



Munich Personal RePEc Archive

**The Timing and Probability of Switching
to Second-line Regimen - An application
to Second-Line Antiretroviral Therapy in
India.**

Felipa, de Mello-Sampayo

ISCTE-IUL

29 December 2014

Online at <https://mpa.ub.uni-muenchen.de/60997/>

MPRA Paper No. 60997, posted 11 Jan 2015 07:43 UTC

The Timing and Probability of Switching to Second-line Regimen - An application to Second-Line Antiretroviral Therapy.

Felipa de Mello-Sampayo¹

ISCTE-IUL Institute University of Lisbon, Portugal

Abstract

Health fluctuations render the outcome of any treatment switch uncertain, so that decision makers might have to wait for more information before optimally switching treatments. This paper sets up a stochastic model that provides an optimal rule for the timing of treatment switch. The results of the model were then tested empirically with patient-based data on first-line and second-line antiretroviral treatment in India. The empirical results support the findings of the analytical model.

JEL Classification: C61, D81, I18.

Keywords: Uncertainty; Utility Function; Health; Antiretroviral Treatment; Treatment Switch

Corresponding author's email: fdmso@iscte.pt

Postal Address: ISCTE-IUL, Economics department, Av Forcas Armadas, 1649-026, Lisbon

¹ Financial support from Fundação para a Ciência e Tecnologia, under UNIDE-BRU, for the project entitled "Health and Economic Growth" PTDC/EGE-ECO/104157/2008 is gratefully acknowledged.

1 Introduction

The rapid scale up of antiretroviral therapy (ART) in developing countries over the past decade has resulted in substantial reductions in morbidity and mortality (Mills, Bakanda, Birungi, Chan, Ford, Cooper and et al., 2006) and increased life-expectancy (Mills, Bakanda, Birungi, Chan, Ford, Cooper and et al., 2011) for people living with Immunodeficiency Virus (HIV). A growing proportion of patients on antiretroviral therapy in developing countries have switched to second-line regimens. As second-line ART has a greater role in the management of HIV infection, the need for evidence to inform the criteria for sequencing and stopping ART is becoming critical.

Failure of antiretroviral therapy can be due to a number of factors, including base-line drug resistance amongst patients prior to starting treatment, the evolution of drug resistance during treatment, duration of time on treatment, and poor adherence to medication. 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 (Cohen, 2000, Harrington and Carpenter, 2000, Lasserre, Moatti and Soubeyran, 2006). Some advocate fighting HIV with a powerful combination of drugs as early as possible in the course of the disease in order 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. Although the disease may progress to an advanced stage more rapidly, other clinicians advocate waiting until the disease reaches a more advanced stage to initiate treatments so that future options can be preserved. The potential resistance induced by treating too early must be balanced against the potentially equally harmful consequence of not applying any treatment. In that case, a patient may deteriorate to such an extent that he will not be able to derive the benefits from a new, more advanced, therapy when it becomes available. This is called the therapeutic opportunity cost effect (Lasserre et al., 2006). 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 (Pindyck, 1988). This result is a prediction of the “option-pricing” approach to the analysis of irreversible investment under uncertainty (Dixit, 1989a,b, Dixit and Pindyck, 1994). Analogously, the benefits associated 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 (Zivin and Neideill, 2009, de Mello-Sampayo, 2014). 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 (Claxton, 1999, Claxton, Eggington, Ginnelly, Griffin, McCabe, Philips, Tappenden and Wailoo, 2005, Sculpher and Claxton, 2005).

Meyer and Rees (2012) analyze the treatment decision at a general level. They find that an increase in the degree of uncertainty over the patient's health state, in most cases, makes waiting more attractive. However, this may not hold if the patient's health state has a tendency to improve. This paper follows the theory very closely and its main value-added consists in the concrete application to a problem in medical decision taking, i.e. to patients with HIV. The model presented in this paper formalizes and solves the therapeutic dilemma about whether or not, and when, to switch a therapy. The model analyses and helps understand the optimal decision process. It establishes and rationalizes some surprising results. The results of the model reveal that optimal treatment switch should take longer the higher the uncertainty regarding the future path of health, the healthier the patients tend to become under each line of treatments, and the more risk averse decision makers tend to be. Conversely, the higher the level of health under the second-line treatment relative to that under the first-line treatment, the higher the discount rate and correlation between health in both treatments, and the sooner should the option of switching to second-line treatment be exercised.

With the aim of empirically testing the “option pricing” model, we use data from Freedberg, Kumarasamy, Losina, Cecelia, Scott, Divi and et al. (2007), who used a detailed HIV simulation model with patient-based data from Chennai, India, and projected the clinical outcomes of alternative HIV treatment strategies. In 2008, an estimated 2.27 million people between the ages of 15-49 years of India's 1160 million population was living with HIV (UNAIDS/WHO, 2010). India carries the largest burden of HIV behind South Africa and Nigeria. As the government and private sector continue to develop HIV treatment programmes in India, the importance of determining optimal strategies for initiation, switching, and stopping ART for this population is paramount and might help to improve treatment guidelines.

The government of India National AIDS Control Organization (NACO), the Association of Physicians of India, and the World Health Organization (WHO)

guidelines recommend starting ART in patients with cluster of differentiation 4 cells (CD4) count less than 350 cells per microliter (μl) and considering ART in patients with CD4 counts 500 cells/ μl . The optimal timing of initiation of ART and the importance of opportunistic infections in that policy, remain to be determined. Further, as second-line ART has a greater role in the management of HIV infection, the need for evidence to establish the criteria for sequencing and stopping ART is becoming critical (Gupta, Pujari, Joshi and Patel, 2007).

HIV Disease progression is characterized by decreases in CD4 count, and CD4 count determines the incidence of opportunistic infections and mortality resulting from HIV disease. In our stochastic model, health level volatility can be understood as the variation in cell count and the trend of health level as increases in CD4 count and thus, decreases in the incidence of opportunistic infections and mortality. The results of the empirical application suggest that the theoretical model can explain ART treatment switch. Indeed, as predicted, treatment switch depends negatively on the volatility of patients' health, and on trend of health, i.e. the greater the variation in CD4 count and the more CD4 count increase, the fewer treatment switches one expects to occur. Treatment switch also depends negatively on the degree of irreversibility. 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. The existence of an option value means that ART first-line regimen may be the better choice when considering lifetime welfare. Conversely, treatment switch depends positively on the discount rate and on the correlation between the patient's health under first and second-line treatments. This means that treatment switch is likelier to succeed in second-line treatments that are similar to the first-line treatments, implying that a decision maker should not rely on treatment switch as a risk diversification tool. Thus, by suggesting that ART treatment switch depends not only on the relative health level but crucially on the uncertainty surrounding the future path of health, the overall empirical results corroborate the implications of our analytical model.

The rest of the paper is organized as follows. In Section 2 we introduce a model that exhibits the main features just discussed: irreversible therapeutic decisions are made under uncertainty in a dynamic setup. Section 3 ascertains from any point within the continuation region, the likelihood that using the second-line treatment will become optimal in the future. In Section 4 the data are described and the empirical application

discussed. In Section 5 we go back to the general medical literature and raise issues that could not be tackled without further work.

2 The Model

Consider two lines of treatment in treating a chronic disease. Patients maximize a utility function of health and wealth, $U(h, c)$, where c denotes consumption (or alternatively, in a single-period setting, wealth), and h denotes health. $U(h, c)$ is increasing (or at least non-decreasing) both in c and in h (Levy and Nir, 2012).

The utility function under treatment i can be written as follows:

$$U_i(h, c) = h_i u(c), \quad (1)$$

where the index i addresses the two possible lines of treatment (first-line of treatment ($i = 1$) or the second-line of treatment ($i = 2$)), h_i denotes health level under the treatment i , c denotes consumption, $u(c)$ is a standard utility of consumption function.

Since the variables that determine the level of health behave differently from treatment to treatment, we must characterize health in the first-line treatment and in the second-line treatment differently. In particular, it is assumed that health is stochastic and follows a geometric Brownian motion:

$$dh_1 = \alpha_1 h_1 dt + \sigma_1 h_1 dz_1, \quad (2)$$

$$dh_2 = \alpha_2 h_2 dt + \sigma_2 h_2 dz_2, \quad (3)$$

where the subscripts 1 and 2 denote first-line and second-line treatment, respectively.

$dz_1 = \varepsilon_t \sqrt{dt}$ and $dz_2 = \epsilon_t \sqrt{dt}$ are the increments of Wiener processes and

$$\varepsilon_t \sim N(0; 1), \quad E(\varepsilon_t, \varepsilon_s) = 0 \text{ for } s \neq t,$$

$$\epsilon_t \sim N(0; 1), \quad E(\epsilon_t, \epsilon_s) = 0 \text{ for } s \neq t,$$

Equations (2) and (3) imply that the current value of the random shock is known, but the future values are log-normally distributed² with a variance growing linearly with the time horizon. $E_t(dh_1; dh_2) = \rho dt$ and ρ is the correlation coefficient between the random shocks affecting health under the first-line and second-line treatments. Health level's variability, σ_i^2 , can be interpreted as the level of the treatment risk and the trend of the health level, $\sigma_i = \mu - \delta_i$, as the growth rate of the health level, where δ_i stands for the healing rate.

Let $\alpha_1 < \alpha_2$ and $\sigma_1 < \sigma_2$, indicating that the second-line treatment represents an enhancement of the first and thus is more innovative and uncertain than the first one.

² See Aitchison and Brown (1957).

Furthermore, if δ_i represents the healing rate from treatment i , it can be interpreted as the therapeutic opportunity costs for waiting to use treatment i . Thus, $\delta_2 < \delta_1$ means that the second-line treatment has a lower opportunity cost of not immediately starting to use the second-line treatment than the first-line treatment. Assuming a power utility function of consumption, Equation (1) is now given by:

$$U_{it}(h, c) = h_{it} \frac{c_t^{1-\theta}}{1-\theta} \quad (4)$$

where $0 < \theta < 1$. Under the utility function given by Equation (4), a rich person and a poor person with the same health state, h , are willing to give up the same proportion of their consumption in order to be cured, i.e. we assume constant relative risk aversion.

Maximization of Equation (4) involves choosing the consumption level according to the following optimal rule, if we assume consumption to be unit-price:

$$c_t = h_{it}^{\frac{1}{\theta}} \quad (5)$$

Equation (5) implies that the optimal consumption increases with the level of contemporaneous health. Substituting Equation (5) into Equation (4) one obtains the utility function when the patient is constantly optimizing over time:

$$U_{it}^*(h, c) = \frac{h_{it}^{\frac{1}{\theta}}}{1-\theta}. \quad (6)$$

Using Ito's lemma, it can be confirmed that U_i^* also follows a geometric Brownian motion: $dU_i^* = \left(\frac{\alpha_i}{\theta} + \frac{1-\theta}{2\theta^2} \sigma_i^2 \right) U_i^* dt + \frac{\alpha_i}{\theta} U_i^* dz_i$.

Proposition 1: *Higher volatility of patient's health bounds the expected utility associated with both lines of treatment.*

Proof:

Analyzing the state variable in Equation (14) in Appendix, the expected utility under treatment i is proportional to h_i . Intuitively, this condition forces intertemporal utility to be bounded by imposing the time preference to be higher than the rate at which the θ th root of health is expected to increase. Health uncertainty affects differently the expected utility under both lines of treatment. Under our model set up, the magnitude of this effect depends on the drift α_i and volatility σ_i of health under treatment i , $i = \{1,2\}$.

Proposition 2: *Treatment switch will occur only if the value associated with using second-line treatment exceeds that of a situation of using the first-line treatment, i.e.*

$$\tilde{u}^* = \frac{\beta_1}{\beta_1 - 1} \quad (7)$$

Proof: See Appendix.

It follows from Equation (7) and the previous assumptions on the parameters that the value of $\tilde{u}^* > 1$, implying that the decision maker will switch treatments only if the expected utility under the second-line treatment exceeds that attained when using the first-line treatment, and that is due to uncertainty of the future path of health state.

In order to obtain the critical value as a function of the ratio of the model's state variables, i.e. the health level before and after treatment switch, Equation (14) in Appendix is used to obtain:

$$\frac{h_2^*}{h_1^*} = \left[\frac{\phi - \xi + \sqrt{(\xi - \phi)^2 - 4\phi \left(\left(\frac{\alpha_1 + \frac{1-\theta}{2\theta^2} \sigma_1^2 \right) - \mu \right)}}{-\delta - \phi + \sqrt{(\xi - \phi)^2 - 4\phi \left(\left(\frac{\alpha_1 + \frac{1-\theta}{2\theta^2} \sigma_1^2 \right) - \mu \right)}} \times \frac{\mu - \frac{\alpha_2 - \frac{1-\theta}{2\theta^2} \sigma_2}{\theta}}{\mu - \frac{\alpha_1 - \frac{1-\theta}{2\theta^2} \sigma_1}{\theta}} \right]^\theta, \quad (8)$$

where $\phi = \left(\frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} - \rho\sigma_1\sigma_2 \right) \frac{1}{\theta^2}$ and $\xi = \left(\frac{\alpha_1}{\theta} + \frac{1-\theta}{2\theta^2} \sigma_1^2 \right) - \left(\frac{\alpha_2}{\theta} + \frac{1-\theta}{2\theta^2} \sigma_2^2 \right)$. For values of the ratio $\frac{h_2}{h_1}$ lower than $\frac{h_2^*}{h_1^*}$ it is optimal not to switch treatments, and conversely, if the value of the ratio is greater than the critical value, the decision maker should use the second-line treatment. It follows that Equation (8) defines the line that divides the h_2 h_1 space into two regions: one where it is optimal to exercise the treatment switch option and another where it is not.

In order to do some comparative statics, the original setup will be simplified and it will be considered that $\alpha_1 = \alpha_2 = \alpha$ and $\sigma_1 = \sigma_2 = \sigma$, which allows one to write Equation (8) as:

$$\frac{h_2^*}{h_1^*} = \left[\frac{\sigma^2(1-\rho) + \sqrt{\sigma^2(1-\rho)(\sigma^2(1-\rho)) - 4 \left(\left(\frac{\alpha + \frac{1-\theta}{2\theta^2} \sigma^2 \right) - \mu \right)}}{-\sigma^2(1-\rho) + \sqrt{\sigma^2(1-\rho)(\sigma^2(1-\rho)) - 4 \left(\left(\frac{\alpha + \frac{1-\theta}{2\theta^2} \sigma^2 \right) - \mu \right)}} \right]^\theta \quad (9)$$

The above equation shows that the value of the ratio $\frac{h_2^*}{h_1^*}$ above which the decision maker should optimally switch treatments is greater than 1. This means that the decision maker will switch treatments only if the health level associated with the second-line treatment strictly exceeds that of a situation of exclusive first-line treatment and that is due to uncertainty of the future path of health.

Moreover, since

$$\frac{\partial(h_2^*/h_1^*)}{\partial\sigma} > 0,$$

the greater the volatility of health (i.e. the higher σ^2), the higher the critical value has to be to make it optimal for the decision maker to switch treatments. The more correlated are the shocks affecting the level of health under both the first-line and second-line

treatments, the less the treatment switch option is worth, and so the lower is the value of the relative health level that triggers treatment switch, i.e.

$$\frac{\partial(h_2^*/h_1^*)}{\partial\rho} < 0.$$

The reason is that the more correlated the shocks are, the more closely both health processes move, and so the lower the uncertainty that results from the switch from the first-line of treatment to the second-line of treatment. It follows from this result that the decision concerning treatment switch is not primarily geared at diversifying the patient's overall risk, quite the contrary. This means that treatment switch is likelier to succeed in second-line treatments that are similar to the first-line treatments, implying that a decision maker should not rely on treatment switch as a risk diversification tool.

The greater the expected trend of health, the more the option of switching treatments is worth, and thus the higher the value that triggers the switch of treatments, i.e.

$$\frac{\partial(h_2^*/h_1^*)}{\partial\alpha} > 0.$$

The healthier the patient is expected to become, the higher the uncertainty that results from the switch from a situation of using the first-line of treatment to one of using the second-line of treatment.

The more risk averse the agents are, the fewer treatment switches one expects to occur, i.e.

$$\frac{\partial(h_2^*/h_1^*)}{\partial\theta} > 0.$$

Treatment switches always exhibit some degree of irreversibility such that an increase in the share of consumption that the patients are willing to give up in order to be cured raises patient's overall risk exposure.

In regard to the discount rate, the greater is the agents' time discount rate, the less she values the switching treatments' option, and so the lower will be the value $h_{-2} = h_{-1}$ that triggers optimal treatment switch, that is

$$\frac{\partial(h_2^*/h_1^*)}{\partial\mu} < 0.$$

This result stems from the fact that a higher time preference increases the decision maker's opportunity cost of not immediately switching treatment. In the extreme case where the decision maker cares only about the present moment, so that $\mu \rightarrow \infty$, then $\lim_{\mu \rightarrow \infty} \frac{\beta_1}{\beta_1 - 1} = 0$ and $\tilde{u}^* = 0$, so that uncertainty is disregarded and the value of the option to switch treatments collapses to zero.

3 The Timing and Probability of Switching to Second-line Regimen

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 using the second-line treatment becomes optimal.

Using standard properties of the Brownian motion and the lognormal distribution, (see Dixit (1993) and Oksendal (2003)) closed-form solutions for the probability $Q\left(\frac{h_2}{h_1}\right)$ and expected time $T\left(\frac{h_2}{h_1}\right)$ for the process h_2/h_1 to hit the barrier $\frac{h_2}{h_1}$ from any point inside the continuation region, can be shown to be given by

$$Q\left(\frac{h_2}{h_1}\right) = \begin{cases} 1 & \text{if } (\alpha_2 - \alpha_1) \leq (\sigma_1^2 - \sigma_2^2) \\ \frac{(\alpha_2 - \alpha_1) + (\sigma_1^2 - \sigma_2^2) \left[\ln\left(\frac{h_2}{h_1}\right) - \ln\left(\frac{h_2}{h_1}\right) \right]}{e^{\sigma_1^2/2 + \sigma_2^2/2 - \sigma_1\sigma_2\rho} \left[\ln\left(\frac{h_2}{h_1}\right) - \ln\left(\frac{h_2}{h_1}\right) \right]} & \text{if } (\alpha_2 - \alpha_1) > (\sigma_1^2 - \sigma_2^2) \end{cases} \quad (10)$$

$$T\left(\frac{h_2}{h_1}\right) = \begin{cases} 1 & \text{if } (\alpha_2 - \alpha_1) \geq (\sigma_1^2 - \sigma_2^2) \\ \frac{\left[\ln\left(\frac{h_2}{h_1}\right) - \ln\left(\frac{h_2}{h_1}\right) \right]}{(\alpha_2 - \alpha_1) + (\sigma_1^2 - \sigma_2^2)} & \text{if } (\alpha_2 - \alpha_1) < (\sigma_1^2 - \sigma_2^2) \end{cases} \quad (11)$$

where $(\alpha_2 - \alpha_1) + (\sigma_1^2 - \sigma_2^2)$ and $\sigma_1^2/2 + \sigma_2^2/2 - \sigma_1\sigma_2\rho$ are respectively, the drift and variance parameters of the process $\frac{h_2}{h_1}$.

Equations (10) and (11) 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 health under the second-line treatment relative to that under the first-line treatment. The greater is the variability of patient's health under the second-line treatment relative to that under the first-line treatment, the higher is the likelihood that $\frac{h_2}{h_1}$ diverges 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, $(\alpha_2 - \alpha_1) + (\sigma_1^2 - \sigma_2^2)$, the more likely long excursions of $\frac{h_2}{h_1}$ 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.

For the set of parameters for which $\frac{h_2}{h_1}$ has a positive drift, i.e. when $(\alpha_2 - \alpha_1) > (\sigma_1^2 - \sigma_2^2)$, there is still a positive probability that treatment switch will become optimal sometime in the future, as given by (10). This is because, in spite of $\frac{h_2}{h_1}$ 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. However, the expected time for this event is infinite, given that there is a positive probability that $\frac{h_2}{h_1}$ never reaches $\frac{h_2^*}{h_1^*}$, which that drives the expectation into diverging.

Treatment switch will become optimal with certainty provided that $(\alpha_2 - \alpha_1) \leq (\sigma_1^2 - \sigma_2^2)$ and it is expected to occur sooner the higher both $\frac{h_2}{h_1}$ and α_2 , and the lower σ_2^2 . For the limiting case where $(\alpha_2 - \alpha_1) = (\sigma_1^2 - \sigma_2^2)$, even though the probability that the treatment switch will occur is one, the expected time for it to occur is infinite. The intuition behind these apparently contradictory results is that if the drift of $\frac{h_2}{h_1}$ is zero, long diversions away from the barrier $\frac{h_2^*}{h_1^*}$ might occur. Thus, since the probabilities for successfully longer hitting times do not fall sufficiently fast, and the expectation, which is the average of the possible hitting times weighted by their respective probabilities, diverges³.

In summary, the higher volatility of health under first-line treatment relative to that under the second-line treatment, and the healthier patients are expected to become under the second-line treatment relative to first-line's, 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 is $\frac{h_2}{h_1}$ to $\frac{h_2^*}{h_1^*}$.

4 Empirical Application

The model presented gives clear indications regarding treatment switch (TS) decisions under health uncertainty. It predicts that the higher the trend and volatility of the patient's health under both lines of treatment and the more risk averse is the decision maker, the more valuable the option of using the second-line treatment will be, and so the fewer switches of treatment there will be. Conversely, the higher the discount factor and the higher the correlation between health under the first and second-line treatments,

³ This argument is presented in Dixit (1993), p. 56.

the more switches of treatment one would expect to observe. Thus, for empirical testing purposes, the reduced form of the model can be written as follows:

$$TS = f \left(\begin{matrix} \sigma_1^2 & \alpha_1 & \sigma_2^2 & \alpha_2 & \rho & \mu & \theta \\ - & - & - & - & + & + & - \end{matrix} \right).$$

The simulations are performed against a benchmark case⁴. The data in the present application consist of the clinical outcomes for HIV-infected patients in India that used first-line and second-line ART therapy (Freedberg et al., 2007).

Freedberg et al. (2007) used a detailed HIV simulation model with patient-based data from Chennai, India, to project the clinical outcomes and cost-effectiveness of alternative HIV treatment strategies, and the impact of more expensive second-line therapy. The analysis showed that a single regimen of ART nearly doubled life expectancy with one-line ART starting at CD4 < 200 cells per microliter (μl). Starting ART earlier with CD4 < 350 cells/ μl further increased life expectancy. With the addition of a second line regimen, the life expectancy of all ART strategies increased.

Figures 1-6 provide a sensitivity analysis⁵ of the trigger value $\frac{h_2^*}{h_1^*}$ with respect to the following parameters of the model: σ_1 , σ_2 , α_1 , α_2 , ρ , μ , and θ . The simulations carried out on the critical values of relative health levels confirm the results of the comparative statics discussed above. Figure (1) reveals that the trigger value is much more sensitive to σ_2 than to σ_1 . This is due to the fact that the higher the risk of the second-line regimen, the higher the risk of treatment switch, and thus the higher the threshold in order to trigger the use of the second-line treatment.

(Insert Figure 1 here)

Figures (2) and (3) illustrate that the positive influence of higher α_1 and α_2 , respectively, on the critical value strengthens as σ_2 increases. Both figures confirm that the healthier the patient becomes, the more exposed to uncertainty she is, making the trigger value to increase.

(Insert Figure 2 here)

(Insert Figure 3 here)

Figure (4) shows that the trigger value rises when σ_2 is high and ρ moves toward minus 1, since this combination entails the highest increase in uncertainty after using the second-line treatment.

(Insert Figure 4 here)

⁴ See Appendix for data description.

⁵ The parameters are calibrated with the values shown in Appendix.

Figure (5) illustrates the impact of σ_2 and μ on the critical value of treatment switch, showing that the dampening influence of higher μ on the critical ratio strengthens as σ_2 increases. Figure (6) shows that if decision makers are risk averse, they will switch treatment only for high trigger values.

(Insert Figure 5 here)

(Insert Figure 6 here)

Figures (7) and (8) illustrate, respectively, the impact of σ_2 and $\frac{h_2}{h_1}$ on the expected probability of treatment switch and on the probability of optimal treatment switch, as given by Equations (10) and (11)⁶.

(Insert Figure 7 here)

(Insert Figure 8 here)

Figure (7) illustrates the impact of σ_2 and $\frac{h_2}{h_1}$ on the probability of optimal treatment switch. When $(\alpha_2 - \alpha_1) > (\sigma_1^2 - \sigma_2^2)$, the probability that $\frac{h_2^*}{h_1^*}$ will be hit in the future is increasing in $\frac{h_2}{h_1}$ and decreasing in σ_2 . This is because, first, the higher σ_2 , the lower is the drift of the process $\frac{h_2}{h_1}$, and second, the higher is $\frac{h_2}{h_1}$, the more likely is that process to be “thrown off-course” by a sequence of positive shocks toward the optimality threshold.

Figure (8) simulates the effect of changes in σ_2 and $\frac{h_2}{h_1}$ on the expected time for optimal treatment switch. Our simulation benchmark values are expressed in annual terms so that the simulations for the expected time for optimal treatment switch can be read in years. It shows that the lower is σ_2 and the higher is $\frac{h_2}{h_1}$, the sooner treatment switch is expected to occur. However, the further away is $\frac{h_2}{h_1}$ from the trigger value, the greater is the impact of an increase of σ_2 on the delay expected before treatment switch becomes optimal.

The empirical results suggest that as ART second-line regimen becomes less efficacious, i.e. CD4 count or life years expectancy decreases, the health uncertainty becomes more dominant. The existence of an option value means that ART first-line regimen may be the better choice when considering lifetime welfare. Thus, under

⁶ The parameters are calibrated with the values shown in Appendix.

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.

5 Conclusion

Our stochastic dynamic model of sequential therapeutic regimes underlines the importance of expectations about efficacy in current and future therapies. The fear that diffusion of ART may spread viral resistance tends to become the most powerful argument in favor of limiting or delaying access to second-line antiretroviral treatment. At the empirical level, there is evidence from pilot experiments in India and African countries that a high proportion of patients on second-line antiretroviral therapy were reported to be failing virologically, with most failures occurring within the first six months after initiation of second-line therapy (Ajose, Mookerjee, Mills, Boulle and Ford, 2012).

Rates of virological failure in second-line therapy are high in resource limited settings and associated with duration of exposure to previous drug regimens and poor adherence. The main concern appears to be poor adherence, rather than drug resistance, from the limited number of studies addressing both factors (Ajose et al., 2012). At the same time, we should not lose sight of the fact that not all second-line virological failures are due to poor adherence, and investigate drug resistance due to duration of exposure to first-line regimen.

Our study indicates several directions for future research. Hontelez et al. (2011)'s study provides a strong argument in favor of immediately adopting the new WHO treatment guidelines. Current WHO guidelines recommend that patients failing virologically be subject to an adherence support intervention, after which a second viral load test should be performed prior to deciding on a regimen change. Future studies should be encouraged to report the results of both the first and second viral load, and the duration of exposure to each regimen, in order to better quantify the proportion of virological failures due to duration of time on treatment. Recent costing studies have concluded that when the benefits of guided regimen switches are considered, viral load monitoring is found to be cost effective and life saving (Ajose et al., 2012). Improving the feasibility and reducing the cost of viral load monitoring are important policy objectives (Keebler et al, 2014).

HIV Disease progression is characterized by decreases in CD4 cell count, and CD4 cell count determines the incidence of opportunistic infections and mortality

resulting from HIV disease. In our stochastic model, health level volatility can be understood as the variation in CD4 cell count and the trend of health level as increases in CD4 cell count and, thus, decreases in the incidence of opportunistic infections and mortality. A main conclusion of our model is that, the switching to second-line regimens will further increase survival, but the switching effectiveness depends on the efficacy of the regimens, on the treatment risk, and degree of irreversibility. These conclusions may help clarify current inconsistencies between recommendations and practical behaviors of HIV/AIDS clinicians and public health experts on the one hand, and the expressed set of preferences and expectations of these same decision-makers, on the other.

The study also has several limitations. Although we used data from an earlier published model, time series data were unavailable for viral load monitoring. However, we examined the model parameters in sensitivity analyses, and found that treatment switch depends negatively on the volatility of patients' health, and on trend of health, i.e. the greater the variation in CD4 counts and the more CD4 counts increase, the fewer treatment switches one expects to occur. Treatment switch also depends negatively on the degree of irreversibility. 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. The existence of an option value means that ART first-line regimen may be the better choice when considering lifetime welfare. Conversely, treatment switch depends positively on the discount rate and on the correlation between the patient's health under first and second-line treatments. This means that treatment switch is likelier to succeed in second-line treatments that are similar to the first-line treatments, implying that a decision maker should not rely on treatment switch as a risk diversification tool.

Our model focus on the impact of the decision to switch treatment in a population that is already HIV-infected. From a public health perspective, negative externalities associated with the diffusion of resistant HIV-strains into the rest of the population, are important factors. In India, antiretroviral therapy will lead to major survival benefits and is cost effective by World Health Organization criteria (Freedberg et al., 2007). Any decisions about stopping therapy in patients who are deriving minimal, but some, benefit should be part of a wider discussion of the strategic use of resources for HIV prevention and treatment by persons living with HIV/AIDS, policy makers, clinicians, and public health officials in India.

References

- Aitchison, J. and Brown, J. (1957), *The Lognormal Distribution*, Cambridge: Cambridge University Press.
- Ajose, O., Mookerjee, S., Mills, E. J., Boule, A. and Ford, N. (2012), “Treatment outcomes of patients on second-line antiretroviral therapy in resource-limited settings: a systematic review and meta-analysis”, *AIDS* 261, 929-38.
- Claxton, K. (1999), “The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies”, *Journal of Health Economics* 18, 342-64.
- Claxton, K., Eggington, S., Ginnelly, L., Griffin, S., McCabe, C., Philips, Z., Tappenden, P. and Wailoo, A. (2005), A pilot study of value of information analysis to support research recommendations for NICE., Technical report, Centre for Health Economics, CHE Research Paper 4, University of York.
- Cohen, O. (2000), “Antiretroviral therapy: time to think strategically”, *The Lancet* 32, 320-22.
- de Mello-Sampayo, F. (2014), “The timing and probability of treatment switch under cost uncertainty: an application to patients with GIST”, *Value in Health* 17, 215-22.
- Dixit, A. (1989a), “Entry and exit decisions under uncertainty”, *Journal of Political Economy* 97, 620-638.
- Dixit, A. (1989b), “Hysteresis, import penetration, and exchange rate pass-through”, *Quarterly Journal of Economics* 104, 205-228.
- Dixit, A. (1993), “The art of smooth pasting”, *Fundamentals of Pure and Applied Economics* 55.
- Dixit, A. and Pindyck, R. (1994), *Investment under Uncertainty*, Princeton: Princeton University Press.
- Freedberg, K. A., Kumarasamy, N., Losina, E., Cecelia, A. J., Scott, C. A., Divi, N. and et al. (2007), “Clinical impact and cost-effectiveness of antiretroviral therapy in India: starting criteria and second-line therapy”, *AIDS* 21, 117-28.
- Gupta, S. B., Pujari, S. N., Joshi, S. R. and Patel, A. K. (2007), “API consensus guidelines for use of antiretroviral therapy in adults (API-ART guidelines). endorsed by the AIDS Society of India”, *Journal of Association of Physicians of India* 54, 57-74.

- Harrington, M. and Carpenter, C. (2000), “Hit HIV-1 hard, but only when necessary”, *The Lancet* 355, 2147-152.
- Lasserre, P., Moatti, J. P. and Soubeyran, A. (2006), “Early initiation of highly active antiretroviral therapies for AIDS: Dynamic choice with endogenous and exogenous learning”, *Journal of Health Economics* 25, 1603-11.
- Levy, M. and Nir, A. R. (2012), “The utility of health and wealth”, *Journal of Health Economics* 31, 379-392.
- Meyer, E. and Rees, R. (2012), “Watchfully waiting: Medical intervention as an optimal investment decision”, *Journal of Health Economics* 31, 349-58.
- Mills, E. J., Bakanda, C., Birungi, J., Chan, K., Ford, N., Cooper, C. L. and et al. (2006), “Population-level effect of HIV on adult mortality and early evidence of reversal after introduction of antiretroviral therapy in Malawi”, *The Lancet* 37, 579-598.
- Mills, E. J., Bakanda, C., Birungi, J., Chan, K., Ford, N., Cooper, C. L. and et al. (2011), “Life expectancy of persons receiving combination antiretroviral therapy in low-income countries: a cohort analysis from Uganda.”, *Annals of Internal Medicine* 155, 209-216.
- Oksendal, B. (2003), *Stochastic Differential Equations: An Introduction with Applications*, Berlin/Heidelberg/New York: Springer.
- Pindyck, R. (1988), “Irreversible investment, capacity choice, and the value of the firm”, *American Economic Review* 78, 969-985.
- Sculpher, M. and Claxton, K. (2005), “Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty - when is there sufficient evidence?”, *Value in Health* 8, 433-46.
- UNAIDS/WHO (2010), Country progress report, India, Technical report, UNAIDS/WHO, Available from <http://www.unaids.org> (accessed May 2013).
- Zivin, J. G. and Neideill, M. (2009), “Medical technology adoption, uncertainty, and irreversibilities: Is a bird in the hand really worth more than two in the bush?”, *Health Economics* 19, 142-53.

Appendix

Proof: (*Proposition 1*)

Since the patient can adjust her consumption according to the present realization and future expectations of health (the state variable), she will do so in order to maximize the following intertemporal utility function:

$$U_{it} = E_h \int_t^\infty \left\{ h_{it} \frac{c^{1-\theta}}{1-\theta} \right\} e^{-\mu(\tau-t)} d\tau, \quad (12)$$

where μ is the risk-free time-discount rate and $i = \{1, 2\}$ denotes the first and second-line of treatment, respectively. It turns out that the patient is an expected utility maximizer. Substituting Equation (5) into Equation (12) one obtains the intertemporal utility function when the patient is constantly optimizing over time:

$$U_{it}^e = E_h \int_0^\infty \frac{h_{it}^{\frac{1}{\theta}}}{1-\theta} e^{-\mu(\tau-t)} d\tau, \quad (13)$$

Bearing in mind that $h_i(t)$ follows a geometric Brownian motion, the properties of the lognormal distribution⁷ can be used to transform Equation (13) into:

$$U_{it}^e = \frac{\frac{h_{it}^{\frac{1}{\theta}}}{1-\theta}}{\mu - \frac{\alpha_i}{\theta} - \sigma_i^2 \frac{1-\theta}{2\theta^2}}, \quad (14)$$

provided that $\mu - \frac{\alpha_i}{\theta} - \sigma_i^2 \frac{1-\theta}{2\theta^2} > 0$, which will be assumed here.

Proof: (*Proposition 2*)

When the patient is using exclusively the first-line of treatment, the decision maker's decision as to whether or not to use the second-line of treatment constitutes an optimal stopping problem for which the relevant Bellman equation is:

$$I(U_1, U_2) = \text{Max} \left\{ U_2 - U_1; \lim_{dt \rightarrow 0} \frac{1}{\mu dt} E[dI(U_1, U_2)] \right\}, \quad (15)$$

where $I(U_1, U_2)$ is the value of the option associated with second-line treatment, $U_2 - U_1$ accounts for the expected agent's value gain that results from using 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 is greater than the first defines the continuation region, in which it is optimal not to exercise the option. In the continuation region the Bellman equation is given by:

$$\mu I(U_1, U_2) = \lim_{dt \rightarrow 0} \frac{1}{dt} E[dI(U_1, U_2)]. \quad (16)$$

Applying Ito's lemma to the RHS of Equation (16) yields the partial differential equation:

⁷ See Aitchison and Brown (1957).

$$\frac{\partial I}{\partial U_1} \alpha_1 U_1 + \frac{\partial I}{\partial U_2} \alpha_2 U_2 + \frac{1}{2} \frac{\partial^2 I}{\partial U_1^2} \sigma_1^2 U_1^2 + \frac{1}{2} \frac{\partial^2 I}{\partial U_2^2} \sigma_2^2 U_2^2 + \frac{\partial^2 I}{\partial U_1 \partial U_2} \rho \sigma_1 \sigma_2 U_1 U_2 - \mu I(U_1, U_2) = 0. \quad (17)$$

The set of boundary conditions that applies to this optimal stopping problem is composed of a value-matching condition,

$$I(U_1, U_2) = U_2^* - U_1^*,$$

and of two smooth-pasting conditions,

$$\frac{\partial I(U_2^*, U_1^*)}{\partial U_2} = 1 \text{ and } \frac{\partial I(U_2^*, U_1^*)}{\partial U_1} = -1.$$

The vector (U_2^*, U_1^*) defines the boundary line that separates the (U_1, U_2) space into a stopping region, where it is optimal to start the second-line treatment and a continuation region, where it is optimal to refrain from switching treatments. The derivation of the option value function $I(U_1, U_2)$ from the partial differential Equation (17) and the corresponding boundary conditions, although possible, are unnecessary. Since the optimal choice regarding switching treatments depends exclusively on the relative value of the utility attained before and after switching treatments has been undertaken, that is, on the ratio $\tilde{u} = (U_2/U_1)$, we can impose homogeneity of degree one of $I(U_1, U_2)$ in (U_1, U_2) ⁸, such that:

$$I(U_1, U_2) = U_1 i(U_2/U_1) = U_1 i(\tilde{u}) \quad (18)$$

This transformation allows one to re-write Equation (17) as a function of \tilde{u} rather than (U_1, U_2) :

$$i(\tilde{u})(\alpha - \mu) + \frac{di(\tilde{u})}{d\tilde{u}} \tilde{u}(\alpha - \mu) + \frac{1}{2} \frac{d^2 i(\tilde{u})}{d\tilde{u}^2} \tilde{u}^2 (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2) = 0 \quad (19)$$

which turns out to be an ordinary differential equation. The corresponding boundary conditions become:

$$i(\tilde{u}) = \tilde{u} - 1, \frac{di(\tilde{u})}{d\tilde{u}} = 1 \text{ and } i(\tilde{u}) - \tilde{u} \frac{di(\tilde{u})}{d\tilde{u}} = -1.$$

To solve the optimal stopping problem given by Equation (19) and the respective boundary conditions, one must search for a solution and test its validity by substituting it into Equation (19). Considering $i(\tilde{u}) = B\tilde{u}^\beta$, one finds out that it constitutes a solution to Equation (19) if and only if it is a root of the following quadratic equation:

$$Q(\beta) = \phi\beta^2 + (\delta - \phi)\beta + \left(\frac{\alpha_1}{\theta} + \frac{(1-\theta)\sigma_1^2}{2\theta^2} - \mu\right) = 0, \quad (20)$$

where $\phi = \frac{\sigma_1^2}{2\theta^2} + \frac{\sigma_2^2}{2\theta^2} - \frac{\rho}{\theta^2} \sigma_1 \sigma_2$ and $\delta = \left(\frac{\alpha_2}{\theta} + \frac{(1-\theta)\sigma_2^2}{2\theta^2}\right) - \left(\frac{\alpha_1}{\theta} + \frac{(1-\theta)\sigma_1^2}{2\theta^2}\right)$.

The roots of $Q(\beta)$ can be shown to be equal to:

⁸ This solution strategy is borrowed from Dixit and Pindyck (1994, p.210).

$$\beta_1, \beta_2 = \frac{\phi - \delta \mp \sqrt{(\phi - \delta)^2 - 4\phi \left(\frac{\alpha_1}{\theta} + \frac{(1-\theta)\sigma_1^2}{2\theta^2} - \mu \right)}}{2\phi}$$
 Since for $-1 < \rho < 1$ the coefficient of β^2 in Equation (20) is positive, $Q(\beta)$ is an upward pointing parabola. Moreover, since $Q(1) = \frac{\alpha_2}{\theta} + \frac{(1-\theta)\sigma_2^2}{2\theta^2} - \mu$ and $Q(0) = \frac{\alpha_1}{\theta} + \frac{(1-\theta)\sigma_1^2}{2\theta^2} - \mu$ are both negative by previous assumptions, it follows that $\beta_1 > 1$ and $\beta_2 < 0$.

The general solution for Equation (20) is then $i(\tilde{u}) = B_1 \tilde{u}^{\beta_1} + B_2 \tilde{u}^{\beta_2}$, which simplifies to $i(\tilde{u}) = B_1 \tilde{u}^{\beta_1}$, since $B_2 = 0$ in order to satisfy the boundary condition⁹, $i(0) = 0$.

Making use of the value-matching and smoothpasting conditions, the expression for the critical utility ratio is obtained and likewise for the constant B_1 as:

$$\tilde{u}^* = \frac{\beta_1}{\beta_1 - 1} \tag{21}$$

$$B_1 = \frac{(\beta_1 - 1)^{\beta_1 - 1}}{\beta_1^{\beta_1}} \tag{22}$$

Now, since $\beta_1 > 1$, Equation (21) implies that $\tilde{u}^* > 1$, meaning that the decision maker will engage in the second-line treatment only if the expected utility after switching treatments exceeds that attained when using the first-line treatment. Equation (21) is the trigger value of demand separating the region in u space where the agent's option of using second-line treatment remains unexercised (i.e. for $\tilde{u} > \tilde{u}^*$) from the one where immediate exercise of that option is perceived as optimal (i.e. for $\tilde{u} \leq \tilde{u}^*$).

Data

The simulations relate to the critical ratio obtained in Equations (7), (9), and (10). These simulations are conducted with reference to a benchmark case.

The values of the parameters considered in the benchmark case, as well as the ranges used in the simulations of the critical ratio, were drawn from Freedberg et al. (2007), who used a detailed HIV simulation model with patient-based data from Chennai, India, to project the clinical outcomes and cost-effectiveness of alternative HIV treatment strategies, and the impact of more expensive second-line therapy. The analysis showed that a single regimen of ART nearly doubled undiscounted life expectancy, from 37.5 life months with no therapy to 70.8 life months with one-line ART starting at CD4 < 200 cells per microliter (μl). Starting ART earlier with CD4 < 350 cells/ μl further increased undiscounted life expectancy to 73.6 life months. With the addition of a second line regimen, the life expectancy of all ART strategies increased. Starting ART at a CD4 count of < 200 cells/ μl provided an undiscounted life expectancy of 101.1. Starting two lines of ART treatment with CD4 < 350 cells/ μl further increased undiscounted life expectancy to 106.5 life months. The input choices of sensitivity analyses are presented in Table 1, below.

⁹ If \tilde{u} reaches zero, the utility will stay at zero thereafter, in which case the treatment switch will have no value.

The parameters from the equations of the critical ratio are defined as:

h_1 : Health level under the first-line treatment is proxied by projected undiscounted life years expectancy with first line of ART regimen, co-trimoxazole prophylaxis in combination with stavudine/lamivudine/nevirapine (Freedberg et al., 2007).

h_2 : Health level under the second-line treatment is proxied by projected undiscounted life years expectancy with second-line protease-inhibitor based regimen (Freedberg et al., 2007).

σ_1 : Volatility under the first-line treatment is proxied by the standard deviation of the natural logarithm of h_1 of the three moments when ART treatment was initiated on the basis of CD4 cell count (200, 250, or 350 cells/ μ l).

σ_2 : Volatility under the second-line treatment is proxied by the standard deviation of the natural logarithm of h_2 of the three moments when ART treatment was initiated on the basis of CD4 cell count (200, 250, or 350 cells/ μ l).

α_1 : Drift under the first-line treatment is proxied by the average of the growth rate of the natural logarithm of h_1 of the three moments when ART treatment was initiated on the basis of CD4 cell count (200, 250, or 350 cells/ μ l).

α_2 : Drift under the second-line treatment is proxied by the average of the growth rate of the natural logarithm of h_2 of the three moments when ART treatment was initiated on the basis of CD4 cell count (200, 250, or 350 cells/ μ l).

ρ : Correlation between both lines of treatment is proxied by the correlation between the natural logarithm of h_1 relative to h_2 of the three moments when ART treatment was initiated on the basis of CD4 cell count (200, 250, or 350 cells/ μ l).

μ : Life years were discounted at a rate of 3% per year, in accordance with Freedberg et al. (2007).

θ : Since the data on the parameter for relative risk aversion are not available, the benchmark value and the range of variation were picked arbitrarily.

Table 1 presents the range as well as the mean of each parameter according to the data set specified above. The baseline values are used to define the benchmark case while the maximum and minimum values bound the range used for the simulations of the critical ratio.

Table 1: Parameter Values

Parameter	Maximum	Minimum	Baseline
σ_1	0.3	0.1	0.25
σ_2	0.5	0.3	0.45
α_1	0.3	0.1	0.2
α_2	0.4	0.1	0.3
ρ	0.9	-0.9	0.5
μ	0.05	0.01	0.03
θ	0.9	0.1	0.5

Figures to be included in the Main Text

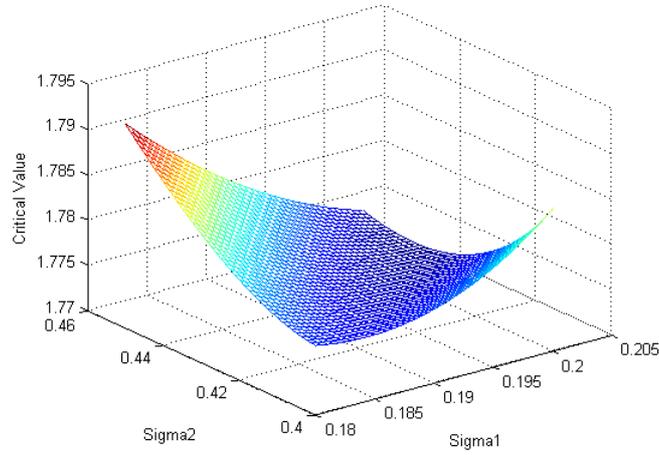


Figure 1. Simulations of the **Critical Value** of relative health levels: h_1/h_2 , where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Sigma1** (σ_1) is the health level's volatility under the first-line ART; **Sigma2** (σ_2) is the health level's volatility under the second-line ART.

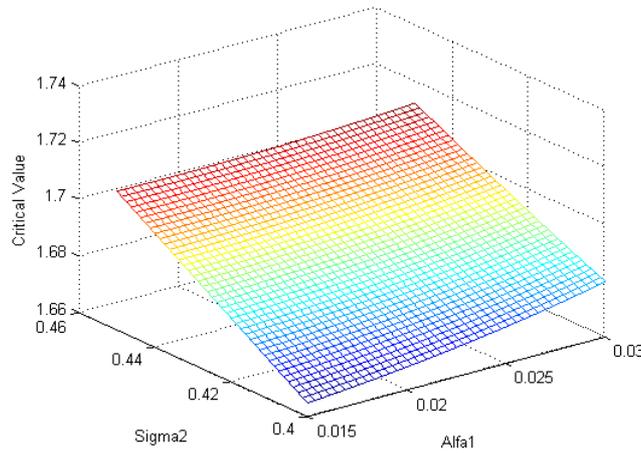


Figure 2. Simulations of the **Critical Value** of relative health levels: h_1/h_2 , where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Alfa1** (α_1) is the trend of the health level, under first-line ART; **Sigma2** (σ_2) is the health level's volatility under the second-line ART.

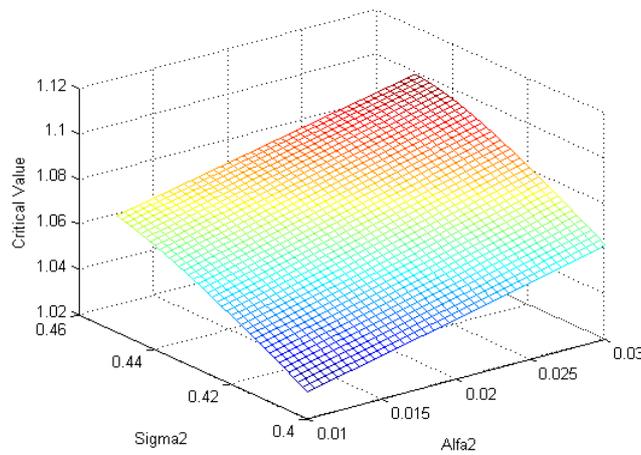


Figure 3. Simulations of the **Critical Value** of relative health levels: h_1/h_2 , where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Alfa1** (α_2) is the trend of the health level, under second-line ART; **Sigma2** (σ_2) is the health level's volatility under the second-line ART.

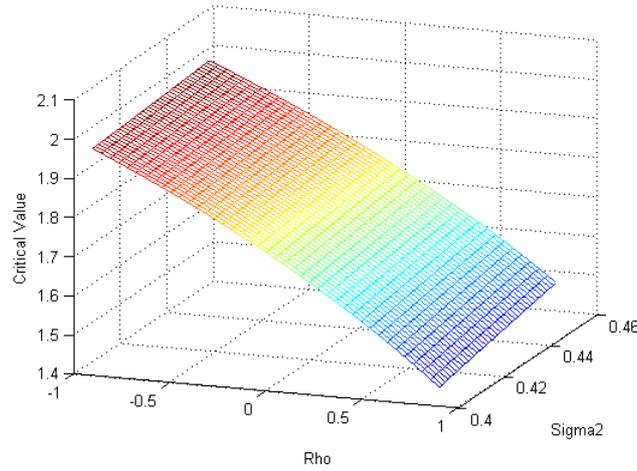


Figure 4. Simulations of the **Critical Value** of relative heath levels: h_1/h_2 , where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Rho** (ρ) is the correlation between heath levels under both lines of treatment; **Sigma2** (σ_2) is the heath level's volatility under the second-line ART.

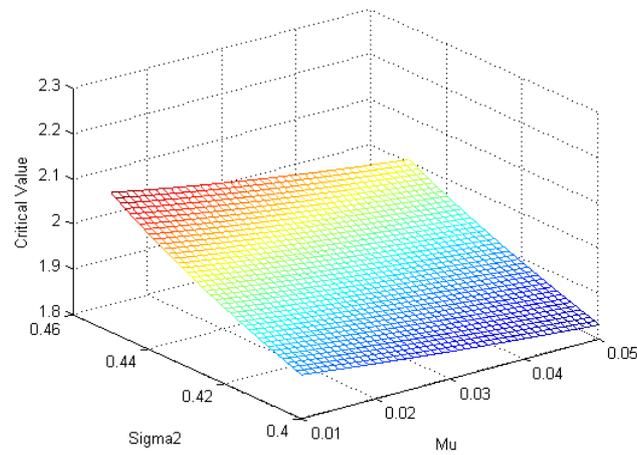


Figure 5. Simulations of the **Critical Value** of relative heath levels: h_1/h_2 , where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Mu** (μ) is the parameter for the discount rate; **Sigma2** (σ_2) is the heath level's volatility under the second-line ART.

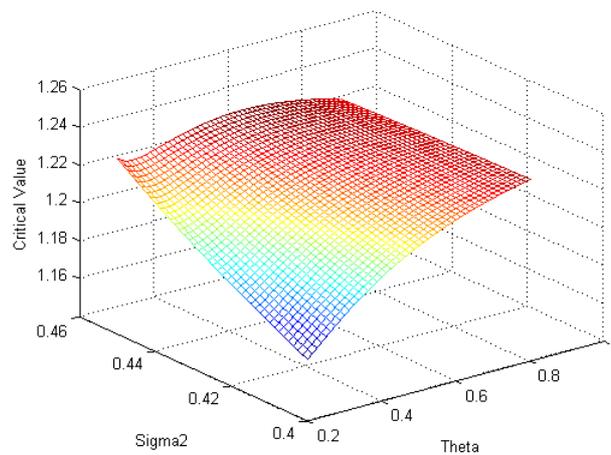


Figure 6. Simulations of the **Critical Value** of relative heath levels: h_1/h_2 , where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Theta** (θ) is the parameter for relative risk aversion; **Sigma2** (σ_2) is the heath level's volatility under the second-line ART.

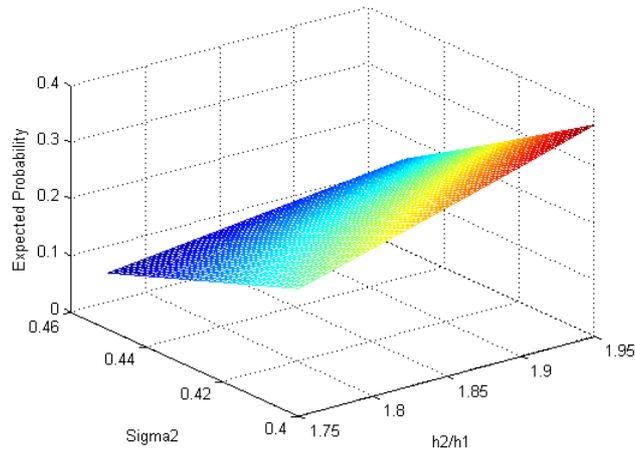


Figure 7. Simulations of the probability of optimal treatment switch. **Expected Probability** of relative heath levels: $Q(h_1/h_2)$, where h_1 is health level under the first-line ART; h_2 is the health level under the second-line ART; **Sigma2** (σ_2) is the heath level's volatility under the second-line ART.

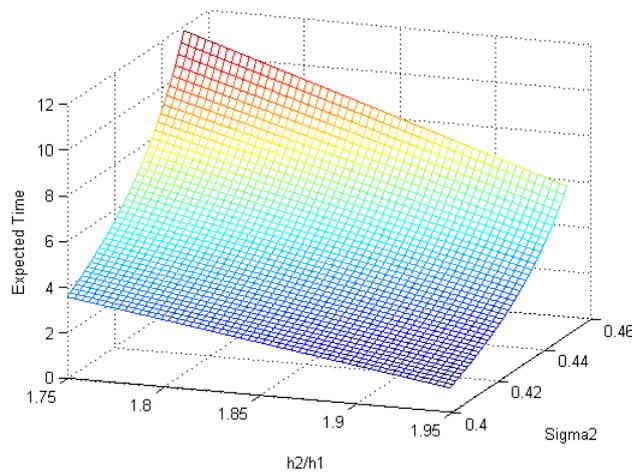


Figure 8. Simulations of the expected time for optimal treatment switch. **Expected Time** of relative heath levels: $T(h_1/h_2)$, where h_1 is health level under the first-line ART; **Sigma2** (σ_2) is the heath level's volatility under the second-line ART.