

Health Care Demand and Impact of Policies in a Congested Public System (Draft - do not quote)

Étienne Gaudette, UQAM

November 15, 2011

Abstract

Increasing health care costs and long waiting times for public care are growing sources of concern for governments in industrialized countries. In this paper, we investigate the impact of waiting times on the demand for care and ask whether allowing congestion to increase in order to contain ever growing health care costs is socially desirable. We develop a stochastic dynamic health care demand model in which agents who differ in terms of age, stock of health capital, wealth and anticipation of future health fluctuations choose whether or not to use a public health care system. We then equate the global health care demand of a distribution of agents to a public supply that is funded through income taxes. Calibrating our model with Quebec data, we find that waiting times act as a weak rationing device and should not be used by welfare-maximizing governments to contain costs. This conclusion follows from the finding that the aggregate demand for care is inelastic with respect to waiting times. On the brighter side, policy simulations reveal that a 50% reduction in Quebec waiting times can be achieved through a variety of fiscal measures without dramatically increasing the costs of care. Such policies would result in major welfare gains, with agents willing to sacrifice over 5% of post-policy consumption flow to move away from the existing *status quo*.

Contact: gaudette.etienne@courrier.uqam.ca.

1 Introduction

Western countries have seen a constant growth of health care spending over the last half century, doubling as a proportion of GDP for over half of OECD members and reaching an average of 8.6% in 2007 (OECD, 2010). While the causes underlying this trend may vary from country to country, it has become a growing source of concern for governments. The public provision of care is a common feature in industrialized nations, with the government of all countries except the United States of America and Mexico assuming over half of the health-related costs.

Simultaneously, waiting times have been increasingly used to contain the demand for health care (Cullis, Jones and Propper, 2000; Gravelle and Siciliani, 2007), and are now considered an important problem in their own right in many industrialized countries. This led to the OECD Project on Waiting Times in the early 2000's, involving 12 countries reporting a high congestion¹, and a number of other studies. Despite these efforts, our understanding of the relationship between congestion, demand for care and costs in public health care systems remains limited and many fundamental questions remain unanswered: Is it desirable to allow waiting times to increase in order to contain ever growing costs of care? Which patients are more likely to be discouraged from using health care as the congestion of the health care system increases? If both the cost

¹These were Australia, Canada, Denmark, Finland, Ireland, Italy, Netherlands, New Zealand, Norway, Spain, Sweden and United Kingdom.

and the congestion of care are high, what type of policy is most beneficial to social welfare and economic activity? To answer these questions, we first develop a theoretical model of health care demand in a public setting characterized by waiting times and zero monetary prices. We then add a macroeconomic equilibrium calibrated with Quebec data in order to quantitatively study the impact of waiting times on consumer demand and predict the expected outcome and desirability of policies aiming at the reduction of congestion and costs of care.

Our paper departs from and adds to the existing literature in three major ways. First, we introduce a health-capital demand model in a dynamic stochastic context that captures key features of public systems. This closes an important gap in the literature. While dynamic health-capital models originated in the 1970's (Grossman, 1972) and have addressed a variety of important questions, the literature has exclusively assumed that health markets were competitive and cleared with monetary prices. This is also the case with the more recent and growing literature which applies numerical methods to estimate the demand of care (Picone, Uribe and Wilson, 1998) and macroeconomic health outcomes (Hall and Jones, 2007; Fonseca et. al., 2009). As noted above, a competitive health care market is at odds with the reality of the vast majority of OECD countries. Even the United States of America's portion of government spending, the lowest among OECD countries, was of 46.5% in 2007 and may rise significantly in the next decade as reforms to increase the generosity of Medicare are being pursued by the Obama administration.

Second, our methodology addresses limitations of the empirical literature on waiting times, which has so far been constrained to the study of waiting lists for specific procedures, such as cataract surgery, hip replacement, and hysterectomy, or groups of similar procedures (Coyte et al., 1994; Anderson, 1997; Martin and Smith, 1999; Martin and Smith, 2003; Siciliani and Hurst, 2004). The interest in these particular procedures is understandable. Since the studied procedures are elective in nature, they generate the longest waiting lists and are thus a natural source of public concern. Two main issues can be raised with such studies. A first issue is that elective surgeries are a limited subset of the quantity of services provided by health care systems, and agents eligible for them are usually similar in terms of age and health. Consequently, these studies are not informative about the general congestion of the system. A second issue concerns the distinction, raised by Lindsay and Feigenbaum (1984), between waiting *queues*, where agents must wait in person at a given location, and waiting *lists*, where agents can use their time freely until receiving the desired good or service in the future. As rationing devices, the latter clear markets through a decrease in the present value of the desired good or service, while the former impose a direct opportunity cost on agents, who are unable to use this time for other purposes. Anecdotal evidence and intuition suggest that waiting time spent in crowded rooms before obtaining care (or being added to a list) is an important component of public care congestion and, plausibly, its most direct rationing mechanism. Some evidence from the medical literature also supports this notion. For instance, a 1990 survey revealed that the number of patients who left Los Angeles County emergency departments without being seen was correlated both with reported waiting times (in queues) and the public nature of facilities (Stock et al., 1994). 7.3% of public hospital patients left without being seen, versus only 2.4% of private hospital patients. Despite this, queues have largely been ignored by the empirical literature. Again, this is understandable, since data on the duration of waiting lists abound, while data on the amount of time between the arrival at a care facility and entry into the waiting list are scarce. Both of these issues are addressed with our approach. We model congestion as reducing the time available for agents to allocate between leisure and consumption, and assess the demand of agents of all age and health profiles in face of the *global* congestion of care.

Third, because of our macroeconomic equilibrium, we take into account general equilibrium effects such as the reaction of workers to differing taxation mechanisms or congestion levels, and are able to predict the

welfare impact of policies.

In our theoretical demand model, we study agents who differ in terms of age, health and financial assets. The depreciation of health capital and the impact of care on health are modeled as stochastic variables that depend on the age and health of agents. Death is modeled as a terminal value that arises when a critically low value of health capital is reached. At any date, agents maximize their discounted lifetime utility by choosing their consumption, working time and whether or not to use the health care system. During a given period, the amount of time available for production and leisure is limited by the health of agents and, if they use health care, the congestion of the public system. In this model, we find that agents choose to consume health care when the *uncertain* expected gain of doing so exceeds the *certain* utility cost endured in the current period, which is directly determined by the congestion of care. The global demand of public care is modeled as the sum of all agents who choose to consume health care, weighted by the unit cost of their health level. Equilibrium arises when congestion is such that the global demand equals the public funds available for health care, obtained from taxes collected from the population of agents.

We devise a calibration strategy that assigns the congestion level, tax rate, and the health law of motion in order to replicate observed health care use, GDP, health care costs, mortality rates and life expectancy. We calibrate our model using Quebec data from the 2005 Canadian Community Health Survey (CCHS). Quebec's case is particularly interesting for two reasons. First, health care expenses represented 11.3% (CIHI, 2008) of its GDP in 2007, which would have placed second behind the United States of America among OECD ratios (had it been a country). Second, indicators and anecdotal evidence suggest that Quebec's health care system has a high congestion. Of the 12% of Quebecers who claimed not to receive the appropriate amount of care in 2005 in the CCHS survey, 47% asserted that waiting times were to blame. As such, Quebec is a natural candidate for public reforms to its health care system *both* to reduce costs and combat the congestion of care.

Using our calibration as a starting point, we simulate the impact on demand of exogenous congestion levels to study elasticity with respect to waiting times, then simulate a quantity of policies by modifying the parameters under the government's control, such as the income tax rate and the generosity of the pension system. Our main findings are that demand (as measured by the number of patients) is inelastic with respect to waiting times, with a simulated elasticity of -0.19 in the neighborhood of the present congestion level. In other words, a 1% increase in waiting times should be expected to reduce the number of users of health care by only 0.2%. When congestion emerges, young, healthy agents are the ones chased away from the system, while agents who stand to benefit the most from health care remain. Since agents with lower health are most expensive to treat, the cost of public care is only marginally reduced by the exit of healthy agents. Thus, congestion is a weak device for containing health-related expenses. Policy simulations indicate that reducing health care funding, which increases congestion, creates large welfare losses as measured by a variety of welfare measures, including the average value of our distribution of agents. The results indicate that a 1% reduction in the tax-to-GDP ratio of Quebec at the expense of health care would create a 260% permanent increase in waiting times, decreasing the welfare of all age and health groups. In contrast, introducing user fees, increasing income tax or decreasing the generosity of other government programs achieve a permanent reduction of congestion (via the increased funding of health care) and lead to important welfare gains.

The remainder of the paper is organized as follows. Section 2 describes the theoretical demand model. Section 3 develops the equilibrium concept and calibration strategy needed to solve our model numerically. Section 4 presents the calibrated demand function and our simulations of exogenous congestion levels and policies. Finally, Section 5 puts our results in a practical policy perspective and suggests future avenues of research.

2 Demand Model

To describe the consumer problem in a public health care setting, we use a discrete dynamic programming approach. The fundamental basis of our demand model is that health care usage is represented as the binary choice α , *i.e.* whether to use ($\alpha = 1$) or not use ($\alpha = 0$) the health care system at a given date. While this discrete-choice approach differs from the literature, the rationale is that the quantity of health care received by a patient is not restricted by the costs of treatment in a public system. Therefore, an individual that has chosen to seek care will usually rely on his physician's diagnosis and accept the treatment chosen for him. As such, the *level* of health care treatment received is irrelevant as a control variable in a public system. Also, the user faces a non-monetary cost associated to the waiting times for health care services. The time spent waiting in queues at diverse stages of health care production cannot be used for leisure or production by agents, and is thus a direct opportunity cost on his utility. To our knowledge, this cost has not yet been modeled in a dynamic health-capital model and constitutes the second major addition of this article to existing literature.

Our model's main concepts and notation draw from Grossman (1972) and Ehrlich and Chuma (1990). At the beginning of a given period, the agent chooses his working time, savings and health care use in order to maximize his expected infinite-horizon discounted utility:

$$\max \sum_{t=0}^{\infty} \beta^t E [u_t (c_t, L_t)], \quad (1)$$

where β is the discounting factor and c_t and L_t are the consumption and leisure *in good health* during the period t . Health does not directly enter the utility function and therefore does not contribute to the *quality* of consumption and leisure. As will be made clear shortly, our conception of time is such that health determines the *quantity* of time available to the agent, which in practice realizes the same effect. We restrict the utility arguments to consumption and leisure to avoid redundancy.

To reflect the impact of waiting times on consumer decisions, the intra-period time distribution is of key importance. A given P -length period is divided as follows:

$$P = LT (H, \alpha\gamma) + AT \quad (2)$$

During the period, the agent faces a certain amount $LT (H, \alpha\gamma)$ of lost time. As mentioned, lost time can either be understood as the literal loss of time due to waiting for health care in a congested facility, or as a reflection of lower quality of time due to the agent being ill, and thus unable to work as productively or enjoy leisure as fully as in full health. In either case, lost time is unavailable for production or leisure during a given period.² Accordingly, LT is decreasing in its first argument, the agent's health level H , meaning that an agent in perfect health faces no lost time, while an agent nearing death loses a much greater portion of the period. If the agent chooses $\alpha = 1$, the second argument is equal to γ , the congestion of the health care system. LT is strictly increasing in this argument, which means that seeking care bears a cost in time for the agent, and that this cost is increasing in the congestion of the public system. What is left of the period, the agent's available time AT , can be freely divided between working time WT and leisure L , as follows:

$$AT = WT + L \quad (3)$$

In effect, the agent's decision regarding time at a given date is limited to his use of health care α and his

²French (2005) uses a similar interpretation of intra-period time distribution.

working time, which jointly determine the remaining leisure time.

The agent's productivity w is considered exogenous to his choices and state variables, and work income is taxed at rate τ in order to fund the public expenses. In addition to his work income, the agent can choose to consume some of his savings A in any period. The amount of savings will not be permitted to be below 0 at any time. For agents in the workforce, it follows that the constraint on consumption c is:

$$c \leq A + w \cdot WT (1 - \tau) \quad (4)$$

In most OECD countries, agents obtain a complex blend of public and private transfers after retirement, depending on large sets of parameters and rules. Since, in this paper, we model agents as self-employed, we reduce the private component of retirement to the choice of savings made by agents during their lifetime: what is left of their savings at the time of their retirement can then be used for consumption, in addition to a fixed exogenous public pension. Accordingly, we model retirement as the following consumption constraint when agents reach the exogenous retirement age a_θ :

$$c \leq A + \theta \text{ if } a \geq a_\theta, \quad (5)$$

where θ is the annual public transfer received by retired agents.³ Applying this constraint, we know that agents will optimally choose $WT = 0$ when they reach age a_θ if we consider a utility function strictly increasing in both consumption and leisure.

Like Grossman, we consider that death occurs when health level H falls below a critical \underline{H} level, after which the agent receives a certain utility \underline{V} . To enforce life to be desirable over death as a basic rule for the remainder of this paper, we assign

$$\underline{V} = u(0,0)/1-\beta \quad (6)$$

For utility functions that are strictly increasing and continuous in both consumption and leisure, this formulation reaches that effect, the value of death being then equal to an infinite sequence of periods in which the agent neither consumes nor enjoys leisure.⁴

With respect to equations (1) to (6), the agent's maximization problem can be written as the following Bellman equation:

$$V(a, H, A) = \begin{cases} \max_{\alpha, WT, c} u(c, L) + \beta E \{V(a', H', A')\} & \text{if } H > \underline{H} \\ \underline{V} & \text{otherwise} \end{cases} \quad (7)$$

$$\text{s.t.} \quad \begin{cases} \alpha \in \{0, 1\} \\ WT \in [0, P - LT(H, \alpha\gamma)] \\ c \in \begin{cases} [0, A + w \cdot WT (1 - \tau)] & \text{if } a < a_\theta \\ [0, A + \theta] & \text{if } a \geq a_\theta \end{cases} \end{cases}$$

³While the study of retirement choices is not a major aim of this paper, including a simple form of retirement in our model is useful for two reasons. First, using Grossman dynamic health care demand models, a vast literature found important dynamic interactions between retirement and health care consumption of agents (Wolf, 1985; French, 2005; Fonseca et al., 2009). Not involving any form of retirement may thus result in severely misguided behavior predictions. Second, it will later enable us to test how modifying pension generosity parameters affects the health care system.

⁴It is worth noting that a variety of common utility functions, among which the Cobb-Douglas, imply $u(0,0) = 0$, and thus yield $\underline{V} = 0$, a seemingly intuitive level for the value of death. However, if the utility function allows for negative values, choosing $\underline{V} = 0$ by default may result in some agents optimally avoiding health care to precipitate death.

It is noteworthy that (7) does not contain the congestion level γ or the tax rate τ as state variables. While they play an active role in the optimization problem, the former is determined by the collective choices of the distribution of agents and the latter is set by the government. By assumption, individual agents cannot affect these parameters or predict their future variations. They are thus considered fixed in the optimization problem. In section 3, we will see how these parameters are determined in the numerical version of the model.

The laws of motion for the agent's three state variables are as follows. First, the movement of the agent's health-capital stock is given by:

$$H' = \begin{cases} H \cdot (1 - \delta(a, \epsilon_\delta) + \alpha\psi(H, \epsilon_\psi)) & \text{if } H > \underline{H} \\ \underline{H} & \text{otherwise} \end{cases} \quad (8)$$

The agent's stock of health-capital decreases at depreciation rate $\delta(a, \epsilon_\delta)$, whose mean is age-dependent. The depreciation rate at a given age is unknown, being affected by the stochastic shock ϵ_δ . Therefore, the agent cannot predict or choose his age of death. This formulation allows for appreciations of health, which happen when δ is negative. If the agent chooses to use health care, he will add ψ to his health. The expected impact of treatment is dependent on the health level of the patient, though the sign of this relationship is debatable.⁵ As with δ , the success of health care in improving the agent's health is an uncertain phenomenon and depends on the realization of shock ϵ_ψ . Finally, an agent whose health has fallen below the death threshold remains deceased in the next period.

Second, the amount of savings in the next period is the total amount of money available to the agent during the period that is not spent on consumption, increased by the real interest rate r :

$$A' = \begin{cases} (1+r)(A + w \cdot WT(1-\tau) - c) & \text{if } a < a_\theta \\ (1+r)(A + \theta - c) & \text{if } a \geq a_\theta \end{cases} \quad (9)$$

Finally, at the end of each period, the agent's age increases by one:

$$a' = a + 1 \quad (10)$$

This constitutes the basic health care demand model used in this paper. Before describing the numerical model and results, we believe it important to briefly analyze the determinants of the choice of α in the theoretical version of the model. Since health care usage is a binary choice in the model, the agent's core problem can be expressed as a comparison of the expected values when using health care or not. Denoting those values V_1 and V_0 , where the indices refer to the choice of α , the agent will choose to use health care if $V_1 > V_0$. This implies, with respect to (7),

$$\beta \cdot (E[V_1'] - E[V_0']) > u(c_0, l_0) - u(c_1, l_1) \quad (11)$$

This inequality can be interpreted as follows: in a public health care system characterized by waiting times, an agent will choose to use care if the *expected dynamic gain* of doing so, on the left-hand side, is larger than the *utility loss* in the current period, on the right-hand side. In this formulation, the gain of using health

⁵As discussed by Chang (1995), on the one hand, it could be easier to treat a patient with a single illness rather than one with multiple illnesses. In this case, the impact of care should be increasing in the level of health. On the other hand, a patient that is suffering from a severe disease, and thus is in very bad health, could receive a greater appreciation of his health capital by having access to care than a patient with a mild disease. In this scenario, the impact of care should be greater with the level of sickness. To our knowledge, no stylized fact currently exists to determine which of these equally plausible interpretations should be retained.

care is uncertain and set in the future, while the cost is immediate and certain. In Appendix A, we further the analytic development of the demand model: we explore both sides of (11) in further detail, clarify the impact of the stochastic terms on death, and find that waiting times are the most plausible and direct lever available to the government to restrict over-consumption of health care. Most interestingly for the remainder of the paper, our analytical development enables us to decompose the expected dynamic gain of health care in two components. Denoting p_α the probability of being alive in the next period conditional on the choice of α , we find:

$$E[V'_1] - E[V'_0] = \underbrace{[p_1 - p_0] \cdot (E[V'_0 | H'_0 > \underline{H}] - V)}_{\text{Life-saving gain}} + \underbrace{p_1 \cdot (E[V'_1 | H'_1 > \underline{H}] - E[V'_0 | H'_0 > \underline{H}])}_{\text{Quality of life gain}} \quad (12)$$

The first component of (12) is the expected gain represented by the difference in probability of being alive in the next period due to the use of health care. When a patient faces a life-threatening illness that can be addressed with a medical treatment, this component will have a high value. In the rest of the paper, we will refer to this component of gain as the *life-saving gain*.⁶ The second component is the difference in the expected value of life in $t + 1$ conditional to the agent being alive. This component will be high in cases where the use of health care increases both the agent's quantity of available time in subsequent periods and his expected longevity excluding the very next period. As mentioned above, an increase in available time can be interpreted as both an increase in the quality of the agent's time or its literal quantity. We will refer to this component as the *quality of life gain*. More discussion on these components is presented in Appendix A.

3 Equilibrium and Calibration

In the remainder of the paper, we aim at numerically answering our research questions, which requires calibrating the model as closely as possible to available data and aggregating the choice of agents to predict the macroeconomic outcome of policies.

Calibrating our model presents two main challenges. First, although there is considerable evidence of congestion in Quebec's health care system, there is to our knowledge no *single* indicator that can be used to effectively calibrate the congestion of the whole system, γ . Waiting times for specific procedures or at the emergency room level, though somewhat informative, are not satisfying measures of congestion, since γ should reflect the congestion of all services offered by the public health care system. Second, the health law of motion (8), while intuitive, is unobservable. The parameters chosen for this function should replicate some observable data, but a strategy is needed to achieve this goal.

Both to obtain a satisfying calibration and the macroeconomic outcome of policies, we add a macroeconomic equilibrium to our basic model. To that effect, we develop a health care supply and a pension system, both funded by income taxes.⁷ We then specify the model's functional forms, basing ourselves on the literature when possible and on a minimal set of intuitions otherwise. Subsequently, we obtain the equilibrium for a distribution of agents consistent with Quebec data found in the Canadian Community Health Survey (CCHS) of 2005. In effect, obtaining the equilibrium reveals the *price* of health care according to the data, which in our case corresponds to the unknown parameter γ . This approach also provides the taxation and

⁶We refrain from naming this component the *life-expectancy gain* because the impact of the use of health care in t on the probability of being alive in $t + n$, $n > 1$, periods, which also affects the difference in life expectancy, is found in the second component.

⁷Thus, we impose a coherence between the funds collected by taxing the distribution's output and the government's expenses. Of course, agents may simultaneously use their savings to supplement the basic public pensions.

wage parameters consistent with Quebec’s GDP and total health care cost for that year. Lastly, since the CCHS data is consistent with Quebec’s population, we are able to choose the health law of motion’s parameters in order to replicate three observed mortality measures as closely as possible. The current section covers these methodological steps and describes our final calibration.

3.1 Health Care Supply and Equilibrium

Obtaining a market equilibrium requires two additional conceptual steps. First, the aggregate demand for health care is a function of the choices to use the system made by all the population’s N individuals, and their level of health, as follows:

$$HC^D = HC^D(\alpha_1, \alpha_2, \dots, \alpha_{N-1}, \alpha_N, H_1, H_2, \dots, H_{N-1}, H_N) \quad (13)$$

The public supply of health care is defined as:

$$HC^S = \sum_{i=1}^N \tau \cdot w \cdot WT_i - \sum_{i=1}^N 1_{a_i \geq a_\theta} \cdot \theta \quad (14)$$

Pensions, like health care, are financed through income taxes. The health care supply described above is determined only by the government’s health care spending, which is equal to the taxes collected minus the cost of public pensions. The rationale behind this formulation is that higher taxes collected for health care should yield a higher quantity of care supplied at an aggregate level. This formulation implies that health care is produced linearly by the public system and presents no fixed cost. This specification deliberately excludes parameters that play an important role in the capacity of a system to supply care, such as the number of hospitals, beds or physicians. While we are conscious that the supply of health care at an aggregate level is a complex phenomenon,⁸ we believe (14) is sufficient for the study of health care demand, the purpose of our paper.

In this setting, the aggregate equilibrium is obtained when, for a given tax rate τ and distribution of agents, the congestion level γ is such that $HC^D = HC^S$. We denote γ^* the congestion level such that the aggregate demand equals the supply of care. As seen in section 2, the congestion parameter directly restrains health care use through the utility cost in the current period and thus lowers HC^D . On the other hand, congestion affects HC^S through two diverging channels. First, congestion discourages health care use. Agents who decide not to use health care have more available time during the period to work, which generates *more* tax revenue. Second, congestion can also lower HC^S , since patients who are not discouraged spend more time waiting in line and have less time available to work, which yields *less* tax revenue. For plausible functional forms and parameters, the marginal impact of congestion is stronger on the demand than on the supply, which enables us to obtain a unique numerical equilibrium.

3.2 Functional Forms

In total, four functional forms need to be specified to solve our problem numerically: three at the consumer demand level and one at the aggregate level. First, we specify the utility function as:

⁸For instance, in the case of Quebec, the different professionals involved in the deliverance of care are all represented by separate unions with long term contracts, restricting the impact of policymakers on the final supply of care. The specific care given to a patient is determined by an independent institution, the Collège des médecins du Québec, which also fixes the criteria that determines who can work as a physician in the province. Also, the aggregate cost of health care is often greater than the resources that were budgeted for a given year, yielding deficits.

$$u(c, L) = \frac{(c^{1-\omega} L^\omega)^{1-\rho} - 1}{1-\rho} \quad (15)$$

This utility function is standard in modern macroeconomic models because of two useful properties. It has a constant Arrow-Pratt relative aversion coefficient of ρ and a constant I-period elasticity of substitution between consumption and leisure.

Second, the time lost at a given period is specified in order to respect the criteria discussed in section 2:

$$LT(H, \alpha\gamma) = P \cdot \left(1 - H + \alpha\gamma \left(\frac{H - \underline{H}}{\bar{H} - \underline{H}} \right) \right) \quad (16)$$

As prescribed earlier, this function is decreasing in health, which means that the sicker the agent, the less time is available to be spent for leisure or work. Also, when using care, the agent sacrifices an amount of time, increasing in γ , in waiting lines, and thus always obtains a higher lost time when $\alpha = 1$. Finally, as observed in the Quebec health care system, patients with severe health problems are prioritized, which means that the waiting time portion of LT increases with health. The congestion level is allowed to assume values between zero and one and can be interpreted in this formulation as the proportion of the period spent in waiting lines by an agent with full health. A

Third, the health law of motion, which is a function of current health and age of the agent, as well as both shocks, is chosen as follows:

$$H'(a, H, \epsilon_\delta, \epsilon_\psi) = \begin{cases} H \cdot \left(1 - \left(\frac{a - \underline{a} - \epsilon_\delta}{\bar{a} - \underline{a}} \right)^{b_1} + \alpha \left(\frac{\bar{H} - H}{\bar{H} - \underline{H}} \cdot \frac{(1 + \epsilon_\psi)}{b_2} \right)^{b_3} \right) & \text{if } H > \underline{H} \\ \underline{H} & \text{otherwise} \end{cases} \quad (17)$$

In (17), \underline{a} and \bar{a} are the minimum and maximum ages considered in the calibration, respectively. While not intuitive at first glance, this functional form presents many attractive features. The health capital depreciation rate is increasing in the age of the agent for neutral values of stochastic term ϵ_δ . For ages nearing \bar{a} , even neutral values of the stochastic term yield an important depreciation of health capital. Also, for values of the stochastic term above $a - \underline{a}$, an agent can see an improvement in his health without the use of health care.

The second term, the production of health care, has similar features. We assume that agents in bad health can expect a higher gain from using health care. However, for neutral values of the stochastic term ϵ_ψ , the agent may not expect to regain full health for values of parameter b_2 above unity. Also, parameters b_1 and b_3 enable non-linearity in the depreciation and production of health. In effect, these parameters and the distribution of the stochastic terms are tools that make this law of motion flexible, which we will later use to replicate essential aspects of our data regarding health.

The only function that remains to be specified is the aggregate demand, which is needed to obtain market equilibrium and isolate the congestion parameter γ . We formulate it as a simple linear transformation of the sum of users, weighted by the level of their sickness:

$$HC^D = b_4 \cdot \sum_{i=1}^N \alpha_i \cdot (\bar{H} - H_i), \quad (18)$$

where α_i is the optimal choice of α for agent i resulting from the solution of our demand model. It is noteworthy that this form implies that the cost of treating a given patient is linearly increasing in his level of sickness $(\bar{H} - H_i)$. The only parameter of this function is the marginal cost, in dollars, of sickness. It enables us to equate the demand to the supply, already specified by (14).

Table 1: Portrait of CCHS 2005 Quebec Respondents

Groups	n		Users ^a (Weighted)
	Unweighted	Weighted	
Age groups (average of 46.3, median of 45)			
18 to 39	8 911	2 243 093	47.6%
40 to 64	11 709	2 815 709	53.6%
65 to 79	4 651	783 272	67.1%
80 and +	1 230	189 741	75.5%
Health groups ^b (average of 0.81, median of 0.8)			
0.45 to 0.60	2 345	449 325	75.7%
0.65 to 0.80	14 448	3 227 621	57.9%
0.85 to 1.00	9 708	2 354 870	44.0%
Total	26 501	6 031 816	53.8%

^aRespondents are classified as users of health care (i.e. choosing $\alpha = 1$) if they claimed visiting a specialist at least once and/or if they visited a general practitioner more than once during the year. See Appendix B for more details on this variable.

^bHealth values of the distribution are obtained through manipulations of two qualitative questions about the perceived health of respondents. Values of 0.45 to 0.60 correspond to the “Poor” (lowest) health label, .65 to .80 to “Fair” and “Good” labels, and 0.85 to 1.00 to “Very good” and “Excellent” labels. See Appendix B for more details on this variable.

3.3 Data

The bulk of the data used for our calibration is extracted from cycle 3.1 of the Canadian Community Health Survey (CCHS), conducted in 2005. This survey is conducted annually by Canada’s national statistics agency, Statistics Canada, with the objective of gathering health-related data at precise geographic levels across the country and contains hundreds of variables. Included in the CCHS data are a subset of variables that enable us to establish the age, health level and health care choices of the respondents in 2005. To obtain these, we make a number of data manipulations, described in Appendix B.

In total, 26,501 respondents from the province of Quebec and over the majority age of eighteen are included in our database. Taking into account the weighting variable of the survey, these respondents correspond to 5.98 million Quebecers, or 96.7% of the total adult population of the province for that year, according to ISQ data. The weighting variable was thus re-weighted in order to represent the full adult population. Table 1 presents a summary of the resulting distribution. Health values of 0.45 to 0.60 correspond to the “Poor” (lowest) health label in the the CCHS self-reported health level questions, .65 to .80 to “Fair” and “Good” labels, and 0.85 to 1.00 to “Very good” and “Excellent” (highest) labels.

3.4 Parametric Choices and Calibration

In order to calibrate our model, we fix a minimal number of parameters according to the existing literature, data and intuition. First, we fix $\omega = .67$, a value effective to replicate aggregate choices of working time (Kydland and Prescott, 1982; Hansen and Imrohoroglu, 1992). Second, in opposition to most macroeconomic applications using this function, we limit ourselves to values of $\rho < 1$, in order for the value of death described by equation (6) to be realistic. With $\rho \geq 1$, this function yields $u(0, 0) = -\infty$, resulting in minus-infinite values of death for the agent, rendering their choices at any period irrelevant. By fixing $\rho \in (0, 1)$, we find $\underline{V} = -1/(1-\rho)(1-\beta)$. We conducted a thorough sensitivity analysis of the risk aversion parameter, presented in Appendix F, which led us to reject values close to unity, which would lead to waiting times of over half the period for healthy agents. For all other values within the tested range, we found very similar quantitative

results. Ultimately, we set ρ to the middle of the possible range, 0.5, which results in an equilibrium value of congestion of 0.067. With regards to (16), this value means that agents in perfect health, who wait the longest for care, lose 6.7% of the period in queues before receiving care. At the opposite of the spectrum, agents with the lowest health level only lose 0.6% of the period in lines, as their treatment is prioritized.

The minimum age we consider is the majority age, after which agents are assumed to be in charge of their health related decisions. In practice, we do not impose a maximal age. When agents reach any age over the parameter $\bar{a} = 99$, they simply have the same health expectations for $a + 1$ as agents of that age, which does not impose an automatic death. According to the final calibration of our model, 18 year-olds in perfect health have a 0.70% chance of reaching ages above 99. The health capital vector is set according to the intuition that full health is represented by unity, and the health of death is set above zero to enable our health depreciation function to reach \underline{H} for reasonable combinations of age and ϵ_δ . We also choose a value of \underline{H} such that LT is strictly positive and below P . This prevents undesirable outcomes as, for instance, agents using health care because the induced waiting time of doing so is nil, their lost time being already equal to the period. Finally, we allow for 15 shocks of both stochastic vectors, in order for agents to obtain varied health outcomes.⁹

The retirement and macroeconomic parameters are drawn from multiple sources, as follows. We obtain the total cost of health care for 2005, of 19.0 billion Canadian dollars, by withdrawing the cost of public medication insurance of \$ 2.4B, as reported by the Régie de l'assurance maladie du Quebec, from the total cost of health assumed by the public sector of \$ 21.4B, as reported by the Canadian Institute for Health Information. We use the income-based, market prices annual gross domestic product of 2005 published by Statistics Canada. For public pensions, we base ourselves on two basic components of Canada's Old Age Security Program, the Old Age Security Pension (OASP) and the Guaranteed Income Supplement (GIS). They are both publicly funded and constitute the minimum pension funds available to all retired Canadians of 65 years of age and over. The age of retirement and generosity of pensions, computed as the sum of the OASP and GIS available to agents with no other pension funds in 2005, are obtained from Services Canada. The taxation parameter is set to the ratio between the cost of the public expenses and the GDP:

$$\tau = \frac{HCS + \sum_{i=1}^N 1_{a_i \geq a_\theta} \cdot \theta}{GDP} \quad (19)$$

The remaining parameters are obtained through our calibration algorithm in order to replicate the percentage of users found in the database, several demographic facets, the total cost of health care and the GDP of Quebec. This procedure is explained in Appendix D. The full calibration is presented in Table 2.

4 Numerical Results

4.1 Determinants of Health Care Demand

The demand for health care resulting from our procedure is an optimal choice of α for every combination of health, age and savings. Figure 1 presents the youngest age for which users choose $\alpha = 1$ for all health capital values, which is the optimal threshold between using health care or not. For a given health level, we find that older agents always display a positive net gain of using care, resulting in all agents above the line in the figure to choose $\alpha = 1$. This demand fits the trends observed in Table 1, in which we saw that health

⁹The specific values of these parameters are of little importance. Other reasonable sets of parameters were found to lead to qualitatively identical results after the execution of our calibration algorithm.

Table 2: Calibration

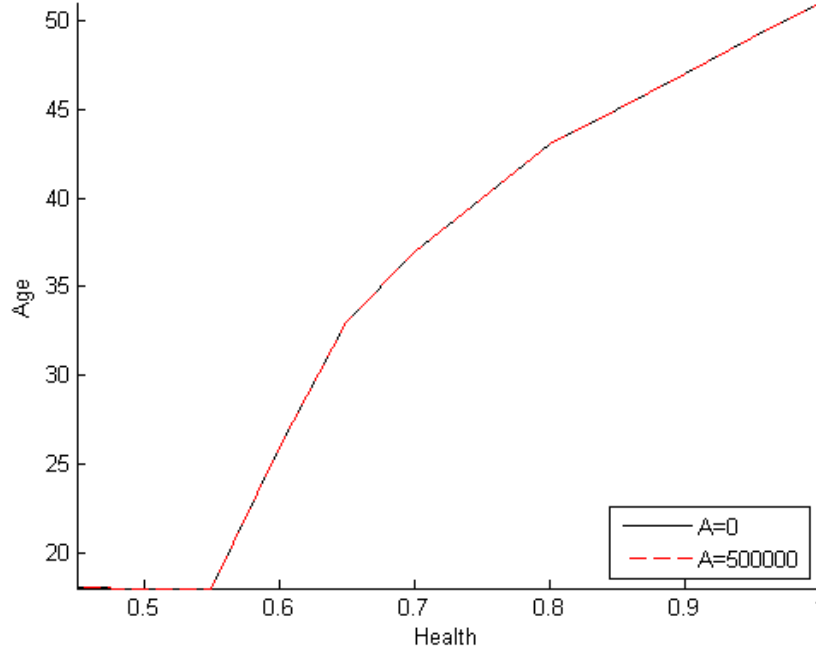
Fixed Parameters	
ω : Leisure preference	0.67
ρ : Arrow-Pratt relative aversion coefficient	0.5
\underline{a} : Minimum age	18
\bar{a} : Maximum age	99
\underline{H} : Minimum health capital (death threshold)	0.4
\bar{H} : Maximum health capital	1.0
\bar{A} : Maximum savings	\$500 000.00
Dimension of vector H	13
Dimension of vector A	51
Dimension of vectors ϵ_δ and ϵ_ψ	15
a_θ : Age of retirement	65
θ : Annual public transfer to retirees	\$10 319.50
GDP	\$ 272 049M
HC^S : Public health spending exc. pharma.	\$ 18 998M
τ : Income tax rate	0.10674
Algorithm-Determined Parameters	
γ^* : Equilibrium congestion rate	0.0671201
w : Hourly productivity	\$ 33.33
b_1 : Health depreciation curve coefficient	1
b_2 : Expected impact of treatment coefficient	15
b_3 : Treatment impact curve coefficient	1
b_4 : Maximum cost of treatment	\$25 562.8
μ_δ : Expected health depreciation shock	58.5
σ_δ : Std. dev. of health depreciation shock	24.1
μ_ψ : Expected treatment shock	-0.55
σ_ψ : Std. dev. of treatment shock	0.109
Standard deviations covered	3

care use was stronger in older and unhealthy individuals in Quebec in 2005.¹⁰

To assess how savings affect health care consumption in a public system, Figure 1 presents the threshold for the lowest and highest savings levels included in our calibration. If wealth was an important decision factor in the demand for care, the figure would display large differences depending on the savings of agents. On the contrary, the youngest age for which agents use care is exactly the same for the poorest and wealthiest agents, and the two lines are indistinguishable. The figure thus reveals that public health care systems are efficient in being equally available to the poor and the rich. While intuitive, this result has several obvious caveats. In our model, wealth differs only through savings and health status, while productivity *in good health* is supposed to be homogeneous and unemployment nonexistent. Also, most workers are not self-employed

¹⁰What our model fails to predict is the occurrence of young, healthy respondents choosing to use care despite long waiting times, and *vice versa*, both of which are intriguingly present in the data. This may occur for a variety of reasons. On the one hand, the distribution's health level is computed from self-declared variables and may lack consistency between different respondents. For instance, some respondents may report feeling in good health while being afflicted by chronic diseases that require continuous care. Similarly, agents may have used health care to respond to a temporary decrease of health level that is untraceable in the data, since we only observe the self-reported health at the end of the period and its yearly variation. That being said, the most important factor behind this discrepancy is plausibly that the real decision framework used by agents is more complex than the one developed in this paper and cannot be replicated perfectly.

Figure 1: Minimum Age for Which Agents Use Health Care



and many of them benefit from a form of income insurance against the use of health care in the form of sick days. Thus, we may incorrectly estimate the difference in usage between the wealthiest and poorest agents.

Turning our attention to the determinants of the choice of health care use, we present in Table 3 the numerical decomposition of the choice of agents, for savings of \$50,000. Different savings values yield very similar numbers to those presented here. We see that the components of expected gain present wide variations among age-health combinations. It is interesting to note that, for all but the lowest health level, the expected gains are higher for agents of 60 years and older. Two factors explain the stronger demand by older agents. First, with our health law of motion, older agents can expect a more severe health depreciation than younger agents and use health care as a prevention measure. Second, the expected gain in longevity from using care is discounted less for older agents than their younger counterparts, resulting in important expected gains of seeking care.

Negative quality of life gains in the presence of positive life-saving gain are found in the table, and both effects are not always correlated. As noted earlier, the quality of life gain component comprises both the increase of available time in future periods and longevity gains past the next period. In Appendix A, we discuss both components of gain in further detail, and explain graphically why they can be uncorrelated depending on the health expectations of agents.

As for the utility loss component, the cost of using health care, we note that its level is small compared to the expected gain for most age-health combinations. This occurs because the risk aversion penalizes agents for death and encourages the use of health care. It is an unsurprising result considering that, according to the CCHS data, 53.8 % of the population chose to use health care in 2005 despite long reported waiting times.

Figure 2 is illustrative of the use of savings by agents in our model. It shows the expected savings of agents by age, calculated from the predicted lives of a theoretical population with our stochastic design (the approach is described in Appendix C). Agents optimally increase their savings until the age of retirement, slowing this trend only in their early 50's, which coincides with the first predicted use of health care for agents

Table 3: Decomposition of Health Care Demand Choice ($A=\$50,000$)

H	a	Expected dynamic gain ^a			Utility loss ^b	Net gain ^c	α^{*d}
		L.-S. Gain	Q. of L. gain	Sum			
0.45	20	10.4	28.0	38.4	0.8	37.6	1
	40	72.7	104.0	177.1	0.8	176.3	1
	60	167.6	-124.2	45.6	0.9	44.7	1
	80	65.4	-41.2	30.1	0.8	29.4	1
	99	11.3	3.3	16.7	0.8	15.9	1
0.55	20	0.0	4.0	4.0	2.3	1.7	1
	40	29.0	-11.8	17.3	2.3	15.1	1
	60	101.4	18.7	121.3	2.7	118.6	1
	80	72.3	34.4	113.0	2.0	111.1	1
	99	23.0	31.8	58.8	4.2	54.6	1
0.65	20	0.0	1.8	1.8	3.6	-1.8	0
	40	12.0	-6.3	5.7	3.6	2.1	1
	60	48.5	-6.5	42.6	4.5	38.1	1
	80	58.3	12.3	75.6	3.4	72.2	1
	99	33.0	21.5	59.9	3.1	56.8	1
0.75	20	0.0	0.4	0.4	4.9	-4.4	0
	40	0.0	13.4	13.4	5.0	8.5	1
	60	0.2	61.9	62.1	6.2	55.9	1
	80	2.4	49.3	51.9	3.5	48.4	1
	99	26.0	-5.1	24.9	3.9	21.0	1
0.85	20	0.0	0.2	0.2	6.1	-5.9	0
	40	0.0	1.5	1.5	6.3	-4.8	0
	60	18.9	2.4	21.5	7.9	13.6	1
	80	38.9	12.9	55.1	4.2	50.9	1
	99	38.6	20.8	65.1	5.8	59.3	1
0.95	20	0.0	0.1	0.1	7.4	-7.4	0
	40	0.0	0.8	0.8	7.6	-6.8	0
	60	0.0	23.8	23.8	9.4	14.5	1
	80	0.0	55.4	55.4	4.8	50.7	1
	99	2.0	24.2	26.5	5.2	21.4	1

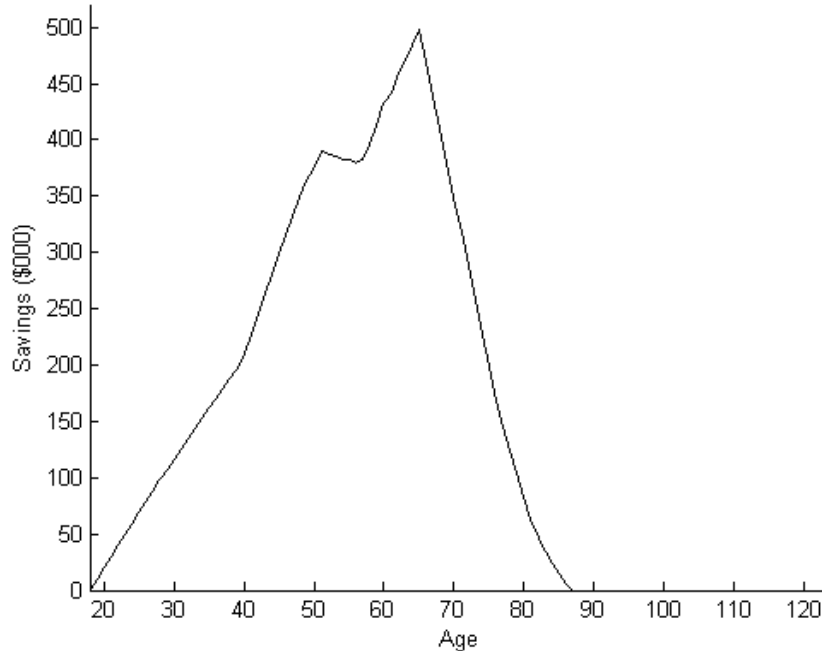
^aThe expected dynamic gain is the left side of inequality (11). Its first component is the *life-saving gain*, driven by the difference in the probability of being alive in the next period when seeking care. The second is the *quality of life gain*, a composite of future life expectancy gains and improvements in the amount of time available to agents during all future periods. They are discussed briefly in Section 2 and in further detail in Appendix A.

^bThe utility loss is the right side of inequality (11) and is the opportunity cost of waiting for care.

^cThe net gain is the difference between the expected gain and utility loss.

^dThe optimal choice of using health care is 1 when the net gain is higher than zero.

Figure 2: Age vs. Expected Savings



in good health in Figure 2. They then use their savings more extensively as they reach retirement to smooth their consumption and utility in their later stages of life. In effect, since we the value of θ is the minimum level of pensions given to elders in Canada, the savings are used by agents as a primary, privately-owned, pension fund. This fund is expected to become extinct soon after agents reach their life expectancy, of 80.6 years old.¹¹

4.2 Elasticity of Demand With Respect to Congestion

A key policy issue for countries affected by a high health care congestion is the elasticity of the demand for care with respect to congestion. If elasticity is close to zero, a direct implication is that waiting times can be easily eradicated by increasing the resources of the system. On the other hand, if the demand is very responsive to congestion, even an important increase in the system's funding and capacity may have little impact on the system's congestion despite the accrued quantity of care received, a plausibly very frustrating outcome for taxpayers. This elasticity is also of key importance to policymakers who may choose to control the demand for public care by allowing congestion to increase: if the elasticity of the demand is close to zero, such a policy would result in an explosion of the congestion level with little actual impact on the cost of the health care system.

Martin and Smith (2003) used a panel data approach to estimate the elasticity of demand with regard to the length of waiting lists for several surgery procedures in the United Kingdom. Their approach yielded an overall value of elasticity of -0.09, with values for individual specialties ranging from -0.24 to 0.38. As noted in the introduction of this paper, such studies have two important flaws, as the accumulation of hours spent

¹¹We recognize that this occurs because of the absence of a bequest motive in our model. If a bequest motive was added to the model, we would observe a fatter tail on the right side of the plot, as agents would aim to leave a portion of their lifetime savings to their heirs. It appears unlikely that this addition would have an important impact on our numerical results, since savings have no noticeable impact on the decision to use health care in public systems.

in waiting rooms before the reception of treatments is not taken into account and could intuitively have a larger impact on demand than the length of waiting lists, and their analysis is restricted in comparison to the vastness of modern health care systems. Our model is thus possibly the best tool developed so far to anticipate a global response to congestion variations.

To that effect, we compute the demand for health care resulting from our calibration when for different values of γ , keeping the rest of the calibration presented in table 2 intact. Figure 3.1 presents the resulting proportion of the total population using health care for 11 values of γ between 0 and 0.1. When congestion is zero, health care services impose no cost in our model, which results in all agents choosing $\alpha = 1$. This situation, while theoretically possible, is implausible, since it would mean receiving treatment instantly upon wishing for it, without any form of transportation costs or duration of treatment. As expected, the number of users is downward sloping in Figure 3.1. In the neighborhood of current congestion γ^* , the elasticity of demand is -0.195, indicating that a 1% exogenous increase in Quebec waiting times would result in only 0.2% less users of care.¹²

Figure 3.2 presents the average profile of health care users for the same congestion values. As the congestion level rises, young agents and healthy agents are less present in the health care system, raising the average age and reducing the average health of patients. This second effect means an increase in the average cost per user, and implies that the elasticity of the costs of care is even lower than the elasticity of demand. In addition to constituting a weak device to ration health care demand and costs, these two results suggest that congestion may induce poor early detection of diseases in young agents, which may have an intricate dynamic impact on both social welfare and public costs as these wait to grow older and more ill before seeking care.

While the elasticity we find is low, it is twice the global elasticity found by Martin and Smith. We believe this arises because, as noted earlier, the congestion of the system as a whole affects the utility of agents in a more direct fashion than waiting lists for specific elective surgeries. Another possible explanation for the larger value we obtain is the absence of a fully specified job market in our model. Since many workers are insured against income losses through a number of sick days, it is possible that we overestimate the impact of congestion on demand. That being said, we also find that the demand for care is inelastic, which leads us to conclude that reforms aiming at increasing the system's capacity will be efficient at reducing congestion even in the absence of productivity gains.

4.3 Predicted Impact of Policies

In Table 4, we present the results of four types of policies that impact the global outcome of the public care system: introducing moderating health care fees and modifying the income tax rate, age of retirement, and pension generosity. We simulate these policies by modifying the corresponding parameters in our model and iterating on γ until a new equilibrium is obtained. Included in these results are the expected impacts on usage, congestion, GDP and health care costs, as well as four welfare measures. These are the mean, median and minimum value, which are informative in an ordinal manner, as well as the consumption adjustment

¹²We find this elasticity by simulating the demand resulting from γ values at the proximity of γ^* :

$$\frac{\ln\left(\sum_{i=1}^N \alpha_i |\gamma^* \cdot 1.1\right) - \ln\left(\sum_{i=1}^N \alpha_i |\gamma^* \cdot 0.9\right)}{\ln(\gamma^* \cdot 1.1) - \ln(\gamma^* \cdot 0.9)}$$

Figure 3: Simulated Impact of Congestion on Health Care Demand

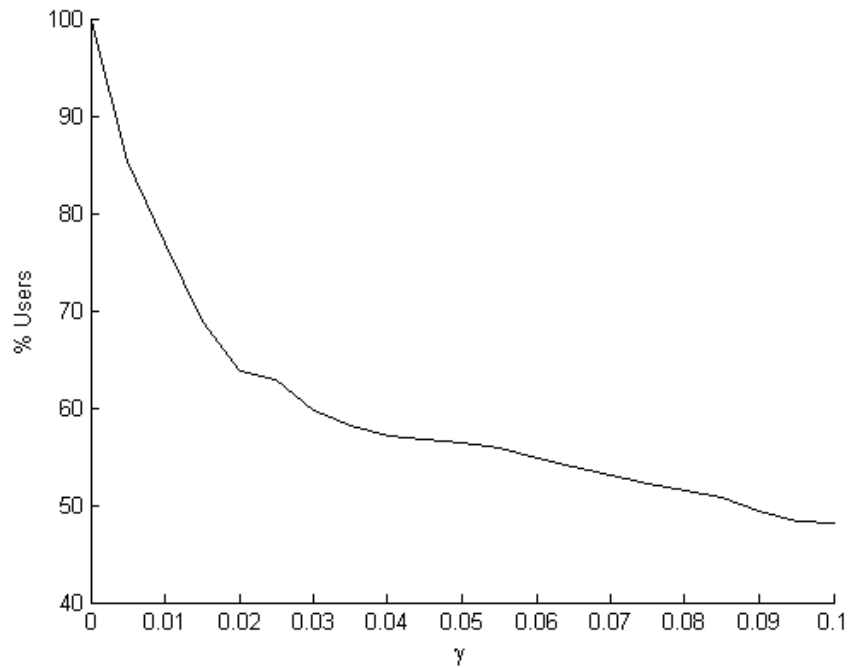


Figure 3.1 Proportion of Users

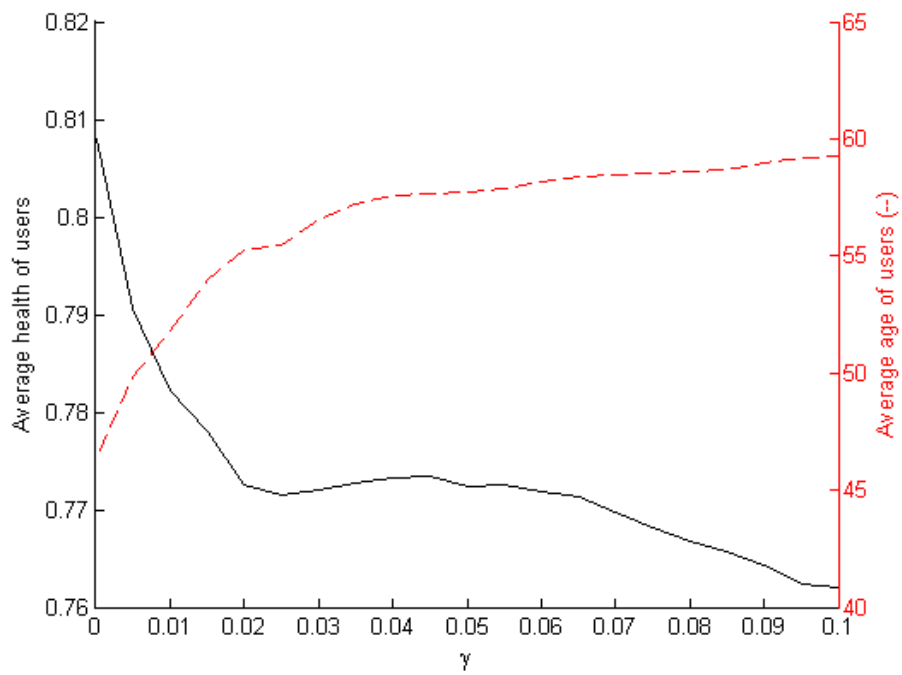


Figure 3.2 Profile of Patients

necessary to obtain the same average value as the base case.¹³ We also exploit the heterogeneity present in our model by presenting how different policies affect the use of care and values of agents of different ages and health levels.¹⁴

It should be noted that, as is generally the case in this type of analysis, from a given initial state, convergence to the equilibrium by applying the given policies is uncertain, since the distribution dynamics are ignored. Even if convergence was certain, the length of the process could be long. This may be especially important considering that increases of revenues for health care are implicitly assumed to be directly converted into new services in our model.

A. Introduction of user fees

Policies 1 to 3 consider the implementation of user fees. We model these as a simple modification of the consumption constraint when agents choose to use health care:

$$c = \begin{cases} A + w \cdot WT(1 - \tau) - Fee - \frac{A'}{(1+r)} & \text{if } a < a_\theta \\ A + \theta - Fee - \frac{A'}{(1+r)} & \text{if } a \geq a_\theta \end{cases} \quad \text{if } \alpha = 1 \quad (20)$$

In Quebec's public health care debate, moderating fees have been described by its proponents as a mean to control the use of health care without relying on waiting lines, but its opponents raised concern that they may prevent poorer citizens from obtaining care. Accordingly, we might expect user fees to decrease congestion through a decreased health care usage and increased revenues available for health care. What we observe, however, is that the lowered congestion is so significant that the percentage of users increases despite the fee for all fee levels considered. Even the imposition of a \$100 fee per use, which would account for less than 2% of health care costs, would result in a 45% decline in congestion. We also observe a simultaneous increase in GDP, caused jointly by the decreased congestion - users of health care have more available time for work during the period - and by a response of health care users to the fee in order to maintain their consumption and savings levels. In all three cases, this policy yields increases in GDP and health care use, and a reduction of congestion.

Interestingly, we also find an increase in social value as measured by all welfare measures when fees are low, meaning that the benefit resulting from a reduction in congestion socially outweighs the fees paid by the individuals. However, as the fee increases, this policy cannot be Pareto optimal, since the minimal value in the distribution lowers. Poor and unhealthy agents prefer being in a system with long waiting lines than paying high fees. Despite this, the health care use and average value of all age and health groups increase even when high fees are charged, suggesting a strong willingness to pay for care with lower waiting lines by almost all agents. Furthermore, according to all welfare measures minus the minimum value, the fee that would lead to the largest social gain is the largest one, which would reduce congestion by almost 90%. We must again remind the reader that our model has no unemployment and that socioeconomic status differs only through the level of savings, since agents have the same wages. Whether these large expected social

¹³This adjustment is obtained by finding the factor k such that

$$\left[\sum_{i=1}^N \sum_{t=0}^{\infty} \beta^t E_i [u(k\hat{c}_t, \hat{L}_t)] \mid \text{Policy} \right] = \left[\sum_{i=1}^N \sum_{t=0}^{\infty} \beta^t E_i [u(c_t, L_t)] \mid \text{Base Case} \right],$$

where (\hat{c}_t, \hat{L}_t) is the optimal consumption and leisure choices made by agents at the equilibrium consistent with the policy scenario, with $k = 1$. We keep these choices fixed, then iterate on k until obtaining an equality.

¹⁴We did not present results per level of savings because they were allocated by us according to the age and health of agents and are thus correlated with those (see Appendix C).

gains would remain without these assumptions is uncertain.

B. Changes in the income tax rate

A more common policy to increase funding for public health care is to increase taxes. Policies 4 and 5 show the predicted impact of increasing the income tax rate τ . As we observed with the introduction of fees, these policies are effective in reducing the congestion of care, and the number of patients increases in reaction to the reduced waiting times. A notable difference from policies 2 and 3 is that the minimum value of our distribution of agents now increases as the policy intensifies. In this policy scenario, poor and sick agents are no longer charged high fees for care and only experience the benefits from the reduced congestion. Thus, if fees were to be implemented, taking the capacity of payers into account appears socially desirable. Finally, we note that the number of users and health care costs are drastically higher with a tax increase of 2.5% than with fees of \$1000, despite achieving a similar congestion reduction. This constitutes the biggest contrast between the results of the two policies and confirms that fees for care do act as a rationing mechanism. Otherwise, the welfare results of policies 4 and 5 indicate that raising taxes is a viable option, as long as the end result is an effective reduction in congestion.

Policy 6 is equivalent to a government deliberately allowing congestion to rise in order to lower the cost of health care, which funds a tax cut of 1% of GDP. As predicted in the last section, such a policy induces very high welfare costs, as equilibrium congestion more than doubles and the percent of users decreases by 26%. All four measures of welfare also display a noticeable reduction compared with the base case, and agents would require a 12% increase of their consumption to be indifferent to the base case. As expected, the percentages of users in younger and healthier agents show the largest declines, and the older agents, who remain heavily in the health care system, show the largest declines of their average value.

In Table F1.iii, presented in appendix, we test the sensitivity of policies 4 and 6 to six different consumer aversion values. We find these policy results to be qualitatively robust, as only the amplitude of the effects is affected by differing values of ρ .

C.-D. Changes in age of retirement and pension generosity

Policies 7 to 11 study the impact of modifying the pension system without modifying the income tax. Thus, if pension costs decrease (increase), funding that was previously directed to retirees is directed to (removed from) the health care system. These simulations are conducted in part as an attempt to determine whether pension generosity or health care should be prioritized when public funds are limited. Also, it has been suggested in Quebec's public spending debate that the longevity of users could be perceived as an opportunity to increase the retirement age and thus generate more production in the economy as a whole, reduce pensions costs and increase public funds (Castonguay and Laberge, 2010). Policies 7 and 8, which increase retirement age to 67 and 70 years old, are conducted in this spirit, and largely support this claim. Such policies would have the highest positive impact on aggregate production and would prove beneficial for all our welfare measures. A somewhat surprising result is that both agents of 65 to 79 and 80+ see an important increase of their average values. This arises because agents are able to work and accumulate savings for five more years, and are able to consume more per period during the rest of their lifetime.

Interestingly, reducing pension generosity without changing the retirement age, as simulated in policy 10, yields a much lower GDP increase than encouraging the elderly to remain in the work force. This policy forces agents to work and save more during their working ages, which further discourages the use of care by young agents. This effect can be best seen when comparing savings and the use of care by younger groups between policies 7 and 10, which otherwise yield a close level of congestion.

Policies 9 and 11 are a complement of policy 6, which allowed congestion to increase in order to fund a tax cut. Instead, these policies explore the outcome of using the reduced costs of care to fund other public programs - in this case allowing for more generous pension programs. Not surprisingly, policy 9, which allows agents to retire at 60 with full pension at the expense of the public care system, yields an explosion in health care congestion, since the funding for care is also affected by a decline in GDP as the workforce is reduced. This policy results in the worse welfare losses of our policy attempts, which is a corollary of Castonguay and Laberge's claim that an increase in retirement age would be socially desirable in Quebec. A 10% increase in pension generosity, as studied in policy 11, is less damaging than lowering the retirement age, but also induces an important congestion increase and a social loss as measured by all but one welfare measures. Only the minimal value measure shows an increase, spurred by the additional pension revenue.

E. Combination of policies

While their results indicate large welfare gains, the political feasibility of reforms 3, 5 and 8 seem weak at best, given their austere nature. However, policy 12 indicates that a similar outcome in terms of additional financing of care, congestion reduction and welfare gains can be obtained by combining fees of \$250, an increase of 0.75% of the tax rate and an increase of two years to the age of pension eligibility. This policy also presents several appealing facets of its own: a higher GDP than policies 3 and 5 because of the increased age of retirement; a lower percentage of users than in policy 5 because of the inclusion of user fees; and a reduced need to increase the age of pension than in policy 8 because of the use of two other financing levers. As the application of each approach is more modest, such a policy *may* be more politically viable.

Summary

To summarize, we find that reducing the funding to a public health care system that displays congestion leads to significant welfare losses, since it causes waiting times to increase dramatically. Our results also suggest that Quebec's prevailing situation is far from socially optimal, as all policies aimed at increasing the funding for care and decreasing congestion produce important welfare gains. For instance, agents would be willing to sacrifice up to 8.3% of the flow of their consumption to be in a system where the tax-to-GDP ratio is 2.5% higher but the waiting times for care are reduced by 90%. While many of the policies that increase health care funding are austere in nature and may be politically infeasible, a combination of several more modest approaches would lead to similar welfare gains and waiting times reduction.

5 Conclusion

In this paper, we build a dynamic stochastic model to study the demand for public health care in systems characterized by zero monetary prices and waiting times. Our theoretical results show that agents choose to seek care in such a system when the expected gain of doing so, which depends on their health movement expectations, outweighs the option cost of spending time in congested health care facilities rather than working or participating in leisure activities.

Our numerical results illustrate the perverse nature health care congestion. On the one hand, users of health care spend a large amount of time waiting for care, lowering both their individual utility during the period and the global production of the economy and public finances. On the other hand, younger and healthier agents are the most discouraged by the utility cost of waiting times, which suggest profound negative implications on the early detection and prevention of diseases and future public health. By simulating how

Table 4: Results of Simulated Policies

A. Introduction of user fees

	Base Case <i>Fee = 0</i>	1 <i>Fee = 100</i>	2 <i>Fee = 500</i>	3 <i>Fee = 1000</i>
Users (%)	53.7	57.7	61.8	72.1
γ^* ^a	100	54.7	36.7	10.9
<i>GDP</i>	100	102.7	102.6	103.1
<i>HC^S</i> (% <i>GDP</i>) ^b	7.0	7.2	7.7	8.6
Fees (% <i>HC^S</i>)	0	1.7	8.6	18.0
Public pensions	100	100	100	100
Working time among workers ($a < a_\theta$)				
Median	100	100	100	100
Mean	100	101.7	101.5	102.1
Savings				
Median	100	89.5	84.2	84.2
Mean	100	98.7	96.5	95.1
Welfare measures				
Median Value	100	100.7	100.9	101.1
Mean Value	100	100.7	100.8	101.2
Minimum Value	100	100.3	99.7	99.0
Δ Consumption ^c	100	96.0	94.5	92.0
Users (%) per age and health group				
Ages 18 to 39	5.2	10.1	16.9	31.0
40 to 64	76.4	80.9	84.4	95.2
65 to 79	100	100	100	100
80 and over	100	100	100	100
<i>H</i> of 0.45 to 0.60	91.2	97.4	98.1	100
0.65 to 0.80	62.3	65.3	72.6	82.1
0.85 to 1.00	34.8	39.6	40.2	53.1
Mean value per age and health group				
Ages 18 to 39	100	100.4	100.5	100.8
40 to 64	100	100.9	101.2	101.6
65 to 79	100	100.8	101.0	101.3
80 and over	100	100.7	100.6	100.6
<i>H</i> of 0.45 to 0.60	100	100.7	100.9	101.2
0.65 to 0.80	100	100.7	100.9	101.2
0.85 to 1.00	100	100.6	100.8	101.1

^aEquilibrium congestion of care.

^bHealth care spending.

^cAdjustment to the consumption flow of agents necessary to obtain the same mean value as the base case (see Footnote 12).

B. Changes in the income tax rate

	Base Case $\Delta\tau = 0$	4 $\Delta\tau = 0.01$	5 $\Delta\tau = 0.025$	6 $\Delta\tau = -0.01$
Users (%)	53.7	63.7	81.0	39.7
γ^*	100	31.2	8.8	260.2
<i>GDP</i>	100	102.0	101.8	97.0
<i>HC^S</i> (% <i>GDP</i>)	7.0	8.1	9.5	5.9
Fees (% <i>HC^S</i>)	0	0	0	0
Public pensions	100	100	100	100
Working time among workers ($a < a_\theta$)				
Median	100	100	100	120
Mean	100	100.8	100.8	101.9
Savings				
Median	100	84.2	89.5	78.9
Mean	100	95.9	95.4	91.8
Welfare measures				
Median Value	100	101.1	101.6	97.5
Mean Value	100	101.0	101.5	98.1
Minimum Value	100	100.8	101.0	98.1
Δ Consumption	100	94.0	91.7	112.2
Users (%) per age and health group				
Ages 18 to 39	5.2	21.2	49.5	4.0
40 to 64	76.4	85.1	99.5	47.5
65 to 79	100	100	100	99.5
80 and over	100	100	100	100
<i>H</i> of 0.45 to 0.60	91.2	100	100	84.0
0.65 to 0.80	62.3	75.3	90.8	49.6
0.85 to 1.00	34.8	41.0	63.9	17.8
Mean value per age and health group				
Ages 18 to 39	100	100.6	100.9	99.2
40 to 64	100	101.4	102.0	97.2
65 to 79	100	101.4	101.8	96.6
80 and over	100	101.3	101.7	96.9
<i>H</i> of 0.45 to 0.60	100	101.1	101.6	97.8
0.65 to 0.80	100	101.1	101.6	98.0
0.85 to 1.00	100	101.0	101.4	98.3

C. Changes in age of retirement

	Base Case $a_\theta = 65$	7 $a_\theta = 67$	8 $a_\theta = 70$	9 $a_\theta = 60$
Users (%)	53.7	61.1	72.3	33.5
γ^*	100	43.6	19.9	363.9
<i>GDP</i>	100	103.3	108.1	96.3
<i>HC^S</i> (% <i>GDP</i>)	7.0	7.6	8.4	5.2
Fees (% <i>HC^S</i>)	0	0	0	0
Public pensions	100	85.8	67.3	143.8
Working time among workers ($a < a_\theta$)				
Median	100	100	100	120.0
Mean	100	101.1	101.0	109.5
Savings				
Median	100	84.2	57.9	84.2
Mean	100	93.5	67.1	96.7
Welfare measures				
Median Value	100	100.9	100.8	96.6
Mean Value	100	101.0	101.1	96.9
Minimum Value	100	100.7	100.9	96.8
Δ Consumption	100	94.4	93.9	121.0
Users (%) per age and health group				
Ages 18 to 39	5.2	16.0	33.2	3.1
40 to 64	76.4	83.5	93.8	37.0
65 to 79	100	100	100	91.5
80 and over	100	100	100	100
<i>H</i> of 0.45 to 0.60	91.2	98.1	100	83.2
0.65 to 0.80	62.3	71.2	83.9	41.3
0.85 to 1.00	34.8	40.2	51.1	13.2
Mean value per age and health group				
Ages 18 to 39	100	100.6	100.9	98.6
40 to 64	100	101.1	100.8	96.0
65 to 79	100	102.5	103.4	90.8
80 and over	100	102.5	104.7	92.4
<i>H</i> of 0.45 to 0.60	100	101.1	101.1	96.3
0.65 to 0.80	100	101.0	101.1	96.7
0.85 to 1.00	100	100.9	101.0	97.2

D. Changes in pension generosity

	Base Case $\theta \cdot 1$	10 $\theta \cdot 0.9$	11 $\theta \cdot 1.1$
Users (%)	53.7	59.0	46.7
γ^*	100	46.9	159.4
<i>GDP</i>	100	102.7	97.7
<i>HC^S</i> (% <i>GDP</i>)	7.0	7.4	6.5
Fees (% <i>HC^S</i>)	0	0	0
Public pensions	100	90.0	110.0
Working time among workers ($a < a_\theta$)			
Median	100	100	100
Mean	100	101.7	100.1
Savings			
Median	100	89.5	78.9
Mean	100	97.0	91.6
Welfare measures			
Median Value	100	100.6	98.7
Mean Value	100	100.6	99.3
Minimum Value	100	98.5	101.2
Δ Consumption	100	96.3	104.6
Users (%) per age and health group			
Ages 18 to 39	5.2	13.8	4.6
40 to 64	76.4	80.9	61.9
65 to 79	100	100	100
80 and over	100	100	100
<i>H</i> of 0.45 to 0.60	91.2	97.9	87.1
0.65 to 0.80	62.3	67.8	56.6
0.85 to 1.00	34.8	39.6	25.4
Mean value per age and health group			
Ages 18 to 39	100	100.4	99.7
40 to 64	100	100.9	98.8
65 to 79	100	100.5	99.3
80 and over	100	99.9	100
<i>H</i> of 0.45 to 0.60	100	100.7	99.2
0.65 to 0.80	100	100.6	99.2
0.85 to 1.00	100	100.6	99.3

E. Combination of policies

	Base Case $fee = 0$ $\Delta\tau = 0$ $a_\theta = 65$	12 $fee = 250$ $\Delta\tau = 0.0075$ $a_\theta = 67$
Users (%)	53.7	72.3
γ^*	100	18.0
<i>GDP</i>	100	103.3
<i>HC^S</i> (% <i>GDP</i>)	7.0	8.7
Fees (% <i>HC^S</i>)	0	4.4
Public pensions	100	85.8
Working time among workers ($a < a_\theta$)		
Median	100	100
Mean	100	100.6
Savings		
Median	100	78.9
Mean	100	90.1
Welfare measures		
Median Value	100	101.0
Mean Value	100	101.3
Minimum Value	100	100.5
Δ Consumption	100	92.2
Users (%) per age and health group		
Ages 18 to 39	5.2	33.2
40 to 64	76.4	93.8
65 to 79	100	100
80 and over	100	100
<i>H</i> of 0.45 to 0.60	91.2	100
0.65 to 0.80	62.3	83.9
0.85 to 1.00	34.8	51.1
Mean value per age and health group		
Ages 18 to 39	100	100.8
40 to 64	100	101.6
65 to 79	100	102.9
80 and over	100	102.8
<i>H</i> of 0.45 to 0.60	100	101.4
0.65 to 0.80	100	101.4
0.85 to 1.00	100	101.3

our model reacts to a variety of congestion levels, we find that the demand for care is inelastic. In the case of Quebec, a 1% increase in waiting times would reduce the demand by only 0.19%. Thus, simulations of attempting to reduce of health care costs by increasing waiting times led to strong overall negative welfare impacts. On the bright side, a low elasticity means that congestion can be decreased successfully without exponentially increasing the costs of health care.

Our policy simulations reveal that, in the case of Quebec, a 1% increase of the global tax rate, a \$500 yearly fee for users of care or an increase of the retirement age to 67 years old would be sufficient to eliminate over half of current congestion, assuming that the increased public income could be transferred linearly into health care production. While such reforms would surely present important political obstacles to implementation in most countries afflicted with long waiting times, our simulations indicate that the social value would be improved if a permanent decrease in waiting times was obtained.

In our view, this paper should be considered a first attempt at using computational methods to gain a better understanding of public health care systems. We note that further work needs to be done to expand on our findings. Because our focus is on the demand for care, we assume that supplementary funding can be translated into additional care without the need for long term investments, training more skilled personnel or any form of decreasing marginal productivity of care. Further research is needed to assess more precisely how health care capacity can or should be increased when more financial resources are introduced in the health care system. Also, we recognize that our analysis presents a very unconstrained modeling of labor: in order to focus our attention on age, health and savings, we assume that agents were homogeneous in productivity and that they can freely choose the precise amount of time to invest in their working activity. Adding unemployment and heterogeneity in wages could have a major impact on policy outcomes, possibly diminishing the social desirability of reforms aimed at reducing congestion. Finally, and perhaps most importantly, the dynamics of the distribution in the years following the application of policies and their eventual impact on long term equilibrium need to be assessed. This will prove particularly helpful in assessing the practical desirability of policies - notably in societies with aging populations.

In the future, another important question that should be addressed is the desirability of a dual public-private system in replacement of a public system showing congestion of care, and the optimal size that should be allowed for the private sector. This question arises naturally as congestion and costs force many OECD public systems to consider partial privatization, and access to care concerns in the United States is answered by reforms increasing the public involvement in health care. From a public system presenting long waiting queues, our results suggest that a dual system could show a massive transition of users to private care, since 72% of our distribution of agents seek health care even in the presence of \$1000 annual fees when congestion is reduced. We conclude by emphasizing the relevance of numerical approaches to gain a better understanding of the mechanisms at work behind complex problems such as those addressed in this paper.

References

- [1] Anderson, G., 1997. "Willingness to Pay to Shorten Waiting Time for Cataract Surgery", *Health Affairs*, 16(5). 181-190.
- [2] Arrow, K.J., 1963. "Uncertainty and the Welfare Economics of Medical Care", *American Economic Review*, 53(5), 941-973.
- [3] Canadian Institute for Health Information (CIHI), 2008. *National Health Expenditure Trends, 1975-2008*. 2011/05 at http://secure.cihi.ca/cihiweb/products/nhex_2008_en.pdf.

- [4] Canadian Institute for Health Information (CIHI), 2011. *Wait Times in Canada - A Comparison by Province*. 2011/05 at http://secure.cihi.ca/cihiweb/products/Wait_times_tables_2011_en.pdf.
- [5] Castonguay, C. and Laberge, M., 2010. "La longévité : une richesse", *CIRANO Research Paper* 2010RP-01.
- [6] Chang, F-R, 1995. "Uncertainty and Investment in Health", *Journal of Health Economics*, 15(3), 369-376.
- [7] Coyte, P.C., et al., 1995. "Waiting Times for Knee Replacement Surgery in the United States and Ontario", *New England Journal of Medicine*, 331(16), 1068-71.
- [8] Cullis, P., J.G. Jones and Propper, C., 2000. "Waiting and Medical Treatment: Analyses and Policies", Chapter 23 in Cuyler, A.J. and Newhouse, J.P (eds), *Handbook on Health Economics*, Amsterdam: Elsevier.
- [9] Ehrlich, I., and Chuma, H., 1990. "A Model of the Demand for Longevity and the Value of Life Extension", *Journal of Political Economy*, 98(4), 761-782.
- [10] Fonseca R., Michaud, P.-C., Galama, T., and Kapteyn, A., 2009. "On the Rise of Health Spending and Longevity", *RAND Working Paper*.
- [11] French, E., 2005. "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour", *Review of Economic Studies*, 72(2), 395-427.
- [12] Gravelle, H. and Siciliani, L., 2007. "Is Waiting-Time Prioritisation Welfare Improving?", *Health Economics*, 17(2), 167-184.
- [13] Grossman, M., 1972. "On the concept of Health Capital and the Demand for Health", *Journal of Political Economy*, 80(2), 223-255.
- [14] Hall, R.E. and Jones, C.I., 2007. "The Value of Life and the Rise in Health Spending", *Quarterly Journal of Economics*, 122(1), 39-72.
- [15] Hansen, G.D., and Imrohoroğlu, A., 1992. "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard", *Journal of Political Economy*, 100(1), 118-142.
- [16] Lindsay, C.M., and Feigenbaum, B., 1984. "Rationing by Waiting Lists", *The American Economic Review*, 74(3), 404-417.
- [17] Kydland, F.E., and Prescott, E.C., 1982. "Time to Build and Aggregate Fluctuations", *Econometrica*, 50(6), 1345-137.
- [18] Martin, S. and Smith, P.C., 1999. "Rationing by Waiting Lists: an Empirical Investigation", *Journal of Public Economics*, 71(1), 141-64.
- [19] Martin, S. and Smith, P.C., 2003. "Using Panel Methods to Model Waiting Times for National Health Service Surgery", *Journal of Royal Statistical Society*, 166(3), 369-387.
- [20] Organisation for Economic Co-operation and Development (OECD), 2010. *OECD Health Data 2010 (database)*, 2011/05 at www.oecd.org/health/healthdata.

- [21] Picone, G., Uribe, M., and Wilson, M.R., 1998. “The Effect of Uncertainty on the Demand for Medical Care, Health Capital and Wealth”, *Journal of Health Economics*, 17(2), 171-185.
- [22] Siciliani, L. and Hurst, J., 2004. “Explaining Waiting-Time Variations for Elective Surgery Across OECD Countries”, *OECD Economic Studies*, No. 38, 95-123.
- [23] Stock, L.M., Bradley, G.E., Lewis, R.J., Baker, D.W., Sipse, J., Stevens, C.D., 1994. “Patients Who Leave Emergency Departments Without Being Seen by a Physician: Magnitude of the Problem in Los Angeles County”, *Annals of Emergency Medicine*, 23(2), 294-298.
- [24] Wolfe, J.R., 1985. “A Model of Declining Health and Retirement”, *Journal of Political Economy*, 93(6), 1258-1267.

A Analytic Developments

In this appendix, we explore the theoretical determinants of health care consumption in a public health care system by analyzing both sides of inequality (11) in further detail. First, we take a closer look at the expected dynamic gain of health care. In the following developments, in order for both shocks to affect the agent’s future health positively, we will assume that $\delta(a, \epsilon_\delta)$ is *decreasing* in ϵ_δ , while $\psi(H, \epsilon_\psi)$ is *increasing* in ϵ_ψ . Denoting $f(\epsilon_\delta)$ the density function and $F(\epsilon_\delta)$ the cumulative density function of the stochastic shock on the health care depreciation rate, ϵ_δ , the expected value when not using care is as follows:

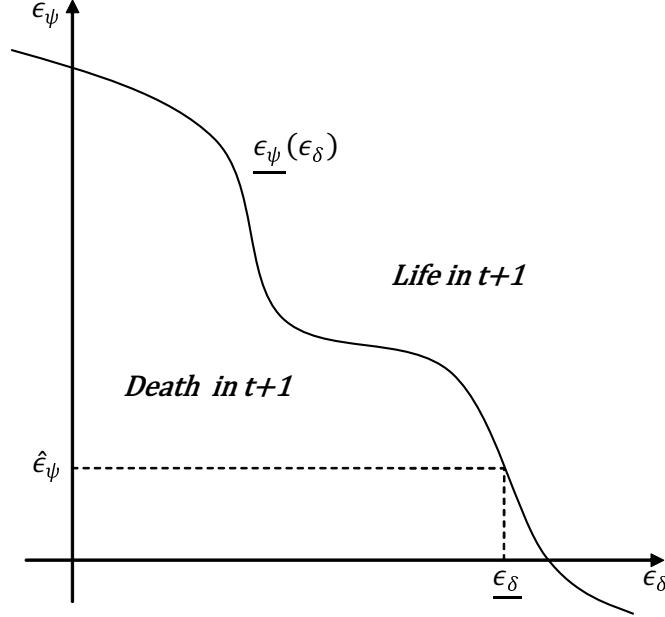
$$E[V'_0] = \int V(a+1, H \cdot (1 - \delta(a, \epsilon_\delta)), A'_o) f(\epsilon_\delta) d\epsilon_\delta \quad (21)$$

For a given H , we can obtain a critical value $\underline{\epsilon}_\delta$ such that $H \cdot (1 - \delta(a, \epsilon_\delta)) = \underline{H}$. By construction, any realization of ϵ_δ equal to or below that value results in death and a \underline{V} value for the agent in the next period. By differentiating, it is easy to show that $\underline{\epsilon}_\delta$ is an implicit function of age and health, increasing in the former and decreasing in the latter. Intuitively, the older and sicker the agent, the more plausible is his death at the end of the period. Using this term, we can rewrite (21) as the probability of being alive in $t+1$ times the expected value conditional to being alive plus the probability of dying times \underline{V} :

$$E[V'_0] = (1 - F(\underline{\epsilon}_\delta)) \cdot E[V(a+1, H'_o, A'_o) | \epsilon_\delta > \underline{\epsilon}_\delta] + F(\underline{\epsilon}_\delta) \cdot \underline{V} \quad (22)$$

With some further work, we can express the expected value when $\alpha = 1$ in a similar fashion. Since the agent is now affected by two stochastic shocks, one on his health depreciation and the second on the success of health care, we cannot find a single critical value of ϵ_ψ such that $H' = \underline{H}$. Instead, $\underline{\epsilon}_\psi$ is a continuum of critical values with respect to the realization of the first shock, ϵ_δ , such that $H \cdot (1 - \delta(a, \epsilon_\delta) + \psi(H, \epsilon_\psi)) = \underline{H}$. Differentiation reveals that $\underline{\epsilon}_\psi$ is an implicit function decreasing in ϵ_δ and increasing in age, while the impact of H can be either positive or negative, as discussed in the previous section. Figure A1 graphically summarizes the concepts just developed. The probability of living in the next period is the probability of picking a couple $(\epsilon_\delta, \epsilon_\psi)$ above the function $\underline{\epsilon}_\psi(\epsilon_\delta)$. In the , $\hat{\epsilon}_\psi$ is the shock on health care production such that the impact of care is exactly zero and only δ affects health, *i.e.* $\psi(H, \hat{\epsilon}_\psi) = 0$. If the agent does not use health care or, equivalently, uses health care but obtains shock $\hat{\epsilon}_\psi$, the probability of being alive is simply the probability of picking a depreciation shock above $\underline{\epsilon}_\delta$.

Figure A1: Life and Death According to Stochastic Realizations



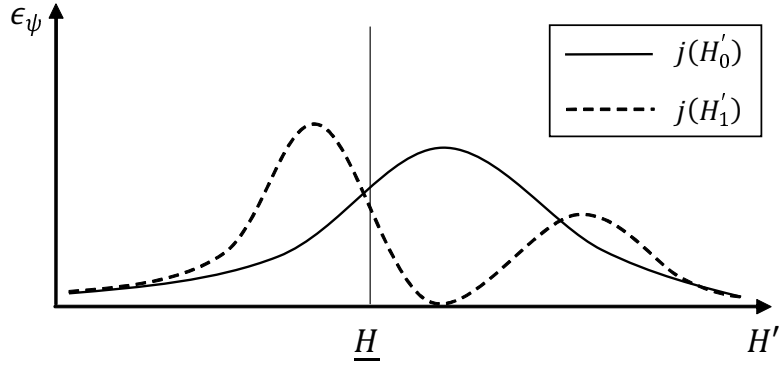
Denoting the density and cumulative density functions of health care production shock $g(\epsilon_\psi)$ and $G(\epsilon_\psi)$, respectively, we find the following expected value when using health care:

$$E[V_1'] = \left(1 - \int G(\underline{\epsilon}_\psi(\epsilon_\delta)) f(\epsilon_\delta) d\epsilon_\delta\right) \cdot E[V(a+1, H_1', A_1') \mid \epsilon_\psi > \underline{\epsilon}_\psi(\epsilon_\delta) \forall \epsilon_\delta] \quad (23)$$

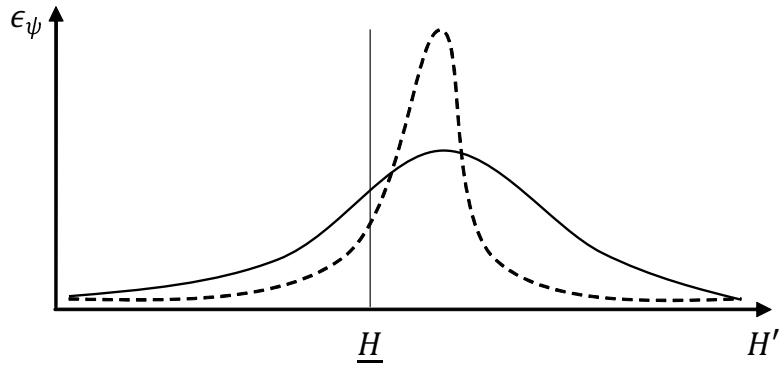
$$+ \int G(\underline{\epsilon}_\psi(\epsilon_\delta)) f(\epsilon_\delta) d\epsilon_\delta \cdot \underline{V}$$

Unsurprisingly, we find once again that the expected value is determined by the probability of dying, the expected value in the next period conditional to being alive, and the value of death. It is by combining (22) and (23) that we can decompose the expected dynamic gain in a *life-saving* and a *quality of life* components, as discussed in the last paragraph of section 2. As should be clear by now, for a given combination of health and age, both components of the dynamic gain of health care are determined by the distributions of health shocks, $f(\epsilon_\delta)$ and $g(\epsilon_\delta)$, whose joint distribution result in a $t+1$ health distribution. Denoting this distribution $j(H'_\alpha)$, we present in Figure A2 a graphical analysis of differing scenarios. The first two graphs present a situation in which one of the component of dynamic gain is positive and the other is negative. In Figure A2.i, health care induces a lower probability of being alive in $t+1$, but a higher expected level of health in the case of survival. For instance, a delicate attempt to remove a tumor might yield such a distribution of future health. Figure A2.ii, on the other hand, presents a case in which care increases the agent's probability of being alive in the next period, but lowers his expected quality of life after the treatment, which could plausibly be observed in the case of the amputation of an infected limb, for instance. Figure A2.iii presents an interesting special case, in which all realizations of ϵ_ψ are superior to $\hat{\epsilon}_\psi$, meaning that the impact of health care usage is strictly positive. Conditional only on the value function being increasing in H , we can show that this situation ensures a positive expected gain of health care. In the absence of this extreme scenario, for a given patient-treatment combination, we must attempt to estimate both of the components of gain to ensure a positive expected impact of health care. In particular, the increase of the expected health level of a patient is insufficient to ensure a gain. These observations are largely in agreement with the 1963 seminal article by Arrow, which emphasizes the importance of uncertainty on individual health care consumption choices.

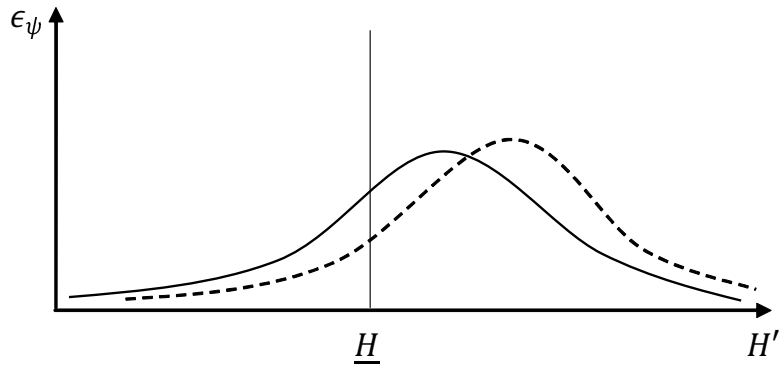
Figure A2: Dynamic Gain Examples



i. Life-saving loss and quality of life gain



ii. Life-saving gain and quality of life loss



iii. Strictly positive impact of health care

In summary, the left-hand side of (11) is mainly determined by the variables affecting the agent's future health. These variables are the depreciation function $\delta(a, \epsilon_\delta)$, the health production function $\psi(H, \epsilon_\psi)$, the agent's age via its impact on the depreciation of health, the agent's level of health in the current period and the distribution of health shocks, $f(\epsilon_\delta)$ and $g(\epsilon_\psi)$. We also note that the important factor in the agent's decision is not the real distribution $j(H'_\alpha)$, but the perceived one, a theoretical justification of social advertisement campaigns informing the population of new treatments and recommending early detection of diseases. This also suggests that public policies would have little success restraining the use of health care through this side of the inequality, unless governments deliberately lessened the perceived quality of care by replacing procedures by less effective ones, an improbable strategy in most countries. Of course, other variables, such as τ and γ , also impact the dynamic gain through the expected values in $t + 1$, but this impact is indirect, since they appear in both V'_1 and V'_0 .

We now focus our attention on the right-hand side of (11), the utility sacrificed in the current period when choosing $\alpha = 1$. To keep the following developments as simple as possible and focus on the most pertinent aspects of the model for policy purposes, we will be omitting the savings variable, which means imposing $A = A' = 0$.¹⁵ Under this assumption, the only way to impact the inter-temporal aspect of his problem is through the use of health care. Once the choice of α has been made, the remaining choice is limited to how much time the agent spends working, which determines directly, through time constraint (3) and consumption constraint (4), the level of consumption and leisure time. Interestingly, this problem is both static and very simple:

$$\begin{aligned} \max_{WT} u(w \cdot WT \cdot (1 - \tau), P - LT(H, \alpha\gamma) - TW) \\ \text{s.t. } WT \in [0, P - LT(H, \alpha\gamma)] \end{aligned} \quad (24)$$

The interior solution to this problem resulting from the first order condition is the common microeconomic result that the ratio of marginal utilities of consumption and leisure must be equal to the ratio of their prices:

$$\frac{u_2(c, L)}{u_1(c, L)} = w \cdot (1 - \tau) \quad (25)$$

For utility functions that present convex indifference curves, the single constraint of (24) is always respected. Now, assuming only that $LT(H, \alpha\gamma)$ is strictly higher when using health care, it is easy to show using differentiation that the first order condition yields both lower consumption and leisure time if $\alpha = 1$, implying that $u(c_0, L_0) - u(c_1, L_1) > 0$. This assumption is intuitive in our context, since we are interested in health care systems presenting long waiting times. Unsurprisingly, still using differentiation, we find that this utility loss is increasing in the accrued lost time incurred by using health care. This is directly relevant for policy purposes, since LT is increasing in its second term, the congestion of health care. Since the only direct impact of this parameter on the consumer problem is an increase of the cost in utility of seeking care, this result leads us to conclude that a mean for governments to restrain health care consumption is to allow the congestion of the system to grow.

All other things being equal, the second parameter determined by the government, τ , has a much lesser impact on the cost of using care. First, it affects the agent's utility in t for both values of α . Second, its impact on the optimal choice of working time depends on the utility function, since income or substitution

¹⁵ As will be obvious in Section 4, savings are mainly used by agents as a tool to smooth their consumption over time, to insure against health shocks, and to use for their consumption during retirement. We believe that such results can be left to numerical analysis.

effects could be dominating. It would thus be necessary to impose additional assumptions on the value function to obtain a clear impact of τ on the utility loss, which we will refrain from. The exact same can be said about the agent’s productivity w , the second determinant of real wage in the model. In reality, τ and w are expected to have an impact on the use of health care, but also on the general equilibrium of the model, since they also affect the working time and consumption of agents, the economy’s global production and the funding of the health care system. Thus, numerical results are more appropriate to inform us about these parameters.

B Data Manipulations

To obtain a distribution of agents compatible with our numerical model, we adapted the CCHS database in several ways. First, in the CCHS, the age of adult respondents is a choice between one of 14 adult age groups, from *18 to 20* to *80 to 101* and over. We use demographic data from the Institut de la Statistique du Québec (ISQ) to establish the proportion of respondents in each age group, then allocate a precise age to the respondents within the age group randomly, thus respecting the age distribution of that year.

Second, as for the health variable, the survey contains two questions useful for our purpose. The first asks respondents their self-perceived level of health at the moment of the survey, the choices being “excellent”, “very good”, “good”, “fair” and “poor”. The result from this question alone is not sufficient to calibrate the health of agents, as we wish to know the health level they had *before* their health care decision. The second question enables us to fill this gap, as it asks the perception of this year’s health in comparison to the last, the choices being “much better”, “somewhat better”, “about the same”, “somewhat worse”, “much worse”. Using both variables, we infer the health level of the respondents at the beginning of the year, before the choice to use health care was made. However, since both variables are expressed in a scale of one to five, this results in large numbers of respondents spread among few health levels. This made obtaining a value of γ^* all but impossible, since either nil or very wide variations in the number of users were then observed when iterating on γ . To avoid this, we allow for 12 different health levels (excluding death), and picked random health values between the bounds of our health variable corresponding to the respondents choices, resulting in a more diverse distribution.

Third, the database contains two variables that allow us to determine the choice of α made by respondents: the number of consultations with general practitioners during the year and with medical doctors in general. Combining those variables, we can easily infer that the number of consultations with specialists corresponds to the total number of consultations minus those with general practitioners. In effect, a large portion of the population of Quebec visits general practitioners on a yearly check-up basis, which, being planned, is neither submitted to the waiting times found in the remainder of the health care system nor is an active pursuing of care. Thus, we interpret as choosing to use health care during the year any respondent who either visited a specialist or had more than one visit with a generalist. Those choosing accordingly represent 53.8% of the population in 2005.

Last, a caveat of the CCHS survey is that it does not include any savings variable. Thus, following each value function solution, we calculate the savings most probable with our model for all age-health combinations, which we then allocate to the CCHS distribution. This approach, which we describe in Appendix C, is opted for instead of a simpler randomization of savings among our distribution because early results suggested a clear and intuitive strategic savings path resulting from dynamic maximization. This savings path is presented and discussed in section 4.

C Mortality, Life Expectancy and Savings Calculation

Following any value function solution, we calculate a *probability tree* from a theoretical starting point where a population Q_{18} of agents are 18 years old and have maximum health capital \bar{H} . For all following ages, we mechanically predict the quantity of agents with any possible age-health-savings profiles, as follows:

$$\begin{aligned}
 Q(a+1, H, A) = & \sum_{H > \underline{H}} \sum_A Q(a, H, A) \cdot 1_{A'=A} \\
 & \cdot [(1 - \alpha^*) \sum_{m=1}^M p_m \cdot 1_{H'_0(H, a, \epsilon_{\delta m})=H'} \\
 & + \alpha^* \sum_{m=1}^M \sum_{n=1}^N p_m \cdot p_n \cdot 1_{H'_1(H, a, \epsilon_{\delta m}, \epsilon_{\psi n})=H'}],
 \end{aligned} \tag{26}$$

where A' and α^* are the optimal savings and health care demand choices of agents with profile $Q(a, H, A)$. For each a , we isolate the number of agents whose health level falls below \underline{H} , and are thus eliminated from the stock of living agents:

$$Q_{\underline{H}}(a) = \sum_A Q(a+1, \underline{H}, A) \tag{27}$$

We do so until the quantity of living agents falls below a sufficiently low threshold, that we set to $1/2 \cdot 10^8$. This means that we consider the theoretical population extinct when individuals have less than 1 chance in 200 million of falling in any given age-health-saving category the next year.

This procedure achieves two distinct purposes. First, it enables the estimation of the mortality per age group and life expectancy resulting from the model, which are necessary to our calibration algorithm (see step 6 of Appendix D). We calculate the mortality rate for a given age group k as the ratio between the quantity of agents that are predicted to die and the quantity of agents that are predicted to be alive during the ages covered by the age group:

$$MR_{model, k} = \frac{\sum_{a=\underline{a}_k}^{\bar{a}_k} Q_{\underline{H}}(a)}{\sum_{a=\underline{a}_k}^{\bar{a}_k} \sum_{H > \underline{H}} \sum_A Q(a, H, A)}, \tag{28}$$

where \underline{a}_k and \bar{a}_k are the lowest and oldest ages included in age group k . Note that agents may be counted as many times as the number of years comprised in the age group, as long as they survive each age. In turn, the life expectancy is computed as follows:

$$LE_{model} = \frac{\sum_a a \cdot Q_{\underline{H}}(a)}{Q_{18}} \tag{29}$$

Second, it enables us to determine which savings agents of a given age and health profile should possess, on average, according to the model's predictions. Since our database contains no information on financial assets, we assign agents these expected assets in order to calculate the optimal choices of our distribution, and the ensuing macroeconomic outcome. For a given age-health combination, the expected asset is the average asset level saved by agents of that profile:

$$\hat{A}(a, H) = \frac{\sum_A Q(a, H, A) \cdot A}{\sum_A Q(a, H, A)} \quad (30)$$

Whenever an age-health combinations was not experienced in the probability tree but present in the CCHS data, such as very young agents in bad health, we chose the average savings of the closest health level with $Q(a, H, A) > 0$.

D Calibration Algorithm

We use a recursive approach to calibrate the unknown parameters of the model. This approach enforces that our calibration replicates the precise number of health care users found in the database, the total cost of health care and the GDP of the province and the mortality rates and health expectancy published by the Institut de la Statistique du Quebec (ISQ). The algorithm we use is as follows:

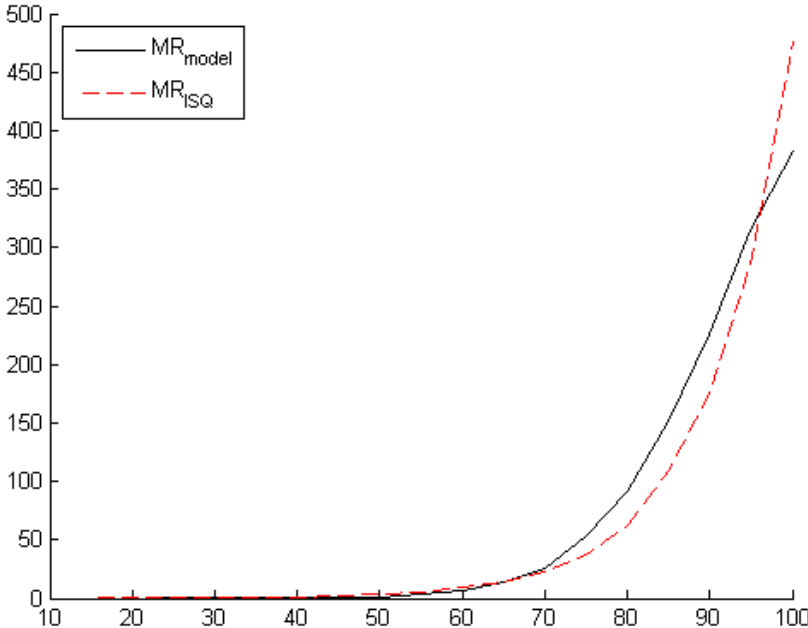
1. As a starting point, we use an *ad hoc* calibration.
2. We solve the Bellman equation (7) by iterating on the value function agents. Since this step is central to our numerical analysis, we describe it in Appendix E. From the solution, we extract the optimal choices of our distribution of agents, which enable us to compute HC^D and HC^S resulting from the current calibration.
3. We iterate on γ until the predicted total number of users is sufficiently close to the 53.8% observed in the CCHS data.
4. We fix $b_4 = \frac{HC^S}{\sum_{i=1}^N \alpha_i \cdot (1-H_i)}$, which enforces $HC^D = HC^S$. This also ensures that the congestion level found in the previous step is exactly γ^* .
5. We observe the GDP and working time of agents and adjust w . The new value to be implemented is $\tilde{w} = GDP_{2005} / \sum_{i=1}^N WT_i$, in order to replicate the observed GDP in the next iteration if the working time decisions of agents are sufficiently close to those in this iteration. Initial results showed that this strategy is efficient when the new γ^* is close to its predecessor, since a property of our utility function is that the individual choice of WT_i is unaffected by the real wage of agents. We then repeat steps 1 to 4 until $GDP = GDP_{2005}$.
6. At this stage, we fix the optimal choices of agents and iterate on large quantities of different plausible combinations of the health function parameters. We use the probability tree scheme presented in Appendix C to calculate the resulting mortality per age group and life expectancy of a given health function. We then systematically restrict the range of the different parameters in order to minimize the following *mortality criterion*: $SSE_{uM}^{0.4} \cdot SSE_{wM}^{0.3} \cdot |Error_{LE}|^{0.3}$. If this criterion is minimized, we believe the health function parameters are effective in approximating the health movements that agents can expect to go through, and thus effective for the purpose of this article. Its first sub-criterion is the sum of squared errors of mortality rates per age group as compared with mortality data from ISQ. Denoting by k the age groups for which ISQ publishes mortality rate, this sub-criterion can be expressed as $SSE_{uM} = \sum_k (MR_{model,k} - MR_{ISQ,k})^2$. Since mortality is much higher in the older age groups, this component really measures the success of replicating the mortality rates of the elderly. The dynamic choices of agents being based in part on their expectations of being alive in the next period, we find important to replicate very well the mortality rates of these later age groups, and

thus give this sub-criterion a larger value. The weighted mortality component, on the other hand, is the sum of the squared errors of mortality rates per age group divided by the observed mortality rates: $SSE_{wM} = \sum_k \left(\frac{MR_{model,k} - MR_{ISQ,k}}{MR_{ISQ,k}} \right)^2$. This weighing ensures that the observed mortality in the younger age groups is also well replicated. While we would expect that a calibration that yields good results for these sub-criteria will also yield a life expectancy close to the one observed in the data, we observed discrepancies of up to 4 years with ISQ estimates when targeting only the first two sub-criteria. This discrepancy could result from the coarseness of the ISQ mortality age groups, which contain 5 years each. By including the absolute life expectancy error as the third sub-criteria, we are able to ensure a minimal gap.¹⁶

7. To confirm the quality of the health function parameters isolated in stage 6, we repeat steps 1 to 6 until the health function parameters yield stability in the mortality criterion.

The model's GDP and health care costs obtained with the calibration are 0.04 % and 0.07 % smaller than those observed in 2005, respectively. The the life expectancy error is of 0.05 years, the SSE_{uM} is of 15 155 and the SSE_{wM} is of 8.4 and resulting in a mortality criterion of 36.2. As can be seen in Figure D1, while the mortality rates of our calibrated model replicate the general trend of the data, our rates are somewhat superior for ages 70 to 90 and lower for agents over 90.

Figure D1: Comparison of Model and Data Mortality Rates



¹⁶The reasons why we fix the optimal choices at this stage are twofold. First, steps 1 to 5 are extremely calculation-heavy and necessitate a prohibitive amount of time to obtain a final solution at existing calculation speed. It would thus be unfeasible, even with a considerably larger computing power than that in our possession, to directly test thousands of different health law of motion parameters, which we are able to do by using this scheme. Second, a feature that proved consistent for a variety of calibrations is that old agents with low health are the most prone to use health care. Differing health functions merely affect the critical health level for which an agent with a given age-savings combination will choose to use health, while globally maintaining this result intact. We discuss this result in section 4.1. What this means for our calibration strategy is that even *ad hoc* health function parameters generate relatively accurate optimal choices following steps 1 to 5, enabling us to test the health parameters without repeating these steps.

E Bellman Equation Solving Algorithm

To solve the dynamic problem represented by Bellman equation (7), we iterate on the value function following Banach's fixed-point theorem. First, we set an *ad hoc* initial value function V . We choose:

$$V(a, H, A) = 0 \quad (31)$$

Using this initial value, we calculate the expected value in $t + 1$ for both choices of α and for all possible choices of savings in the next period:

$$E[V'_o(a, H) | A'] = \begin{cases} \sum_{m=1}^M p_m \cdot V(a + 1, H'_0(H, a, \epsilon_{\delta m}), A') & \text{if } a < \bar{a} \\ \sum_{m=1}^M p_m \cdot V(\bar{a}, H'_0(H, a, \epsilon_{\delta m}), A') & \text{if } a = \bar{a} \end{cases} \quad (32)$$

$$E[V'_1(a, H) | A'] = \begin{cases} \sum_{m=1}^M \sum_{n=1}^N p_m \cdot p_n \cdot V(a + 1, H'_1(H, a, \epsilon_{\delta m}, \epsilon_{\psi n}), A') & \text{if } a < \bar{a} \\ \sum_{m=1}^M \sum_{n=1}^N p_m \cdot p_n \cdot V(\bar{a}, H'_1(H, a, \epsilon_{\delta m}, \epsilon_{\psi n}), A') & \text{if } a = \bar{a} \end{cases} \quad (33)$$

Note that the level of A in the current period is irrelevant to this calculation. We then find the optimal savings and working time in $t + 1$ for *both* choices of α using expected values (32) and (33), as follows:

$$V_\alpha(a, H, A) = \max_{WT, A'} u(c, L) + E[V'_\alpha(a, H, A) | A'] \quad (34)$$

$$\text{s.t.} \quad \begin{cases} c = \begin{cases} A + w \cdot WT(1 - \tau) - (1 + r)A' & \text{if } a < a_\theta \\ A + Pension - A' & \text{if } a \geq a_\theta \end{cases} \\ L = P - LT(H, \alpha\gamma) - WT \\ u(c, L) = -\inf & \text{if } c < 0 \end{cases}$$

This process enables us to isolate the value functions for both possibilities of α , assuming that the expected values computed previously were right. We can then find the value function and optimal choice of α :

$$V(a, H, A) = \max_\alpha V_\alpha(a, H, A) \quad (35)$$

Replacing our initial *ad hoc* function with this new one, we repeat steps (32) to (35) until the supremum of the point-by-point distance between two successive value functions is nil.

F Relative Risk Aversion Sensitivity Analysis

Table F1 presents how the main results of our paper would have been affected by adopting different Arrow-Pratt consumer relative risk aversion values. We present the results for 7 alternative values of ρ within the span of possible values, $(0, 1)$, keeping all the other calibration parameters of Table 2 intact. The first table presents the global equilibrium obtained with all parameters, the second presents the elasticity at the

proximity of γ^* and the third reproduces the results of policies 4 and 6 from section 4.3.

For the whole table, column 7, with $\rho = 0.9$, presents the largest difference of results in comparison with our base results (column 4, in bold characters). With this high risk aversion value, agents are extremely fearful of death and avid consumers of care, which leads to equilibrium waiting times escalating to almost 60% of the period - 7 months - for users in good health. This leads to small GDP and funds for care in contrast to the other values of ρ , as users of care wait astronomic times in queues, during which they cannot work. Ironically, this reduced funding for care and large waiting times also mean that such a risk aversion leads to less users of care and a higher elasticity of demand at the proximity of γ^* . This implausibly high congestion rate and the extreme welfare impacts of policies increasing the tax rate are such that we outright reject this value of risk aversion for the purpose of our paper.

For risk aversion values of 0.1 to 0.9, we see quantitative variations, but qualitative results remain largely similar. Interestingly, we note that the elasticity of demand is small and negative at the proximity of γ^* for all these values in table F1.ii, and table F1.iii indicates that the impact of modifying the tax rate are qualitatively very similar. The variation that we observe among values is intuitive, as for lower values of ρ , agents are less avid users of care, which generates lower equilibrium congestion and a lesser welfare impact of policies. The results from our sensitivity analysis comfort our choice of $\rho = 0.5$ and indicate that the results presented in the remainder of the paper would only be affected quantitatively by other values of γ .

Table F1: Sensitivity of Results to Relative Risk Aversion Values (ρ)

i. Sensitivity of global outcome

ρ		1	2	3	4	5	6	7
		0.1	0.25	0.4	0.5	0.6	0.75	0.9
Users (%)		58.3	53.5	53.7	53.7	54.7	54.2	41.1
γ^*		0.034	0.057	0.063	0.067	0.067	0.092	0.581
<i>GDP</i>		285.7	273.6	272.4	271.9	274.0	271.6	242.1
<i>HC^S</i> (\$B)		20.5	19.2	19.0	19.0	19.2	19.0	15.8
Median value		107 005	30 754	9 314	4 300	2 042	709	226
Mean value		98 681	28 514	8 577	3 955	1 877	647	191
Minimum value		5 513	1 757	564	254	97	-31	-204
Users (%)	Ages 18 to 39	11.9	6.6	5.4	5.2	5.2	4.7	3.8
	40 to 64	80.8	74.8	76.1	76.4	78.5	77.9	50.7
	65 to 79	100	100	100	100	100	100	98.8
	80 and over	100	100	100	100	100	100	100
	<i>H</i> of 0.45 to 0.6	98.1	93.7	91.8	91.2	91.2	89.0	82.9
	0.65 to 0.8	66.4	62.5	62.3	62.3	62.8	62.5	51.9
	0.85 to 1.0	39.4	33.5	34.5	34.8	36.8	36.3	18.2

ii. Sensitivity of demand elasticity at the proximity of γ^*

ρ		1	2	3	4	5	6	7
		0.1	0.25	0.4	0.5	0.6	0.75	0.9
Users (%)	γ^*	58.3	53.5	53.7	53.7	54.7	54.2	41.1
	$\gamma^* \cdot 0.9$	59.0	54.9	54.8	54.5	55.4	55.0	43.1
	$\gamma^* \cdot 1.1$	56.9	51.1	51.8	52.4	53.8	53.4	38.5
Elasticity		-0.184	-0.358	-0.283	-0.195	-0.144	-0.149	-0.563

iii. Sensitivity of policy results

ρ		1	2	3	4	5	6	7
		0.1	0.25	0.4	0.5	0.6	0.75	0.9
$\Delta\tau = 0$ (Base Case)	Users (%)	58.3	53.5	53.7	53.7	54.7	54.2	41.1
	γ^*	0.034	0.057	0.063	0.067	0.067	0.092	0.581
	Mean value	98 681	28 514	8 577	3 955	1 877	647	191
$\Delta\tau = 0.01$ (Policy 4)	Users (%)	67.9	64.8	62.9	63.7	64.7	65.3	66.8
	γ^*	0.014	0.017	0.026	0.021	0.020	0.023	0.027
	Mean V (%BC)	100.9	101.4	101.0	101.0	100.9	100.9	106.4
	Δ Consumption	97.1	94.7	95.2	94.0	93.8	91.1	43.8
$\Delta\tau = -0.01$ (Policy 6)	Users (%)	48.6	45.4	42.2	39.7	39.2	37.2	36.0
	γ^*	0.066	0.092	0.138	0.175	0.215	0.335	0.682
	Mean V (%BC)	98.5	98.7	98.2	98.1	97.8	97.4	98.9
	Δ Consumption	105.2	105.5	109.4	112.2	117.5	131.9	115.4