The Optimal Prevention of Epidemics[1]

Hippolyte d'ALBIS[2]
Toulouse School of Economics (LERNA)

Emmanuelle AUGERAUD-VÉRON[3]
LMA, University of La Rochelle

[1]We are grateful to Jean-Pierre Amigues, Eric Benoît, Raouf Bouccekine, Martial Dupaigne, Albert Marcet and participants in seminars at Universities of Exeter, Paris and Toulouse, at European University Institute and ETH Zurich for their valuable comments and suggestions. The usual disclamer applies.

[2]Correspondence: LERNA, Université de Toulouse I, 21 allée de Brienne, 31000 Toulouse, France. E-mail: hippolyte.d-albis@univ-tlse1.fr

[3]Correspondence: LMA, Université de La Rochelle, Avenue Michel Crépeau, 17042 La Rochelle, France. E-mail: eaugerau@univ-lr.fr

Abstract

This article studies the optimal intertemporal allocation of resources devoted to the prevention of deterministic epidemics that admit an endemic steady-state. In a stylized 'yeoman-farmer' economy, the dynamics of the optimal prevention depends on the interplay between the epidemiological characteristics of the disease, the labour productivity and the intergenerational equity. A minimal level of labour productivity is shown to be necessary to reduce in the long run the prevalence rate of the epidemic. If this threshold is not reached, the prevention is then at best temporary, simply slowing down the spread of the epidemic disease. However, it may not optimal to undertake temporary prevention. Conversely, if labour productivity is sufficiently high, permanent allocation of resources to prevention is feasible but not necessarily optimal. If it is the case, the prevention monotonically increases with time for low initial prevalence rate, while it is decreasing or hump-shaped otherwise. Finally, paths that yield to the eradication of the epidemic disease are considered. Upon existence, such paths are optimal if the pure discount rate is sufficiently low.

JEL Classification: I18, C61

Keywords: Epidemics, Dynamic Optimal Control.

1 Introduction

Epidemic diseases constitute a major health issue for which there is a very large consensus on the legitimacy of governments interventions. Yet barely no analysis has been undertaken to determine the socially optimal allocation of resources to prevention of epidemics. In this article, we derive a welfare criterion from individual preferences and give precise fundations on a generally held belief, according to which if prevention is socially desirable it should begin as soon as possible. In fact, this statement does not hold in general.

Epidemics prevention is a legitimate topic for economics since, as argued by Bloom and Canning [6], there is little doubt that resource constraints play an important role in the spread of epidemics. Moreover diseases importantly affect labor and capital markets and thus growth. However, Gersovitz and Hammer [15] pointed out that it is only recently that economists have entered the field. Most articles adopt a positive approach focusing on private behaviors, such as individual choices on self-exposure to the risk (as in Geoffard and Philipson [17] and Kremer [21]), health expenditure (Momota et al. [24]), human capital accumulation (Bell and Gersbach [4] and Corrigan et al. [10]) or fertility (Young [30]). Some papers analyzes the effect of public policies on private behavior (e.g. Geoffard and Philipson [18]), while others consider optimal policy correcting for the obvious externalities the epidemics generate (e.g. Gersovitz and Hammer [16] or Francis [13]).

Even though only few papers in economics adopt a normative perspective, there exists a rich mathematical epidemiology literature, that dates back from Bernouilli [5]. Following Sethi [27] and Wickwire [29], it is common to use optimal control techniques to define the desirable timing of vaccination, screening or health promotion campaigns. In most cases these studies use as a criterion a convex combination of the dynamic costs of the control and of the number of infected individuals. Moreover, the time horizon is usually finite and, in analytical models, the problem is linear with respect to the control. Based on this approach Behncke [3] finds that the optimal solution is, in general, such that the prevention effort is maximal on some initial time interval and then set to zero. When the case of disease eradication is considered, the problem is more complex since terminal conditions are free (Barrett and Hoel [2]). Some economic studies (in particular Gersovitz and Hammer [15] and Francis [13]) modify the criterion and use the present discounted value of total income net of the costs of the disease and of the control. To conclude we can safely state that most of the literature relies on economic calculus and ignore welfare.

We propose an optimal control model in the tradition of Ramsey [26], Cass [8] and Koopmans [22] in which the whole population is affected by an epidemic disease whose dynamics is general and admits an endemic steady-state. The social welfare function is the present discount value of the product of individual utility and the size of the population. We notably show how this criteria relies on preferences. Introducing population is to avoid the problem stressed by the optimal population literature (see notably Dasgupta [11]) about maximizing the welfare of alive individuals only. Our work thus extends Delfino and Simmons [12], Boucekkine et al. [7] and Goenka and Liu [19], who consider specific epidemic dynamic processes, and Gersovitz and Hammer [16], who proceed by simulations. A second important assumption we make is about the production structure which is of the 'yeoman farmer' type and allow us to completely solve the model despite a general formulation for the epidemic dynamics. The dynamics of the optimal prevention then depends on the interplay between the epidemiological characteristics of the disease, labour productivity and intergenerational equity.

We find that it may be optimal to reduce the prevalence rate of the epidemic in the long run only if labour productivity is above some minimal level. If this threshold is not reached, prevention is then at best temporary, simply slowing down the spread of the epidemic disease. However, it may not be optimal to undertake temporary prevention. When instead labour productivity is sufficiently high, permanent allocation of resources to prevention is feasible though not necessarily optimal. If permanent prevention is socially optimal, the prevention effort monotonically increases with time for low initial prevalence rate, and is hump-shaped or decreasing otherwise. Hence our paper establishes that under a welfare criterion for social intertemporal optimization a "the-sooner-the-better" strategy may not be the optimal one, in contrast to Behncke [3]. This statement is however reversed when we consider paths that yield to the eradication of the epidemic disease. They are characterized by an increasing prevention for a finite interval of time and, once the epidemics is eradicated, the prevention is zero. We show that upon existence, such paths are optimal if the pure discount rate is sufficiently small. In that case, it consequently is socially desirable that prevention should begin as soon as possible.

We begin by presenting the dynamics of the population affected by an epidemic in section 2. The epidemiological assumptions are put forward and discussed using a standard example. In section 3 we set up the social planner's problem, then prove the existence of a solution and characterize it. The dynamics of the optimal prevention is analyzed in section 4. In the following section we ask whether it is socially optimal to eradicate the epidemic disease. We show that, upon existence, eradication is optimal if the pure discount rate is sufficiently low. Section 6 concludes.

2 The population dynamics

This section presents the general characterization of the epidemic dynamics that will be used in the paper. The dynamics is constrained by some general assumptions that are satisfied, for instance, in a classical compartmental epidemic model.

2.1 The general framework

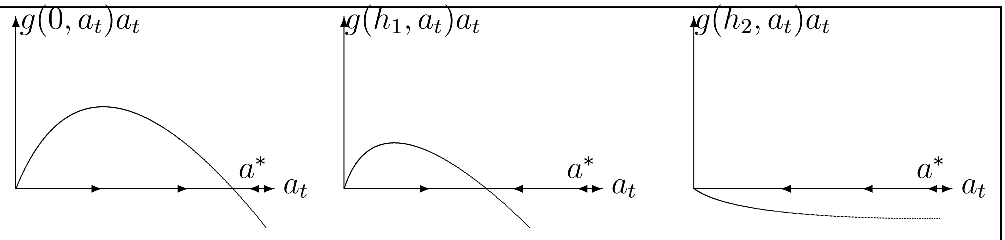
Time is continuous and indexed by $t \in \mathbb{R}^+$. The population at time t, whose size is denoted P_t, is affected by an epidemic disease and is thus decomposed in two classes of individuals: the susceptible, who are healthy, and the infected. The number of individual of each class is respectively denoted $S_t > 0$ and $I_t \geq 0$, and satisfy: $P_t = S_t + I_t$. It will be convenient to define the relative share of infected individuals with respect to share of susceptible ones as follows: $a_t = I_t/S_t$. This ratio is monotonically increasing transformation of the prevalence rate of the epidemic, given by: $a_t/\left(1 + a_t\right)$, and will be named as the prevalence index throughout the remaining of the paper. Moreover, we consider the following general law of motion for a_t:

$$\dot{a}_t = g\left(h_t, a_t\right) a_t, \tag{1}$$

where the dot indicates the first derivative with respect to time and where h_t stands for the per capita expenditures devoted to the epidemic's control. These expenditures will be interpreted as prevention campaigns that are supposed to modify individual's behaviors and consequently to reduce the spread of the epidemic (see, for instance, Castilho [9]). Note that these expenditures can not stand for vaccination or screening expenditures, that yield a third class within the population. The growth rate of the prevalence index, characterized by function g, is supposed to satisfy:

H1. $g : \mathbb{R}_+^2 \to \mathbb{R}$, is C^1, $g_1'(h_t, a_t) < 0$, $g_2'(h_t, a_t) < 0$. Moreover, there exist $a^* > 0$ and $h^* > 0$ such that $g(0, a^*) = g(h^*, 0) = 0$.

The growth rate of the prevalence index is thus supposed to decrease with a_t which, of course, does not imply a monotonic relationship between $\dot{a}_t$ and a_t. In fact, Assumption H1 fully describes the kind of epidemic we are dealing with. If there is no expenditure, the epidemic lasts forever but stabilizes within the population: the prevalence index converges to the steady-state a^*. Prevention campaigns may modify this dynamics. To fix ideas, let h_t be an exogenous constant. If this constant belongs to $(0, h^*)$, the prevalence index converges to another stable steady-state characterized by a lower prevalence rate, and if the constant is larger than h^*, the epidemic vanishes as the prevalence index converges to zero. Figures 1 illustrate such dynamics with three different exogenous h_t: 0, $h_1 \in (0, h^*)$ and $h_2 > h^*$.



Figures 1: The dynamics of a_t

In the remaining of the paper, h_t shall constitute the control optimally chosen by a social planner. Remark that it is easy to extend our approach to other kinds of epidemic dynamics. An alternative dynamics without expenditures could be obtained by assuming: $a^* < 0$. Such an epidemic would not exhibit an endemic steady-state. The unique stable steady-state being the one with a zero prevalence rate. Preventions campaigns may then also be launched to accelerate the convergence process. Similarly, assuming that $a^* \to +\infty$, allow to consider an epidemic whose prevalence monotonically increases if there is no control of it.

The population growth rate writes:

$$\frac{\dot{P}_t}{P_t} = \frac{1}{1 + a_t} \left(\frac{\dot{S}_t}{S_t} + \frac{\dot{I}_t}{I_t} a_t \right). \tag{2}$$

For computational reasons, we assume that the growth rate is characterized by a function denoted $n(.)$ that satisfies:

H2. $\dot{P}_t/P_t = n(a_t)$. $n : \mathbb{R}_+ \to [\underline{n}, n(0)]$, is C^2, $n'(a_t) \leq 0$.

Assuming that the population growth rate does not depend on (S_t, I_t, h_t) is rather strong but is widely used in the epidemiological literature, notably in the example developed below. Moreover, the assumption concerning the sign of $n'(a_t)$ is not only more realistic but also, as it will be discussed throughout the paper, more meaningful.

Our framework generalizes the rare works in economics that have analytically studied the optimal dynamics of an epidemic: Delfino and Simmons [12], Gersovitz and Hammer [15] and Barrett and Hoel [2] study a dynamics similar to (1), but for a function g which is specified, while Boucekkine et al. [7] consider a shock on the initial stock of the population describing an epidemic with instantaneous effects and no endemic steady-state. Our assumptions are now confronted to a two-class version of the widely used SIR model (Kermack and Mac Kendrick [20]).

2.2 An example

Consider first an epidemic dynamics without control. The natural growth rate of the susceptible population is given by $\beta - \mu$, where $\beta > 0$ and $\mu > 0$ respectively stand for the birth and the death rates, while the growth rate of the infected population is $\beta - \mu - \gamma$; parameter $\gamma > 0$ measuring the over-mortality yield by the disease. Both vertical and horizontal contamination are considered: first a proportion $\pi \in [0, 1]$ of the children of infected people are born healthy, while the others are infected. Moreover, as in May and Anderson [23], it is assumed that the incidence of the epidemic follow a law in frequency: contamination is proportional, up to a parameter $\sigma > 0$, to the density of infected individuals in the total population. The dynamics of each subpopulation is therefore given by:

$$\dot{S}_t = (\beta - \mu) S_t + \beta\pi I_t - \sigma S_t \frac{I_t}{S_t + I_t}, \quad (3)$$

$$\dot{I}_t = [\beta(1-\pi) - \mu - \gamma] I_t + \sigma S_t \frac{I_t}{S_t + I_t}. \quad (4)$$

It can be easily shown that this system does not generically admit a steady-state except: $S_t = I_t = 0$. Hence, the dynamics of the epidemic is better understood using the variable a_t defined as: $a_t = I_t/S_t$. Using (3) and (4), the dynamics of a_t solves:

$$\dot{a}_t = [\sigma - \beta\pi - \gamma - \beta\pi a_t] a_t, \quad (5)$$

which is a logistic equation. Therefore, if $\pi > 0$, the dynamics of a_t writes:

$$a_t = \frac{\frac{(\sigma-\beta\pi-\gamma)}{\beta\pi} a_0}{a_0 + \left(\frac{(\sigma-\beta\pi-\gamma)}{\beta\pi} - a_0\right) e^{-(\sigma-\beta\pi-\gamma)t}}. \quad (6)$$

Equation (5) admits (i) two steady-states if $\sigma > \beta\pi + \gamma$: namely, $\hat{a} = 0$, which is unstable and $a^* = (\sigma - \beta\pi - \gamma)/\beta\pi$, which is stable, (ii) one steady-state if $\sigma \leq \beta\pi + \gamma$: namely, $\hat{a} = 0$, which is stable. Consequently, if the contamination coefficient σ is sufficiently low, the epidemic ultimately disappears. Conversely, if σ is high, the epidemic survives as the prevalence index stabilizes. If there is no vertical transmission (i.e. if $\pi = 0$), the only possible steady state is $\hat{a} = 0$, whose stability depends on the sign of $\sigma - \gamma$. Using (3) and (4), the population growth rate $n(a_t)$ solves:

$$n(a_t) = \beta - \mu - \frac{\gamma a_t}{1 + a_t}, \quad (7)$$

which decreases in a_t; moreover, $\beta - \mu - \gamma \leq n(a_t) < \beta - \mu$. The higher bound for $n(.)$ being the population growth rate without epidemic.

Suppose now that some expenditures may affect the epidemic dynamics through the contamination rate. Let h_t be the per capita expenditures and the contamination coefficient at time t be a function that writes $\sigma(h_t)$ and satisfies $\sigma'(h_t) < 0$. The dynamics of a_t is now given by:

$$\dot{a}_t = [\sigma(h_t) - \beta\pi - \gamma - \beta\pi a_t] a_t. \tag{8}$$

It immediate to check that Assumption H1 is satisfied if $\sigma(0) > \beta\pi + \gamma$ and that Assumption H2 is always satisfied.

3 The optimal control problem

This section establishes the optimal control problem we are going to study. It is an infinite horizon framework with an economic structure and the population dynamics described in section 2. We first present and discusses the social welfare function and then prove the existence of an optimal solution.

3.1 The social welfare function

The social welfare function we introduce is derived from the aggregation of individual's preferences. Each individual is supposed to belong to a dynasty of altruistic individuals. Without epidemics, the growth rate of the dynasty is the constant $n(0)$. However, at each point of time, the epidemic disease may kill the dynasty. Denote by λ_t the probability as of time $t = 0$ that the dynasty is still alive at time t. If alive at time t, the utility of a dynasty member depends on consumption c_t and is independent of the health status. The utility function is then $u(c_t)$. If not alive at time t, the utility is supposed to be the constant $u(0) \gg -\infty$. The expected utility of the dynasty at time $t = 0$ is therefore:

$$\int_0^{+\infty} e^{-(\rho - n(0))t} \left[\lambda_t u(c_t) + (1 - \lambda_t) u(0)\right] dt, \tag{9}$$

where ρ is the pure discount rate which satisfies $\rho > n(0)$. Moreover, function u satisfies the following assumption:

H3. $u : \mathbb{R}^+ \to \mathbb{R}^+$, $u \in C^3$, $u' > 0$, $u'' < 0$ and $\lim_{c \to 0} u'(c) = +\infty$.

To obtain the social welfare function, we assume that the population is composed by a continuum of identical dynasties, whose total size at time $t = 0$ is P_0. By the law of large numbers, the probability λ_t is, at the aggregate level, the ratio between the size of the population and the size that would prevail without epidemic. Thus: $\lambda_t = P_t / P_0 e^{n(0)t} = e^{\int_0^t [n(a_s) - n(0)] ds}$. The social welfare function at time $t = 0$ is therefore simply obtained by multiplying the function (9) by the initial size of the population P_0, and rearranging to obtain:

$$P_0 \int_0^{+\infty} e^{-\int_0^t [\rho - n(a_s)] ds} \left[u(c_t) - u(0)\right] dt + \frac{P_0 u(0)}{\rho - n(0)}. \tag{10}$$

Maximizing this latter function is, in fact, equivalent to maximizing:

$$\int_0^\infty e^{-\rho t} P_t u\left(c_t\right) dt. \tag{11}$$

The social planner function is thus the discounted value of the product of the size of the population, P_t, and of the instantaneous utility of each individual. For a given path of consumption, a larger population hence increases the social welfare. Consequently, the assumption $n'\left(a_t\right) < 0$, implies that reducing the number of infected individuals increases welfare, everything being equal.

3.2 The social planner's program

The social planner faces the resource constraint of the economy. There is one material good produced using labor and it is assumed that the productivity of an infected individual is lower than the one of an susceptible individual. Production per capita writes: $\alpha f\left(a_t\right)$ where f is a non increasing function and $\alpha > 0$ is a measure of the productivity of labor. The larger α, the wealthier the economy. For instance, one may consider that the production function is linear with respect to labor, with the productivity being equal to α for susceptible individuals and to $\eta\alpha$ for infected individual (with $0 \leq \eta \leq 1$). Then, in this example: $f\left(a_t\right) = \left(1 + \eta a_t\right) / \left(1 + a_t\right)$.

The produced good can be used for consumption or for the expenditures devoted to the prevention campaign. The resource constraint written in per capita units is therefore:

$$c_t + h_t = \alpha f(a_t). \tag{12}$$

Moreover, consumption, expenditures and the prevalence index should be non negative. The program of the social planner is to maximize (11) subject to (1) and (12). It writes:

$$\begin{aligned} &\max_{h_t} \int_0^\infty e^{-\int_0^t \theta(a_s)ds} u(\alpha f(a_t) - h_t)\, dt, \\ &s.t. \left| \begin{array}{l} \dot{a}_t = g(h_t, a_t)\, a_t, \\ 0 \leq h_t \leq \alpha f(a_t),\ a_t \geq 0 \text{ and } a_0 \in (0, a^*) \text{ given.} \end{array} \right. \end{aligned} \tag{13}$$

where $\theta(a_t) = \rho - n(a_t)$. If there is no epidemic (i.e. if $a_0 = 0$), the problem is trivial: the optimal consumption is constant and equals the production $\alpha f(0)$. Conversely, for $a_0 > 0$, the problem is formally similar to an optimal growth model with endogenous discounting. To reduce the length of the proofs, we solved the case $a_0 \in (0, a^*)$, but the derivation of the case $a_0 > a^*$ is a straightforward extension.

The intertemporal trade-off is the following: an increase in h_t yields a reduction of both the immediate per-capita consumption and the prevalence of the epidemic. The latter implies first an increase in future per-capita production and therefore the prevention campaign can be understood as an investment. Moreover, reducing prevalence leads to a modification of the spread between the discount rate and the population growth rate. For $n'(a_t) < 0$, an increase in h_t implies a reduction of the spread, meaning a more equal treatment between generations as it increases the weight associated to the utility of future generations.

3.3 Characterization of the solution

The program (13) is a non autonomous problem with endogenous discounting but can be equivalently analyzed as a exogenous discounting problem using the virtual time method described by Uzawa [28]. The following results are then derived.

Lemma 1 *There exists an optimal solution to program (13).*

Proof. See the Appendix.

Lemma 2 *An optimal path is necessarily a solution of the following system:*

$$\begin{cases} \dot{a}_t = \frac{g(h_t,a_t)a_t}{\theta(a_t)} \\ \dot{h}_t = \frac{\Phi(h_t,a_t,c_t)}{-\theta(a_t)\left[\frac{u''(\alpha f(a_t)-h_t)}{u'(\alpha f(a_t)-h_t)}+\frac{g''_{11}(h_t,a_t)}{g'_1(h_t,a_t)}\right]} \end{cases} \quad \text{if } h_t > 0, \tag{14}$$

$$\dot{a}_t = \frac{g(0,a_t)a_t}{\theta(a_t)} \quad \text{if } h_t = 0,$$

with $c_t = \alpha f(a_t) - h_t$ *and:*

$$\begin{aligned} \Phi(h_t, a_t, c_t) &= \left[-\alpha f'(a_t)\frac{u''(c_t)}{u'(c_t)} + \frac{g''_{12}(h_t,a_t)}{g'_1(h_t,a_t)} + \frac{\theta'(a_t)}{\theta(a_t)}\right] g(h_t,a_t)\, a_t \\ &\quad + \left[-\alpha f'(a_t) + \frac{u(c_t)}{u'(c_t)}\frac{\theta'(a_t)}{\theta(a_t)}\right] g'_1(h_t,a_t)\, a_t \\ &\quad - g'_2(h_t,a_t)\, a_t + \theta(a_t). \end{aligned} \tag{15}$$

Proof. See the Appendix.

From Lemma 2, it is straightforward to derive the optimal dynamics of consumption. Moreover, while the optimal prevention may be equal to zero, the optimal consumption is always positive, by the assumptions made on the utility function.

4 The dynamics of the optimal prevention

This section studies the system of equations (14) that characterizes the dynamics of the optimal prevention. Paths that indefinitely exhibit a positive prevalence are first studied and then compared those that lead to the eradication of the epidemic. Geometrical illustration using phase diagrams are also provided.

4.1 Steady-states and dynamics

Let us first consider paths that converge to the steady-states of (14). Define a 'corner steady-state' as a steady-state for which the optimal prevention is zero and an 'interior steady-state' as a steady-state for which the optimal prevention is positive.

Given Assumption H1, the pair $(0, a^*)$ is a corner steady-state. It satisfies the following properties:

Lemma 3 *The corner steady-state* $(0, a^*)$ *satisfies* $h_t = 0$ *in its neighborhood and is locally stable.*

Proof. See the Appendix.

Whatever its initial dynamics, the optimal prevention is thus equal to zero after a finite date if the prevalence index converges to the steady-state without prevention. The intuition is that $(0, a^*)$ is not a steady-state for the interior dynamics of system (14). Optimal prevention may hence not converge to zero but only reach zero in a finite time. Moreover, it is never optimal to reach a^* in a finite time. As a consequence $h_t = 0$ in the neighborhood of $(0, a^*)$, which rules out local indeterminacy.

Let us now study interior steady-states, for which prevention is positive. Using (14), such a steady-state is a pair (h, a) that satisfies $h \in (0, \alpha f(a))$, $a \in (0, a^*)$ and solves:

$$g(h, a) = 0, \tag{16}$$

$$-[g_2'(h, a)\,a - \theta(a)] = -\left[-\alpha f'(a) + \frac{u(c)}{u'(c)}\frac{\theta'(a)}{\theta(a)}\right] g_1'(h, a)\,a, \tag{17}$$

where $c = \alpha f(a) - h$. Then, a necessary condition for existence of an interior steady-state is the positivity of equation (17)'s RHS, which rewrites as follows:

$$\frac{d}{da}\left(\frac{u(\alpha f(a) - h)}{\theta(a)}\right) < 0. \tag{18}$$

Condition (18) means that the discounted welfare of a generation in the long run should be increased by a reduction of the epidemic. This condition is always satisfied if a reduction of the epidemic implies a more equal treatment between generations (i.e. if $n'(a) \leq 0$). Otherwise, the increase in utility implied by a marginal decreases of a (measured by $-\alpha f'(a) u'(c) / \theta(a)$) should be larger than the negative impact on the endogenous discount (given by $\theta'(a) u(c) / (\theta(a))^2$). Necessary and sufficient conditions (16) and (17) may be rewritten in the following way:

$$\frac{u'(c)}{\theta(a)} = \frac{d\dot{a}}{dh} \times \frac{d}{da}\left(\frac{u(\alpha f(a) - \eta(a))}{\theta(a)}\right), \tag{19}$$

where $\eta(a)$ is the implicit relation between h and a derived from (16). The contemporaneous desutility yield by a marginal increase in prevention should equal the benefits of the reduction of the epidemic.

The following lemma studies the existence of interior steady-states.

Lemma 4 *i) There exists $\bar{\alpha} > 0$, such that there is no interior steady-states if $\alpha \leq \bar{\alpha}$. ii) There exists $\hat{\alpha} > \bar{\alpha}$, such that there exist interior steady-states if $\alpha \geq \hat{\alpha}$. iii) Upon existence, interior steady-states are locally unstable.*

Proof. See the Appendix.

Lemma 4 shows the importance of labor productivity, or equivalently, of the level of wealth per capita, on the prevalence index in steady-state: there are thresholds below which there is an high prevalence and no prevention, and above which there is lower prevalence with some prevention. The intuition of this result hinges on the concavity and on the Inada condition imposed on the utility function: when the average production is low, resources are exclusively devoted to consumption since a marginal decrease of it yields a large desutility and since the marginal impact of prevention is independent of the level of productivity. Consequently, an interior steady-state is more likely to exist if the labor productivity is increased: the immediate marginal desutility of prevention (i.e. the LHS of (19)) is then lowered while its impact on future generations' discounted utility (i.e. the RHS of (19)) is increased.

Importantly, interior steady-states are locally unstable. As it will be done below, it is possible to characterize, under further conditions, saddle-path steady-states. Hence, from Lemmas 3 and 4, we conclude there are at most two kinds of paths that converge to a steady-state. First, we have a family of paths converging to the corner steady-state $(0, a^*)$. Second, upon existence of a saddle-path steady-state, we have the stable arm. The paths are candidates for optimality.

Suppose there is no interior steady-state. Paths converging to the corner steady-state have been shown to stop prevention campaign after a finite date. According to Lemma 3, we indeed have: $h_t = 0$ in the long run. From this date, the epidemic follows its own dynamics without intervention. The question of optimality among these paths hinges, in fact, on the optimal choice of the date, denoted T, at which the prevention campaign stops. Pursuing the campaign slows down the spread of the epidemic but is costly. To derive the optimal date, we rewrite problem (13) as a problem with free terminal conditions (a_T, T). We derive the following result:

Lemma 5 *If α is sufficiently small or if a_0 is sufficiently large, the optimal path satisfies $h_t = 0$ for all $t \geq 0$. Conversely, if α is sufficiently large and a_0 is sufficiently small, the optimal path may exhibit positive h_t for a finite interval of time.*

Proof. See the Appendix.

Lemma 3 showed that if there is no interior steady-state, the prevention campaigns are at best temporary, slowing down an epidemic that necessarily converges to its endemic steady-state. With Lemma 5, we claim that prevention campaigns are launched if the labor productivity is sufficiently large. Along the considered paths, the growth rate of the discount rates reduces, which increase welfare. However, the dynamics of prevention is decreasing with time, which may implies that consumption is increasing with time if the impact of the prevalence index on the production per capita is low enough. There is few chances that a path that exhibit an increase in both the discount rate and consumption be optimal. Choosing $h_t = 0$ for all $t \geq 0$, implies on the contrary that consumption continuously decreases. Conversely, when labor productivity is large enough, the increase of the prevalence index strongly reduces the production per capita and may imply a consumption path that decline with time. Such a behavior is more likely to be optimal. Lemma 5 also suggest that the initial prevalence and labour productivity play opposite roles. Indeed, production per capita is supposed to decrease with the prevalence rate.

Suppose now there exists at least one saddle-path steady-state. Let us compare the unique path that converges to this steady-state to the family of those which converge to the corner steady-state.

Lemma 6 *Suppose there exists a saddle-path steady-state, denoted by* $(\bar{h}, \bar{a})$*. The stable arm that converges to* $(\bar{h}, \bar{a})$ *may not be optimal.*

Proof. See the Appendix.

Lemma 6 shows that the existence of an interior steady-state does not implies that the stable arm is necessarily optimal. In the proof, we show that using a particular case where the long run cost of the prevention is higher that the benefit in terms of production of having a lower share of infected individuals. Then, there exist sets of initial condition such that the intertemporal utility yield by the stable arm is lower than the one yield by the path converging to the corner steady-state. Conversely, there exists initial conditions, upon which the saddle-path is always preferred to those converging to the corner steady-state.

4.2 A graphical resolution

We now propose a geometrical representation of the results that have been previously established by drawing the phase diagram associated to the system of equations (14). To reduce the length of the proofs, an additional set of restrictions is assumed:

H4. $g''_{11}(h,a) \geq 0$, $g''_{12}(h,a) = 0$, and $-u'''(c)/u''(c) \geq -u''(c)/u'(c)$.

Assumption H4 is sufficient to ensure that function (15) is differentiable everywhere with respect to h_t. The epidemic dynamics considered in Assumption H1 is now constrained by further assumptions on the impact of prevention on $\dot{a}_t$: it is now convex and proportional to a_t. Moreover, the utility function restricts to a representative individual with absolute risk aversion lower than absolute prudence, a property satisfied by standard utility functions including those with harmonic absolute risk aversion.

Lemma 7 *i) The* $\dot{a} = 0$ *locus is downward slopping in the plane* (a, h) *and is such that* $\dot{a} > 0$ *below the locus. ii) The* $\dot{h} = 0$ *locus is defined by a function* $h = \chi(a)$ *that satisfies* $\lim_{a\to 0}\chi(a) = -\infty$ *and is such that* $\dot{h} > 0$ *above the locus. iii) As* α *increases, the* $\dot{h} = 0$ *moves upward in* $\{(a,h) \in \mathbb{R}^2,\ g(h,a) > 0\}$.

<u>Proof.</u> See the Appendix.

Lemma 7 gives enough information to draw the phase diagram associated to system (14). Depending on the existence of interior steady-states, optimal paths can hence be represented. Possible phase diagrams are given by figure 2a if there is no interior steady-states and by figure 2b if there are two interior steady-states.

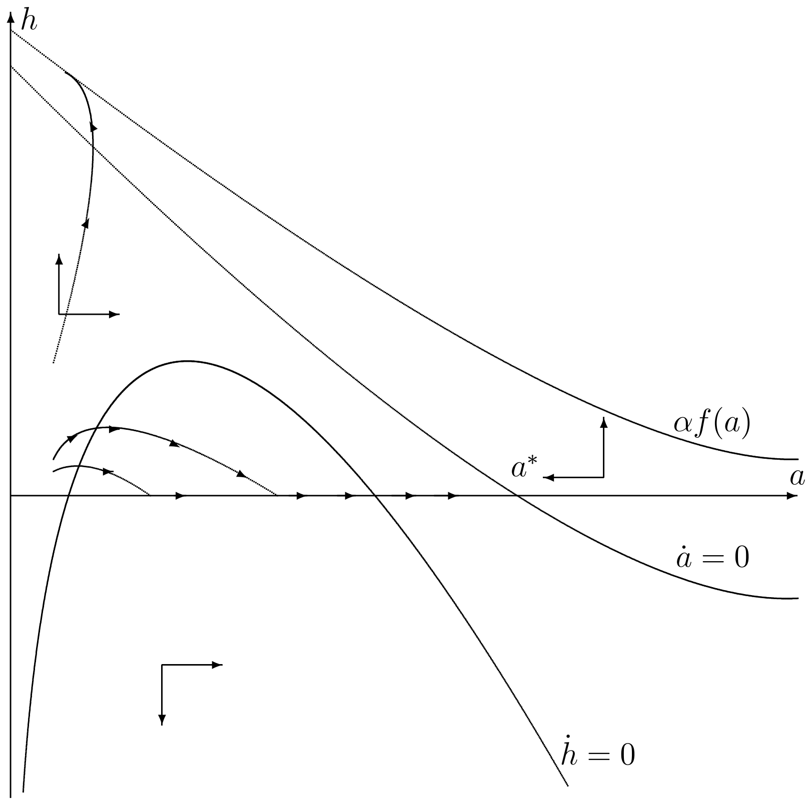


Figure 2a: No interior steady-state

Note that we have represented the upper limit for h, given by function $\alpha f(a)$, above the $\dot{a} = 0$ locus, but it as well can be below for some value of a. For an initial condition $a_0 \in (0, a^*)$, there is hence a family of feasible paths: they converge toward the corner steady-state $(0, a^*)$ and provided that the $\dot{h} = 0$ locus is above the horizontal axis, a positive level of prevention is possible for a finite interval of time. The dynamics of the epidemic may be slowed down for a while but, ultimately, the prevalence index reaches the long-run level with no intervention. Another family of paths is drawn in Figure 2a: they move to the vertical axis with a high level of prevention. These paths are however not feasible since the prevention monotonously increases with time and reaches in finite time the upper limit given by $\alpha f(a)$. The consumption is there equal to zero and, consequently, the path is not optimal.

Labor productivity, which has been shown in the previous section to be crucial for the existence of interior steady-states, has an impact on the dynamics. Geometrically, as α increases, the $\dot{h} = 0$ locus and the constraint $\alpha f(a)$ move up. Since the $\dot{a} = 0$ locus is left unchanged, interior steady-states are more likely to appear, as stated in lemma 4 and drawn in Figure

2b.

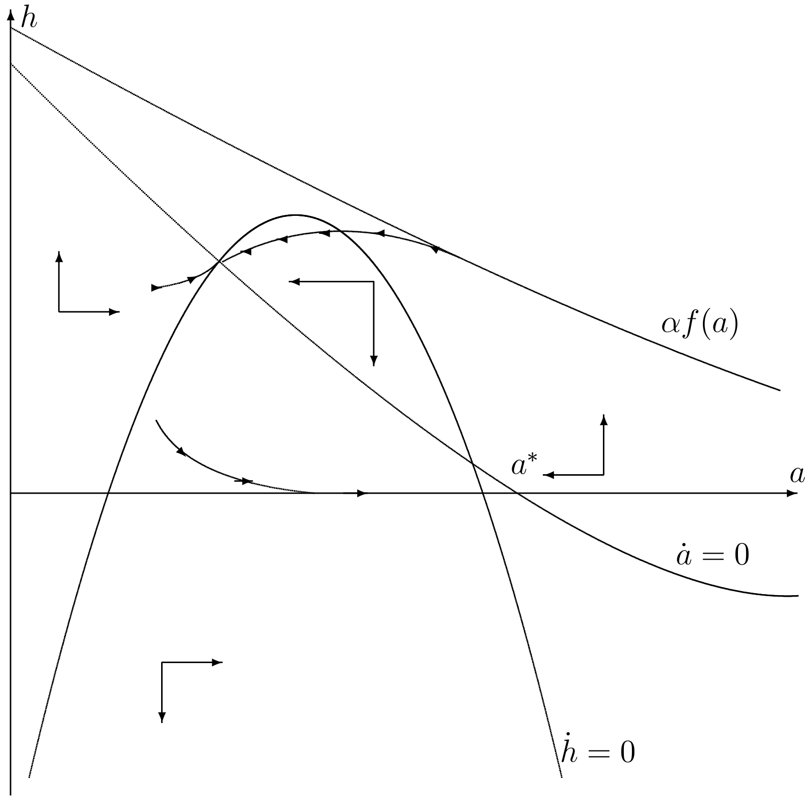


Figure 2b: Two interior steady-states

Two interior steady-states are represented in the phase diagram of Figure 3: a first one with a higher level of prevention expenditures and a lower prevalence, which is saddle-path and a second one which is a repulsive cycle. Hence, in addition to the families of paths that have been considered in the case without interior steady-states, there is a unique path that converges to the saddle-path steady-state. Remark that if there exist more than two interior steady-states, the phase diagram would exhibit alternatively saddle-paths and cycles. If the stable arm converging to the steady-state $\left(\bar{h}, \bar{a}\right)$ is optimal, the prevention monotonically increases with time if $a_0 < \bar{a}$ and can be an hump shaped function of time if $a_0 > \bar{a}$.

In this phase diagram analysis, we have insisted on the fact that on the neighborhood of the vertical axis, the growth rate of the prevention is always positive (i.e. $\dot{h} > 0$). This behavior is induced, using (14), by the assumption of a "not too concave" relationship between the epidemic growth rate and the prevention. More precisely, it is true if:

$$\lim_{h \to 0} \frac{g''_{11}(h,0)}{g'_1(h,0)} < -\frac{u''(\alpha f(0))}{u'(\alpha f(0))}, \tag{20}$$

which is satisfied in Assumption H4. Consequently, any path that may converge to the vertical axis, and thus to the eradication of the epidemic, necessarily reaches in finite time the resource constraint and is therefore not optimal. Remark that assuming the opposite inequality in (20) would not change the statement about the impossibility to eradicate, but simply the phase diagram. In the next section, we propose a slight change in our model that allows to consider the eradication of the epidemic.

4.3 The eradication of the epidemic

To consider the eradication of the epidemic as a possible output of the model, we assume an other dynamics of the prevalence index. We propose to replace equation (1) by:

$$\begin{aligned} &\dot{a}_t = g(h_t, a_t - a_{\min})(a_t - a_{\min}) && \text{if } a_t > 0, \\ &\dot{a}_t = 0 && \text{if } a_t = 0, \end{aligned} \tag{21}$$

where $a_{\min} > 0$. Assumption H1 is left unchanged except for the domain of function g which is now $g : \mathbb{R}_+ \times [-a_{\min}, +\infty[\to \mathbb{R}$. Equations (21) introduce a positive threshold below which the epidemic disappear in a finite time, even without control. Above this threshold, the dynamics has the same qualitative property as the one we have considered before. A minimal prevalence within the population is thus necessary for the epidemic to survive and spread within the population. This threshold may be reach by the immigration of infected individuals, for instance. The dynamics of a_t is thus represented in Figure 3, in the case of no prevention.

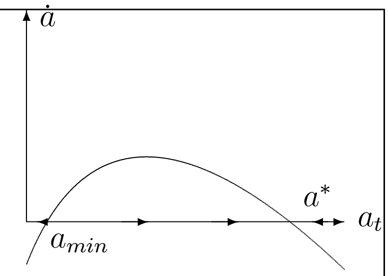


Figure 3: The dynamics of a_t

As we consider a simple translation in function (1), Lemmas 1 to 6 still hold. However, the phase diagram is modified and a new feasible path appears. This path yields to the eradication of the epidemic in finite time. Figure 4a represents a possible phase diagram when there is no interior steady-state.

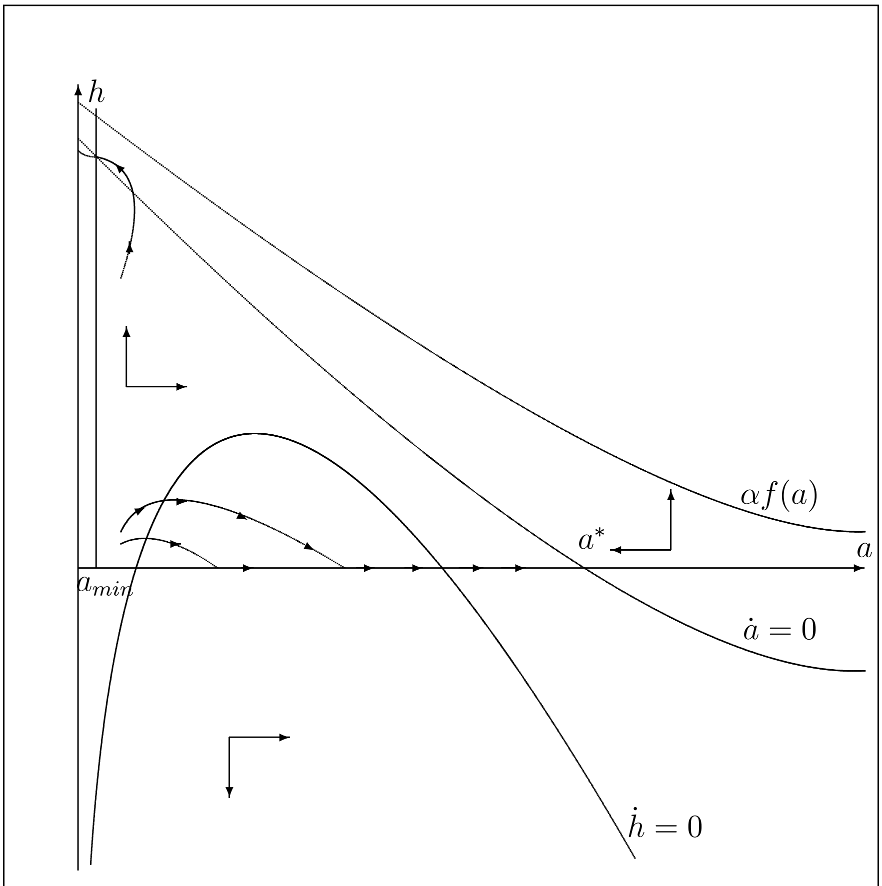


Figure 4a: No interior steady-state

The family of paths that converges to the corner steady-state $(0, a^*)$ is still represented in Figure 4a. Moreover, there is a path that reaches the vertical axis in finite time, along which prevention monotonically increases. Moreover, the path is unique as it necessarily goes through the intersection between the vertical line at $a_{\min}$ and the $\dot{a} = 0$ locus. A necessary condition for the existence of such a path is simply that those coordinates are feasible, meaning that the $\breve{h}$ such that $g\left(\breve{h}, a_{\min}\right) = 0$ satisfies the following condition: $\breve{h} < f(a_{\min})$. A path leading to eradication also appears when there are interior steady-states, as it is shown by Figure 4b.

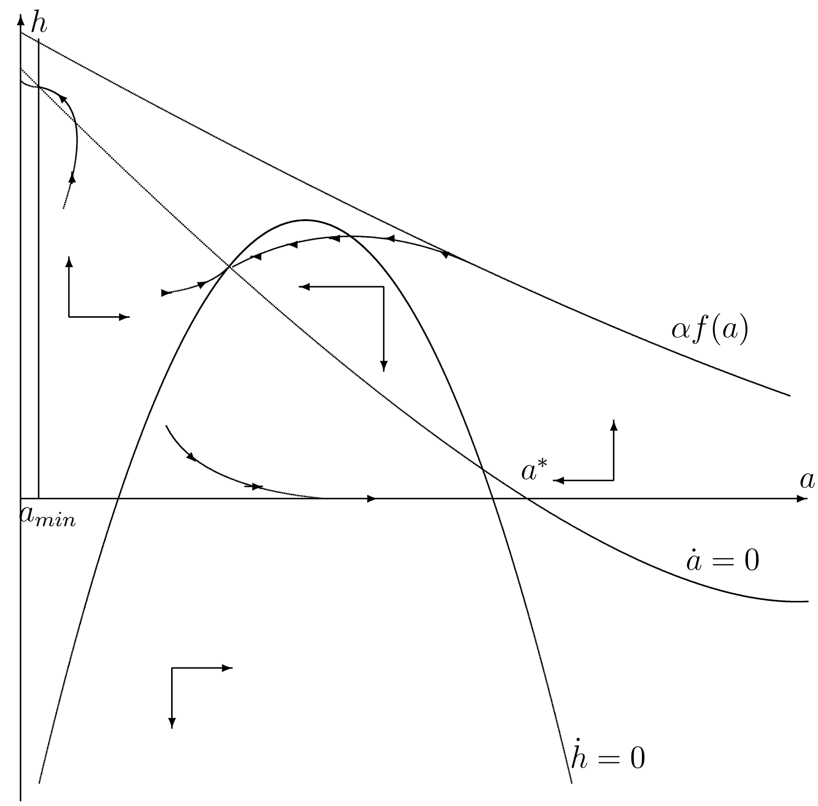


Figure 4b: Two interior steady-states

In the following lemma, we discuss on the optimality of the eradication of the epidemic.

Lemma 8 *Suppose there exists a path that leads to the eradication of the epidemic. There is a threshold for the pure discount rate, denoted $\bar{\rho}$, such that for $\rho < \bar{\rho}$, this path is optimal.*

Proof. See the Appendix.

When eradication is achieved, the discounted utility of generation t, (i.e. $u(c_t)/\theta(a_t)$) is the highest possible since there is no prevention and the prevalence index is zero. This however implies that generations that have lived before the eradication have supported large reduction of their welfare due to the necessarily high levels of expenditures devoted to prevention. The path that yields to eradication is thus optimal if the pure discount rate is sufficiently small.

5 Conclusion

In this article, we have exhibited the relative role of resource constraints and individual preferences on the dynamic of the optimal prevention. Resources constraints are crucial for defining which paths are feasible, while preferences, and notably the discount rate, are used to characterize optimality. In the limit case of no pure discounting, as in Ramsey [26], "the-sooner-the-better" strategy is always optimal, provided there is an minimal threshold for the epidemics.

Possible extensions of the present work may include the traditional decomposition of the population in three classes, a production economy and/or some endogenous growth factors. The size of the dynamic system would then increase making the analytical resolution of the model unlikely. More importantly, future researches should incorporate delay and age structure effects as they are crucial for the HIV/AIDS epidemic. The next step is to incorporate individual behaviors and decentralize the optimum.

Appendix

Proof of Lemma 1. The program (13) is a non autonomous problem with endogenous discounting. Let us rewrite it as an exogenous discounting problem using the virtual time method[1] notably used by Uzawa [28]. This method is based on a change of time scale. First define: $\delta_t = \int_0^t \theta(a_s)\, ds$; since δ is invertible from $\mathbb{R}^+$ to $\mathbb{R}^+$, it is possible to characterize $t = \tau(\delta)$; moreover, $d\delta = \theta(a_t)\, dt$. The social planner's program (13) is then equivalent to:

$$\max_{\tilde{h}_\delta} \int_0^\infty e^{-\delta} \frac{u\left(\alpha f(\tilde{a}_\delta) - \tilde{h}_\delta\right)}{\theta(\tilde{a}_\delta)} d\delta,$$

$$s.t. \left| \begin{array}{l} \frac{d\tilde{a}_\delta}{d\delta} = \frac{g\left(\tilde{h}_\delta, \tilde{a}_\delta\right)\tilde{a}_\delta}{\theta(\tilde{a}_\delta)}, \\ 0 \leq \tilde{h}_\delta \leq \alpha f(\tilde{a}_\delta),\ \tilde{a}_\delta \geq 0 \text{ and } \tilde{a}_0 \in (0, a^*) \text{ given.} \end{array} \right. \quad (22)$$

where $\left(\tilde{a}_\delta, \tilde{h}_\delta\right) \equiv \left(a_{\tau(\delta)}, h_{\tau(\delta)}\right)$. Since there is no ambiguity, we will nevertheless keep the usual notations (a_t, h_t).

The existence problem is then standard. To apply classical results, it is sufficient to show that feasible paths are uniformly bounded. Observe first that the application $a \longmapsto g(h,a)\, a/\theta(a)$ is locally lipschitz and that

$$|\dot{a}| \leq \frac{|g(0,a)\, a|}{\rho - n(0)}. \quad (23)$$

Using Assumption H1, a_t belong to $[0, a^*]$ which is a compact, non empty and invariant set. Moreover since h is bounded, u is bounded and the objective is

[1] Francis and Kompas [14] propose a nice presentation of the method and of its conditions of application.

well defined. Finally since $h \longmapsto u(\alpha f(a) - h)/\theta(a)$ is concave with respect to h, we apply Theorem 3.6 of Balder [1] to conclude. $\square$

<u>Proof of Lemma 2.</u> Let us consider problem (22). Within the domain, the current Hamiltonian writes:

$$\mathcal{H}(h_t, a_t, \mu_t) = \frac{u(\alpha f(a_t) - h_t)}{\theta(a_t)} + \mu_t \frac{g(h_t, a_t) a_t}{\theta(a_t)}, \tag{24}$$

where μ_t is the current costate variable. Necessary conditions are therefore:

$$-u'(\alpha f(a_t) - h_t) + \mu_t g_1'(h_t, a_t) a_t = 0, \tag{25}$$

and:

$$\begin{aligned} \dot{\mu}_t &= -\alpha f'(a_t) \frac{u'(\alpha f(a_t) - h_t)}{\theta(a_t)} + \frac{\theta'(a_t)}{\theta(a)} \frac{u(\alpha f(a_t) - h_t)}{\theta(a)} \\ &\quad -\mu_t \left(\frac{g_2'(h_t, a_t) a_t + g(h_t, a_t)}{\theta(a_t)} + \frac{\theta'(a_t)}{\theta(a_t)} \frac{g(h_t, a_t) a_t}{\theta(a_t)} - 1 \right). \end{aligned} \tag{26}$$

By differentiating (25) with respect to t and using (26), we obtain the expression for $\dot{h}_t$.

Note that a path for which there exists at least one h_t that satisfies $h_t = \alpha f(a_t)$ is not optimal. Define indeed $\hat{h}_t \equiv \alpha f(a_t) - \varepsilon$, then with (24) compute the following:

$$\begin{aligned} &\mathcal{H}\left(a_t, \hat{h}_t, \mu_t\right) - \mathcal{H}(a_t, \alpha f(a_t), \mu_t) \\ = &\frac{u(\varepsilon) - u(0)}{\theta(a_t)} + \mu_t \frac{\left[g\left(\hat{h}_t, a_t\right) - g(\alpha f(a_t), a_t)\right] a_t}{\theta(a_t)}. \end{aligned} \tag{27}$$

Use the Inada condition (Assumption H3) to conclude that the optimal path always satisfies $h_t < \alpha f(a_t)$, which implies $c_t > 0$. □

Proof of Lemma 3. In the neighborhood of $(0, a^*)$ the interior solution of h_t solves:

$$\dot{h}_t\Big|_{(0,a^*)} = \frac{-\left[g_2'(0,a^*)\,a^* - \theta(a^*)\right] + \left[-\alpha f'(a^*) + \frac{u(\alpha f(a^*))}{u'(\alpha f(a^*))}\frac{\theta'(a^*)}{\theta(a^*)}\right] g_1'(0,a^*)\,a^*}{-\theta(a^*)\left[\frac{u''(\alpha f(a^*))}{u'(\alpha f(a^*))} + \frac{g_{11}''(0,a^*)}{g_1'(0,a^*)}\right]}. \tag{28}$$

Thus, $\dot{h}_t$ can be positive or negative, depending on parameters values. If $\dot{h}_t > 0$, it is not optimal to consider a h such that $(h, a) \in V$ and for which h and a both increase. If $\dot{h}_t < 0$, there are two kind of trajectories. The first ones reach $h = 0$ in finite time t_0 and, given (14), satisfy $h_t = 0$ for all $t \geq t_0$. The second ones satisfy the interior dynamic of (14) and the corner conditions $(a_0, a_{t_1} = a_0)$ where $t_1 \ll +\infty$. It can be easily shown that the second kind of trajectories are not optimal. □

Proof of Lemma 4. As a preliminary, use (16) as an implicit equation to define $h = \eta(a)$, which, given Assumption H1, satisfies $\eta'(a) < 0$ and $\eta(a^*) = 0$, and replace it in (17) to define the function $\phi(a, \alpha)$ such that:

$$\begin{aligned} \phi(a,\alpha) &= -\left[g_2'(\eta(a), a)\,a - \theta(a)\right] \\ &\quad + \left[-\alpha f'(a) + \frac{u(\alpha f(a) - \eta(a))}{u'(\alpha f(a) - \eta(a))}\frac{\theta'(a)}{\theta(a)}\right] g_1'(\eta(a), a)\,a. \end{aligned} \tag{29}$$

Function $\phi(a, \alpha) \in C^2(D(\alpha) \times \mathbb{R}^+, \mathbb{R})$ where:

$$D(\alpha) = \left\{a \in \mathbb{R}^+ / \alpha f(a) - \eta(a) > 0\right\}. \tag{30}$$

Then, an interior steady-state is a $a \in (0, a^*)$ such that $\phi(a, \alpha) = 0$.

To prove claim *i)*, use Lemma 2 and Assumption H3 to establish the following limit:

$$\lim_{\alpha \to 0} \phi(a, \alpha) = -\left[g_2'(0, a)\, a - \theta(a)\right] > 0, \tag{31}$$

and conclude using the continuity of $\phi(.,.)$ with respect to α.

Claim *ii)*: since $\alpha \longmapsto \alpha f(a)$ is strictly increasing and since $\lim_{\alpha \to \infty} \alpha f(a) = \infty$, there exists $\acute{\alpha}$ such that $D(\acute{\alpha}) = \mathbb{R}^+$. Suppose $\alpha \geq \acute{\alpha}$. Under Assumptions H1 and H3, $\phi(a, \alpha)$ decreases with α, $\phi(0, \alpha) = \theta(0) > 0$ and:

$$\begin{aligned} \phi(a^*, \alpha) &= -\left[g_2'(0, a^*)\, a^* - \theta(a^*)\right] \\ &\quad + \left[-\alpha f'(a^*) + \frac{u(\alpha f(a^*))}{u'(\alpha f(a^*))} \frac{\theta'(a^*)}{\theta(a^*)}\right] g_1'(0, a^*)\, a^*, \end{aligned} \tag{32}$$

is negative for sufficiently large α.

Claim *iii)*: the stability property is obtained by computing the trace of the Jacobian matrix of system (14) on the neighborhood of any interior steady-state. Since:

$$\left.\frac{\partial \Phi(h_t, a_t)}{\partial h_t}\right|_{\Phi(h_t, a_t) = 0} = 1 - \frac{g_2'(h_t a_t)\, a_t}{\theta(a_t)}, \tag{33}$$

$$\left.\frac{\partial g(h_t, a_t)\, a_t}{\partial a_t}\right|_{g(h_t, a_t) = 0} = \frac{g_2'(h_t a_t)\, a_t}{\theta(a_t)}, \tag{34}$$

the trace is consequently equal to 1. The steady-states are not stable. □

<u>Proof of Lemma 5.</u> The idea of the proof is to incorporate in our problem free terminal conditions. We proceed in three steps. 1/ We define the date

$T \geq 0$ such that the optimal prevention satisfies $h_t = 0$ for all $t \geq T$. Then we characterize some properties of the dynamics of a_t that are useful for the next steps. 2/ We rewrite the social planner problem letting (T, a_T) be choices variables, and derive the necessary conditions for an optimal solution. 3/ Assuming there is no saddle-path steady-state, we characterize the optimal pair (T, a_T) and consequently exhibit the optimal path within the family of candidates that converge to the corner steady-state.

1/ For all $t \geq T \geq 0$, the dynamics of the epidemic satisfies the following differential equation:

$$\dot{a}_t = \frac{g(0, a_t)}{\theta(a_t)} a_t, \tag{35}$$

with $a_T > 0$ given. The general solution of equation (35) is a function of (a_T, t, T), that can be denoted $\eta(a_T, t, T)$. Our first claim is that the solution of (35) can be written as a function of $(a_T, t - T)$ that we denote $\hat{a}(a_T, t - T)$. This claim can be shown by observing that $\eta(a_T, T, T) = \hat{a}(a_T, 0) = a_T$ and that:

$$\frac{\partial \eta(a_T, t, T)}{\partial t} = \frac{\partial \hat{a}(a_T, t - T)}{\partial t}. \tag{36}$$

Then, it is sufficient to apply the property of uniqueness of the solution of the differential equation. Our second claim is that:

$$\frac{\partial \hat{a}(a_T, z)}{\partial a_T} = \frac{g(0, \hat{a}(a_T, z))\, \hat{a}(a_T, z)}{g(0, a_T)\, a_T} \frac{\theta(a_T)}{\theta(\hat{a}(a_T, z))}. \tag{37}$$

To prove that claim, observe that by definition of the dynamics:

$$\frac{\partial \hat{a}(a_T, z)}{\partial z} = \frac{g(0, \hat{a}(a_T, z))}{\theta(\hat{a}(a_T, z))} \hat{a}(a_T, z). \tag{38}$$

Thus:

$$\frac{\partial^2 \hat{a}(a_T, z)}{\partial a_T \partial z} = \frac{\partial \hat{a}(a_T, z)}{\partial a_T} \times \frac{d}{da}\left(\frac{g(0, a)\, a}{\theta(a)}\right)\Bigg|_{a=\hat{a}(a_T, z)}. \tag{39}$$

Integrating the latter equation yields:

$$\frac{\partial \hat{a}(a_T, z)}{\partial a_T} = e^{\int_{a_T}^{\hat{a}(a_T, z)} \frac{d}{da}\left(\frac{g(0,a)a}{\theta(a)}\right)\big|_{a=\hat{a}(a_T, u)} du}. \tag{40}$$

Define $v = \hat{a}(a_T, u)$ and thus $dv = \frac{\partial \hat{a}(a_T, u)}{\partial u} du = \frac{g(0, \hat{a}(a_T, u))}{\theta(\hat{a}(a_T, u))} \hat{a}(a_T, u)\, du$. Apply the change of variables to obtain:

$$\frac{\partial \hat{a}(a_T, z)}{\partial a_T} = e^{\int_{a_T}^{\hat{a}(a_T, z)} \frac{\frac{d}{da}\left(\frac{g(0,a)a}{\theta(a)}\right)\big|_{a=v}}{\frac{g(0,v)v}{\theta(v)}} dv} = e^{\left[\ln \frac{g(0,v)v}{\theta(v)}\right]_{a_T}^{\hat{a}(a_T, z)}}. \tag{41}$$

2/ The problem (22) can be rewritten as follows:

$$\max_{h_t, a_T, T} \int_0^T e^{-t} \frac{u(\alpha f(a_t) - h_t)}{\theta(a_t)} dt + e^{-T} G(a_T),$$

$$s.t. \left| \begin{array}{l} \dot{a}_t = \frac{g(h_t, a_t)}{\theta(a_t)} a_t, \\ 0 \leq h_t \leq \alpha f(a_t), \ a_t \geq 0 \text{ and } a_0 \in (0, a^*) \text{ given}, \\ \chi(a_T, T) = 0. \end{array} \right. \tag{42}$$

where $G(a_T)$ is the continuation value which, using the first claim of step 1, writes:

$$G(a_T) = \int_0^\infty e^{-z} \frac{u(\alpha f(\hat{a}(a_T, z)))}{\theta(\hat{a}(a_T, z))} dz, \tag{43}$$

and where $\chi(a_T, T) = 0$ is a terminal condition that says that a_T is a solution of (42) that satisfies $h_T = 0$.

The Lagrangian $L(a_t, h_t, \mu_t, a_T, T, \lambda) \equiv \mathcal{L}$ writes:

$$\begin{aligned} \mathcal{L} &= \int_0^T e^{-t} \frac{u(\alpha f(a_t) - h_t)}{\theta(a_t)} dt + e^{-T} G(a_T) \\ &+ \int_0^T e^{-t} \mu_t \left(\frac{g(h_t, a_t) a_t}{\theta(a_t)} - \dot{a}_t \right) dt + e^{-T} \lambda \chi(a_T, T), \end{aligned} \tag{44}$$

where μ_t and λ are the current costate variables. Equivalently, we have:

$$\begin{aligned} \mathcal{L} &= \int_0^T e^{-t} \frac{u(\alpha f(a_t) - h_t)}{\theta(a_t)} dt + e^{-T} G(a_T) + \int_0^T e^{-t} \mu_t \frac{g(h_t, a_t)}{\theta(a_t)} a_t dt \\ &+ \int_0^T e^{-t} \dot{\mu}_t a_t dt - \int_0^T e^{-t} \mu_t a_t dt + \mu_0 a_0 - e^{-T} \mu_T a_T + e^{-T} \lambda \chi(a_T, T). \end{aligned} \tag{45}$$

First order conditions are, for all $t \in [0, T]$: (1), (25), (26), and the condition on a_T that write, using the second claim of step 1:

$$\mu_T = \frac{\theta(a_T)}{g(0, a_T) a_T} \int_0^\infty e^{-z} \frac{g(0, \hat{a}(a_T, z)) \hat{a}(a_T, z)}{\theta(\hat{a}(a_T, z))} \eta(a_T, z) dz + \lambda \chi_1'(a_T, T), \tag{46}$$

with:

$$\eta(a_T, z) = \frac{\alpha f'(\hat{a}(a_T, z)) u'(\alpha f(\hat{a}(a_T, z)))}{\theta(\hat{a}(a_T, z))} - \frac{u(\alpha f(\hat{a}(a_T, z))) \theta'(\hat{a}(a_T, z))}{\theta^2(\hat{a}(a_T, z))} \tag{47}$$

Using integration by parts, observe that:

$$\int_0^\infty e^{-z} \frac{g(0, \hat{a}(a_T, z)) \hat{a}(a_T, z)}{\theta(\hat{a}(a_T, z))} \eta(a_T, z) dz = -\frac{u(\alpha f(a_T))}{\theta(a_T)} + G(a_T) \tag{48}$$

and, therefore, (46) rewrites:

$$\mu_T = \frac{\theta(a_T)}{g(0, a_T) a_T} \left[-\frac{u(\alpha f(a_T))}{\theta(a_T)} + G(a_T) \right] + \lambda \chi_1'(a_T, T), \tag{49}$$

Finally, the first order condition on T writes:

$$\begin{aligned} \mathcal{L}'_T &\leq 0 \text{ if } T = 0, \\ \mathcal{L}'_T &= 0 \text{ if } T \in \left]0, \infty\right[, \\ \mathcal{L}'_T &\geq 0 \text{ if } T = \infty, \end{aligned} \tag{50}$$

where $\mathcal{L}'_T$ is the partial derivative of the Lagrangian with respect to T, which writes:

$$\mathcal{L}'_T = e^{-T} \left[\frac{u\left(\alpha f\left(a_T\right)\right)}{\theta\left(a_T\right)} - G\left(a_T\right) + \mu_T \frac{g\left(0, a_T\right)}{\theta\left(a_T\right)} a_T + \lambda \chi'_2\left(a_T, T\right) \right], \tag{51}$$

which, using (25) evaluated at T, and (46), rewrites:

$$\begin{aligned} \mathcal{L}'_T &= e^{-T} \left[\frac{u\left(\alpha f\left(a_T\right)\right)}{\theta\left(a_T\right)} - G\left(a_T\right) + \frac{u'\left(\alpha f\left(a_T\right)\right)}{g'_1\left(0, a_T\right)} \frac{g\left(0, a_T\right)}{\theta\left(a_T\right)} \right] + e^{-T} \frac{\chi'_2\left(a_T, T\right)}{\chi'_1\left(a_T, T\right)} \\ &\quad \times \left[\frac{u'\left(\alpha f\left(a_T\right)\right)}{g'_1\left(0, a_T\right) a_T} + \frac{u\left(\alpha f\left(a_T\right)\right)}{g\left(0, a_T\right) a_T} - \frac{\theta\left(a_T\right)}{g\left(0, a_T\right) a_T} G\left(a_T\right) \right]. \end{aligned} \tag{52}$$

The sign of $\mathcal{L}'_T$ is not immediate.

3/ Suppose there is no interior steady-state. Paths that are candidates for optimality converge to the corner steady-state $(0, a^*)$. Moreover, Lemma 3 has shown that there exists a $T < \infty$, such that $h_t = 0$ for all $t \geq T$. Condition (50) characterizes this optimal T. We thus look at the sign of (52). Remark first that applying the implicit functions theorem implies that:

$$\frac{da_T}{dT} = -\frac{\chi'_2\left(a_T, T\right)}{\chi'_1\left(a_T, T\right)}. \tag{53}$$

Let us now define κ such that:

$$\frac{da_T}{dT} = \frac{g\left(0, a_T\right)}{\theta\left(a_T\right)} a_T - \kappa. \tag{54}$$

We now claim that $\kappa > 0$ if $T > 0$. To prove it, we consider da_t/dt on the neighborhood of a_T for which we know that h_t is small and equal to ε. Hence:

$$\left.\frac{da_t}{dt}\right|_{t=T^-} = \frac{g(\varepsilon, a_t)\, a_t}{\theta(a_t)} = \frac{g(0, a_t)\, a_t}{\theta(a_t)} + \varepsilon \frac{g_1'(0, a_t)\, a_t}{\theta(a_t)} \tag{55}$$

Since the optimal da_t/dt should be lower with positive h_t than without, and since $g_1'(0, a_t) < 0$, we obtain that $\kappa > 0$. Using the same argument, observe that $\kappa = 0$ if $T = 0$. Observe now that (52) rewrites:

$$\mathcal{L}_T' = e^{-T}\kappa \frac{\theta(a_T)}{g(0, a_T)\, a_T}\left[\frac{u'(\alpha f(a_T))}{\theta(a_T)} \frac{g(0, a_T)}{g_1'(0, a_T)} + \frac{u(\alpha f(a_T))}{\theta(a_T)} - G(a_T)\right] \tag{56}$$

Using (43), (50) and (56), we define $\varphi(x, \alpha)$ such that:

$$\varphi(x, \alpha) = \frac{u'(\alpha f(x))}{\theta(x)} \frac{g(0, x)}{g_1'(0, x)} + \frac{u(\alpha f(x))}{\theta(x)} - \int_0^\infty e^{-z} \frac{u(\alpha f(\hat{a}(x, z)))}{\theta(\hat{a}(x, z))} dz, \tag{57}$$

and conclude that the optimal T is such that $T = 0$ if $\varphi(x, \alpha) < 0$ and $T \in (0, \infty)$ if there exist $\hat{x} \in [0, a^*]$ such that $\varphi(\hat{x}, \alpha) = 0$. Since $\hat{a}(0, z) = 0$, one has $\varphi(0, \alpha) < 0$; moreover $\varphi(a^*, \alpha) = 0$ and $\varphi_1'(a^*, \alpha) > 0$. Conclude there are generically a even number of $\hat{x}$ within $(0, a^*)$. Since $\lim_{\alpha \to 0} \varphi(x, \alpha) = -\infty$, $\lim_{\alpha \to +\infty} \varphi(x, \alpha) > 0$ and $\varphi_2'(x, \alpha) > 0$, we conclude that if α is sufficiently large, there exist a_T such that $h_t > 0$ for all $t < T$. Finally, a temporary prevention campaign is then launched if $a_0 < a_T$. □

<u>Proof of Lemma 6.</u> Let us denote by $(\bar{h}, \bar{a})$ the interior steady-state and by

$(0, a^*)$ the corner steady-state. To prove the lemma, we compute the intertemporal utilities for each paths in a particular case.

Suppose that $a_0 = \bar{a} < a^*$. Consider two paths that are candidates for optimality: the first one is given by: $h_t = \bar{h}$ and $a_t = \bar{a}$, and the second one is given by $h_t = 0$ and $a_t = \hat{a}(\bar{a}, t)$ (as defined in the first step of the proof of Lemma 5). The intertemporal utility yield by the first path is:

$$\int_0^{\infty} e^{-t} \frac{u\left(\alpha f(\bar{a}) - \bar{h}\right)}{\theta(\bar{a})} dt = \frac{u\left(\alpha f(\bar{a}) - \bar{h}\right)}{\theta(\bar{a})}. \tag{58}$$

The intertemporal utility yield by the second path is denoted $U(\bar{a})$ and satisfies:

$$U(\bar{a}) = \int_0^{\infty} e^{-t} \frac{u(\alpha f(\hat{a}(\bar{a}, t)))}{\theta(\hat{a}(\bar{a}, t))} dt. \tag{59}$$

Since $U'(\bar{a}) < 0$, conclude that $U(\bar{a}) > U(a^*)$. Hence, the first path is not optimal if:

$$\frac{u\left(\alpha f(\bar{a}) - \bar{h}\right)}{\theta(\bar{a})} < \frac{u(\alpha f(a^*))}{\theta(a^*)}, \tag{60}$$

It is easy to check that the last inequality may not be satisfied for a pair $\left(\bar{h}, \bar{a}\right)$ that satisfies: (16), (17). For instance use the following functions:

$$g(h, a) = e^{-h} - [\beta\pi(1 + a) + \gamma + \delta], \tag{61}$$

$$f(a) = \frac{1 + \lambda a}{1 + a}, \tag{62}$$

$$u(c) = \sigma c^{\frac{1}{\sigma}}, \tag{63}$$

and the following parameters: $\alpha = 1.5$, $\mu = 0.01$, $\gamma = 0.01$, $\pi = 0.9$, $\beta = 0.5$, $\rho = 0.5$, $\lambda = 0.6$, $\sigma = 1/0.69$. $\square$

Proof of Lemma 7. We consider successively the $\dot{a}=0$ locus and the $\dot{h}=0$ locus.

Claim *i)*. The $\dot{a}=0$ locus for all $a>0$ is given by the implicit function $g(h,a)=0$. Given Assumption H1, the locus is downward slopping in the plane (a,h), and is such that $\dot{a}>0$ below the locus and $\dot{a}<0$ above.

Claim *ii)*. Using the definition of $\dot{h}$ in (14) and Assumption H4, the $\dot{h}=0$ locus is given by the function $\psi(h,a,\alpha)=0$ where:

$$\begin{aligned}\psi(h,a,\alpha) &= \left[-\alpha f'(a)\frac{u''(\alpha f(a)-h)}{u'(\alpha f(a)-h)}+\frac{\theta'(a)}{\theta(a)}\right]g(h,a)\,a \\ &\quad+\left[-\alpha f'(a)+\frac{u(\alpha f(a)-h)}{u'(\alpha f(a)-h)}\frac{\theta'(a)}{\theta(a)}\right]g_1'(h,a)\,a \\ &\quad-g_2'(h,a)\,a+\theta(a),\end{aligned} \tag{64}$$

which, using Assumptions H1, H3 and H4, satisfies $\psi_1'(h,a,\alpha)>0$ for all (h,a) below the $\dot{a}=0$ locus (i.e. for $g(h,a)>0$). It is then possible to define the function χ such that $h=\chi(a,\alpha)$, which satisfies: $\lim_{a\to 0}\chi(a,\alpha)=-\infty$. Using (14), compute, on the neighborhood of $(0,0)$, the interior solution of h_t to obtain:

$$\left.\dot{h}_t\right|_{(0,0)}=-\left[\frac{u''(\alpha f(0))}{u'(\alpha f(0))}+\frac{g_{11}''(0,0)}{g_1'(0,0)}\right]^{-1}, \tag{65}$$

which is positive given the convexity of g (Assumption H4). Conclude that $\dot{h}>0$ above the locus and that $\dot{h}<0$ below.

Claim *iii)*. Define $R=\{(a,h)\in\mathbb{R}^2,\ g(h,a)>0\}$. Using Assumptions H1, H3 and H4, we have $\psi_3'(h,a,\alpha)<0$ for all (h,a) such that $g(h,a)>0$. As

$\psi_1'(h, a, \alpha) > 0$ on R, the claim is proved. □

Proof of Lemma 8. A path that leads to eradication in a finite time T can be characterized by $\{h_t^e, a_t^e\}$ for all $t < T$ and by $h_t = a_t = 0$ for all $t \geq 0$. This yields the following intertemporal utility:

$$\int_0^T e^{-t} \frac{u\left(\alpha f\left(a_t^e\right) - h_t^e\right)}{\theta\left(a_t^e\right)} dt + e^{-T} \frac{u\left(\alpha f\left(0\right)\right)}{\theta\left(0\right)}. \tag{66}$$

For $\rho \to n(0)$, one has by definition $\theta(0) \to 0$, and therefore the intertemporal utility is infinite while intertemporal utilities yield by any other path is finite. We conclude by continuity. □

References

[1] E. J. Balder, An existence result for optimal economic problems, Journal of Mathematical Analysis and Applications 95, (1983) 195-213.

[2] S. Barrett and M. Hoel, Optimal disease eradication, working paper, University of Oslo, 2000.

[3] H. Benhcke, Optimal control of deterministic epidemics, Optimal Control Applications and Methods 21, (2000) 269-285.

[4] C. Bell and H. Gersbach, Growth and enduring epidemic diseases, working paper, CER at ETH Zurich, 2007.

[5] D. Bernoulli, Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir, Memoire de mathématiques et de physique, Académie Royale des Sciences (1760) 1-45.

[6] D. E. Bloom and D. Canning, Epidemics and Economics, PGDA working paper No. 9, Harvard School of Public Health.

[7] R. Boucekkine, B. Diene and T. Azomahou, The growth economics of epidemics I: Theoretical considerations, Mathematical Population Studies, forthcoming.

[8] D. Cass, Optimum growth in an aggregative model of capital accumulation, Review of Economic Studies 32 (1965), 233-240.

[9] C. Castilho, Optimal control of an epidemic through educational campaigns, Electronic Journal of Differential Equation 125, (2006) 1-11.

[10] P. Corrigan, G. Glomm and F. Méndez, AIDS crisis and growth, Journal of Development Economics 77, (2005) 107-124.

[11] P. Dasgupta, On the concept of optimal population, Review of Economic Studies 36, (1969) 295-318.

[12] D. Delfino and P. J. Simmons, Positive and normative issues of economic growth with infectious disease, working paper, University of York, 2000.

[13] P. J. Francis, Optimal tax/subsidy combinations for the flu season, Journal of Economic Dynamics and Control 28, (2004) 2037-2054.

[14] J. Francis and T. Kompas, Uzawa's transformation and optimal control problems with variable rates of time preference, working paper, 1998.

[15] M. Gersovitz and J. S. Hammer, The economical control of infectious diseases, Economic Journal 114, (2004) 1-27.

[16] M. Gersovitz and J. S. Hammer, Tax/subsidy policies toward vector-borne infectious diseases, Journal of Public Economics 89, (2005) 647-674.

[17] P.-Y. Geoffard and T. Philipson, Rational epidemics and their public control, International Economic Review 37, (1996) 603-624.

[18] P.-Y. Geoffard and T. Philipson, Disease eradication: Private versus public vaccination, American Economic Review 87, (1997) 222-230.

[19] A. Goenka and L. Lin, Infectious diseases and economic growth, mimeo, National University of Singapore, 2008.

[20] W. O. Kermarck and A. G. Mac Kendrick, A contribution to the mathematical theory of epidemics, Proc. Royal Soc. London 115, (1927) 700-721.

[21] M. Kremer, Integrating behavioral choice into epidemiological models of AIDS, Quarterly Journal of Economics 111, (1996) 549-573.

[22] T. C. Koopmans, On the concept of optimal economic growth, in "The Econometric Approach to Developpement Planning", North Holland, Amsterdam, 1965.

[23] R. M. May and R. M. Anderson, The transmission dynamics of Human Immunodeficiency Virus, in S. A. Levin, T. G. Hallam and L. J. Gross, Eds., Applied Mathematical Ecology, Springer-Verlag, Berlin, 1989.

[24] A. Momota, K. Tabata and K. Futagami, Infectious disease and preventive behavior in an overlapping generations model, Journal of Economic Dynamics and Control 29, (2005) 1673-1700.

[25] R. Naresh, A. Tripathi and S. Omar, Modelling the spread of AIDS epidemic with vertical transmission, Applied Mathematics and Computation 178, (2006) 262-272.

[26] F. P. Ramsey, A mathematical theory of savings, Economic Journal 38, (1928) 543-559.

[27] S. P. Sethi, Optimal control of some simple deterministic epidemic models, The Journal of the Operational Research Society 29, (1978) 129-136.

[28] H. Uzawa, Time preference, the consumption function, and optimal asset holding, in J. N. Wolfe (ed.), Capital and Growth Papers in Honor of Sir John Hicks, The University of Edimburg Press, 485-504, 1968.

[29] K. H. Wickwire, Mathematical models for the control of pests and infectious diseases, A survey. Theoretical Population Biology 11, (1977) 182-238.

[30] A. Young, The gift of the dying: the tragedy of AIDS and the welfare of future African generations, Quarterly Journal of Economics 120, (2005) 423-466.