The Timing and Probability of Treatment Switch under Cost Uncertainty: An Application to Patients with Gastrointestinal Stromal Tumor

Felipa de Mello-Sampayo, PhD*
ISCTE-IUL – Economics, Lisbon, Portugal

ABSTRACT

Background: Cost fluctuations render the outcome of any treatment switch uncertain, so that decision makers might have to wait for more information before optimally switching treatments, especially when the incremental cost per quality-adjusted life year (QALY) gained cannot be fully recovered later on. Objective: To analyze the timing of treatment switch under cost uncertainty. Methods: A dynamic stochastic model for the optimal timing of a treatment switch is developed and applied to a problem in medical decision taking, i.e. to patients with unresectable gastrointestinal stromal tumour (GIST). Results: The theoretical model suggests that cost uncertainty reduces expected net benefit. In addition, cost volatility discourages switching treatments. The stochastic model also illustrates that as technologies become less cost competitive, the cost uncertainty becomes more dominant. With limited substitutability, higher quality of technologies will increase the demand for those technologies disregarding the cost uncertainty. The results of the empirical application suggest that the first-line treatment may be the better choice when considering lifetime welfare. Conclusions: Under uncertainty and irreversibility, low-risk patients must begin the second-line treatment as soon as possible, which is precisely when the second-line treatment is least valuable. As the costs of reversing current treatment impacts fall, it becomes more feasible to provide the option-preserving treatment to these low-risk individuals later on. Keywords: cost uncertainty, decision analysis, economic evaluation, health economics, switch treatments.

Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

For some diseases, patients receive a sequence of treatments. These may involve different drugs or different dosages of the same drugs. The decision regarding whether to move a patient to the next treatment in a sequence may be based on patient characteristics or patient history, and therefore subject to variability. If it is accepted that adoption decisions should be made with consideration of the associated decision uncertainty, then we may say that models submitted to decision makers should do two things: estimate expected net benefit (NB) and characterize decision uncertainty. If this dual purpose of models is accepted, failure to fulfill the latter requirement will limit its value for decision making and leave the decision maker without a key element of information.

The decision to adopt a particular technology should be based on the expected NB so that when comparing mutually exclusive treatment strategies for a particular disease, the optimal strategy is simply the one with the highest expected NB [1]. Nevertheless, decisions based on the expected NB are appropriate only if there is also some consideration of whether current evidence is sufficient for allocating health care resources, based on an assessment of the consequences of decision uncertainty [2]. If the decision uncertainty and the consequences of adopting a suboptimal treatment strategy are large, the decision maker may require further evidence on which to base the adoption decision [3].

For example, adopting some medical technologies restricts the use of certain medical technologies in the future, and explains the lack of consensus about when to start therapy in HIV patients [4,5]. Some advocate fighting HIV with a powerful combination of drugs as early as possible in the course of the disease to prevent the disease from progressing. Others are concerned that starting therapy at early stages may lead to the development of viral resistance to these drugs and related compounds and the disease may progress to an advanced stage more rapidly, while other clinicians advocate waiting until the disease reaches a more advanced stage to initiate treatments so that future options can be preserved. This problem of current decisions affecting future options has received considerable theoretical attention in the literature on economic investments. The higher the uncertainty about future outcomes, the more individuals will gain from waiting for more information before committing to investment (or dis-investment) whenever there are significant sunk costs [6]. This result is a prediction of the “option-pricing” approach to the analysis of irreversible investment under uncertainty [7–9]. Analogously, benefits associated

* Address correspondence to: Felipa de Mello-Sampayo, ISCTE-IUL, Economics, Av Forcas Armadas, c.c. 187, Lisbon 1649-026, Portugal. E-mail: fdmso@iscte.pt.

1098-3015/$36.00 – see front matter Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2013.12.008
with actions that preserve treatment choices in the future, above and beyond the direct value associated with those actions, are referred to as the option value of the intervention [10].

For many physicians the observation that current medical treatment decisions have repercussions for the treatment of health conditions in the future is an obvious one that is often considered in their clinical decision making. Such considerations form no part of health care technology assessment calculations, leading to potentially significant mischaracterizations of the treatment value. While it is difficult to systematically assess the size of the bias induced from ignoring option values, the only empirical study in the health domain found an increase in consumer willingness-to-pay of approximately 53% when option values were considered [11].

Using the option-pricing approach for the analysis of irreversible treatment choices under uncertainty is important because the health sector is one in which there is tremendous uncertainty about the demand for future medical technologies. When we begin treating a population of individuals, we do not know what additional conditions they will develop in the future. Because new diseases are constantly emerging, we do not even necessarily know the nature of these future conditions. Higher life expectancy prospects for new conditions to arise, especially those associated with aging such as cancer and dementia, make the option value of the interventions a key variable of the valuation equation. Ignoring option values during the drug approval and reimbursement setting process could result in disincentives to create socially valuable technologies. Finally, unlike many private investment decisions, decisions taken by national health systems may be effectively irreversible for political reasons. Palmer and Smith [12] focus on the timing of health investments and whether it makes sense to delay the adoption of a new technology in anticipation of the exogenous arrival of new information about its value. While the prospects for delaying investments have potentially important implications for decision making, delay is often not feasible in this setting, especially on the time scale under which we expect new information to arrive. When analyzing situations in which current treatment decisions have irreversible implications for the treatment of future diseases, and decision makers are choosing between competing interventions with differing temporal consequences, Zivin and Neidell [10] find that irreversibility raises the value of treatment modalities that preserve future treatment options. Introducing some reversibility, however, can either increase or decrease the option value, depending on the distribution of patient types. These authors also examine the relationship between these values and the biological and economic parameters that characterize any given set of technologies. Meyer and Rees [13] analyze the treatment decision at a general level. They determine optimal threshold values for initiating the intervention, and derive comparative statics results with respect to model parameters. In particular, an increase in the degree of uncertainty over the patient’s health state, in most cases, makes waiting more attractive. This may not hold, however, if the patient’s health state has a tendency to improve.

This article follows the theory very closely to develop a dynamic stochastic model for the optimal timing of a treatment switch. Its main value addition consists in the concrete application to a problem in medical decision taking, that is, to patients with unresectable gastrointestinal stromal tumor (GIST). In the stochastic model, we assume two lines of treatment in treating a chronic disease and we consider the problem of a patient who is using the first-line treatment but the decision maker is contemplating switching to a second-line treatment that consists of higher doses of the drug used in the first-line treatment and then provides a more advanced drug. The patient will use the new line treatment only if such a move is deemed beneficial in the medium and long term. That, in turn, will depend on the perceived evolution of cost. The higher the uncertainty regarding the cost of a new treatment, the more likely it is that a favorable situation will turn into an unfavorable one, and the more the patient will gain from waiting for more information before committing to the new treatment whenever the incremental cost per quality-adjusted life-year (QALY) gained cannot be fully recovered later on.

With the aim of empirically testing this study’s option-pricing model, an empirical application uses data from a modeling exercise that compared alternative treatment pathways for patients with unresectable GIST who failed to respond to imatinib 400 mg/d [14]. The study of Hislop et al. [14] assessed the effectiveness and cost-effectiveness of imatinib at escalated doses of 600 and 800 mg/d following progression of disease at a dose of 400 mg/d, compared with sunitinib, or the provision of best supportive care (BSC) only for patients with unresectable and/or metastatic GISTs. Several studies have reported further disease control after progression on an initial imatinib dose of 400 mg/d with dose escalation of imatinib to 800 mg/d, and this has also become common practice [15,16]. However, it should be noted that current National Institute for Health and Care Excellence guidelines for imatinib do not actually recommend dose escalation for patients with unresectable and/or metastatic GISTs who progress on an initial dose of 400 mg/d [17] but suggest that clinical decisions be made on an individual case-by-case basis, reflecting uncertainty regarding optimal practice.

Three studies [18–20] compared imatinib with BSC. The study by Wilson et al. [18] used the manufacturer submissions (Novartis model) and compared imatinib and BSC, but in the imatinib group allowed for escalation of doses from 400 to 600 mg/d for those who failed to respond or were intolerant to imatinib at the 400 mg/d dose. The study by Mabasa et al. [20] noted that patients included from retrospective cohorts in their analysis were given imatinib 400 mg/d until disease progression, and later were allowed escalated doses of between 600 and 800 mg/d. Six of 56 patients in the imatinib group of patients considered in this economic evaluation were then allowed to switch to sunitinib therapy. The economic evaluation by Huse et al. [19] considered imatinib at 400 mg/d. Two studies [21,22] compared both imatinib and sunitinib with BSC for patients who had failed or become resistant to imatinib 400 mg/d.

The empirical application of this study assumes that patients in the first-line treatment are being treated with 400 mg/d and the second-line treatment consists of dose escalation of imatinib to 600 mg/d followed by sunitinib. Empirical results suggest that the existence of an option value means that the first-line treatment may be the better choice when considering lifetime welfare. Thus, under irreversibility, low-risk patients must begin the second-line treatment as soon as possible, which is precisely when the second-line treatment is least valuable. As the costs of reversing current treatment impacts fall, it becomes more feasible to provide the option-preserving treatment to these low-risk individuals later on.

This article is organized as follows. The following section develops the stochastic option-pricing model, specifying the two feasible treatments and examining the effect of cost shocks on both the timing of treatment switching and the NB of each treatment. The next section presents the probability and expected time of treatment switch. The discussion of the results with an empirical application is presented in the following section. Conclusions are discussed in the last section.

The Model

In this section, we develop a model to illustrate the role that uncertainty and irreversibility can play in determining the decision regarding whether to move a patient to the next treatment
in a sequence. Suppose there are two lines of treatment in treating a chronic disease, denoted by L1 and L2, respectively. The first-line treatment consists of exclusively using one technology denoted by T1. The second-line treatment consists of using a higher dosage of T1 and then starting with a different technology, denoted by T2. We start constructing a one-sector model that allows health capital to augment through public investment. Then, we determine the NB of using Ti, i = {1, 2}. Finally, we define each line of treatment and show that when the patient is using L1, the decision as to whether to use the second-line treatment, L2, constitutes an optimal stopping problem.

Assume a one-sector model that allows health capital to augment through public investment. In particular, an individual’s health status is determined by public health measures such as provision of clinical facilities, sanitation, inoculation, and disease control programs (see Chakraborty [23]). Population individuals are endowed with one unit of labor that is inelastic supplied to firms, and receive wage income at the rate w0. Public health expenditure in period t is financed through a constant tax x ∈ (0, 1) on labor income so that health investment per person equals rw0. Such investment augments private health capital through a constant returns technology:

\[ h_t = rw_t \]  

(1)

Decision makers observe and decide the viability, utility, and characteristics of health care goods and services only after using those products or services. Thus, the quality of health care goods or services can be ascertained only on their consumption. In such cases, a drop in price is often interpreted by the prospective consumer as a drop in quality or utility of the product or service. Indeed, it is possible for the demand curve for medical care to be upward sloping, even though medical care is a noninferior good, a relationship that has some empirical support [24–26]. Housing is another example of a noninferior good whose own demand can be upward sloping (see Dusansky and Koç [27]). Under this hypothesis, the demand for medical care is given by

\[ T_i = B_i M^s_i, \]  

(2)

where Ti, i = {1, 2} is the total quantity demanded of a health care good or service at time t, Bi is the benefit gain at time t, and φ is the parameter for the elasticity of demand. We consider that medical care operates where a patient’s demand for treatments is inelastic, 1 < φ < 1. An example of a perfectly inelastic demand would be a lifesaving drug that people will pay any price to obtain. Even if the price of the drug were to increase dramatically, the quantity demanded would remain the same. In this model set-up, the mechanism that leads health to be a Giffen good also involves a wealth consideration; when the price of treatment falls, the patient is effectively wealthier, he or she can afford more treatments generally, and so, he or she needs fewer treatments of this kind. From Equation 2, the benefit function of a representative patient is given by

\[ B_i = T_i / M^s_i. \]  

(3)

The benefits increase with the number of health care goods or services demanded, but it is inversely related to the per capita health investment. The higher a patient’s health capital, the fewer health care goods or services he or she needs for the same benefit. The cost of the technology, Ci, i = {1, 2}, is a nonlinear function of a patient’s particular characteristic, x, that evolves over time:

\[ C_i = rx^k, \]  

(4)

where r is the interest rate, k is the capital invested in the technology, and x is the level of the state variable that represents the random shock of the cost side at time t. For analytical tractability, the state variable is assumed to evolve according to a geometric Brownian motion:

\[ dx = ax dt + o dx \]  

(5)

where \( dz = \sigma \sqrt{dt} \) is the increment of a Wiener process and \( \sigma^2 N(0,1), E(\xi_t, \xi_s) = 0 \) for \( s \neq t \).

Equation 5 implies that the current value of the random shock is known, but the future values are log-normally distributed with a variance growing linearly with the time horizon.

The value of each technology is expressed in terms of NBs, that is, the value of health benefits generated by Ti, i = {1, 2}, at time t, \( \{ (B_i) \} \), minus the cost of that technology, \( \{ (c_i) \} \). The NB from treatment with T1, at time t, is denoted by NB1t and the NB from using T2 is denoted by NB2t. The NB for the patient using Ti, i = {1, 2}, is as follows:

\[ NB_i = B_i - C_i \]  

(6)

Under uncertainty, the decision to adopt a particular technology is based on the expected present discounted value of the NB, so that when comparing mutually exclusive treatment strategies for a particular disease, the optimal strategy is simply the one with the highest expected present discounted value.

**Proposition 1**: Higher volatility of a patient’s particular characteristic reduces the expected NB associated with both technologies.


Comparing the state variables in Equations 17 and 18 in Appendix A, the expected NB is proportional to \( x^d, \delta = 2(e^\gamma - 1) / \gamma \), under both technologies. Recalling that medical care operates where the patient’s demand for health care good or service is inelastic (\( \phi < 1 \)), it follows that \( \delta < 0 \). Consequently, cost uncertainty reduces the expected NB under both technologies. Under our model set-up, the magnitude of this adverse effect is identical for both technologies.

**Two Lines of Treatment**

We assume two lines of treatment in treating a chronic disease. The first-line treatment, L1, consists of exclusively using T1 and the second-line treatment, L2, consists of using a higher dosage of T1 and then starting with T2. When the patient is exclusively using T1, the decision as to whether to use the second-line treatment constitutes an optimal stopping problem for which the relevant Bellman equation is as follows:

\[ V^2(x,t) = \text{Max} \left\{ V^2; \text{NB}_i + \lim_{\text{dt} \to 0} \text{E}_t \left[ \frac{1}{\text{dt}^2} \right] \right\} \]  

(7)

where \( V^2(x,t) \) is the option value of intervention associated with using the second-line treatment, \( V^2 \) accounts for the expected value gain that results from switching treatment and starting the second-line treatment, and the second term in curly brackets yields the time-discounted expected increment in the value of the option that arises from keeping the option unexercised for an additional lapse of time, dt. The range of values for which the second term in curly brackets, in Equation 7, is greater than the first defines the continuation region, in which it is optimal not to exercise the option.

During the second-line treatment, patients are offered a higher dosage of T1 for the first \( \bar{T} \) periods, after which point they switch to T2. Letting \( \bar{T} \) be the parameter for the time at which T2 will be used during the second-line treatment, the value of the optimal treatment switch, \( V^*(x) \), is given by the NB after using T2 minus the incremental cost per quality-adjusted life-year (ICQALY) gain of using a higher dosage of T1, that is, the value of QALY
wasted in $T_1$ dose escalation, as follows:

$$V^2 = E_x \left\{ \int_0^\infty NB_{x,t} e^{-\alpha t - b} dt \right\} - \int_0^t IC_{QALY} e^{-\alpha t - b} dt$$  \hspace{1cm} (8)$$

where $t$ denotes the time at which $T_2$ will be used, and $IC_{QALY}$ stands for the incremental cost of quality-adjusted life-year (QALY) gain of using a higher dosage of $T_1$. QALY measures the cost of disease burden, including both the quality and the quantity of life lived. $IC_{QALY}$ measures the QALY value wasted when using a higher dosage of $T_1$ instead of using the more advanced technology $T_2$. The first term in Equation 8 reflects the expected returns under $T_2$, and the second term reflects the expected penalty associated with the higher dosage of $T_1$.

**Proposition 2:** Treatment switch occurs only if the relative benefit associated with $T_2$ exceeds the present value of the relative cost of $T_2$.

**Proof:** See Appendix A in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2013.12.008.

In Equation 26 in Appendix A, $\hat{x}$ is the critical value, that is, the value above and beyond the direct value of the second-line treatment. When $\hat{x} > 0$, the value associated with using the second-line treatment exceeds that of a situation of using the first-line treatment. It follows from Equation 26 and the assumption on the parameters that the value of $x$ is greater than 0 if

$$\phi_{x,\beta} \left( e^{-\alpha x} \right) > \phi_1,$$

implying that the decision maker will switch to second-line treatment only if the relative benefit associated with $T_2$ exceeds the present value of the relative cost of $T_2$ weighted geometrically by the elasticity of demand, and that is due to the uncertainty of treatment’s cost.

Moreover, because

$$\frac{\partial \hat{x}}{\partial \sigma^2} > 0 \text{ and } \lim_{\sigma \to \infty} \hat{x} = \infty,$$

the greater the volatility of the cost (i.e., the higher $\sigma^2$), the higher the critical value has to be to make it optimal for the decision maker to switch to the second-line treatment. The higher the expected trend of the treatment’s cost, the less the option of using the second-line treatment is worth, and thus the lower the value that triggers the use of $T_2$, that is,

$$\frac{\partial \hat{x}}{\partial \alpha} < 0,$$

The reason for this is that the more expensive one expects technology to become, the lower the uncertainty that results from the switch from a situation of using the first-line treatment to one in which the patient uses the second-line treatment.

With regard to the discount rate, the greater the agent’s time discount rate, the less he or she values the option, and thus the lower the value $x$ that triggers optimal treatment switch; that is,

$$\frac{\partial \hat{x}}{\partial \beta} < 0.$$

This result stems from the fact that a higher time preference increases the decision maker’s opportunity cost of not immediately using the second-line treatment. In the extreme case in which the decision maker cares only about the present moment, so that $\mu \to \infty$, then

$$\lim_{\mu \to \infty} \frac{\beta_1}{\beta_1 - \beta} = 0 \text{ and } \hat{x} = 0,$$

so that uncertainty is disregarded and the value of the second-line treatment option collapses to 0.

The greater the tax rate, the higher the patient’s health capital; thus, the more decision makers value using the second-line treatment option, and thus the higher the value $\hat{x}$ that triggers optimal $T_2$ use; that is,

$$\frac{\partial \hat{x}}{\partial \phi} > 0.$$

Last, the lower the relative cost of $T_2$, the higher the relative benefit of $T_2$ and the sooner $T_2$ will be used, the lower the threshold for using the second-line treatment.

$$\frac{\partial \hat{x}}{\partial \phi} < 0, \text{ } \frac{\partial \hat{x}}{\partial \phi} > 0 \text{ and } \frac{\partial \hat{x}}{\partial \alpha} < 0.$$

**The Timing and Probability of Switching to Second-line Treatment**

Before proceeding to the empirical application, it would be interesting to ascertain, from any point within the continuation region, the likelihood that using the second-line treatment will become optimal in the future. It is important for the decision maker to know the expected time that will transpire until the decision of using the second-line treatment becomes optimal.

Using standard properties of the Brownian motion and the lognormal distribution (see Dixit [28] and Øksendal [29]), closed-form solutions for the probability $Q(x)$ and the expected time $T(x)$ for the process $x$ to hit the barrier $\hat{x}$ from any point inside the continuation region are given by

$$Q(x) = \begin{cases} 1 & \text{if } a \leq \frac{1}{2} \sigma^2 \\ \frac{\left( e^{-\beta_1 x} - \frac{b}{\alpha} \right)}{x} & \text{if } a > \frac{1}{2} \sigma^2 \end{cases}$$  \hspace{1cm} (9)$$

$$T(x) = \begin{cases} \infty & \text{if } a \geq \frac{1}{2} \sigma^2 \\ \frac{\ln(x/\alpha)}{\beta_1/2} & \text{if } a < \frac{1}{2} \sigma^2 \end{cases}$$  \hspace{1cm} (10)$$

where $a = (\sigma^2/2)$ and $\sigma^2$ are, respectively, the drift and variance parameters of the process $x$.

Equations 9 and 10 indicate that the probability and expected time until using the second-line treatment to become optimal depend on the variability and trend of the patient’s particular characteristic. The greater is the variability, $\sigma^2$, the higher is the likelihood that $x$ diverges away from the threshold that triggers the use of the second-line treatment, and so the lower the probability that using the second-line treatment will ever become optimal. Similarly, the higher the drift, $a$, the more likely long excursions of $x$ away from the critical ratio become, and so the more time the system is expected to take until hitting the threshold beyond which using the second-line treatment is optimal.

Using the second-line treatment will become optimal with certainty provided that $a < (\sigma^2/2)$, and it is expected to occur sooner the higher $x$ and the lower $\sigma^2$. For the limiting case in which $a = (\sigma^2/2)$, even though the probability that the patient will start using the second-line treatment in the future is 1, the expected time for it to occur is infinite. The intuition behind this apparently contradictory result is that if the drift of $x$ is 0, long diversions away from the barrier $\hat{x}$ might occur. Thus, because the probabilities for successfully longer hitting times do not fall sufficiently fast, the expectation, which is the average of the possible hitting times weighted by their respective probabilities, diverges. This argument is presented in Dixit [28].

For the set of parameters for which $x$ has a positive drift, that is, when $a > (\sigma^2/2)$, there is still a positive probability that using the second-line treatment will become optimal sometime in the future, as given by Equation 9. This is because in spite of $x$ drifting away from the critical ratio, there is the possibility that a combination of positive shocks might just bring the system toward the threshold barrier. The expected time for this event
is infinite, however, as given by Equation 10, given that there is a positive probability that \( x \) never reaches \( \bar{x} \) that drives the expectation into diverging.

### Empirical Application

The model presented gives clear indications regarding treatment switch decisions under cost uncertainty. It predicts that the higher the volatility of the patient’s particular characteristic, the sooner \( T_2 \) is used, the higher the tax rate and the higher the relative cost of \( T_2 \), the more valuable the option of using the second-line treatment will be, and so the fewer switches of treatment will be observed. Conversely, the higher the trend of the patient’s particular characteristics, the higher the discount factor and the higher the relative benefit of \( T_2 \), the more switches of treatment one would expect to observe. Thus, for empirical testing purposes, the reduced form of Equation 12 can be written as follows:

\[
treatment\ switch = f\left(\frac{\alpha^2}{\sigma^2} - \alpha - \mu - \bar{t} + \tau + \pi + \phi\right)
\]

These results are extended for Equation 26 in Appendix A by using simulations. The simulations are performed against a benchmark case. The data in the present application consist of the cost-effectiveness of imatinib for GISTs [18]. The data are described in detail in Appendix B. Current guidelines at the time of the assessment recommended an initial dose of 400 mg/d, with the option of proceeding to a higher dose in the event of a poor response or disease progression, and withdrawal of treatment in the absence of benefit after 8 weeks. Nevertheless, because of a paucity of data, the best starting dose of imatinib and best treatment pattern were highly uncertain. The model had four health states: progressive disease, treatment with 400 mg imatinib, treatment with 600 mg imatinib, and death. Patients in the imatinib treatment group began with 400 mg/d. For those patients who failed to respond to 400 mg imatinib, a random number was generated to determine whether they would be moved to 600 mg, or straight to the progressive disease state. The probability of receiving 600 mg was based on the number of patients who had responded after crossing over from 400 to 600 mg imatinib in a clinical trial. We assume that patients start with 400 mg/d and the second-line treatment consists of dose escalation of imatinib to 600 mg/d followed by sunitinib.

**Figures 1 to 5** provide a sensitivity analysis of the trigger value \( \bar{x} \) with respect to the following parameters of the model: \( \sigma, \alpha, \mu, \bar{t}, \tau, \pi, \) and \( \phi \). The parameters are calibrated with values...
shown in Appendix B in Supplemental Materials found at http://dx.doi.org/%2010.1016/j.jval.2013.12.008. The simulations carried out on critical values of cost shock confirm results of comparative statics discussed above. Figure 1 reveals that the trigger value is much more sensitive to $\sigma$ than to $\alpha$. This is because the higher the cost uncertainty, the higher the risk of treatment switch, and thus the higher the threshold to trigger the use of the second-line treatment. Figure 2 illustrates that the dampening influence of higher $\mu$ on the critical value strengthens as $\sigma$ increases.

Figure 3 shows that the trigger value rises when both $\varphi$ and $\pi$ increase. It illustrates that as $T_2$ becomes less cost competitive, the uncertainty about using the second-line treatment decision becomes more dominant. With limited substitutability, higher quality of $T_2$ will increase the demand for $T_2$ disregarding the cost uncertainty of treatment switch. The accentuated curvature of the surface graphed in Figure 3, in which the critical value rises very quickly as both $\varphi$ is high and $\pi$ is low, indicates that the lower the cost-quality relationship, the more the uncertainty of use of the second-line treatment decision becomes dominant.

Figure 4 illustrates that as soon as one is expected to use the new technology, the uncertainty about the use of the second-line treatment decision becomes more dominant. Figure 5 illustrates the impact of $\sigma$ and $t$ on the critical value of treatment switch, showing that the lower is the patient’s health capital, the higher is the uncertainty regarding the treatment switch decision.

Figures 6 and 7 illustrate, respectively, the impact of $\sigma$ and $x$ on the expected probability of treatment switch and on the probability of optimal treatment switch, as given by Equations 9 and 10.

In Figure 6, when $\alpha > (\sigma^2/2)$, the probability that $\tilde{x}$ will be hit in the future is increasing in $x$ and decreasing in $\alpha$. This is because, first, the lower is $\sigma^2$, the less valuable is the option to switch treatment, and so the more treatment switches will be observed, and second, the higher is $\alpha$, the more likely it is that the process will be “thrown off-course” by a sequence of positive shocks toward the optimality threshold.

In Figure 7, our simulation benchmark values are expressed in year terms, so that simulations for the expected time for optimal treatment switch can be read in years. Figure 7 shows that the lower is $\sigma^2$ and the higher is $x$, the sooner the treatment switch is expected to occur. The further away $x$ is from the trigger value, however, the greater is the effect of an increase in $\sigma^2$ on the delay expected before treatment switch becomes optimal.

In summary, the lower the volatility of cost, the more likely treatment switch is to become optimal and the sooner it is expected to occur. Moreover, treatment switch becomes likelier and is expected sooner, the closer is the system to the critical threshold, that is, the closer $x$ is to $\tilde{x}$.

Empirical results suggest that the existence of an option value means that the first-line treatment may be the better choice when considering lifetime welfare. Thus, under irreversibility, low-risk patients must begin the second-line treatment as soon as possible, which is precisely when the second-line treatment is least valuable.

**Conclusions**

Our stochastic dynamic model of sequential therapeutic regimes underlines the importance to characterize uncertainty. Growth in the availability of treatments for chronic diseases that require permanent intervention, along with general increases in life expectancy, suggests that the effect of omitting option values from evaluations will only become larger. In this article, we addressed these intertemporal dependencies by explicitly modeling the cost uncertainty.
The optimal timing for treatment switch becomes increasingly important for patients who might switch treatments. This article sets up a stochastic model that provides an optimal rule for the timing of treatment switch. In the real world, cost fluctuations among other factors render the outcome of any treatment switch uncertain, so that decision makers might have to wait for more information before optimally switching treatments, especially when the incremental cost per QALY gained cannot be fully recovered later on. Because in most cases decision makers are not compelled to switch treatments at any specific moment, they hold an option to switch treatments that should be exercised only when it is optimal to do so.

Viewed from the perspective of real option theory, this article sheds new light on some debates about switching treatments. Our theoretical model suggests that cost uncertainty reduces the expected NB. In addition, cost volatility discourages switching treatments. The stochastic model also illustrates that as technologies become less cost competitive, the cost uncertainty becomes more dominant. With limited substitutability, higher quality of technologies will increase the demand for these technologies disregarding the cost uncertainty. Several key insights emerge. Irreversibility raises the value of the option-preserving treatment. The existence of an option value means that a seemingly poorer treatment may be the better choice when considering lifetime welfare. Optimal decision making requires a careful comparison of the "costs" of a less effective treatment for a condition today with the "benefits" of more effective treatments in the future.

The intuition for these results is deepened when we recognize that one of the principal features driving our results is that patients have particular characteristics that make technology’s cost uncertain. Thus, under irreversibility, low-risk patients must begin the option-preserving treatment as soon as possible, which is precisely when the second-line treatment is least valuable. As the costs of reversing current treatment impacts fall, it becomes more feasible to provide the option-preserving treatment to these low-risk individuals later on.

These results are critical given the expiration of the patent for imatinib in 2015 for the United States and in 2016 for the European Union. Because imatinib offers the most effective response rate and durability for the initial treatment of GIST, the ability for society to obtain less costly generic imatinib may substantially change the enthusiasm for consumers and insurance payers alike to consider alternative treatments as a viable treatment option unless undoubted clinical benefit can be established. This dynamic interplay between drug manufacturers, insurance providers, hospitals, and policymakers will continue to unfold over the next several years.

There are limitations to this model that temper the application of its conclusions to health care decision making. This model considered the two most commonly used drug regimens in GIST and did not include more recently studied medications, such as regorafenib [30], or cytoreductive surgery [31]. These drugs may change survival probabilities and costs because they are more costly than BSC when second- and third-line treatments fail.

Our basic framework of option valuation highlights potentially important macro-level strategies to improve social welfare through medical technologies. Research investments that focus on transforming irreversibility from complete to partial could generate large social benefits. Investments in the development of alternative treatments for future diseases are also important, but the return to such investments will depend on the degree of irreversibility.

The greatest obstacle to translating theory to practice is the intensive data requirement, which in some cases would require coordination across firms whose products might interact.

Source of financial support: The author has no other financial relationships to disclose.