

Irreversible Treatment Decisions under Consideration of the Research and Development Pipeline for New Therapies

November 2, 2009

Steven M. Shechter, PhD (corresponding author)
Operations and Logistics Division
Sauder School of Business
University of British Columbia
2053 Main Mall
Vancouver, BC V6T 1Z2
Canada
Phone: 604-822-8340
Fax: 604-822-9574
steven.shechter@sauder.ubc.ca
Member of IIE

Oguzhan Alagoz, PhD
Department of Industrial and Systems Engineering
University of Wisconsin
3222 Mechanical Engineering Building
1513 University Way
Madison, Wisconsin, 53706
USA
alagoz@engr.wisc.edu
member of IIE

Mark S. Roberts, MD
Section of Decision Sciences and Clinical Systems Modeling
Department of Medicine
Department of Industrial Engineering
University of Pittsburgh
200 Meyran Avenue, 2nd Floor
Pittsburgh, Pennsylvania 15213
USA
robertsm@msx.upmc.edu

Abstract

This paper addresses a topic not considered in previous models of patient treatment: the possible downstream availability of improved treatment options coming out of the medical research and development (R&D) pipeline. We provide clinical examples in which a patient may prefer to wait and take the chance that an improved therapy comes to market rather than choose an irreversible treatment option that has serious quality of life ramifications and would render future treatment discoveries meaningless for that patient. We then develop a Markov Decision Process model of the optimal time to initiate treatment, which incorporates uncertainty around the development of new therapies and their effects. After deriving structural properties for the model, we provide a numerical example that demonstrates how models that do not have any foresight of the R&D pipeline may result in optimal policies that differ from models that have such foresight, implying erroneous decisions in the former models. Our example quantifies the effects of such errors.

1 Introduction

Medical decision making is the process of using decision analytic techniques to make sound health care decisions for both patients (e.g., whether to initiate some treatment or wait) and society (e.g., how to allocate a budget to different health care programs). Patient-level models are intended to help clinicians and patients determine the best course of treatment, given the probabilistic progression of the patient's health and the values placed on different outcomes. Typically, disease progression and treatment outcomes are estimated based on treatment options currently available, and it is implicitly assumed that the set of available treatment options remains static. However, research and development (R&D) in medicine often leads to new significant medical discoveries in relatively short periods of time. Therefore, the present paper seeks to address a topic that is lacking in previous models of patient treatment: consideration of how the possible downstream availability of improved treatment options

coming out of the R&D pipeline may affect treatment decisions at the present time.

R&D in medicine has dramatically changed the landscape of therapeutic options available to patients over the years, and many of these options present themselves to patients during the course of their disease. In 2005, \$55 billion was spent on R&D of new pharmaceuticals and biotechnology [18]. As a result of these enormous investments, there are approximately 650 new medicines in development for treating cancer, 150 for heart disease and stroke, and ongoing development to treat many other diseases [18]. However, there is considerable uncertainty in which drugs will come to market: approximately 70% of drugs make it from Phase I to Phase II trials, 33% of those make it to Phase III trials, and 25-30% of those successfully complete Phase III trials [16]. The average duration of Phase I, II, and III clinical trials has been estimated to be 22, 26, and 31 months, respectively [10].

How should a physician recommend which treatment a patient should take and when to take it, given the possibility that improved treatments in the R&D pipeline may become available in the near future? Specifically, this paper explores irreversible treatment decisions, in the sense that opting to take some course of treatment now precludes the possibility of taking some other new treatments later (or at least renders such options irrelevant for the patient). For example, patients with slow-growing prostate cancer may choose to have a radical prostatectomy (removal of the entire prostate) and risk the negative side effects of this surgery (such as erectile dysfunction and urinary incontinence). If a patient chooses this option, then any new treatment discoveries for prostate cancer are useless to this patient. On the other hand, another reasonable option for the patient is to undergo a period of “watchful waiting” [5], in which the cancer progression is monitored. During this time, improved therapies with fewer side effects may be discovered, and the patient may pursue these treatments instead of the current form of radical prostatectomy. For example, as demonstrated with the 2007 Edelman prize from the Institute for Operations Research and the Management Sciences, significant advances have been made with respect to treating prostate cancer with brachytherapy [15]. Moreover, in the past 15 years there have been

significant technological advances even in the way radical prostatectomy itself is performed, leading to improved patient outcomes [20]. Another example can be found in whether or not a woman at high risk for breast cancer should choose to have a prophylactic mastectomy. Women with the breast-cancer susceptibility genes BRCA1 or BRCA2 have a 55 to 85% risk of invasive breast cancer by the time they reach 70 years of age, with increased risk starting at the age of 25 [17]. Many women choose to have the drastic surgery, though perhaps they might wait longer if new treatments or improved diagnostic testing have a reasonable chance of becoming available in the near future.

The treat vs. wait decision has recently been modeled using a real options approach, which borrows ideas from pricing financial options [11]. Our Markov decision process (MDP) approach differs in two important respects: 1. we explicitly model patient disease progression and values of treatment as a function of the patient’s health state, and 2. our model considers the possibility of new treatments becoming available for patient use. The question we pose also has similarities to the question of when to replace equipment in an environment of uncertain technological change (for a review of this literature, see [21]). The paper from this literature perhaps most similar to ours takes an MDP approach to the optimal time to replace a deteriorating machine with an existing technology, while considering the possibility of an improved technology becoming available [13]. Assumptions of their model that differ from ours are that: 1. replacement with the current technology takes the system to the best state, 2. there is no limit to the number of times the system can replace the old technology “like with like”, and 3. the terminal reward for replacing a machine with the new technology is independent of the state of the replaced machine. In our patient context, patients can choose irreversible treatment one time, and the value of initiating any kind of treatment is dependent on the patient’s health state at the time of initiation.

The question we explore has some interesting connections to the work by Alagoz et al. [2, 4] on optimizing transplant decisions for a patient with end-stage liver disease. For example, Alagoz et al. [2] consider a patient with an available living-donor liver who is also

on a waiting list for the possibility of a cadaveric organ. If the patient takes the living-donor liver, it precludes the possibility of obtaining a better cadaveric liver in the future. This is analogous to our consideration of a patient who has some currently available irreversible therapy but might also wait for the possibility of an improved therapy to become available. The liver model takes an infinite horizon approach in which there is a possibility of a cadaveric liver offer at every decision period. We take a combined finite and infinite horizon modeling approach in which a new treatment option is under a clinical trial and therefore might only arrive at some specific time in the future, after the trial finishes (we assume below that the trial finishes N periods from now and that the new therapy becomes available with probability q ; we discuss the modeling details more in Section 2.2). We also consider the “naive” patient who is unaware of the possibility of a new therapy becoming available. The modeling approach we take for that patient is similar to the approach taken in Alagoz et al. [4], in which a patient does not join the waiting list for cadaveric organs and only has a living-donor liver available to accept. The difference is that the naive patient of our model may in fact end up taking the new therapy should it become available, whereas the patient of this liver model may be ineligible to receive a cadaveric organ (or so far down the waiting list as to make it virtually impossible to ever receive one).

We are unaware of any medical decision making models that explicitly consider the dynamics of new treatment development. We hypothesize that models that neglect the development of new therapies may result in different estimates of treatment outcomes, and more importantly, may result in incorrect treatment decisions when compared to models that do consider the real presence of medical R&D. Our goal in this paper is to develop a modeling framework that considers new therapy development, explore structural properties of this model, and explore differences between models that do and do not consider medical R&D.

2 A Markov Decision Process Approach

MDP models provide a natural modeling framework for optimizing treat vs. wait decisions because these models are well suited to determining optimal policies for decisions that are made periodically and under uncertainty [19]. Although traditionally applied to problems in inventory control and machine maintenance, MDPs have increasingly been applied to modeling medical treatments [23]. In particular, MDPs have recently been used to analyze treat vs. wait decisions for liver transplantation, dialysis, and HIV therapy [2, 3, 4, 14, 24]. Our model extends previous treat vs. wait MDP models by creating a framework for incorporating the possibility of new treatment development and exploring structural properties of these models.

We consider the situation in which a patient diagnosed with a degenerative disease may currently treat it with some treatment T_1 , which extends length of life but is irreversible and has significant implications for quality of life. We let $R_{T_1}(h)$ be the remaining expected quality-adjusted lifetime for the patient upon initiating T_1 when in health state h . At the same time, a new treatment T_2 is in a clinical trial, which will be completed N time periods from now (e.g., twelve months). There is a probability q that the treatment passes the trial and comes to market as a better treatment option than T_1 (that is, $R_{T_2}(h) > R_{T_1}(h)$ for all h), and probability $1 - q$ that it fails and the patient remains with the one treatment option of T_1 . By irreversible, we mean that taking T_1 precludes the patient from taking T_2 should it become available. One can imagine clinical situations where this would hold, such as mastectomy precluding the choice of lumpectomy (followed by radiation therapy), or radical prostatectomy precluding the possibility of brachytherapy.

2.1 Model Components

We describe our modeling framework more formally with the following components:

T_1 : the (irreversible) treatment always available to the patient.

T_2 : the better treatment which is under clinical trials and may become available in the future.

N : the period of time (in months) until which either T_2 passes the clinical trial and comes to market or fails the clinical trial.

q : the probability that T_2 passes the clinical trial.

$s = (h, m)$: the state of the patient, represented by the health state ($h \in \{0, 1, \dots, H\}$), and the number of therapies available ($m \in \{1, 2\}$). Patient death is indicated by $h = 0$, and we associate higher values of h with better health states. If $m = 1$, then only treatment T_1 is currently available, whereas if $m = 2$, then both T_1 and T_2 are currently available.

$a[(h, m)]$: the decision taken when the patient is in state (h, m) prior to initiating any therapy. If $m = 1$, then $a \in \{W, T_1\}$, where W indicates the patient will wait another period and T_1 indicates the patient will start taking therapy T_1 . If $m = 2$, then $a \in \{W, T_1, T_2\}$. Note that although T_1 is available when $m = 2$, we shall assume T_2 dominates T_1 , and hence it is not considered when T_2 is available.

$r(h)$: the expected quality-adjusted life time associated with waiting one period in health state h .

$R_{T_1}(h)$: the expected total remaining quality-adjusted life time, obtained when the patient initiates treatment T_1 from health state h .

$R_{T_2}(h)$: the expected total remaining quality-adjusted life time, obtained when the patient initiates treatment T_2 from health state h .

$p(h'|h)$: the probability that the patient's health state goes from h at time t to h' at time $t + 1$ when the patient has not initiated any treatment and waits another period. Patient death is represented by $p(0|h)$, and we assume the probability structure ensures that state 0 is reachable from every state. Also, state 0 is absorbing, so that $p(0|0) = 1$. If $a(h, m) \in \{T_1, T_2\}$, the patient obtains the appropriate terminal reward (according to which therapy is taken) and moves with certainty to state 0, an absorbing state of 0 reward. Details on how one might go about estimating transition probabilities and terminal rewards can be

found in models of end-stage liver disease [1, 4] and HIV [24].

$v_t[(h, m)]$: the value vector that gives the optimal expected remaining quality-adjusted lifetime, when the patient is in state (h, m) , has not yet initiated therapy, is aware of the R&D pipeline, and has t periods to go until the uncertainty regarding T_2 is resolved.

$w^*(h)$: the value vector that gives the perceived optimal expected remaining quality-adjusted lifetime, when the patient is in health state (h) , has not yet initiated therapy, and is unaware of the R&D pipeline.

Note that while intuitively, patient aging from one period to the next should lead to decreasing values of $R_{T_1}(h)$ as well as increasing mortality rates as expressed through $p(0|h)$, we are considering diseases for which waiting a long time is not realistic, and hence these values would not change appreciably prior to the patient starting some therapy. Moreover, we think it is plausible that patients and their physicians would integrate information regarding a clinical trial into a treatment decision for a progressive disease *only* when the time to trial completion is in the short term (i.e., the patient develops the disease some time during a Phase III trial). Therefore, our model is only appropriate for shorter time horizons, during which the effects of aging are secondary to both the dynamics of disease progression (which may very well be significant in the short term) and the differential effects of treatments. Note that the time stationarity assumption of these parameters has been made in treat vs. wait models of liver transplantation [2, 3, 4] and HIV [24]. Also, our model factors in patient age in that it drives the initial estimates of the reward and transition probability matrices.

We assume that the probability of trial success and estimates of the benefits of T_2 are fixed and do not change as the time until trial resolution draws closer. That is, we assume that no information gathering takes place by the patient during the clinical trial, which is appropriate if intermediary results are kept private. Initial estimates might be obtained by historical trial success rates as well as information that may be available from earlier phases of the clinical trial under consideration. In Section 4, we discuss how our model may be adapted to consider updates regarding the Phase III trial's evolution.

2.2 Solution Approach

First let us consider the perspective of the patient unaware of the R&D pipeline (sometimes referred to as the “naive” patient). The following optimality equations may be used to represent this model:

$$w(h) = \max \left\{ r(h) + \sum_{j=0}^{j=H} p(j|h)w(j), R_{T_1}(h) \right\} \quad \forall h. \quad (1)$$

In other words, the patient considers the quality of life without treatment ($r(h)$), the probability of transitioning to other health states over the next month ($p(j|h)$), the expected quality-adjusted lifetime associated with the treatment option T_1 ($R_{T_1}(h)$), and solves the resulting MDP (using, for example, standard MDP solution techniques such as the value or policy iteration algorithm [19]). The patient is unaware of the possibility of T_2 arriving in the near future, so this does not enter into the optimality equations. Also, because we suppose the patient cannot wait too long before initiating therapy, and because we assumed the model parameters do not vary over this time period, the optimality equations above are based on an infinite time-horizon model. Moreover, because we do not know how long a patient will survive, it is not clear what an appropriate horizon length would be a priori if we chose a finite horizon modeling approach.

Note that the resulting value vector does not necessarily indicate the patient’s actual expected remaining quality-adjusted lifetime; it just indicates what the patient *perceives* is the case. More importantly, the patient makes decisions based on this perspective. In reality, if the patient has not initiated T_1 and T_2 then becomes available, we assume the patient would then have knowledge of it and would choose that over T_1 . Before this time, however, this patient does not consider this possibility. Below, we will consider the *actual* value vector for the patient, which estimates the patient’s quality-adjusted lifetime under the reality that should the improved therapy come to market, the patient may then opt for it if he or she has not already taken T_1 .

While there are good arguments for discounting both quality-adjusted life years and costs when they are combined in a cost-effectiveness analysis [12], we have not incorporated a discount factor in the optimality equations of (1). This is because when one focuses solely on health outcomes, it is less intuitive to discount life years (the discounting of dollars, on the other hand, is never controversial). Moreover, there is no technical need for discounting in this setting, as the convergence of MDP solution algorithms is guaranteed by the fact that all patients eventually die [6].

Next let us consider the perspective of the patient aware of the R&D pipeline (sometimes referred to as the “knowledgeable” patient). In this case, we consider value functions indexed by the time left until the uncertainty regarding T_2 is resolved. We start from the time of trial resolution and work backwards: if T_2 passes the trial, then the patient solves the following:

$$v_0[(h, 2)] = \max \left\{ r(h) + \sum_{j=0}^{j=H} p(j|h)v_0[(j, 2)], R_{T_2}(h) \right\} \quad \forall h. \quad (2)$$

Note that we are making the assumption that if T_2 becomes available, then it is a better treatment option for the patient than T_1 (i.e., $R_{T_2}(h) \geq R_{T_1}(h)$ for all h). If T_2 fails the clinical trial, then the patient solves the following:

$$v_0[(h, 1)] = \max \left\{ r(h) + \sum_{j=0}^{j=H} p(j|h)v_0[(j, 1)], R_{T_1}(h) \right\} \quad \forall h. \quad (3)$$

Note that the equations of (3) are identical to those of (1) for the patient unaware of the possibility of T_2 . The reason we model this stage of the decision process as an infinite horizon MDP is that the time until the next major event (death) is random and can happen at any period. The fact that the absorbing state of death is reachable from all other states guarantees the finite solution of this problem. This infinite-horizon approach to medical decision making has also been applied to models of end-stage liver disease ([2, 3, 4]) and HIV therapy ([24]). Note that while there may be other clinical trials ongoing, we assume

that their estimated completion times are sufficiently far enough away so as not to affect treatment decisions at the conclusion of the trial under consideration.

Stepping back one month before the trial ends, the patient solves the following (at this time, only T_1 is available):

$$v_1[(h, 1)] = \max \left\{ \begin{array}{l} r(h) + (1 - q) \sum_{j=0}^{j=H} p(j|h)v_0[(j, 1)] + q \sum_{j=0}^{j=H} p(j|h)v_0[(j, 2)], \\ R_{T_1}(h) \end{array} \right\} \quad \forall h. \quad (4)$$

The above considers the probability q that T_2 becomes available.

Then, for $t = 2, \dots, N$, the patient solves the following:

$$v_t[(h, 1)] = \max \left\{ r(h) + \sum_{j=0}^{j=H} p(j|h)v_{t-1}[(j, 1)], R_{T_1}(h) \right\} \quad \forall h. \quad (5)$$

Note that the probability of T_2 passing the trial is only explicitly considered in the calculation of $v_1[(h, 1)]$, and it is then implicitly considered through the recursion of (5) (as indicated above, this setup assumes that there is no information gathering over time which may alter one's estimate of the probability of trial success). Also, the optimality equations of (4) and (5) represent finite horizon MDPs (solved by backwards induction), with terminal rewards obtained as the solution of infinite horizon MDPs (represented by (2) and (3) and solved with algorithms such as policy iteration).

Now that we have set up the solution approach for the knowledgeable patient, we can understand how to calculate the *actual* expected quality-adjusted lifetime for the naive patient rather than the perceived values from solving the optimality equations of (1). By solving those, the patient comes up with what is believed to be an optimal (stationary) treatment policy (denoted by π , which may not be unique). In following this policy, the patient may die before the clinical trial resolves; however, if the patient survives until that time and T_2 becomes available, we suppose the patient then becomes aware of this fact and may consider

taking the new therapy. Therefore, we let $w_{act,N}^\pi(h)$ be the (actual) expected quality-adjusted lifetime for the naive patient who acts according to a myopic treatment policy, π , starting from the time N periods before the clinical trial resolves. Note that because the possibility of an improved therapy becoming available cannot make a patient worse off, we have the following value function ordering: $w^*(h) \leq w_{act,N}^\pi(h) \leq v_N[(h, 1)]$. Also note that the expected value of information (VOI) to the naive patient is given by $v_N[(h, 1)] - w_{act,N}^\pi(h)$. In other words, this is the gain in expected quality-adjusted lifetime by becoming aware of the possibility of T_2 arriving in N periods. This notion of VOI is different than what is typically discussed in the medical decision making literature, which focuses on the value of eliminating uncertainty around parameter estimates (for a review of VOI in health risk management, see [25]). In our framework, there is an actual probability, q , of a new therapy becoming available, and the naive patient simply does not even consider the possibility of such an event.

2.3 Structural Properties

For the results that follow, we need only mild assumptions, which guarantee convergence of our infinite horizon MDP [6]: 1. rewards are bounded, and 2. with probability 1, every treatment policy reaches an absorbing state of zero reward. These are clearly satisfied in our patient treatment model, as the rewards are expressed in terms of patient lifetimes and the state of death is unavoidable.

The following indicates that for a patient in a given health state and aware of the possibility of T_2 becoming available, it is better that the trial is resolved sooner rather than later (note that the time index of v is the number of time periods before trial resolution). The intuition behind this is that the patient has less time to possibly die before having the chance at a better treatment. All proofs appear in the Appendix.

Proposition 1 *For each h and $t \geq 1$, $v_{t+1}[(h, 1)] \leq v_t[(h, 1)]$.*

The following says that the value function of the knowledgeable patient converges to

that of the naive patient. Note that in combination with Proposition 1, the convergence is monotonic. The practical value of this result is that it suggests that patients and physicians should put more effort into estimating q and R_{T_2} if a trial is closer to finishing than if it is further away.

Theorem 1 *As $N \rightarrow \infty$, $v_N[(h, 1)] \rightarrow w^*(h)$ for all h .*

A common structural property sought in MDP models is that of a control-limit policy. For example, prior to the trial resolution, a control-limit policy in health states would say that for some health state and worse, it is optimal to initiate T_1 and otherwise it is optimal to wait. These results have been presented in an earlier reference [4]. Instead, we present the following, which establishes a control limit policy in time.

Theorem 2 *If for some h' and t' it is optimal for the patient knowledgeable of the possibility of T_2 to take T_1 , then it is optimal to take T_1 in state h' for all $t > t'$.*

Similarly, because the naive patient's perceived value function is no more than that of the knowledgeable patient, we have the following corollary:

Corollary 1 *If it is optimal for the naive patient to wait in some health state h' , it is optimal for the knowledgeable patient to wait in health state h' for all $1 \leq t \leq N$.*

The following implies that if at any number of periods to go until trial resolution we find that the knowledgeable patient's optimal value function matches the naive patient's perceived value function in each health state, then we no longer need to proceed with the backwards induction algorithm; we will then know the optimal value and policy for longer time periods to go.

Theorem 3 *If for some t' , $v_{t'}[(h, 1)] = w^*(h)$ for all h , then*

1. $v_t[(h, 1)] = v_{t'}[(h, 1)]$ for $t' < t \leq N$, and
2. $a_t^*[(h, 1)] = a_{t'}^*[(h, 1)]$ for $t' < t \leq N$.

3 Numerical Example

Here we consider a numerical example based on an MDP model of the optimal time to initiate HIV therapy [24]. The model considers an HIV patient whose health state is categorized by four different ranges of CD4 count (the white blood cell attacked by the HIV virus): 0-49 ($h = 1$), 50-199 ($h = 2$), 200-349 ($h = 3$), and ≥ 350 ($h = 4$). Monthly decisions are considered, in which the patient could initiate therapy (according to the standard of care for which therapy to start with) or wait another period and then reassess what to do. We suppose that in twelve months, a clinical trial will be completed and the patient may then have the possibility of taking a new and improved treatment, T_2 . The objective is to maximize the patient’s expected total quality-adjusted life years. Note that while HIV therapy does not fit the notion of irreversibility we have discussed so far (i.e., the surgical examples), it is possible that initiating certain HIV therapies now can render other, never-before-taken therapies ineffective later. This is because when patients start one therapy, the virus may build up resistance to that therapy as well as others (this phenomenon is known as cross-resistance [9]). In our example, we assume total cross-resistance would develop to the new therapy if the patient initiates T_1 , making the new therapy completely ineffective for the patient.

Based on the HIV model in [24], we start with a baseline reward structure for $r(h)$ and $R_{T_1}(h)$ (in quality-adjusted life years) as given in Table 1. We also suppose the transition probabilities between health states are as given in Table 2. Note that in this example, the patient’s health cannot improve over time. Although this is a common assumption made in disease progression models, it is not required by any of our structural properties; we assumed it for this example to facilitate solving the Bellman’s equations of equations (2) and (3). We refer the reader to [24] for details on how one might go about generating transition probability matrices and rewards for these types of models.

First let us consider a naive patient who neglects the possibility of new therapy development and solves the model “as-is.” Applying the policy iteration MDP solution algorithm

Table 1: Rewards

Reward	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
$r(h)$	0	.0733	.0758	.0808	.0820
$R_{T_1}(h)$	0	3.93	7.43	11.73	16.86

Table 2: Transition probability matrix

h	0	1	2	3	4
0	1	0	0	0	0
1	.100	.900	.000	.000	.000
2	.034	.063	.903	.000	.000
3	.014	.013	.081	.892	.000
4	.001	.002	.003	.038	.956

[19], we obtain the optimal policy and value vector shown in Table 3. The intuition behind the policy of initiating T_1 from every health state is that the patient’s deteriorating health progression is such that the patient would be worse off waiting one period and then initiating T_1 from the next health state. This intuition is formalized into necessary and sufficient conditions for obtaining such a treatment policy in [24].

Table 3: Optimal policy and value vector

h	1	2	3	4
$a^*[(h, 1)]$	T_1	T_1	T_1	T_1
$w^*(h)$	3.93	7.43	11.73	16.86

Because the naive patient initiates T_1 for all health states, $w_{act,t}(h) = R_{T_1}(h)$ for all h and t . Therefore, for any health state we find in which it is optimal for the knowledgeable patient to wait t months before the trial is resolved, the difference between $v_t[(h, 1)]$ and $R_{T_1}(h)$ will represent the cost of not considering the real possibility that T_2 arrives to the market. In other words, this is the expected value of information to the naive patient.

We now examine how different combinations of q and R_{T_2} may cause a patient who is aware of the clinical trial to wait in some health states and time periods. We vary R_{T_2} by

considering R_{T_1} as a baseline set of values and apply fixed percentage increases over R_{T_1} . For example, a 5% increase would set $R_{T_2}(h) = 1.05R_{T_1}(h)$ for all h . Note that under every percentage increase, if T_2 becomes available, it is optimal to take it from every health state, just as it is optimal to take T_1 from every health state if T_2 fails the clinical trial (the formal proof of this is based on results found in [24]). Interestingly, under more general increases of R_{T_2} over R_{T_1} , it is possible that a patient whose only treatment option is T_1 would take it in every health state, whereas a patient with the option of taking the improved T_2 does not necessarily take it from every health state. A simple example demonstrates this. Consider a 3-state example with the reward structure given in Table 4 and transition probability matrix given in Table 5.

Table 4: Rewards

Reward	$h = 0$	$h = 1$	$h = 2$
$r(h)$	0	.083	.083
$R_{T_1}(h)$	0	6	7
$R_{T_2}(h)$	0	7	7.5

Table 5: Transition probability matrix

h	0	1	2
0	1	0	0
1	.2	.8	0
2	.1	.4	.5

In solving the Bellman's equations in (3), one finds that if only T_1 will be available, it is optimal to take it from both health states $h = 1$ and $h = 2$. However, in solving the Bellman's equations in (2), one finds that even though R_{T_2} is greater than R_{T_1} , if T_2 is available then it is optimal to take it from health state $h = 1$ but to wait from state $h = 2$. The intuition for this result is that the patient with T_1 faces a bigger drop-off in reward (compared to the patient with T_2) when initiating it from $h = 1$ compared to $h = 2$. Therefore the patient

in state $h = 2$ with T_1 does not want to risk waiting further and possibly deteriorating to health state $h = 1$. The patient with T_2 , on the other hand, finds that overall it is better to spend some time waiting in health state $h = 2$ before using the benefit of the treatment from health state $h = 1$.

When $q = .10$ and the increase over R_{T_1} is 10%, then even when there is just one period until the trial is resolved, the patient still initiates T_1 from every health state. In other words, given that the naive patient initiates T_1 from every health state for the reasons discussed above, the knowledgeable patient is still not enticed to wait when the probability of trial success and the improvement of the new therapy are fairly low. As a result of Theorem 3, if for some $1 \leq t' < 12$ we find that $v_t[(h, 1)]$ matches $w^*(h)$ for all h , we will not present all twelve months of results since they will be identical for $t > t'$. We varied q between .1 and 1 (in increments of .1) and the percentage increase of R_{T_2} over R_{T_1} between 10% and 100% in increments of 10%; we present a few of the combinations here.

Table 6 shows the optimal values and actions for the model across a range of combinations of q and R_{T_2} . There are several interesting results to note. For the cases of $(q, R_{T_2}$ increase) of $(.30, 10\%)$, $(.30, 50\%)$, and $(.10, 50\%)$, it is optimal to initiate T_1 from every health state when there are twelve months to go before the trial is completed. In other words, the knowledgeable patient acts no differently than the naive patient. The intuition behind this is that the combination of probability of trial success and improvement of R_{T_2} over R_{T_1} is such that the patient is not tempted to wait when it will take so long to learn if the trial is successful. Although we presented the perspective of a patient deciding what to do twelve months before a trial is resolved, because we assume the fixed probability q of T_2 coming to market, one can use the table to answer the question “what should the patient do after discovering that a clinical trial is under way, which will be resolved in t months with a probability of success of q ?” Every case presented shows some time period and health state for which the patient prefers to delay taking T_1 , which indicates an incorrect decision for the naive patient.

The first three cases consider $q = .30$, alongside increasing R_{T_2} vectors. We chose $q = .30$ as a starting point because as noted in Section 1, by some estimates, a drug in a Phase III clinical trial has a 30% chance of success. With that probability of success, if the benefit of R_{T_2} is expected to be twice that of R_{T_1} (the 100% scenario), then a knowledgeable patient in health state 4 would choose to wait twelve months prior to trial resolution, and expect 18.35 more quality-adjusted life years. This is an expected gain of 1.49 quality-adjusted life years over the naive patient who chooses to initiate T_1 . Suppose the trial is two months away from completion and R_{T_2} would be a 50% increase over R_{T_1} . Then the knowledgeable patient would wait in health states 2, 3, and 4, with expected gains in quality-adjusted life years over the naive patient of .20, .53, and 2.07, respectively.

The second three cases fix the percentage increase of R_{T_2} over R_{T_1} to 50% and vary the probability of trial success. One can observe from the table that for a given health state, the “wait” region expands as q increases (for example, the patient in health state 3 waits in time period 1 in the (.10, 50%) scenario, waits in time periods 1 through 4 in the (.50, 50%) scenario, and waits in time periods 1 through 8 in the (1, 50%) scenario). It is easy to prove that for a fixed R_{T_2} vector, waiting with a lower q implies waiting with a higher q . Similarly, for a fixed q , if a scenario B has an R_{T_2} vector that is at least as large as an R_{T_2} vector of another scenario A, then waiting in A implies waiting in B. This was observed with the first three cases discussed in the preceding paragraph.

The final case we present is when there is a 100% chance that R_{T_2} is twice as good as R_{T_1} . Interestingly, a patient in health state 1 still chooses to take T_1 when 8 months remain in the trial. However, if the trial is 7 months or less from completion, the patient would wait for the much better T_2 . While there is no question that the much better therapy will come to market, in this example the only question is whether or not the patient will survive long enough without therapy to see this occur. Patients in health state 1 face dim enough prospects of surviving 8 or more periods without therapy that they cannot afford to wait; they expect to live longer by taking T_1 at that time instead.

Table 6: Optimal values and actions, at different periods before trial resolution, for various combinations of q and R_{T_2} (as a percentage increase over R_{T_1}). The value in each cell gives the optimal expected remaining quality-adjusted life years. A darker gray cell indicates the optimal action is to start taking therapy T_1 while a lighter gray cell indicates the optimal action is to wait. Once it is optimal to initiate T_1 from a state, then we blank out all the cells to the left to indicate they have the same values and action (in accordance with Theorem 2).

q, R_{T_2} increase	h	t																		
		12	11	10	9	8	7	6	5	4	3	2	1							
.30, 10%	1												3.93							
	2												7.43							
	3												11.73							
	4										16.86	16.98	17.17							
.30, 50%	1											3.93	4.14							
	2										7.43	7.63	8.08							
	3										11.73	12.26	12.86							
	4		16.86	17.05	17.25	17.46	17.68	17.92	18.16	18.41	18.67	18.93	19.17							
.30, 100%	1											3.93	4.28	4.67						
	2								7.43	7.65	8.12	8.61	9.12							
	3							11.73	11.94	12.55	13.18	13.84	14.53							
	4	18.35	18.61	18.88	19.17	19.47	19.78	20.11	20.44	20.76	21.07	21.37	21.65							
.10, 50%	1												3.93							
	2												7.43							
	3											11.73	11.75							
	4								16.86	16.90	17.09	17.29	17.51							
.50, 50%	1											3.93	4.12	4.50						
	2										7.43	7.81	8.28	8.77						
	3									11.73	12.08	12.68	13.32	13.98						
	4	17.78	18.02	18.26	18.52	18.79	19.07	19.36	19.67	19.98	20.27	20.56	20.83							
1, 50%	1											3.93	4.12	4.50	4.91	5.34				
	2						7.43	7.79	8.27	8.79	9.33	9.91	10.51							
	3				11.73	11.79	12.39	13.04	13.72	14.43	15.17	15.95	16.76							
	4	20.71	21.08	21.47	21.87	22.30	22.71	23.12	23.52	23.91	24.28	24.64	24.97							
1, 100%	1												3.93	4.14	4.52	4.94	5.41	5.93	6.51	7.15
	2	7.43	7.51	7.96	8.46	9.01	9.60	10.24	10.91	11.62	12.37	13.16	13.99							
	3	12.50	13.18	13.91	14.67	15.48	16.34	17.23	18.16	19.13	20.15	21.21	22.31							
	4	27.11	27.71	28.31	28.91	29.51	30.09	30.66	31.22	31.77	32.29	32.79	33.27							

4 Conclusions and Future Research

This paper seeks to address an issue that is lacking in current models of optimal patient treatment, namely the real possibility of new therapies coming out of the R&D pipeline. MDPs provide a framework for balancing the immediate and future effects of decisions and therefore allow for forward thinking rather than myopic decision making. However, an MDP model of treatment is still myopic if it considers that the actions available at the current time period will be the same actions available for every future time period, especially when decisions are made in an environment of R&D.

We developed an MDP model of whether a patient should undergo an irreversible treatment or wait for the possibility that an improved therapy currently under clinical trial will come to market N periods from now. Structural properties of the solution provide insight into the optimal decisions and their effects. For example, a knowledgeable patient in some health state who decides to wait will also wait in that health state for all time periods closer to the trial resolution. Results also relate behavior of the naive patient to that of the knowledgeable patient: if the naive patient would wait to take an available therapy in some health state, so would the patient aware of the clinical trial. Other structural properties were presented in Section 2.3. **Our assumptions that rewards and transition probabilities do not change over the time horizon of the problem involved a tradeoff between increasing model detail and demonstrating these analytical properties of the solutions. However, we believe our assumptions capture the key dynamics that would present itself even if we considered time-based model parameters: modest near-term gains when waiting (as represented by $r_t(h)$), real near-term mortality risks when waiting ($p_t(0|h)$), larger gains obtained by taking an existing therapy ($R_{T_1,t}(h)$), and even larger gains achievable by taking an improved therapy should it become available ($R_{T_2,t}(h)$). Therefore, we believe similar structural results would arise in this setting as well.**

Our numerical examples yielded intuitive outcomes supported by our theoretical results. The examples demonstrate that knowledge of the possibility of a new treatment becoming available can have considerable effects on optimal treatment policies when compared to decisions made without regard to new treatment development. Our examples also quantify the costs of suboptimal decisions made by a naive decision maker. This underscores the motivation for our work: that models of treatment that do not consider the real possibility of new treatment discovery might lead to errors in medical decision making.

Future research directions include building upon the current work by extending its applicability in a number of ways. First, we assumed that once a therapy is taken, it precludes the

use of downstream therapies. Although we provided examples where patients are faced with serious decisions in which this applies, there are certainly many situations for which this is too strong of an assumption. In our numerical example, we supposed that a patient initiating T_1 would develop total cross-resistance to the possible new therapy T_2 . However, it is possible that an entirely new class of drugs is developed for which resistance to T_1 does not confer resistance to the new T_2 , or perhaps the virus only builds partial resistance to T_2 (leaving T_2 partially effective). In the former case, there is no opportunity cost to the patient who starts with T_1 , as it does not alter the ability nor the benefit of taking T_2 at a later time. The case of partial resistance would lie somewhere between the notion of complete irreversibility and the case of T_1 having no effect on T_2 : the patient initiating T_1 could still take T_2 later, but at a reduced benefit. Situations in which taking a therapy now does not preclude taking an improved therapy later are not well suited to the optimal stopping time framework of our model in which we associate estimates of expected remaining quality-adjusted lifetimes with initiating treatment. The new modeling framework would require explicitly modeling disease progression after initiating therapy and the possible therapy switching decisions involved as new treatments become available. We also felt it was more interesting to first focus on cases for which there is a significant opportunity cost of taking certain therapies at the current time: that a patient cannot then use therapies availing themselves at later times.

We can extend our model to relax other assumptions fairly easily. For example, we assumed a fixed vector of rewards, $R_{T_2}(h)$, which is known to the patient and physician. Our model can consider the case where the reward vector may take on several values with some probability distribution. In that case, the optimality equations of (4) would use the expected value of $v_0[(j, 2)]$ under the wait action. Our model may also be adapted to consider a distribution over the time until trial resolution (N), rather than a fixed and known time. In that case, the finite horizon solution algorithm would start at the last possible trial completion time and assume with certainty that the trial is resolved at that time until it steps back to the time period before the second-to-last possible trial completion time. At

that point, the algorithm “sees” that the trial may be resolved in one period or it may be resolved at that last possible period. The probabilities would be based on the Bayesian updates to the prior distribution of all possible trial completion times. For example, suppose the trial might be resolved in 10, 12, or 14 months, each with probability $1/3$. Starting at time 13 months, there is certainty that the trial will finish in one more month (though the success or failure of the trial still remains uncertain with success probability q). When the algorithm steps back to time 11 months, then it finds a probability of $1/2$ that the trial finishes in one more period, and probability $1/2$ that it finishes in 3 more periods. Then at time 9 months, the algorithm considers the three different possible completion periods with the original prior distribution.

We also assumed that estimates of q are not updated over the course of the decision horizon due to a lack of detailed information on the trial’s evolution. However, if the state of the trial could be observed periodically, we could model its dynamics and effect on q . For example, one period before the trial ends, we could consider the system to be in state (s_p, s_c) , where s_p describes the state of the patient and s_c describes the state of the clinical trial. Each s_c would then correspond to an estimate of q , and for the time leading up to that period, we could consider a transition probability matrix over possible states of the trial.

Our model implicitly assumes patients are risk-neutral for health outcomes. Because various risk attitudes (both risk-averse and risk-seeking) have been observed with different types of patients [8, 22], it may be worthwhile to cast the treat vs. wait decision in an expected utility framework. Risk aversion has been considered in an MDP model of the optimal time to accept a liver transplantation [7]. Although the authors of that paper obtained results comparing the optimal value functions between patients with different risk preferences, it would also be interesting to explore structural properties relating the different risk preferences to the patients’ optimal treatment policies. It is not clear how various risk considerations may affect treat vs. wait decisions in our model, and we leave further consideration of risk sensitivity for future work.

Because medical R&D presents significant improvements in therapy for many patients over the course of their disease, the modeling insights and results presented here are intended to motivate patients and their physicians to consider how R&D pipeline dynamics might affect important decisions to be made at the present time. There are various sources from which one can glean information about the state of clinical trials for many drugs and diseases (see for example, www.clinicaltrials.gov and www.phrma.org). These sources may serve as starting points for estimating the likelihood of improved therapies coming to market, along with their potential effects.

5 Acknowledgments

We would like to thank Turgay Ayer, Fatih Erenay, Mahesh Nagarajan, Martin Puterman, and Greg Werker for their helpful comments. We also thank three anonymous referees and the associate editor for all of their thoughtful feedback. Steven Shechter was supported by the Natural Sciences and Engineering Research Council Discovery Grant (341415-07). Oguzhan Alagoz was supported by the National Science Foundation Grant (CMMI-0700094).

6 Appendix

Proof of Proposition 1 Let $t \geq 2$, and suppose $v_t[(h, 1)] \leq v_{t-1}[(h, 1)]$ for all h . Then by the optimality equations given in (5), it follows that $v_{t+1}[(h, 1)] \leq v_t[(h, 1)]$ for all h . Therefore, if we can show that $v_2[(h, 1)] \leq v_1[(h, 1)]$ for all h , the result follows by induction.

$$v_2[(h, 1)] = \max \left\{ r(h) + \sum_{j=0}^{j=H} p(j|h)v_1[(j, 1)], R_{T_1}(h) \right\} \quad \forall h.$$

Rewriting $v_1[(h, 1)]$, we have:

$$v_1[(h, 1)] = \max \left\{ r(h) + \sum_{j=0}^{j=H} p(j|h)[(1-q)v_0[(j, 1)] + (q)v_0[(j, 2)], R_{T_1}(h) \right\} \quad \forall h.$$

Therefore, if we show that $v_1[(j, 1)] \leq (1-q)v_0[(j, 1)] + (q)v_0[(j, 2)]$, we are done. Case 1: $v_1[(j, 1)]$ is obtained by taking T_1 (so that $v_1[(j, 1)] = R_{T_1}(j)$). Then the result follows because $v_0[(j, 1)] \geq R_{T_1}(j)$ (by the optimality equations of (3)), and $v_0[(j, 2)] \geq R_{T_1}(j)$ (by the optimality equations of (2) and the assumption that $R_{T_2}(j) \geq R_{T_1}(j)$). Case 2: $v_1[(j, 1)]$ is obtained by waiting. Then we have:

$$v_1[(j, 1)] = r(j) + \sum_{k=0}^{k=H} p(k|j)[(1-q)v_0[(k, 1)] + (q)v_0[(k, 2)]]. \quad (6)$$

Note that by the optimality equations of (3) and (2), we have the following

$$v_0[(j, 1)] \geq r(j) + \sum_{k=0}^{k=H} p(k|j)v_0[(k, 1)], \text{ and} \quad (7)$$

$$v_0[(j, 2)] \geq r(j) + \sum_{k=0}^{k=H} p(k|j)v_0[(k, 2)]. \quad (8)$$

Combining (7), (8), and (6), we have:

$$(1-q)v_0[(j, 1)] + (q)v_0[(j, 2)] \geq r(j) + \sum_{k=0}^{k=H} p(k|j)[(1-q)v_0[(k, 1)] + (q)v_0[(k, 2)] = v_1[(j, 1)],$$

which completes the proof.

Proof of Theorem 1 Note that the optimality equations of (5) have the same structure as those in (1) (the only difference being that the former consider time-indexed value functions). To solve the equations of (1), we can apply value iteration and initiate the algorithm with the

value function obtained by solving (2), (3), and then (4). As the solution of (4) is also used when stepping back in the solutions of (5), the value iterates of the infinite horizon model of (1) coincide with the time-indexed values in (5). The convergence of value iteration for this undiscounted problem is guaranteed by observing that this problem fits the framework of a stochastic shortest path problem and applying the results of [6] (in our problem we actually want the longest path, but the results still apply).

Proof of Theorem 2 Suppose $v_{t'}[(h', 1)] = R_{T_1}(h')$. By Proposition 1, we know that $v_t[(h', 1)] \leq v_{t'}[(h', 1)] = R_{T_1}(h')$ for all $t > t'$. But by the optimality equations, $v_t[(h', 1)] \geq R_{T_1}(h')$. So $v_t[(h', 1)] = R_{T_1}(h')$, and thus it is optimal to take T_1 in state h' with $t > t'$ periods to go until the trial is resolved.

Proof of Corollary 1 We prove the contrapositive. Suppose there is a t' , $1 \leq t' \leq N$, such that it is optimal for the knowledgeable patient to take T_1 when in state h' . Then $v_{t'}[(h', 1)] = R_{T_1}(h')$. By Proposition (1) and Theorem (1), we have that $w^*(h') \leq v_{t'}[(h', 1)]$. However, by the optimality equations, it also holds that $w[(h', 1)] \geq R_{T_1}(h')$. Therefore $w^*(h') = R_{T_1}(h')$ and it is optimal for the naive patient to take T_1 when in state h' .

Proof of Theorem 3 The first part of the theorem follows from Theorem 1, noting the similarities of (1) and (5), and the result from MDP theory that says that the max operator of (1) acting on the optimal value vector returns the optimal value vector [19]. The second part follows from the MDP result that conserving decision rules are optimal [19].

References

- [1] O. Alagoz, C.L. Bryce, S. Shechter, A. Schaefer, C.C.H. Chang, D.C. Angus, and M.S. Roberts. Incorporating biological natural history in simulation models: Empirical estimates of the progression of endstage liver disease. *Medical Decision Making*, 25(6):620–632, 2005.
- [2] O. Alagoz, L. M. Maillart, A. J. Schaefer, and M. S. Roberts. Choosing among living-donor and cadaveric livers. *Management Science*, 53(11):1702–1715, 2007.
- [3] O. Alagoz, L. M. Maillart, A. J. Schaefer, and M. S. Roberts. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research*, 55(1):24–36, 2007.
- [4] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. The optimal timing of living-donor liver transplantation. *Management Science*, 50(10):1420–1430, 2004.
- [5] American Cancer Society, 2007. Retrieved August 3, 2007 www.cancer.org/docroot/CRI/content/CRI_2_4_4X_Expectant_Therapy_Watching_and_Waiting_30
- [6] D.P. Bertsekas and J.N. Tsitsiklis. *Dynamic Programming and Optimal Control*. Athena Scientific, Belmont, Massachusetts, 2001.
- [7] A. Bhandari, A.J. Schaefer, and M.S. Roberts. Optimal liver acceptance for risk-sensitive patients, 2007. Working paper.
- [8] D.J. Cher, J. Miyamoto, and L.A. Lenert. Incorporating risk attitude into Markov-process decision models: Importance for individual decision making. *Medical Decision Making*, 17:340–350, 1997.
- [9] C. Cohen. Introduction: Why do we need a new class of HIV medications?, 2003. Retrieved August 1, 2007, www.thebody.com/fuzeon/intro.html.

- [10] J.A. DiMasi, R.W. Hansen, and H.G. Grabowski. The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, 22:151–185, 2003.
- [11] T. Driffield and P.C. Smith. A real options approach to watchful waiting: theory and an illustration. *Medical Decision Making*, 27:178–188, 2007.
- [12] M.R. Gold, D.L. Patrick, G.W. Torrance, D.G. Fryback, D.C. Hadorn, M.S. Kamlet, N. Daniels, and M.C. Weinstein. Identifying and valuing outcomes. In M.R. Gold, J.E. Siegel, L.B. Russell, and M.C. Weinstein, editors, *Cost-Effectiveness in Health and Medicine*, pages 82–134. Oxford University Press, New York, New York, 1996.
- [13] W.J. Hopp and S.K. Nair. Markovian deterioration and technological change. *IIE Transactions*, 26(6):74–82, 1994.
- [14] C.P. Lee, G.M. Chertow, and S.A. Zenios. Optimal initiation and management of dialysis therapy, 2007. Working paper.
- [15] E.K. Lee and M. Zaider. Operations research advances cancer therapeutics, 2008. To appear in *Interfaces*.
- [16] L.J. Martinez. An overview of the drug approval process, 2002. Retrieved October 15, 2007, www.thebody.com/content/art16845.html.
- [17] H. Meijers-Heijboer, B. van Geel, W.L.J. van Putten, S.C. Henzen-Logmans, C. Seynaeve, M.B.E. Menke-Pluymers, C.C.M. Bartels, L.C. Verhoog, A.M.W. van den Ouweland, M.F. Niermeijer, C.T.M. Brekelmans, and J.G.M. Klijn. Breast cancer after prophylactic bilateral mastectomy in women with a BRCA1 or BRCA2 mutation. *New England Journal of Medicine*, 345(3):159–164, 2001.
- [18] Pharmaceutical Research and Manufacturers of America, 2007. Retrieved July 30, 2007 www.phrma.org.

- [19] M.L. Puterman. *Markov Decision Processes*. John Wiley and Sons, New York, New York, 1994.
- [20] T.S. Quang, K.E. Wallner, P.R. Herstein, B. Marie Palo, J.B. Walker, and S. Sutlief. Technologic evolution in the treatment of prostate cancer, 2007. Retrieved May 15, 2008 www.cancernetwork.com/prostate-cancer/article/10165/62318.
- [21] S. Rajagopalan, M.R. Singh, and T.E. Morton. Capacity expansion and replacement in growing markets with uncertain technological breakthroughs. *Management Science*, 44(1):12–30, 1998.
- [22] E.B. Rasiel, K.P. Weinfurt, and K.A. Schulman. Can prospect theory explain risk-seeking behavior by terminally ill patients? *Medical Decision Making*, 25:609–613, 2005.
- [23] A.J. Schaefer, M.D. Bailey, S.M. Shechter, and M.S. Roberts. Modeling medical treatment using Markov decision processes. In M. Brandeau, F. Sainfort, and W. Pierskalla, editors, *Operations Research and Health Care: A Handbook of Methods and Applications*, pages 597–616. Kluwer Academic, 2004.
- [24] S.M. Shechter, M.D. Bailey, A.J. Schaefer, and M.S. Roberts. The optimal time to initiate HIV therapy under ordered health states. *Operations Research*, 56(1):20–33, 2008.
- [25] F. Yokota and K.M. Thompson. Value of information literature analysis: A review of applications in health risk management. *Medical Decision Making*, 24:287–298, 2004.