $2^{(i+1)}$. This amounts to the expected present value of $V^{i+1}(x, z(z_i, \xi'))$ with delay ξ' where $\xi' = \xi + \hat{t}$ for $\xi \sim \mathrm{Exp}(\alpha_i)$. This equals

$$\mathbb{E}_{\xi'} \left(\mathrm{e}^{-\rho\xi'} V^{i+1}(x, z(z_i, \xi')) \right) = \mathrm{e}^{-\rho\hat{t}} \mathbb{E}_\xi \left(\mathrm{e}^{-\rho\xi} V^{i+1}(x, z(z_i, \xi + \hat{t})) \right).$$

Multiplying by the probability of reaching the time to administer round i, and by the probability of transitioning directly to state $2^{(i+1)}$ given that the patient receives round i, γ_i/α_i and substituting in $z(z_i, \xi + \hat{t}) = z_i \mathrm{e}^{-\zeta(\xi+\hat{t})} + \hat{z}$ we get

$$\frac{\gamma_i}{\alpha_i} \mathrm{e}^{-(\omega_i+\rho)\hat{t}} \int \mathrm{e}^{-\rho\xi} V^{i+1}(x, z_i \mathrm{e}^{-\zeta(\xi+\hat{t})} + \hat{z}) f(\xi) \mathrm{d}\xi.$$

(5), the payoff received upon a transition to state $1^{(i+1)}$ followed by a transition to state $2^{(i+1)}$. With $\xi_1 \sim \mathrm{Exp}(\alpha_i)$ and $\xi_2 \sim \mathrm{Exp}(\delta_{i+1})$, the former amounts to

$$\begin{aligned} &\mathbb{E}_{\xi_1,\xi_2} \int_{\xi_1+\hat{t}}^{\xi_1+\xi_2+x_i(z_i)} \mathrm{e}^{-\rho t} - z_i \mathrm{e}^{-(\rho+\zeta)t} - \hat{z} \mathrm{e}^{-\rho t - \zeta(t-\hat{t})} \mathrm{d}t \\ &\quad = \alpha_i \mathrm{e}^{-\rho\hat{t}} \left(\frac{1}{(\alpha_i + \rho)(\delta_{i+1} + \rho)} - \frac{z_i \mathrm{e}^{-\zeta\hat{t}} + \hat{z}}{(\alpha_i + \rho + \zeta)(\delta_{i+1} + \rho + \zeta)} \right). \end{aligned}$$

Multiplying by the probability of reaching the time to administer round i, and by the probability of transitioning to state $1^{(i+1)}$ from $2^{(i)}$, β_i/α_i, we get

$$\beta_i \mathrm{e}^{-(\omega_i+\rho)\hat{t}} \left(\frac{1}{(\alpha_i+\rho)(\delta_{i+1}+\rho)} - \frac{z_i \mathrm{e}^{-\zeta\hat{t}} + \hat{z}}{(\alpha_i+\rho+\zeta)(\delta_{i+1}+\rho+\zeta)} \right).$$

Finally, upon reaching state $2^{(i+1)}$ from $1^{(i+1)}$ the patient receives the present expected value of $V^{i+1}(x, z(z_i, \xi'))$ with a delay of ξ' where $\xi' = \xi_1 + \xi_2 + \hat{t}$. Substituting $\xi = \xi_1 + \xi_2$ we get

$$\mathbb{E}_{\xi'} \left(\mathrm{e}^{-\rho\xi'} V^{i+1}(x, z(z_i, \xi')) \right) = \mathrm{e}^{-\rho\hat{t}} \mathbb{E}_{\xi} \left(\mathrm{e}^{-\rho\xi} V^{i+1}(x, z(z_i, \xi + \hat{t})) \right).$$

Multiplying by the probability of reaching the time to administer round i, and by the probability of transitioning directly to state $1^{(i+1)}$ (from which reaching state $2^{(i+1)}$ is certain) given that the patient receives round i, β_i/α_i and substituting in $z(z_i, \xi + \hat{t}) = z_i \mathrm{e}^{-\zeta(\xi+\hat{t})} + \hat{z}$ we get

$$\frac{\beta}{\alpha_i} \mathrm{e}^{-(\omega_i+\rho)\hat{t}} \int \mathrm{e}^{-\rho\xi} V^{i+1}(x, z_i \mathrm{e}^{-\zeta(\xi+\hat{t})} + \hat{z}) g(\xi) \mathrm{d}\xi,$$

as $g(\cdot)$ is the density function of $\xi_1 + \xi_2$ by definition.

Summing up components (1) through (5) and adding the cost of one round of therapy, c with delay $\hat{t}$ multiplied by the probability of paying it gives the formula stated by the proposition.

Lemma 3.2

1. (10) is obtained from (9) by setting $\hat{t} = \infty$.
2. To calculate positive component of the payoff (without toxicity and costs), we substitute $\hat{t} = \hat{z} = z_i = c = 0$ into (9) to obtain

$$\begin{aligned} V^i(x,0) = \frac{1}{\alpha_i+\rho} + \frac{\lambda_i}{\rho(\alpha_i+\rho)} + \frac{\gamma_i}{\alpha_i} \int V^{i+1}(x,0) \mathrm{e}^{-\rho\xi} f(\xi) \mathrm{d}\xi \\ + \frac{\beta_i}{(\alpha_i+\rho)(\delta_{i+1}+\rho)} + \frac{\beta_i}{\alpha_i} \int V^{i+1}(x,0) \mathrm{e}^{-\rho\xi} g(\xi) \mathrm{d}\xi. \end{aligned}$$

By point 1, we may substitute $V^{i+1}(x,0) = B_i(\rho)$. Evaluating the integrals gives

$$\begin{aligned} &= \frac{1}{\alpha_i+\rho} + \frac{\lambda_i}{\rho(\alpha_i+\rho)} + \frac{\gamma_i}{\omega_i+\rho} \cdot \frac{1}{\alpha_i+\rho} + \frac{\beta_i}{(\alpha_i+\rho)(\delta_{i+1}+\rho)} + \frac{\beta_i}{\alpha_i+\rho} \cdot \frac{\delta_{i+1}}{\delta_{i+1}+\rho} \cdot \frac{1}{\omega_i+\rho} \\ &= \frac{1}{\alpha_i+\rho} \left(1 + \frac{\lambda_i}{\rho} + \frac{\gamma_i}{\omega_i+\rho} + \beta_i \left(\frac{1}{\delta_{i+1}+\rho} + \frac{\delta_{i+1}}{(\delta_{i+1}+\rho)(\omega_i+\rho)} \right) \right) = A_i(\rho). \end{aligned}$$

By similar calculations the payoffs from toxicity equal $(z_i + \hat{z})A_i(\rho + \zeta)$, while the cost is a lump-sum $-c$. Adding these together gives (11).

3. Calculating the positive components amounts to substituting $\hat{z} = z_i = c = 0$ into (9). This yields

$$V^i(x,0) = B_i(\rho)(1 - \mathrm{e}^{-(\omega_i+\rho)\hat{t}}) + \mathrm{e}^{-(\omega_i+\rho)\hat{t}} A_i(\rho)$$

where the second component follows from the calculations of the positive component of 2. The toxicity can be deduced as

$$-z_i B_i(\rho+\zeta)(1-\mathrm{e}^{-(\omega_i+\rho+\zeta)\hat{t}}) - \mathrm{e}^{-(\omega_i+\rho)\hat{t}}(z_i\mathrm{e}^{-\zeta\hat{t}}+\hat{z})A_i(\rho+\zeta).$$

Adding these together with the lump-sum cost $-c$, factoring in the delay and the probability of paying the cost leads to (12) as stated.

Proposition 3.3

We take a treatment strategy $x \in X_{i+1}$ and evaluate it in state $2^{(i)}$ given toxicity level z_i. To find the optimal $x(i,z_i)=\hat{t}$ we differentiate $V^{i+1}(x,z_i)$ (deduced from Lemma 3.2) with respect to $\hat{t}$ to give

$$\frac{\partial V^i(x,z_i)}{\partial \hat{t}} = \mathrm{e}^{-(\omega_i+\rho)\hat{t}} - z_i\mathrm{e}^{-(\omega_i+\rho+\zeta)\hat{t}} + \frac{\mathrm{e}^{-(\omega_i+\rho)\hat{t}}}{B_i(\rho)}\left(\hat{z}A_i(\rho+\zeta) - A_i(\rho) + c\right) + \frac{\mathrm{e}^{-(\omega_i+\rho+\zeta)\hat{t}}}{B_i(\rho+\zeta)}A_i(\rho+\zeta).$$

Multiplying by $\mathrm{e}^{(\omega_i+\rho+\zeta)\hat{t}}$ and rearranging, the sign of the derivative is the same as that of

$$\mathrm{e}^{\zeta\hat{t}}\overbrace{\left(1-\frac{A_i(\rho)}{B_i(\rho)}+\frac{\hat{z}A_i(\rho+\zeta)+c}{B_i(\rho)}\right)}^{d_1} + z_i\overbrace{\left(\frac{A_i(\rho+\zeta)}{B_i(\rho+\zeta)}-1\right)}^{-d_2} = d_1\mathrm{e}^{\zeta\hat{t}} - d_2 z_i.$$

There are four cases: 1. If d_1 and d_2 are both negative, then the derivative equals zero if

$$\hat{t} = \frac{1}{\zeta}\ln\left(z_i\frac{d_2}{d_1}\right),$$

provided that $z_i > d_1/d_2$. If so, then $\partial(V^i(x,z_i))^2/\partial\hat{t}^2$ is negative due to d_1 being negative, hence $\hat{t}$ is indeed a maximizer, and $z_i\mathrm{e}^{-\zeta\hat{t}} = d_1/d_2 = \overline{z}$, thus the patient waits until toxicity falls to $\overline{z}$. If $z_i < d_1/d_2$, then the first derivative is always negative, hence taking the next round immediately is optimal.

2. If $d_1 > 0$ and $d_2 < 0$, then the first derivative is positive for all $\hat{t}$, hence $\hat{t}=\infty$ is optimal.

3. If $d_1 < 0$ and $d_2 > 0$, then the first derivative is negative for all $\hat{t}$, hence $\hat{t}=0$ is optimal.

4. If d_1 and d_2 are both positive, then if $z_i < \overline{z}$, then the first derivative is positive for all $\hat{t}$, meaning that $\hat{t}=\infty$ is optimal. If $z_i > \overline{z}$, then the first derivative starts negative at $\hat{t}=0$, then turns positive and remains positive as $\hat{t}$ approaches infinity, meaning that either $\hat{t}=0$ or $\hat{t}=\infty$ is optimal. Comparing the payoffs, we get that $\hat{t}=0$ is best if and only if

$$z_i > \frac{B_i(\rho)-A_i(\rho)+\hat{z}A_i(\rho+\zeta)+c}{B_i(\rho+\zeta)-A_i(\rho+\zeta)} = z',$$

which is a stronger condition than $z_i > \overline{z}$.

Approximation method of Example 3.5

All transition parameters with the exception of the cure rate, λ_i, are independent if i. We assume a maximum number of treatments, N, that is, we set $\hat{t}_N = \infty$.

$$\tilde{V}^i(x,z_i) = \sum_{k=i}^{N}\left(b(\rho,k)-b(\rho+\zeta,k)\tilde{Z}_k\right)\mathrm{e}^{-(\omega+\rho)\tilde{T}_k} + \sum_{k=i}^{N-1}\left(a(\rho,k)-a(\rho+\zeta,k)\tilde{Z}_{k+1}\right)\mathrm{e}^{-(\omega+\rho)\tilde{T}_{k+1}}. \tag{14}$$

The components in (14) are as follows: We denote by $\hat{t}_k$ the time of delay before treatment round k with $\hat{t}_N = \infty$. The series T_k denotes the times at which the patient's toxicity increases as a result of the kth round of treatment, which takes place time $\hat{t}_k$ after the patient enters $2^{(k)}$. T_i is taken to be 0, while for $k > i$ we have

$$T_k = \sum_{j=i}^{k-1} \phi_k + \sum_{j=i}^{k} \hat{t}_j,$$

with ϕ_k being the random variable denoting the length of the kth round of therapy from its initiation (i.e. when toxicity increases) to its termination, conditional on the fact that the patient proceeds to state $2^{(k+1)}$.

To get an approximation, we replace T_k in (14) by its expected value, $\tilde{T}_k$, leading to an unbiased estimate of it. Given the patient's strategy, the waiting times $\hat{t}_j$ are fixed, while the expected value of ϕ_k is given by

$$\frac{1}{\lambda_k + \beta + \gamma + \mu} + \frac{\beta}{\delta(\beta + \gamma)},$$

of which the first component is the expected time spent in state $2^{(k)}$ while waiting for the kth round to take effect and the second is the expected time spent in state $1^{(k+1)}$, waiting for progression to state $2^{(k+1)}$, leading to $T_{i+1} = \hat{t}_i$

$$\tilde{T}_k = \sum_{j=i}^{k-1} \left(\frac{1}{\lambda_j + \beta + \gamma + \mu} + \frac{\beta}{\delta(\beta + \gamma)} \right) + \sum_{j=i}^{k} \hat{t}_j.$$

The estimate $\tilde{Z}_k$ denotes the approximation of the patient's toxicity at the time of receiving the kth therapy, i.e. at time T_k. For simplicity and computational ease, we approximate the patient's toxicity level at the time of entering state $2^{(k)}$ by substituting the expected time into the toxicity equation (6), giving a slightly biased estimate of the patient's toxicity:[8]

$$\tilde{Z}_k = z(z_i, \tilde{T}_k).$$

The two major components in (14) are

$$a(\rho, k) = \left(1 + \frac{\lambda_k}{\rho} + \frac{\beta}{\delta + \rho}\right) \left(\frac{\gamma^k}{\prod_{j=1}^{k}(\alpha_j + \rho)}\right) \left(\frac{\beta}{\gamma}\frac{\delta}{\delta + \rho} + 1\right)^k \tag{15}$$

and

$$b(\rho, k) = \frac{1}{\omega + \rho}\left(1 - e^{-(\omega+\rho)\hat{t}_{k+1}}\right) \left(\frac{\gamma^k}{\prod_{j=1}^{k-1}(\alpha_j + \rho)}\right) \left(\frac{\beta}{\gamma}\frac{\delta}{\delta + \rho} + 1\right)^k. \tag{16}$$

To get a visual intuition in deriving (14), from Figure 1, imagine that we fix the maximum number of treatments at N, reducing the model to a finite series of states. We descend N layers in the figure, then calculate all the possibilities to arrive at either state 0 or state 3 after at most

[8] In Example 3.5's parametrization, the bias in $\tilde{Z}_2$ is around 0.005, amounting to 1% of $\hat{z}$ with the estimate being lower, hence the second waiting time is slightly underestimated; the first waiting time's toxicity is unaffected by the bias, while all subsequent rounds have barely measurable payoff-effects.

N treatments by simply counting the number of paths. Each new layer can be reached one of two ways, either a direct transition from state $2^{(i)}$ to $2^{(i+1)}$ with rate γ, or an indirect one from $2^{(i)}$ to $1^{(i+1)}$ at rate β, then from $1^{(i+1)}$ to $2^{(i+1)}$ at rate δ.

The approximations of Table 8 are therefore results of numerically maximizing (in Wolfram Mathematica) equations of the form (14), subject to $\hat{t}_k \geq 0$, and entering $\lambda_k = \lambda^{k+1}$ into (15).

References

Andersen, P.K., Hansen, L.S. and Keiding, N., 1991. Assessing the influence of reversible disease indicators on survival. Statistics in Medicine, 10: 1061-1067.

Axelrod, R., and Axelrod, R.M., 1984. The evolution of cooperation (Vol. 5145). Basic Books (AZ).

Beeler, W.H., Bellile, E.L., Casper, K.A., Jaworski, E., Burger, N.J., Malloy, K.M., Spector, M.E., Shuman, A.G., Rosko, A., Stucken, C.L., and Chinn, S.B., 2020. Patient-reported financial toxicity and adverse medical consequences in head and neck cancer. Oral Oncology, 101: 104521.

Bellman, R., 1957. A Markovian decision process. Journal of Mathematics and Mechanics, 679-684.

Billingham, L.J. and Abrams, K.R., 2002. Simultaneous analysis of quality of life and survival data. Statistical Methods in Medical Research, 11: 25-48.

Bjelic-Radisic, V., Cardoso, F., Cameron, D., Brain, E., Kuljanic, K., da Costa, R.A., Conroy, T., Inwald, E.C., Serpentini, S., Pinto, M., and Weis, J., 2020. An international update of the EORTC questionnaire for assessing quality of life in breast cancer patients: EORTC QLQ-BR45. Annals of Oncology, 31: 283-288.

Blackwell, D., 1962. Discrete dynamic programming. The Annals of Mathematical Statistics, 33: 719-726.

Blackwell, D., 1965. Discounted dynamic programming. The Annals of Mathematical Statistics, 36: 226-235.

Chaikh, A., Docquière, N., Bondiau, P.Y., and Balosso, J., 2016. Impact of dose calculation models on radiotherapy outcomes and quality adjusted life years for lung cancer treatment: do we need to measure radiotherapy outcomes to tune the radiobiological parameters of a normal tissue complication probability model? Translational Lung Cancer Research, 5: 673.

Cleeland, C.S., Allen, J.D., Roberts, S.A., Brell, J.M., Giralt, S.A., Khakoo, A.Y., Kirch, R.A., Kwitkowski, V.E., Liao, Z., and Skillings, J., 2012. Reducing the toxicity of cancer therapy: recognizing needs, taking action. Nature Reviews Clinical Oncology, 9: 471-478.

Cleophas, T.J., and Van Ouwerkerk, B., 2007. The sense and nonsense of exponential models in clinical research. Clinical Research and Regulatory Affairs, 24: 25-37.

Cooper, N.J., Abrams, K.R., Sutton, A.J., Turner, D. and Lambert, P.C., 2003. A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer. Journal of the Royal Statistical Society: Series A (Statistics in Society), 166: 389-405.

Cooper, N.J., Sutton, A.J., Abrams, K.R., Turner, D. and Wailoo, A., 2004. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. Health Economics, 13: 203-226.

Delgado-Guay, M., Ferrer, J., Rieber, A.G., Rhondali, W., Tayjasanant, S., Ochoa, J., Cantu, H., Chisholm, G., Williams, J., Frisbee-Hume, S., and Bruera, E., 2015. Financial distress and its associations with physical and emotional symptoms and quality of life among advanced cancer patients. The Oncologist, 20: 1092.

Delisle, M., Singh, S., Howard, J., Panda, N., Weppler, A.M., and Wang, Y., 2020. Refusal of colorectal cancer surgery in the United States: Predictors and associated cancer-specific mortality in a Surveillance, Epidemiology, and End Results (SEER) cohort. Surgery Open Science, 2: 12-18.

Dias, L.M., Bezerra, M.R., Barra, W.F., and Rego, F., 2021. Refusal of medical treatment by older adults with cancer: a systematic review. Annals of Palliative Medicine, 10(4).

Duffy, S.W., Chen, H.H., Tabar, L. and Day, N.E., 1995. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. Statistics in Medicine, 14: 1531-1543.

Eftimie, R., Bramson, J.L., and Earn, D.J.D., 2011. Interactions between the immune system and cancer: a brief review of non-spatial mathematical models. Bulletin of Mathematical Biology, 73: 2-32.

Ell, K., Xie, B., Wells, A., Nedjat-Haiem, F., Lee, P.J., and Vourlekis, B., 2008. Economic stress among low-income women with cancer: effects on quality of life. Cancer: Interdisciplinary International Journal of the American Cancer Society, 112: 616-625.

Foryś U., and Mokwa-Borkowska, A., 2005. Solid tumour growth analysis of necrotic core formation. Mathematical and Computer Modelling, 42: 593-600.

Frenkel, M., 2013. Refusing treatment. The Oncologist, 18: 634.

Fudenberg, D., and Maskin, E., 1986. The folk theorem in repeated games with discounting or with incomplete information. Econometrica: Journal of the Econometric Society, 533-554.

Gajewski, T.F., Schreiber, H., and Fu, Y.X., 2013. Innate and adaptive immune cells in the tumor microenvironment. Nature Immunology, 14: 1014-1022.

Gatenby, R.A., Silva, A.S., Gillies, R.J., and Frieden, B.R., 2009. Adaptive therapy. Cancer Research, 69: 4894-4903.

Giani, C., Valerio, L., Bongiovanni, A., Durante, C., Grani, G., Ibrahim, T., Mariotti, S., Massa, M., Pani, F., Pellegriti, G., and Porcelli, T., 2021. Safety and quality-of-life data from an Italian expanded access program of lenvatinib for treatment of thyroid cancer. Thyroid, 31: 224-232.

Gilbar, O., 1991. The quality of life of cancer patients who refuse chemotherapy. Social Science & Medicine, 32: 1337-1340.

Glasziou, P.P., Cole, B.F., Gelber, R.D., Hilden, J. and Simes, R.J., 1998. Quality adjusted survival analysis with repeated quality of life measures. Statistics in Medicine, 17: 1215-1229.

Gluzman, M., Scott, J.G., and Vladimirsky, A., 2020. Optimizing adaptive cancer therapy: dynamic programming and evolutionary game theory. Proceedings of the Royal Society B, 287: 20192454.

Hansen, A.R., Ala-Leppilampi, K., McKillop, C., Siu, L.L., Bedard, P.L., Abdul Razak, A.R., Spreafico, A., Sridhar, S.S., Leighl, N., Butler, M.O., and Hogg, D., 2020. Development of the Functional Assessment of Cancer Therapy-Immune Checkpoint Modulator (FACT-ICM): A toxicity subscale to measure quality of life in patients with cancer who are treated with ICMs. Cancer, 126: 1550-1558.

He, C., Huang, X., Zhang, Y., Cai, Z., Lin, X., and Li, S., 2020. Comparison of survival between irreversible electroporation followed by chemotherapy and chemotherapy alone for locally advanced pancreatic cancer. Frontiers in Oncology, 10: 6.

Huijer, M., and van Leeuwen, E., 2000. Personal values and cancer treatment refusal. Journal of Medical Ethics, 26: 358-362.

Kay, R., 1986. A Markov model for analysing cancer markers and disease states in survival studies. Biometrics, 855-865.

Le Lay, K., Myon, E., Hill, S., Riou-Franca, L., Scott, D., Sidhu, M., Dunlop, D. and Launois, R., 2007. Comparative cost-minimisation of oral and intravenous chemotherapy for first-line treatment of non-small cell lung cancer in the UK NHS system. The European Journal of Health Economics, 8: 145-151.

Llorca, J., and Delgado-Rodríguez, M., 2001. Competing risks analysis using Markov chains: impact of cerebrovascular and ischaemic heart disease in cancer mortality. International Journal of Epidemiology, 30: 99-101.

McNulty, J., and Khera, N., 2015. Financial hardship – an unwanted consequence of cancer treatment. Current Hematologic Malignancy Reports, 10: 205-212.

Mokhatri-Hesari, P., and Montazeri, A., 2020. Health-related quality of life in breast cancer patients: review of reviews from 2008 to 2018. Health and Quality of Life Outcomes, 18: 1-25.

Orlando, P.A., Gatenby, R.A. and Brown, J.S., 2012. Cancer treatment as a game: integrating evolutionary game theory into the optimal control of chemotherapy. Physical Biology, 9: 065007.

Ortega-Gutiérrez, R.I., Montes-de-Oca, R. and Lemus-Rodríguez, E., 2016. Uniqueness of optimal policies as a generic property of discounted Markov decision processes: Ekeland's variational principle approach. Kybernetika, 52: 66-75.

Patnaik, A., Doyle, C., and Oza, A.M., 1998. Palliative therapy in advanced ovarian cancer: balancing patient expectations, quality of life and cost. Anti-cancer Drugs, 9: 869-878.

Shumay, D.M., Maskarinec, G., Kakai, H. and Gotay, C.C., 2001. Why some cancer patients choose complementary and alternative medicine instead of conventional treatment. The Journal of Family Practice, 50: 1067-1067.

Simes, R.J., 1985. Treatment selection for cancer patients: application of statistical decision theory to the treatment of advanced ovarian cancer. Journal of Chronic Diseases, 38: 171-186.

Smith, G.L., Shih, Y.T., and Frank, S.J., 2021. Financial toxicity in head and neck cancer patients treated with proton therapy. International Journal of Particle Therapy, 8: 366-373.

Staňková, K., Brown, J.S., Dalton, W.S. and Gatenby, R.A., 2019. Optimizing cancer treatment using game theory: A review. JAMA Oncology, 5: 96-103.

Suh, W.N., Kong, K.A., Han, Y., Kim, S.J., Lee, S.H., Ryu, Y.J., Lee, J.H., Shim, S.S., Kim, Y., and Chang, J.H., 2017. Risk factors associated with treatment refusal in lung cancer. Thoracic Cancer, 8: 443-450.

Sun, J.M., Zhou, W., Choi, Y.L., Choi, S.J., Kim, S.E., Wang, Z., Dolled-Filhart, M., Emancipator, K., Wu, D., Weiner, R., and Frisman, D., 2016. Prognostic significance of PD-L1 in patients with non–small cell lung cancer: a large cohort study of surgically resected cases. Journal of Thoracic Oncology, 11: 1003-1011.

Terpos, E., Mikhael, J., Hajek, R., Chari, A., Zweegman, S., Lee, H.C., Mateos, M.V., Larocca, A., Ramasamy, K., Kaiser, M., and Cook, G., 2021. Management of patients with multiple myeloma beyond the clinical-trial setting: understanding the balance between efficacy, safety and tolerability, and quality of life. Blood Cancer Journal, 11: 1-13.

Traeger, L., Penedo, F.J., Gonzalez, J.S., Dahn, J.R., Lechner, S.C., Schneiderman, N., and Antoni, M.H., 2009. Illness perceptions and emotional well-being in men treated for localized prostate cancer. Journal of Psychosomatic Research, 67: 389-397.

Urru, S.A.M., Gallus, S., Bosetti, C., Moi, T., Medda, R., Sollai, E., Murgia, A., Sanges, F., Pira, G., Manca, A., and Palmas, D., 2018. Clinical and pathological factors influencing survival in a large cohort of triple-negative breast cancer patients. BMC Cancer, 18: 1-11.

Vaughn, J., Ammerman, C., and Stein, J., 2020. Delay discounting as a predictor of adjuvant endocrine therapy adherence among breast cancer survivors.