UNIVERSITÉ DU QUÉBEC À MONTRÉAL

TEMPS D'ATTENTE, VIEILLISSEMENT, MALADIES CHRONIQUES ET COÛTS DES SOINS DE SANTÉ : ENSEIGNEMENTS DU MODÈLE DE CYCLE DE VIE

THÈSE

PRÉSENTÉE COMME

EXIGENCE PARTIELLE DU

DOCTORAT EN ÉCONOMIQUE

PAR

ÉTIENNE GAUDETTE

DÉCEMBRE 2013

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

WAITING TIMES, AGING, CHRONIC CONDITIONS AND HEALTH CARE COSTS: TEACHINGS FROM THE LIFE-CYCLE MODEL

THESIS

PRESENTED AS A

PARTIAL REQUIREMENT FOR THE

PH.D. IN ECONOMICS

BY

ÉTIENNE GAUDETTE

DECEMBER 2013

REMERCIEMENTS

J'aimerais exprimer un grand merci à mes directeurs Stéphane Pallage et Alain Delacroix pour leur aide et leur support dans la rédaction de cette thèse. Stéphane, merci de m'avoir fait découvrir les méthodes utilisées dans mes recherches. Merci également pour ton ouverture d'esprit, ta curiosité et le plaisir de discuter avec toi au fil des ans. Alain, merci pour ta disponibilité, ton attention aux détails et les heures passées à réviser avec moi concepts et méthodologie. Nos rencontres ont beaucoup enrichi le résultat présenté aujourd'hui. Merci également à mes mentors officieux, Pierre-Carl Michaud et Raquel Fonseca, pour leur soutien et leurs précieux commentaires. Finalement, merci à Kimberly McCuaig et Sean Horan pour leurs révisions linguistiques.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

RÉSUMÉ

Cette thèse étudie plusieurs défis auxquels sont confrontés les systèmes de santé modernes avec des modèles calibrés inspirés des théories du capital-santé et du cycle de vie.

Dans le premier chapitre, nous développons le premier modèle dynamique de demande de soins de santé dans un cadre public. Dans ce modèle, les agents choisissent d'utiliser les soins si la désutilité encourue en file d'attente est dominée par les gains dynamiques permis par l'impact des soins sur la santé. Nous intégrons ensuite cette modélisation de la demande dans un modèle macroéconomique, que nous calibrons à l'aide de données québécoises de 2005. À l'aide de simulations, nous questionnons la pertinence, du point de vue du bien-être social, de permettre à de longs temps d'attente d'émerger afin de réduire les coûts de santé. Nous trouvons que les temps d'attente constituent un mécanisme faible de rationnement de la demande et mènent à d'importants coûts sociaux. En contrepartie, toutes les politiques simulées menant à des réductions de temps d'attente génèrent des gains substantiels de bien-être.

Dans le second chapitre, nous étendons notre méthodologie afin d'étudier les impacts du vieillissement de la génération du "baby-boom" sur les systèmes publics de santé. Nous faisons évoluer la distribution d'agents selon les prévisions des démographes de l'Institut de la statistique du Québec (ISQ) et anticipons l'évolution de l'utilisation des soins et des temps d'attente de 2005 à 2050. Nous trouvons que la politique actuelle d'augmentation de 5% par an du budget de la santé mènera à un allongement important des temps d'attente d'ici 2030, et ce malgré une augmentation marquée de la part des coûts de santé dans l'économie. D'ici à 2050, nous estimons qu'il sera nécessaire de doubler la part de l'économie allouée au système de santé afin de maintenir les temps d'attente à leurs niveaux de 2005.

Dans le dernier chapitre, nous nous questionnons sur l'évolution à venir de la santé, de la longévité et des coûts des soins des États-Unis d'Amérique (É-U). Nous développons un modèle de demande de soins dont le point focal est l'incertitude pour les agents de contracter une maladie chronique. Nous calibrons ce modèle sur l'évolution des données américaines de 1985 à 2005. Selon nos simulations, les dépenses de santé atteindront 22% du PIB des É-U d'ici 2050 si les tendances récentes continuent au même rythme. Nous trouvons que la prévalence de maladies chronique continuera de croître durant cette période, mais n'entraînera pas d'augmentations importante de coûts des soins. Ce sont plutôt les progrès technologiques en santé et l'augmentation des revenus qui entraîneront la hausse des dépenses.

Mots-clé: Systèmes publics de santé, demande de soins de santé, coûts des soins de santé, temps d'attente, vieillissement, maladies chroniques.

ABSTRACT

This thesis aims to shed light on challenges faced by modern health care systems by focusing on the study of the demand for care.

In the first chapter, we build the first theoretic health-capital demand model in a public health care setting. In doing so, we close an important gap in the existing literature: public health care systems are the norm rather than the exception in Western countries. To replicate key features of public health care systems, our model showcases the importance of waiting times, rather than monetary prices, on the health care demand problem of individuals. We then embed this demand model in the definition of a macroeconomic equilibrium, calibrated with 2005 Québec data. Our simulations reveal that waiting times for care constitute a weak rationing device and lead to important welfare losses if used to contain costs. Furthermore, all policies that lead to waiting times reductions result in major welfare gains.

In the second chapter, we extend the framework developed in the first chapter to anticipate the impact of the aging baby-boom generation on costs, waiting times and health care use in public systems. Our methodology updates the distribution of agents according to the Institut de la statistique du Québec (ISQ) demographic scenarios, and projects policy responses and outcomes for Québec's health care system over the period 2005-2050. Our results indicate that Québec's current budgeting approach of increasing health spending by 5% each year will result in much longer waiting times by 2030. If the government aimed to maintain waiting times at current levels, the share of health spending in the economy would need to double by 2050.

In the last chapter, we aim to forecast the evolution of health care costs, longevity and chronic conditions prevalence in the US. We develop a dynamic demand model which focuses on the uncertainty of developing a chronic condition. Looking ahead, we find that spending will continue to increase steadily if current trends continue at the same pace. By 2050, they would reach 22% of GDP. While chronic conditions prevalence is most likely to continue escalating in the years to come, our simulations suggest that it will not generate important additional costs. According to our results, income growth and technological innovations will be the main drivers of the increase in medical costs.

Keywords: Public health care systems, Demand for care, Health care costs, Waiting times, Aging, Chronic conditions.

INTRODUCTION

Recent improvements in our capacity to treat illness and extend life have been nothing short of phenomenal. Over the period 1960-2010, life expectancy at birth in Organisation for Economic Co-operation and Development (OECD) countries has leaped by over 10 years, to 79.7 years. If such gains are extremely valuable, they are also quite costly. Over the same period, resources allocated to health care grew from under 5% to 9.5% of Western countries' gross domestic product (GDP). On average, over 70% of their health expenditures are public, and represent an important and growing share of public spending (OECD, 2012; OECD, 2013). Simultaneously, the demand for care is increasing at an even faster pace. As a result, in many industrialized countries, long waiting times for care have emerged and are now an important facet of health care delivery.

Major trends suggest that further pressure will be added to health care systems in the decades ahead. Since the onset of the 2010s, the first-borns of the *baby-boom* generation have been reaching 65 years old. As this large generation grows older, it is reasonable to expect a massive increase in the utilization of medical resources. Simultaneously, recent important trends, such as the rise in obesity and chronic conditions prevalence, could add to the demand for treatments and weigh on many countries' fiscal outlook. In this context, it will be difficult to both supply a high quality of care and contain the expansion of costs.

This thesis aims to shed light on these challenges by focusing on the study of the demand for health care. We use a calibrated life-cycle modeling approach, drawing notably from the theoretical health-capital literature (Grossman, 1972; Wolfe, 1985; Ehrlich and Chuma, 1990) and its recent computational branch (Picone, Uribe and Wilson, 1998; Hall and Jones, 2007; Fonseca et al., 2009).

In the first chapter, we look at the demand for care in a public setting and ask whether it is desirable to allow waiting times to increase in order to contain ever growing costs of care.

If both the cost and the congestion of care are high, we aim to identify which policy stance is most beneficial to social welfare and economic activity. To that effect, we build the first theoretic health-capital demand model in a public health care setting. In doing so, we close an important gap in the existing literature: public health care systems are the norm rather than the exception in Western countries.[1] To replicate key features of public health care systems, our model showcases the importance of waiting times, rather than monetary prices, on the health care demand problem of individuals.

We then embed this demand model in the definition of an equilibrium, which arises when congestion of the health care system is such that the global demand equals the supply for care. The former is modeled as the sum of all agents among a distribution who choose to consume health care, weighted by the unit cost of their health level. The latter is modeled as the sum of public funds obtained from taxes collected from the distribution.

Calibrating our macroeconomic model with 2005 Québec data, we find that waiting times for care constitute a weak rationing device and lead to important welfare losses if used to contain costs. Our policy simulations reveal that a 50% permanent reduction in Québec waiting times can be achieved through a variety of fiscal measures that only moderately increase spending. While the choice of reform has important repercussions, all such policies result in major welfare gains, with agents willing to sacrifice over 5% of their post-policy consumption to leave the existing *status quo*.

In the second chapter, we use the framework developed in the first chapter to anticipate the impact of the aging baby-boom generation on costs, waiting times and health care use in public systems. Our methodology updates the distribution of agents according to the Institut de la statistique du Québec (ISQ) demographic scenarios, and projects policy responses and outcomes for Québec's health care system over the period 2005-2050. We include three main policy stances in our analysis: increasing the system's budget at a steady pace (Québec's current budgeting approach); adapting health care spending to maintain a steady congestion level, which holds access to care steady; and maintaining the share of health care costs to GDP constant.

Our results reveal a clear three-way arbitrage between providing a timely access to care, containing the rise of health care costs and adopting the newest available medical technology.

[1] Among the 34 OECD member countries, the share of health spending assumed by the public sector was greater than that of the private sector for 32 countries in 2010. The exceptions were the United States of America and Mexico (OECD, 2012).

Aging populations increase the demand for and average cost of treatments while reducing the quantity of working age person per elderly, which restrains ability to pay. As such, aging acts as an exacerbating factor on this arbitrage.

In the last chapter, we turn our attention to the demand for medical care in the United States, in which the share of health spending assumed by the public sector is the lowest among OECD countries. Current health care costs projections by the Congressional Budget Office (CBO) use exogenous cost increases applied to an otherwise standard microsimulation model. A main drawback of such methods is that it often leads to unconstrained growth, which is doubtful in the very long run. With co-authors Pierre-Carl Michaud and Raquel Fonseca, we aim to forecast the evolution of American health care costs, longevity and chronic conditions prevalence in a setting which is consistent with economic theory.

We develop a dynamic demand model close to that of Picone et al. (1998), which intrinsically allows agents to curb medical spending if it produces a lower marginal flow of expected utility than contemporary consumption. In contrast to the Québécois of our first two chapters, the agents of this model are constrained by monetary prices in their consumption of medical treatments. In our framework, three underlying forces shape the future environment of the American population. First, health depreciation increases over time, due to an increase in sedentary lifestyles, which results in a trend in the prevalence of chronic conditions in the population. Second, medical innovations improve the productivity of health care in transforming money into health. Finally, the real income of agents grows over time.

Looking ahead, we find that spending will continue to increase steadily if current trends continue at the same pace. By 2050, they would reach 22% of GDP. In our simulations, income growth is responsible for roughly three quarters of the increase of per capita spending, against one quarter for medical innovations. A main result of our simulations is that chronic conditions prevalence is most likely to continue increasing in the years to come, but will not constitute an important generator of medical spending. While curbing obesity and sedentary lifestyles would certainly lead to better quality of life, less chronic conditions and higher life expectancy for Americans, we do not find that it would reduce aggregate health care costs.

1 HEALTH CARE DEMAND AND IMPACT OF POLICIES IN A CONGESTED PUBLIC SYSTEM

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, ask whether allowing congestion to increase in order to contain ever growing health care costs is socially desirable, and predict the expected outcome of policies aiming at the reduction of congestion and costs of care. We develop a macroeconomic model in which agents who differ in terms of age, health capital and wealth choose whether or not to use a public health care system funded through income taxes. Calibrating our model with Québec data, we find that the aggregate demand for care is quite inelastic with respect to health care congestion. Waiting times for care thus constitute a weak rationing device and should not be used by welfare-maximizing governments to contain costs. On the brighter side, policy simulations reveal that a 50% permanent reduction in Québec waiting times can be achieved through a variety of fiscal measures that only moderately increase health care costs. While the choice of reform has important repercussions on the equilibrium, all such policies result in major welfare gains, with agents willing to sacrifice over 5% of their post-policy consumption to leave the existing *status quo*.

1.1 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 congestion[1], 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 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 an equilibrium concept calibrated with Québec data in order to quantitavely 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

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

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 for 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 treatments (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 they are elective in nature, they generate the longest waiting lists and are thus a natural source of public concern. However, two mains 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 quite 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 for 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 Québec data from the 2005 Canadian Community Health Survey (CCHS). Québec'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 Québec's health care system has a high congestion. Of the 12% of Québécois 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, Québec 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 Québec 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 similar permanent reduction of congestion (via the increased funding of health care) and lead to important welfare gains. However, the outcome differs depending on the choice of policy. For instance, raising income taxes induces the largest increase in the number of users as it incorporates no replacement to waiting times as a disencentive to using health care. As a result, health care spending need to grow more as a percentage of GDP to reduce waiting times than is the case with the introduction of user fees. That said, applying high fees is found to decrease the welfare of the poorest and sickest agents, who prefer the status quo. Alternatively, increasing the retirement age by five years produces a large GDP increase and a reduction in the cost of pensions but may be difficult to implement in practice, since Québécois have planned their life-cycle savings according to the existing scheme. We find that a milder combination of these policies also yields a permanent decrease in congestion and moderates the undesirable aspects of individual policies; it may thus be the most attractive and politically feasible policy scenario.

The remainder of the paper is organized as follows. Section 1.3 describes the theoretical demand model. Section 1.4 develops the equilibrium concept and calibration strategy needed to solve our model numerically. Section 1.5 presents the calibrated demand function and our simulations of exogenous congestion levels and policies. Finally, Section 1.6 puts our results in a practical policy perspective and suggests future avenues of research.

1.2 Québec's Health Care System

Under Canada's constitution, provinces are responsible for health care management. Therefore, there is not one national health care system, but ten independant provincial systems reporting to their provincial governments and a relatively small federal program covering First Nations and the military. In practice, however, the federal government exerts a strong influence on provincial health care delivery through the Canada Health Act (CHA) of 1984. This law states requirements that must be met by the provincial health care systems in order to qualify for the Canada Health Transfer (CHT), Canada's largest federal transfer program. In 2005-2006, the base year of our calibration, \$5.0B were transferred to Québec through the CHT, or 23% of the province's public health care spending (Department of Finance Canada, 2012). Among the law's requirements are universality, comprehensiveness, accessibility and public administration of care. In effect, the law ensures access to medically necessary health services without charges at the point of services for Canadians. The provinces can however define the basket of goods deemed necessary, which allows for differences in the range of services covered between provinces and over time (Madore, 2005).

Québec's health care supply is managed by the Ministère de la santé et des services sociaux (MSSS) at the provincial level and by 95 Centres de santé et de services sociaux (most of which include a hospital center) at the local level (MSSS, 2008). The public health insurance plan is administered by the Régie de l'assurance maladie du Québec (RAMQ). This insurance covers medical care free-of-charge, as well as dental services for children under 10 years of age and optometric services for persons under age 18 and age 65 and over (RAMQ, 2013a). The medical services covered include both those rendered by general practitioners and medical specialists. They include examinations, consultations, diagnostic procedures, therapeutic procedures, psychiatric treatments, surgery, radiology, anesthesia and basic hospital services (RAMQ, 2013b). Since 1997, a prescription drug insurance coverage is also required for all Québec residents, either through a private employer plan or through a mandatory basic public plan administered by the RAMQ (RAMQ, 2013c).

1.3 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:

$$\max \sum_{t=0}^{\infty} \beta^t E \left[u \left(c_t, L_t \right) \right], \tag{1.1}$$

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.

During a given period, the agent faces a certain amount $LT \left(H, \alpha\gamma \right)$ of lost time, in which H is his health and γ the congestion level of the health care system. 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 can be freely divided between working time WT and leisure L. Thus, a P-length period is divided as follows:

$$P = LT\left(H, \alpha\gamma\right) + WT + L \tag{1.2}$$

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

$$c \leq A + w \cdot WT\left(1 - \tau\right) \tag{1.3}$$

Since, in this paper, we model agents as self-employed, we reduce private pension plans 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:

$$c \leq A + \theta \text{ if } a \geq a_\theta, \tag{1.4}$$

where θ is the annual public transfer received by retired agents.[3] Applying this constraint, we

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

[3] While the study of retirement choices is not a major aim of this paper, including a simple form of

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 *value of death* $\underline{V}$. To enforce life to be desirable over death as a basic rule for the remainder of this paper, we assign

$$\underline{V} = {}^{u(0,0)}/_{1-\beta} \tag{1.5}$$

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 (1.1) to (1.5), the agent's maximization problem can be written as the following Bellman equation:

$$V(a, H, A) = \begin{cases} \max_{\alpha, WT, c} u(c, L) + \beta E\{V(a', H', A')\} & \text{if } H > \underline{H} \\ \underline{V} & \text{otherwise} \end{cases} \tag{1.6}$$

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

It is noteworthy that γ or the tax rate τ are fixed in equation (1.6). 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

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 misguided results. Second, it will later enable us to test how modifying pension generosity parameters affects the health care system.

[4] It 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.

agents cannot affect these parameters or predict their future variations. In section 1.4, 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 agent's stock of health-capital decreases at an uncertain depreciation rate δ each year, and may also be affected by the health care impact of treatment ψif he chooses to seek care. We formulate that the mean of the depreciation rate $\delta(a, \epsilon_\delta)$ is age-dependent and affected by a stochastic shock, ϵ_δ. Therefore, the agent cannot predict or choose his age of death. This formulation allows for appreciations of health without using health care, which happen when δ is negative. If the agent chooses to use health care, he will add $\psi(H, \epsilon_\psi)$ 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. The movement of the agent's health-capital stock is given by:

$$H' = \begin{cases} H \cdot (1 - \delta(a, \epsilon_\delta) + \alpha\psi(H, \epsilon_\psi)) & \text{if } H > \underline{H} \\ \underline{H} & \text{otherwise} \end{cases} \tag{1.7}$$

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:

$$A' = \begin{cases} (1+r)(A + w \cdot WT(1-\tau) - c) & \text{if } a < a_\theta \\ (1+r)(A + \theta - c) & \text{if } a \geq a_\theta \end{cases} \tag{1.8}$$

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

---

[5] As 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.

$$a' = a + 1 \tag{1.9}$$

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, 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 (1.6),

$$\beta \cdot \left(E\left[V_1'\right] - E\left[V_0'\right]\right) > u\left(c_0, L_0\right) - u\left(c_1, L_1\right) \tag{1.10}$$

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 care is uncertain and set in the future, while the cost is immediate and certain. In Appendix A.1, we further the analytic development of the demand model: we explore both sides of (1.10) 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, 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:

$$\begin{aligned} E\left[V_1'\right] - E\left[V_0'\right] &= \underbrace{\left[p_1 - p_0\right] \cdot \left(E\left[V_0' \mid H_0' > \underline{H}\right] - \underline{V}\right)}_{\text{Life-saving gain}} \\ &\quad + \underbrace{p_1 \cdot \left(E\left[V_1' \mid H_1' > \underline{H}\right] - E\left[V_0' \mid H_0' > \underline{H}\right]\right)}_{\text{Quality of life gain}} \end{aligned} \tag{1.11}$$

The first component of (1.11) is the expected gain caused 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. We thus name this component 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 name to this component as the *quality of life gain*. More discussion on these components is presented in Appendix A.1.

1.4 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 Québec'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 (1.7) 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 simple macroeconomic equilibrium to our demand 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 Québec data found in the Canadian Community Health Survey (CCHS) of 2005.

---

[6] We 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.

[7] Thus, 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.

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 wage parameters consistent with Québec's GDP and total health care cost for that year. Lastly, since the CCHS data is consistent with Québec'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.

1.4.1 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:

$$HC^D = HC^D\left(\alpha_1, \alpha_2, ..., \alpha_{N-1}, \alpha_N, H_1, H_2, ..., H_{N-1}, H_N\right) \tag{1.12}$$

The public supply of health care is defined as:

$$HC^S = \sum_{i=1}^{N} \tau \cdot w \cdot WT_i - \sum_{i=1}^{N} 1_{a_i \geq a_\theta} \cdot \theta \tag{1.13}$$

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 (1.13) is sufficient for the study of health care demand, the

[8] For instance, in the case of Québec, 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.

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 1.3, 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.

1.4.2 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:

$$u(c, L) = \frac{\left(c^{1-\omega} L^{\omega}\right)^{1-\rho} - 1}{1-\rho} \tag{1.14}$$

This utility function is standard in modern macroeconomic models because of two useful properties. It has a constant Arrow-Pratt relative risk aversion coefficient of ρ and a constant 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 1.3:

$$LT(H, \alpha\gamma) = P \cdot \left(1 - H + \alpha\gamma \left(\frac{H - \underline{H}}{\overline{H} - \underline{H}}\right)\right) \tag{1.15}$$

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. When using the health care system, 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 Québec 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.

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:

$$H'\left(a, H, \epsilon_{\delta}, \epsilon_{\psi}\right) = \begin{cases} H \cdot \left(1 - \left(\frac{a-\underline{a}-\epsilon_{\delta}}{\overline{a}-\underline{a}}\right)^{b_1} + \alpha\left(\frac{\left(\overline{H}-H\right)}{\left(\overline{H}-\underline{H}\right)} \cdot \frac{(1+\epsilon_{\psi})}{b_2}\right)^{b_3}\right) & \text{if } H > \underline{H} \\ \underline{H} & \text{otherwise} \end{cases} \tag{1.16}$$

where $\underline{a}$ and $\overline{a}$ are the minimum and maximum ages considered in the calibration and $\overline{H}$ is the health capital corresponding to full health. 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 $\overline{a}$, 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 - for whom $\left(\overline{H} - H\right)$ is high - 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 (1.12), which is needed to obtain equilibrium and isolate the congestion parameter γ. We formulate it as a simple linear transformation of the sum of users, weighted by the severity of their sickness:

$$HC^D = d \cdot \sum_{i=1}^{N} \alpha_i \cdot \left(\overline{H} - H_i\right), \tag{1.17}$$

where α_i is the optimal choice of α for agent i resulting from the solution of our demand model. It

Table 1.1 Portrait of CCHS 2005 Québec Respondents

	Group	**n** Unweighted	Weighted	**Users**[9] (Weighted)
Age groups	18 to 39	8 911	2 243 093	47.6%
(avg. of 46.3)	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[10]	0.45 to 0.60	2 345	449 325	75.7%
(avg. of 0.81)	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%**

is noteworthy that this form implies that the cost of treating a given patient is linearly increasing in his sickness level $\left(\overline{H} - H_i\right)$. The only parameter of this function, d, is the marginal cost, in dollars, of treating illness. It enables us to express the aggregate demand for care in dollar terms, and thus equate the demand to the supply, already specified by (1.13).

1.4.3 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 A.2.

In total, 26,501 respondents from the province of Québec 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 Québécois, 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 1.1 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.

1.4.4 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 Imrohoroğlu, 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 (1.5) 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 sensitivity analysis of the risk aversion parameter, presented in Appendix A.6, 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 (1.15), 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 $\overline{a} = 99$, 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.[11]

For public pensions, we base ourselves on two basic components of Canada's Old Age Se-

[11]The 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.

curity 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 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 Québec. 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 Québec, from the total cost of health assumed by the public sector of \$ 21.4B, as reported by the Canadian Institute for Health Information. The GDP of 2005 is \$ 272.0B and corresponds to the income-based, market prices annual GDP, as published by Statistics Canada. The taxation parameter is set to the ratio between the cost of the public expenses and the GDP, in order to enforce a balanced budget:

$$\tau = \frac{HC^S + \sum_{i=1}^{N} 1_{a_i \geq a_\theta} \cdot \theta}{GDP} \tag{1.18}$$

This calibration procedure is explained in Appendix A.4. The full calibration is presented in Table 1.2.

1.5 Numerical Results

1.5.1 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. For a given health level, we find that older agents always display a positive net gain of using care, resulting a smooth age threshold determining health care use. Figure 1.1 presents this optimal threshold for all health capital values . All agents below the line, which are both young and in good health, optimally choose $\alpha = 0$. All agents on the line or above it choose $\alpha = 1$. This demand fits the trends observed in Table 1.1, in which we saw that health care use was stronger in older and unhealthy individuals in Québec

Table 1.2 Calibration

Fixed Parameters	
ω: Leisure preference	0.67
ρ: Arrow-Pratt relative aversion coefficient	0.5
$\underline{H}$: Minimum health capital (death threshold)	0.4
$\overline{H}$: Maximum health capital	1.0
$\overline{A}$: Maximum savings	\$500,000.00
$\underline{a}$: Minimum age	18
$\overline{a}$: Maximum age	99
a_θ: Age of retirement	65
θ: Annual public transfer to retirees	\$10,319.50
τ: Income tax rate	0.10674
Dimension of vector H	13
Dimension of vector A	51
Dimension of vectors ϵ_δ and ϵ_ψ	15
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
d: 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

in 2005.[12]

To assess how savings affect health care consumption in a public system, we computed the threshold for the lowest and highest savings levels included in our calibration. We found that they were perfectly indissociable: the youngest age for which agents use care is exactly the same for the poorest and wealthiest agents. This result suggests 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 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.

Figure 1.2 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 A.3). 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 in good health in Figure 1.2. 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 θ is set to 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.[13]

---

[12] What 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.

[13] We recognize that this occurs because of the absence of a bequest motive in our model. If such a feature 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.

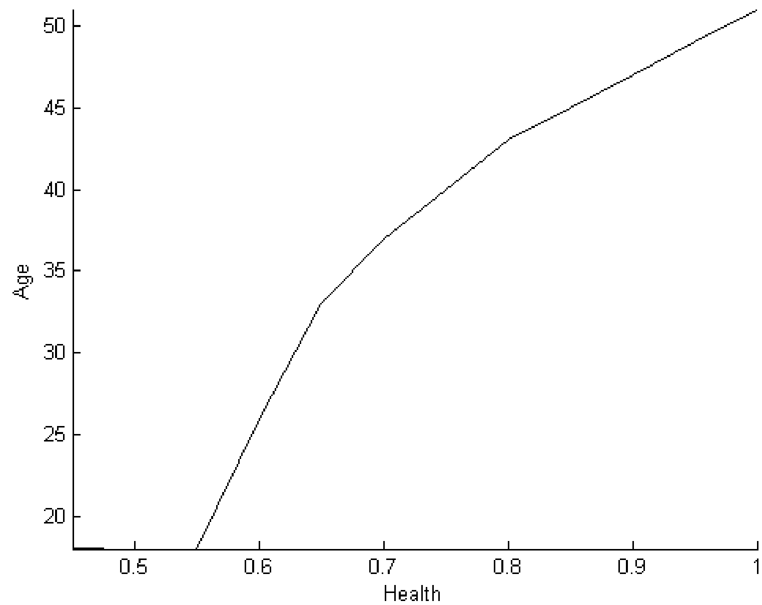


Figure 1.1 Minimum Age for Which Agents Use Health Care (for all savings values)

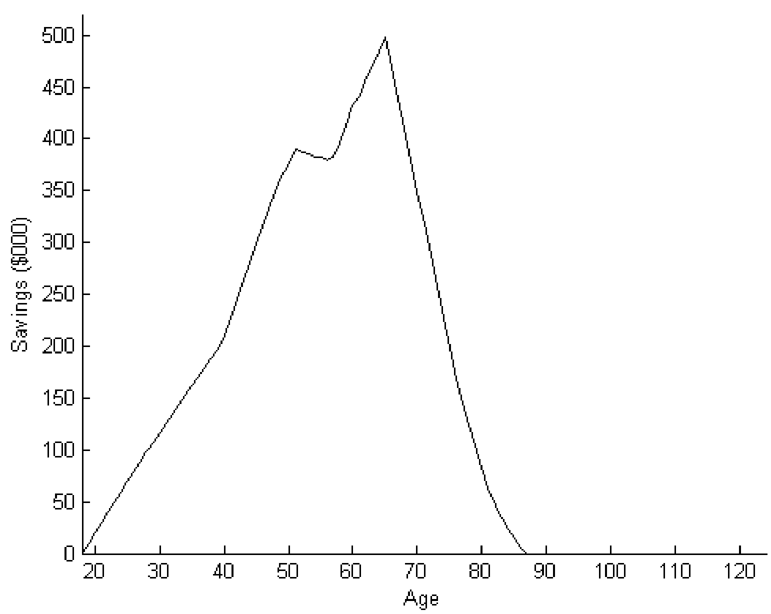


Figure 1.2 Age vs. Expected Savings

1.5.2 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 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 1.2 intact. Figure 1.3.1 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 1.3.1. In the neighborhood of current congestion γ^*, the elasticity of demand is -0.195, indicating that a 1% exogenous increase in Québec waiting times would result in only 0.2% less users of care.[14]

Figure 1.3.2 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.

1.5.3 Predicted Impact of Policies

In Table 1.3, 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 necessary to obtain

---

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

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

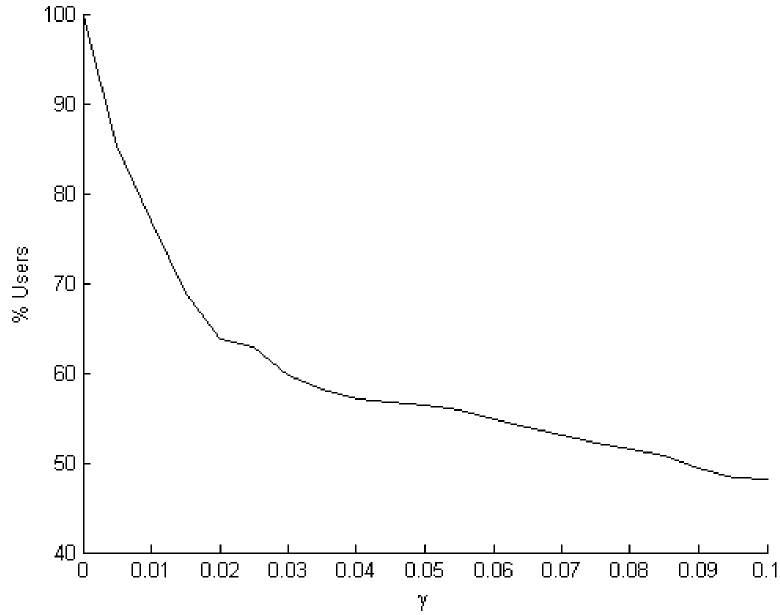


Figure 1.3.1 Proportion of Users

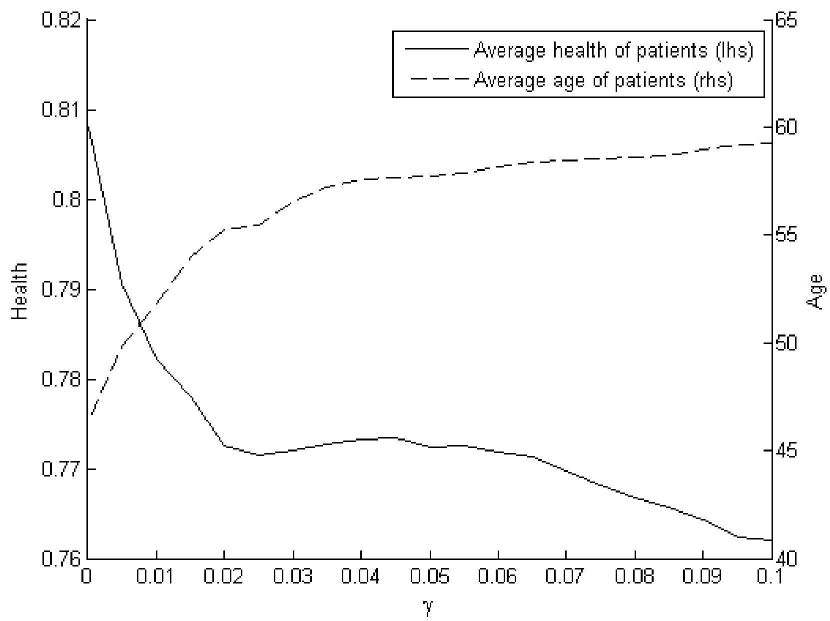


Figure 1.3.2 Profile of Patients

Figure 1.3 Simulated Impact of Congestion on Health Care Demand

the same average value as the base case.[15] 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.[16]

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:

$$c = \begin{cases} A + w \cdot WT\left(1-\tau\right) - Fee - \frac{A'}{(1+r)} & \text{if } a < a_\theta \\ A + \theta - Fee - \frac{A'}{(1+r)} & \text{if } a \geq a_\theta \end{cases} \quad \text{if } \alpha = 1 \tag{1.19}$$

In Québec'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

[15]This adjustment is obtained by finding the factor k such that

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

where $\left(\hat{c}_t,\hat{L}_t\right)$ 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.

[16]We 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 A.3).

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 is not 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 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 A.1.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 Québec'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 Québec. 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 Québec'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.

Table 1.3 Results of Simulated Policies

1.3.A. Introduction of user fees

	Base Case $Fee = 0$	**1** $Fee = 100$	**2** $Fee = 500$	**3** $Fee = 1000$
Users (%)	53.7	57.7	61.8	72.1
γ^{*} [a]	100	54.7	36.7	10.9
GDP	100	102.7	102.6	103.1
HC^S (% GDP)[b]	7.0	7.2	7.7	8.6
Fees (% HC^S)	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
H 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
H 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

[a] Equilibrium congestion of care.

[b] Health care spending.

[c] Adjustment to the consumption flow of agents necessary to obtain the same mean value as the base case (see Footnote 15).

1.3.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
GDP	100	102.0	101.8	97.0
HC^S (% GDP)	7.0	8.1	9.5	5.9
Fees (% HC^S)	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
ΔConsumption	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
H 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
H 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

1.3.C. Changes in age of retirement

	Base Case	**7**	**8**	**9**
	$a_\theta = 65$	$a_\theta = 67$	$a_\theta = 70$	$a_\theta = 60$
Users (%)	53.7	61.1	72.3	33.5
γ^*	100	43.6	19.9	363.9
GDP	100	103.3	108.1	96.3
HC^S (% GDP)	7.0	7.6	8.4	5.2
Fees (% HC^S)	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
ΔConsumption	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
H 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
H 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

1.3.D. Changes in pension generosity

	Base Case	**10**	**11**
	$\theta \cdot 1$	$\theta \cdot 0.9$	$\theta \cdot 1.1$
Users (%)	53.7	59.0	46.7
γ^*	100	46.9	159.4
GDP	100	102.7	97.7
HC^S (% GDP)	7.0	7.4	6.5
Fees (% HC^S)	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
ΔConsumption	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
H 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
H 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

1.3.E. Combination of policies

	Base Case	**12**
	$fee = 0$	$fee = 250$
	$\Delta\tau = 0$	$\Delta\tau = 0.0075$
	$a_\theta = 65$	$a_\theta = 67$
Users (%)	53.7	72.3
γ^*	100	18.0
GDP	100	103.3
HC^S (% GDP)	7.0	8.7
Fees (% HC^S)	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
ΔConsumption	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
H 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
H 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

1.6 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 of health care congestion. On the one hand, users of health care spend a large amount of time waiting in queues 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. In future work, it would be interesting to extend our model to study the effects of waiting times on prevention and population health dynamics in more detail.

By simulating how our model reacts to a variety of congestion levels, we find that the demand for care is quite inelastic. In the case of Québec, 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 important health care costs increases.

Our policy simulations reveal that, in the case of Québec, 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 realized.

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 are homogeneous in productivity and 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 for specific subgroups. For instance, the introduction of user fees would likely prove less desirable for unemployed agents, while tax increases to finance health care supply may be exceedingly costly for agents that are both healthy and poor. 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 evaluating 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 access to care concerns in the United States is answered by reforms increasing the public involvement in health care and congestion and costs force many OECD public systems to consider partial privatization. In recent years, for-profit private clinics started to emerge in Québec to supply some services, such as diagnostic testing for colorectal cancer, in competition with the public system. While the stringent provisions of the Canada Health Act restrict the extent of services provided by private competitors, their numbers have been growing steadily. From a public system presenting long waiting queues, our results suggest that an unrestricted 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.

2 WAITING TIMES FOR MEDICAL CARE IN AN AGING POPULATION CONTEXT: THE CASE OF QUÉBEC

We study the impact of population aging on Québec's public health care system, focusing on waiting times, utilization and costs. To that effect, we build an integrated model that incorporates demographic changes and projects policy responses and health care demand over the period 2005-2050. We calibrate our model with Québec data and use demographic projections by the Institut de la Statistique du Québec as inputs. Our simulations indicate that Québec's current budgeting approach of increasing health spending by 5% each year will result in much longer waiting times by 2030. To maintain waiting times at current levels, we find that the share of health care spending in the economy would need to double by 2050. As the baby-boom generation gets older and consumes more care, it will be difficult for policymakers to meet the population's expectations of a moderate fiscal burden, a timely delivery of care and an access to the most effective medical technology. Taking on the challenge of aging will constitute a balancing act between these objectives.

2.1 Introduction

So far, two distinct literature branches have studied the impact of aging on health expenditures. The first has debated the "red herring" argument that longevity gains have but a moderate impact on health expenditure throughout the life-cycle. According to this argument, since a large share of health related costs occur at the proximity of death, the longevity component of aging merely delays these costs, with the notable exception of long-term care costs (Zweifel et al.,1999; Seshamani and Gray, 2004; Häkkinen et al., 2008). The second branch, which is more relevant to this article, has attempted to project the impact of the large baby boom generation reaching old age over the coming decades on population health and costs by extrapolating observed spending per age data. It has found that this transition will cause substantial increases in spending and pose a financial challenge (Shoven et al., 1994; Reinhardt, 2000; Bagust et. al. 2002; Keehan et. al., 2008). For instance, Keehan et al. (2008) combine actuarial projections and empirical cost models to predict US health care costs, and predict a 3% increase in the US share of GDP spent on health between 2007 and 2017.

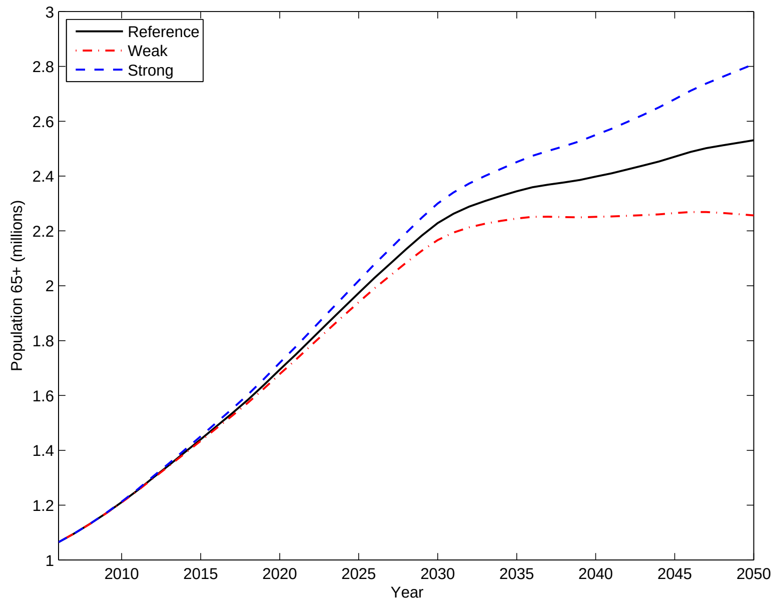
In addition to health care costs, the added demand for care brought forward by aging baby boomers is likely to affect waiting times for care considerably in many countries. In a study of waiting lists for several surgery procedures in the United Kingdom, Martin and Smith (2003) found that the elasticity of demand with regard to the length of lists is quite low, with an overall elasticity of -0.09. Similarly, in our first chapter, we built a public health care model calibrated with Québec data and found an elasticity with regard to waiting queues of -0.19. Given this low elasticity of demand, efforts to moderate the expansion of costs will likely result in important increases in waiting times and a severe deterioration of the access to health care in many western countries.

In this paper, our main objective is to provide insight about the *full* impact of the aging baby-boom generation on public-funded health care systems. In addition to costs and their share of GDP, we attempt to anticipate the evolution of waiting times and access to care. Secondarily, we aim to anticipate the repercussions of different policy stances on these outcomes and quantify the scope of uncertainty associated with demographic forecasts.

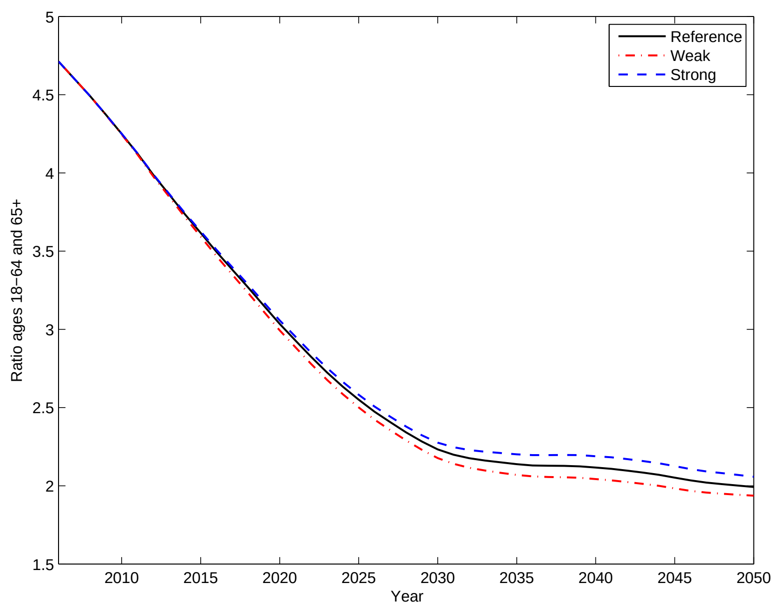
We build an integrated model that incorporates demographic changes and endogenously projects policy responses and outcomes for Québec's health care system over the period 2005-2050. The case of Québec is particularly relevant for such analysis for many reasons. First, it has a fully public-funded universal health care system, as imposed by the Canada Health Act. Second, its health care system already presents long waiting times for care, which have been a policy concern for over a decade. For instance, 47% of Québécois who claimed not to receive the appropriate amount of care in 2005 in the Canadian Community Health Survey (CCHS) asserted that waiting times were to blame, more than any other cause. Third and foremost, Québec will experience a particularly severe aging of its population in the decades to come. According to the Institut de la Statistique du Québec (ISQ), its elderly population will more than double between 2006 and 2030 (Figure 2.1). Meanwhile, the working age population per individual aged 65 and over will shrink from 4.7 to 2.2.

Our model has four modules. First, the Population module updates the distribution according to the Institut de la statistique du Québec (ISQ) demographic projections. We include the ISQ's three main scenarios (Reference, Weak and Strong) in our analysis to account for the uncertainty associated with future demography. The Policy module then determines the income tax rate chosen by the government to fund its programs, which determines the health care supply in a given year. We include three policy stances: 1) Increasing the system's budget at a steady pace (Québec's current budgeting approach); 2) Adapting health care spending to maintain a constant congestion level, which enables a steady access to care; and 3) Maintaining the share of health care costs in the GDP constant. The Health Care module projects the distribution's annual demand of health care, based on the life-cycle calibrated demand model set forward in our first chapter. Simultaneously, this module finds the equilibrium congestion of care that equates the demand and supply. Finally, the Results module computes the equilibrium outcomes for the year.

Along with demography, two main trends have an important impact on our results: labor productivity gains and technological progress in the provision of health care. The former increases the aggregate ability to pay for care and alleviates aging's impact on the fiscal outlook. The latter has more ambiguous implications. Medical progress is generally seen as increasing the effectiveness of treatments, but at higher costs. For our three main policy scenarios, we impose this interpretation. However, new discoveries can also reduce costs for existing treatments (Jones, 2002). As aging imposes a fiscal pressure on governments, policymakers could ration the adoption of pricey new treatments and favor the adoption of those that reduce costs. We thus include a fourth policy scenario, which combines this rationing of health care technology with maintaining the share of health spending in the economy.



i. Population Aged 65 +



ii. Ratio of Population Aged 18-64 and 65+

Figure 2.1 Overview of Population Aging in Québec, ISQ Demographic Scenarios

Our first policy simulation reveals that an institutional status quo of increasing the health care's system nominal funding by 5% annually will be insufficient to maintain waiting times at or below their current level over the next decades. For all demographic scenarios, waiting times would increase by over 75% by 2025. Our second policy simulation shows that maintaining waiting times constant in the population aging context of Québec would be quite costly, and double the share of health care spending to GDP by 2050. Finally, our last two policy simulations indicate that maintaining health care costs at their 2005 share of GDP, if appealing from a public finance perspective, would lead to an explosion of waiting times unless the government aggressively rations the adoption of technological progress.

Aging populations, such as that of Québec in the decades to come, increase health care demand and the average cost of treatment while reducing the quantity of working age person per elderly, which restrains GDP growth and ability to pay. As such, aging creates a tremendous pressure on public health care systems. Despite large differences among the ISQ projection scenarios parameters, they lead to very similar results for the majority of our simulations. This arises because the inevitable aging of Québec's baby-boom generation is such a dominating trend that all demographic scenarios lead to similar age groups shares in the total population for the years to come.

The remainder of the paper is organized as follows. Section 2.2 describes our methodological framework. Section 2.3 presents our calibration and discusses our parametric choices. Section 1.5 presents our simulation results. Finally, Section 2.5 puts our results in a practical policy perspective.

2.2 Methodology

2.2.1 Overview

This section describes the model used to simulate the impact of population aging on Québec's public health care system. Our starting point is the year 2005, in which the initial distribution and outcome are obtained from the calibrated model of our first chapter. For the year 2006 onward, our model consists of four distinct modules, synthesized in Figure 2.2.

First, the Population module updates the 2005 distribution according to ISQ's demographic scenarios. Second, the Policy module determines the taxation rate and medical tech-

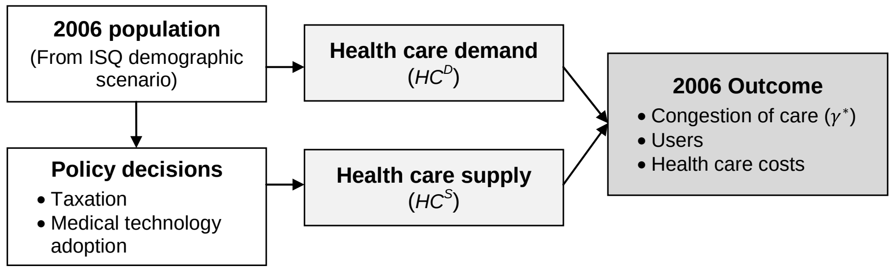


Figure 2.2 Model Overview for Year 2006

nology adoption. These are chosen by the government following a policy rule after observing the population and predicting its demand for medical care. Third, the Health Care module projects the health care demand of the population by solving a life-cycle model, and finds the equilibrium waiting times (the price of using a public health care system) such that medical services provided by the public system clear. Finally, the Policy outcome module extracts the equilibrium results for that year. The same process is then repeated until year 2050, the last year of our simulations. We now describe each module in further detail.

2.2.2 Population Module

The Population module updates the distribution used in the first chapter, which replicate Québec's 18+ population in 2005, for years 2006 to 2050. The initial distribution was created to replicate the number of Québécois of each age, health level and savings (the three state variables of our life-cycle health care demand problem). The age and health distribution was obtained through manipulations of the 2005 Canadian Community Health Survey (CCHS) data, weigthed to match ISQ population data. The savings of the respondents, which was not asked in the CCHS questionnaire, was extrapolated from the solution of the optimization problem. Table 2.1 presents a summary of the initial distribution.

ISQ demographers produce nine population scenarios for long term analysis, three of which are considered *reference* scenarios. The "A - Reference" scenario consists of mortality, fertility, international and inter-provincial migration assumptions based on average trends. The "D - Weak" and " E - Strong" scenarios present more conservative and optimistic assumptions in terms of the total population they produce, respectively. They are intended to reflect the order of magnitude of the uncertainty of future demographic trends in the province (ISQ, 2009).

Table 2.1 Portrait of 2005 Distribution

Age	N (Thousands)	Average Health[a]	Average Savings ($)
18 to 19	193	0.81	5,117
20 to 29	1,040	0.83	38,954
30 to 39	1,010	0.83	118,876
40 to 49	1,274	0.81	242,111
50 to 59	1,116	0.80	355,146
60 to 69	744	0.79	435,431
70 to 79	465	0.77	228,690
80 to 89	163	0.76	35,904
90 and +	26	0.77	0
18 and +	**6 032**	**0.81**	**215 941**

[a]Health values of the distribution are obtained through manipulations of two qualitative questions about the perceived health of CCHS respondents. Values of 0.45 to 0.60 correspond to the "Poor" (lowest) health label, 0.65 to 0.80 to "Fair" and "Good" labels, and 0.85 to 1.00 to "Very good" and "Excellent" labels. See Appendix A.2 for more details on this variable.

The main assumptions of these scenarios are presented in Table 2.2. The extent of population divergence between the scenarios is quite large, with over 3 million separating the Weak and Strong scenarios in year 2051. In our simulations, we use the Reference scenario as a baseline.

In year t, let us denote $f_t(a, H, A)$ the number of Québécois of age a, health H and financial assets A, and $N_t(a)$ the total population of age a. $N_{2005}(a)$ is obtained directly from the initial distribution. For 2006 onward, $N_t(a)$ is taken from one of ISQ's three reference scenario. This module updates the distribution with the ratio of the population of all ages in

Table 2.2 Main ISQ Demographic Projections Parameters

	Reference Scenarios		
	Weak	**Reference**	**Strong**
Life expectancy in 2051 (men / women)	83.0 / 86.5	85.5 / 89.0	88.0 / 90.5
Total fertility rate of females in 2013	1.50	1.65	1.85
Net international migration in 2015	30,000	40,000	50,000
Net inter-provincial migration in 2013	-16,000	-10,000	-4,000
Total population in 2051 (millions)	**7.86**	**9.17**	**10.96**

Source: ISQ (2009).

$t+1$ and t:

$$f_{t+1}(a,H,A) = f_t(a,H,A) \cdot \frac{N_{t+1}(a)}{N_t(a)} \tag{2.1}$$

With this updating procedure, the ISQ scenarios are the only drivers of population change, while the savings and health distribution among a given age group remains static over time. In this article, the distribution of agents serves as an input to predict the optimal use of care given a congestion level. One of our key results of the first chapter was that wealth has no effect on the optimal medical care consumption in a public setting. Because of this, maintaining financial assets constant has no impact on our simulated results.[1] However, health plays a key role in medical consumption decisions, but its future movements are uncertain. Looking at historic CCHS data, we find no evidence of a trend in the self-reported health of Québécois (see Appendix B.1). Thus, if health was trended over time, it would raise the question of whether our results are driven by these predictions and whether they are plausible. In Section 1.5, we will discuss the implications of this choice when we analyze our simulated results.

The algorithm used in the first chapter to weight the 2005 distribution led to unrealistic discrepancies for some number of age groups with ISQ's 2006 projections. To minimize the impact of these distribution differences on our early results, we smooth implausible discrepancies between $N_t(a)$ and $N_{t+1}(a+1)$. This allows for the distribution to quickly catch up with ISQ's demographic projections, while preventing bumpy results for 2006. An overview of the divergence and the methodology used for this smoothing are presented in Appendix B.2.

2.2.3 Policy Module

The government taxes work income at rate τ_t to fund two programs: the public health care system and the public pensions for the population over 65. Each year, once the population is updated, the Policy module determines the taxation rate and medical technology to be adopted for according to a policy rule.

We assume that everyone aged 18 to 64 has the same productivity in a given year (w_t) and works the same number of hours throughout a year ($\overline{WT}$). Thus, the size of the economy

---

[1] Financial assets do play an important role when considering the welfare impact of policies such as tax hikes and changes in the generosity of pensions, which are not the topic of this article.

is directly determined by the demographic projection and labour productivity:

$$GDP_t = \sum_{a=18}^{64} \sum_{H} \sum_{A} f_t(a, H, A) \cdot w_t \cdot \overline{WT} \tag{2.2}$$

Similarly, public pension costs PP_t are given by the elderly population in t:

$$PP_t = \sum_{a \geq 65} \sum_{H} \sum_{A} f_t(a, H, A) \cdot \theta \tag{2.3}$$

where θ is the annual public transfer received by retired agents. As in the first chapter, public pensions correspond to the minimum public transfers available to all retired Canadians of 65 years of age. Such transfers correspond to the main components of Canada's Old Age Security Program: the Old Age Security Pension (OASP) and the Guaranteed Income Supplement (GIS). Both of these programs' rates are reviewed quarterly with the Canadian consumer price index (CPI) to maintain a constant purchasing power over time. We thus hold θ at its 2005 level for our simulations. With GDP and public pension costs given by (2.2) and (2.3), the health care supply HC_t^S (expressed in 2005 dollars) is directly determined by the income tax rate τ_t:

$$\frac{HC_t^S}{GDP_t} + \frac{PP_t}{GDP_t} \equiv \tau_t^{HC} + \tau_t^{\theta} = \tau_t \tag{2.4}$$

Since income tax is the only source of government revenue in our model, τ_t^{HC} and τ_t^{θ} are the share of GDP allocated to health care and public pensions, respectively.

We assume that the government perfectly anticipates the GDP, the population distribution and its demand for care. Thus, by choosing tax rate τ_t, it determines the supply of care HC^S and can reach a number of policy objectives. We consider three main policy scenarios in this module, each reflecting a plausible approach to health care financing in the decades ahead: 1) increasing health care costs at a steady pace - Québec's status quo -, 2) funding health care to maintain a stable congestion of the health care system, and 3) maintaining the ratio of health care costs and GDP constant. With equality (2.4), all policy stances impose a balanced budget in the years to come. We discuss the possibility of using deficits as a financing source in the concluding section of this article.

In the first three policy scenarios, medical improvements are adopted automatically by the health care system, which increases the cost of treating patients.[2] To better understand the role of technological improvements on the provision and costs of health care, we include a fourth policy scenario, which combines scenario 3 with a rationing of the adoption of medical technology.

We present each scenario in further detail in Section 2.4, before describing their simulated outcome in Québec's demographic context.

2.2.4 Health Care Module

The Health Care module is the core of our model. It determines aggregate supply and demand for health care, as well as aggregate conditions that reveal waiting times and health care consumption in year t.

Aggregate supply

The health supply in year t is given directly from the distribution and the government's choice of τ_t described in the policy module. With respect to equations (2.2) to (2.4) , the aggregate supply HC_t^S is given by:

$$HC_t^S = \tau_t \cdot GDP_t - PP_t \tag{2.5}$$

Aggregate demand

The aggregate demand prediction is based on the dynamic demand model set forward in the first chapter. This life-cycle model proved useful to understand the motivations of health care users and predict the impact of healthcare financing policies in a static population context. Results obtained with the calibrated model indicated that agents optimally choose to demand health care when they are older and in bad health, while wealth and retirement have no impact on their demand. Also, the elasticity of demand with regard to waiting times is quite low. We found a simulated elasticity of -0.19, which means that a 1% increase in waiting times should be expected to reduce the number of users of health care by only 0.2%.

---

[2] Accordingly, the impact of medical treatments on the health of patients (medical productivity b_t), and the marginal cost of care (d_t), increase steadily with t at rates z_b and z_d. We discuss these parameters and the role of medical technology in further detail in Section 2.2.4, Section 2.3.3 and Appendix B.3.

This demand model combines the well-known health-capital concept of Grossman (1972) with key elements of comprehensive public health care systems. In particular, it presents zero monetary prices for care and introduces a non-monetary cost to users in the form of lengthy waiting times. Also, it models health care demand as a binary choice $\alpha \in \{0, 1\}$, rather than a continuous choice. The rationale behind this feature is that patients of public health care are not rationed by costs,[3] and therefore automatically accept the best treatment available, as prescribed by their physician. Their choice is thus reduced to seeking medical care ($\alpha = 1$) or not ($\alpha = 0$). Medical care is desirable because a good health increases the agent's expected longevity and quality of life, but comes at a (non-monetary) cost. Agents must wait a portion of the period - the year t - in line, and thus sacrifice income gains and leisure.[4] This sacrifice is increasing in the congestion of the health care system, summarized by variable γ_t. It is also increasing in the agent's current health H_t, since medical care is allocated in priority to patients in bad health. The model's equations are presented in detail in Appendix B.3.

It is of note at this point that we do not incorporate the life-cycle model's working time predictions in our calculation of GDP and HC^S. As mentioned in the conclusion of our previous article, the modeling of labor presents a number of important weaknesses. Notably, the model has no unemployment and makes no account of sick days, which insure many Québec workers against the income losses of waiting times. For these reasons, the model likely overstates the impact of waiting times on labour decisions and we deem more plausible, at the macro-economic level, to hold working times per worker constant than to use the model's predictions. We however maintain the life-cycle model unchanged in the Health Care module since, for workers without sick days (mainly low wage and self-employed workers), lost wages due to waiting in line have an important impact on their decision to visit a physician.

In this paper, the dynamic demand model provides the demand of each individual conditional to a given congestion level, $\alpha\,(a, H, A \mid \gamma_t)$. We then compute $HC_t^D(\gamma_t)$, the aggregate demand for care (expressed in 2005 Canadian dollars) given the distribution in t and congestion

[3] In the case of Québec, the user costs are zero if the services provided are considered medically necessary for the purpose of maintaining health, preventing disease or diagnosing or treating an injury, illness, or disability, in accordance with the Canada Health Act.

[4] The model explicitly emphasizes the deterring effect of waiting *queues* (rather than lists) on the demand of public care. Queues induce a direct opportunity cost, since agents waiting in lines cannot earn work income or enjoy leisurely activities.

γ_t:

$$HC_t^D(\gamma) = \sum_a \sum_H \sum_A f_t(a, H, A) \cdot \alpha(a, H, A \mid \gamma_t) \cdot d_t \cdot \left(\overline{H} - H\right) \tag{2.6}$$

where $\overline{H}$ represents full health and d_t determines the unit cost of treating illness $\left(\overline{H} - H\right)$. The parameter d_t enables us to express the aggregate demand for care in dollar terms, and enables us to equate the demand to the supply, already specified by (**2.5**).

Equilibrium

In this 100% publicly-funded health care framework, waiting times constitute the only *price* faced by users. In a given year, if the aggregate demand for care is high relative to the supply, the health care system becomes more congested, and waiting times increase until an equilibrium is reached. Accordingly, we postulate that the equilibrium is obtained when the overall congestion of the system is such that $HC_t^D(\gamma_t) = HC_t^S$. In this module, we iterate on γ_t, solve the life-cycle demand model and compute the aggregate demand $HC_t^D(\gamma_t)$ until such an equilibrium is obtained. Despite a low elasticity of demand, the equilibrium exists and is unique for each year, since HC_t^S is exogenous with regard to waiting times. We denote γ_t^* the equilibrium congestion level.[5]

2.2.5 Policy Outcome Module

The fourth and last module computes the results of interest of our simulations for each year, ISQ reference scenario and policy. First, the equilibrium congestion of care γ_t^* is a key variable, as it enables us to predict how future waiting times will evolve with aging, in comparison with contemporary waiting times. Second, we compute the share of total population that use health care over time, which is indicative of access to care. As discussed earlier, the basic dynamic demand model found that the demand for care is higher in older agents. If we observe a decrease in the share of population that choose $\alpha = 1$ as the elderly population increases, for instance, this would mean a significant worsening of access to care conditions. Finally, we compute the evolution of aggregate health care costs, both in constant dollars and as a share of GDP. In the context of an aging population, these provide a complementary view, because the size of the economy is directly linked to demographic projections.

---

[5] An exception to this occurs when $HC_t^S > HC_t^D(\gamma_t)$ for all positive values of γ_t, i.e. when the supply for care is sufficiently high to eradicate waiting times. In such cases, we assign $\gamma_t^* = 0$.

2.3 Parametric Choices and Calibration

2.3.1 Health Care Module Parameters

The bulk of the parametric choices that need to be specified in our model are at the Health Care module level, and are directly obtained from our first chapter. In that effort, we devised a calibration strategy that allowed us to simultaneously replicate essential 2005 moments: the precise number of health care users in the 2005 edition of the CCHS survey, the total cost of health care and GDP of the province, and the mortality rates and health expectancy published by the ISQ.

The calibration was then found to correctly reproduce the higher usage of health care by agents that are older and in bad health. These were observed in the CCHS data, but were not directly imposed by the calibration strategy or the model. It also produced an elasticity of demand with regard to waiting times of -0.195, within the scope of the results of Martin and Smith (2003), which estimate the elasticity of demand with regard to the length of waiting lists using panel data methods. [6]

Since we use 2005 as the starting point of our analysis, we use the same parameters in this paper, which are presented in Table 2.3. The working time of the 18-64 population, $\overline{WT}$, is set to 1,614 hours per year, the average working time found in the first chapter.

2.3.2 Labor Productivity Growth

Two sets of parameters remain to be established for our dynamic simulations. The first is the annual growth of labor productivity parameter w_t. It has important repercussions on our simulated results. Large productivity growth over the years to come would increase the ability to pay for health care for all demographic scenarios. It would thus allow Québec to significantly increase the supply for care while maintaining relatively low health care to GDP ratios. As such, it would compensate for the growing health care demand that will likely be induced by the aging

[6] It should be noted that waiting lists do not ration demand in the same way as queues, as was formalized by Lindsay and Feigenbaum (1984). On the one hand, queues force individuals to wait in person at a given location, while on the other hand, waiting lists allow individuals to 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. Because of this fundamental difference, elasticities with respect to waiting lists and waiting queues do not have the same interpretation.

Table 2.3 Calibration

Policy Module Parameter[a]	
$\overline{WT}$: Working time in hours	1613.5
Dynamic Demand Model Parameters[a]	
γ^*_{2005}: Initial equilibrium congestion rate	0.0671201
b_{2005}: Expected impact of treatment coefficient	0.0666
d_{2005}: Initial cost of treatment parameter	\$25,562.8
w_{2005}: Initial hourly productivity	\$ 33.33
τ_{2005}: Initial income tax rate	0.10674
ω: Leisure preference	0.67
ρ: Arrow-Pratt relative aversion coefficient	0.5
$\underline{H}$: Minimum health capital (death threshold)	0.4
$\overline{H}$: Maximum health capital	1.0
$\overline{A}$: Maximum savings	\$500,000.00
$\underline{a}$: Minimum age	18
$\overline{a}$: Maximum age	99
a_θ: Age of retirement	65
θ: Annual public transfer to retirees	\$10,319.50
Dimension of vector H	13
Dimension of vector A	51
Dimension of vectors ϵ_δ and ϵ_ψ	15
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
Dynamic Trends Parameters	
z_b: Medical effectiveness trend	1.018
z_d: Medical cost trend	1.018
z_w: Labor productivity trend	1.0096

[a]Obtained from the first chapter.

population. On the contrary, a low productivity growth would mean that health expenditure will need to increase significantly as a share of the economy to allow the province to provide timely treatments to its elderly population.

According to Statistics Canada, productivity per hour worked has grown by an annualized rate of 0.96% over the period 1981-2010 in Québec. Our simulations use this value as a baseline, and assign $w_t = z_w^{t-2005} \cdot w_{2005}$ with $z_w = 1.0096$. However, this value is low in comparison with neighboring economies: over the same period, the productivity growth of Ontario and Canada as a whole were of over 1.2%. Thus, Québec's Department of Finance has projected that a form of catching up with the other provinces will likely occur, and that labor productivity will grow by at this higher rate on average over the decades to come (CCEFP, 2009). In Appendix B.5, we present a sensitivity analysis with $z_w = 1.012$.

2.3.3 Technological Progress

The second set of parameters concerns technological change in health care. Recent efforts by Hall and Jones (2007) and Fonseca et al. (2009) model aggregate health outcomes using structural models. They found that technological progress are essential components of the increase of longevity and aggregate health spending in the US. Both modeled technological growth as improving the impact of medical treatments on health, and found an endogenous increase in longevity and medical demand. Fonseca et al. developed a dynamic life-cycle model with endogenous death similar to the demand component of our Health Care module. As a part of their calibration effort, they found that an annual health productivity improvement of 1.8% is consistent with the observed increase in American medical spending and longevity growth.

Despite their strikingly different health care structures, it is not implausible that medical productivity growth has been similar in Québec and the neighbouring United States, since longevity at birth has grown by about 5 years in both jurisdictions over the period 1980 to 2009.[7] Assuming that technological growth will continue at the same rate, we impose this interpretation of medical progress directly in our model. Parameter b_t, which linearly dictates the maximum health attainable by using health care,[8] is thus given by $b_t = z_b^{t-2005} \cdot b_{2005}$, with $z_b = 1.018$.

---

[7]According to ISQ and U.S. National Center for Health Statistics data.

[8]See Section "Health" of Appendix B.3.

In addition to increasing the desirability of treatments and the demand for them, technological improvements may also affect costs directly. In particular, Chandra and Skinner (2011) point out that many new "gray area" treatments have uncertain medical value and may increase costs without enabling significant health improvements. According to them, the rapid adoption of such treatments by the United States may explain the unusually fast increase of medical spending in comparison with other western countries. In contrast, Jones (2002) notes the continual reduction of the cost of existing treatments, which can act as a counterbalancing force to new treatments in terms of aggregate health spending. He estimates the reduction in quality-adjusted cost of existing treatments to be of about one or two percent.

The net impact of future medical technology on Québec's health care system's unit cost is uncertain. However, given the steady increase of health expenditure per capita over the last half century, the average cost of treatment, dictated by parameter d_t, is likely to keep increasing by at least as much as the productivity parameter over the foreseeable future. We thus also increase parameter d_t by $z_d = 1.018$ annually. We conduct a sensitivity analysis on this assumption in Appendix B.5, in which we allow costs to increase at both a slower and faster pace than b_t.

2.4 Numerical Results

Policy 1: Increasing Health Care Costs at Steady Pace

In our first policy scenario, the funds allocated to health care increase at a steady rate $\mu > 1$ each year. With this policy, medical facilities' budgets grow steadily over time in nominal terms without regard to efficiency of care, demand or GDP growth. Whenever GDP growth is less than μ, the share of health spending in the economy increases. Despite its flaws, this approach to funding health care is appealing because of the simplicity of its application and the possibility for health care managers to fully anticipate future revenues.

This "historic" approach to health care budgeting has been the prevailing policy in Québec since the inception of the public health care system. According to a recent analysis produced by Québec's Department of Finance, this budgeting approach is set to continue in the foreseeable future, at a pace of 5% in nominal terms (Finances Québec, 2010).

As mentioned in Section 2.2.3, we impose this policy by choosing the right income tax for each year in the Policy module. With regards to equations (2.2) to (2.4), the share of

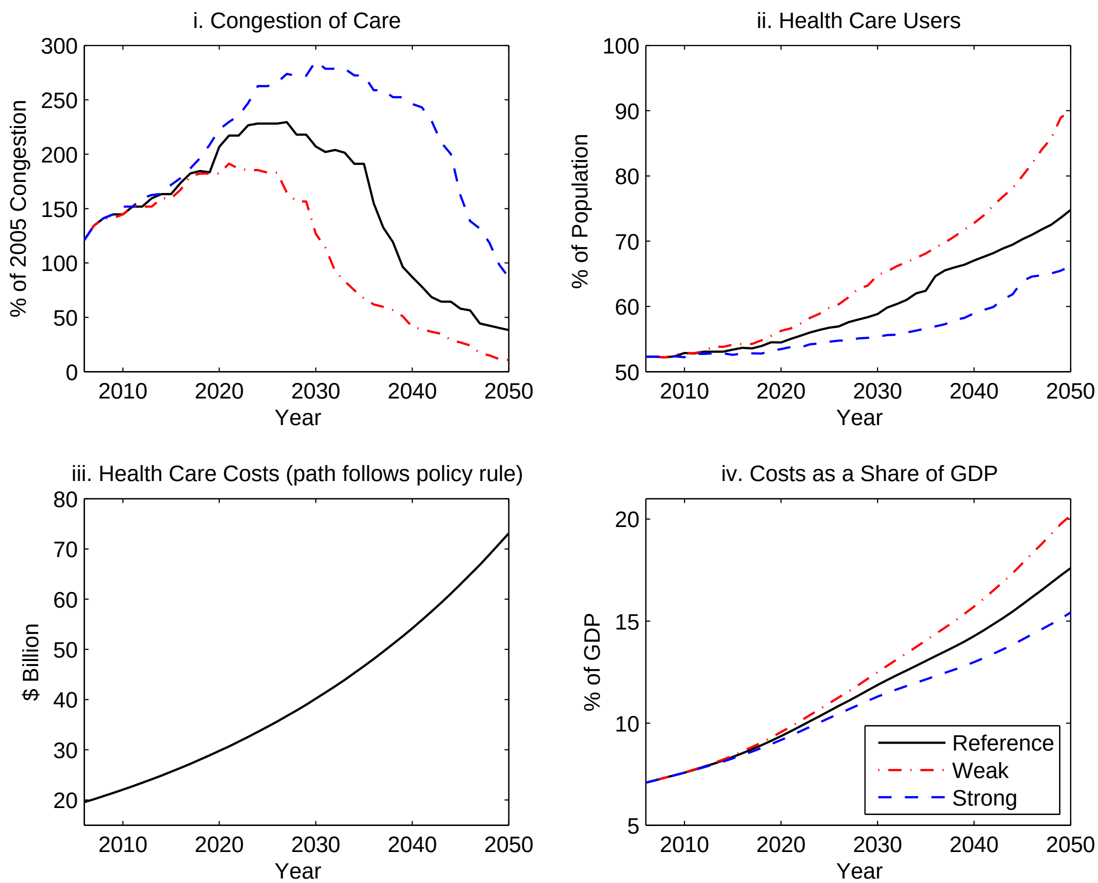


Figure 2.3 Predicted Outcome of Increasing Health Care Costs at a Steady Pace

income tax allocated to care coherent with this policy is $\tau_t^{HC} = \frac{\mu HC_{t-1}^S}{GDP_t}$. We set $\mu = 1.0304$, a value consistent with a 5% nominal budget increase and an annual inflation rate of 1.9%, the annualized CPI growth rate over the 2002 to 2011 period.

Figure 2.3 sums the simulated aggregate results obtained with this policy in four charts. Chart 2.3.i presents the equilibrium congestion level of the health care system expressed in comparison with the initial congestion. The policy simulation conducted with ISQ's Reference demographic scenario is the solid line, while the simulations conducted with the Weak and Strong scenarios are the dotted red and blue lines, respectively. Chart 2.3.ii presents the percentage of the population that optimally chooses to use the health care system over time. Chart 2.3.iii exhibits the total costs of the public health care system. For this policy, health care costs increase steadily regardless of demographic outcomes, following our policy rule. Finally, the Chart 2.3.iv displays the evolution of the share of the economy allocated to public health care.

We find that Québec's current policy of increasing nominal health care budgets will result in considerably longer waiting times for care in the medium term. These will occur despite a steadily increasing share of GDP being spent on health care. By 2025, with the ISQ's Reference demographic scenario, waiting times will have increased by 128%. Despite these increases, the share of the population that chooses to use health care during this period is expected to grow, driven by the rapid increase in the 65+ population presented in Figure 2.1. This is consistent with the results of our previous effort, in which we found that the elasticity of demand with regard to waiting times is quite low and that demand is particularly high in older agents. After 2025, the slower increase in the elderly population will allow budget increases to effectively ease the pressure on the health care system. Waiting times will then decline and fall below 2005 levels by 2040.

As previously mentioned, we do not allow the health of our distribution to evolve over time. Nevertheless, as medical technology continues to progress, the health distribution could improve, notably among the elderly. In the Appendix B.1, we find some indication of the health of older Québécois improving in the CCHS data. An improving population health among the older age groups would mediate the pressure brought forward by Québec's aging population, without reducing the number of users of health care. As we detailed in our 2012 article, the health care module predicts that older agents optimally consume health care even in good health. As a consequence, our projected utilization of care would not decrease if we imposed a better health among the elderly in our simulations. However, such a change would mean less expensive treatments and a decrease in aggregate demand, through equation (2.6). A lower demand would lead to a reduced congestion in Figure 2.3.i and, in turn, to a slightly higher utilization of care than predicted by Figure 2.3.ii.

The Weak and Strong demographic scenarios demonstrate the large uncertainty of the outcomes associated with a budgeting policy that does not adjust to aggregate demand. Since the projected health care budget is the same for all demographic scenarios, future waiting times, health care provision and the share of health care in the economy are quite uncertain. For instance, the Strong demographic scenario, which presents a larger population and more demand for health care, produces a more severe and persistent congestion increase (to over 275% of the 2005 level in 2030). Lower spending per capita would allow for a slower increase of the health care to GDP ratio but, because of longer waiting times, less Québécois would use public health care. At the opposite, the Weak population scenario would result in lower waiting times and more patients, but a 2.6% higher share of GDP spent on health care than the Reference scenario by 2050.

To avoid such uncertainty, the government could adopt policies centered on population needs or costs, which is the rationale behind policy scenarios 2 and 3.

Policy 2: Funding Health Care to Maintain a Steady Congestion Level

Our second policy scenario allocates spending in order to maintain waiting times for care constant at their 2005 level. This scenario allows us to project health care costs and the share of GDP that would be needed to maintain a similar access to care for future users in an aging population context.

This policy can be achieved in our framework through the anticipation by the government of the aggregate demand for care. Denoting $HC_t^D(\gamma_{2005})$ the demand for care of the population at year t given the 2005 congestion of health care, the Policy Module simply chooses τ_t^{HC} such that $HC_t^S = HC_t^D(\gamma_{2005})$. The simulated results of this policy are presented in Figure 2.4.

Targeting congestion of care shows the government modulating health care costs to adapt to the aggregate demand. Thus, the increase in the share of the population that uses health care of chart 2.4.ii resembles the increase in the elderly population presented in Figure 2.1. For the Reference demographic scenario, this share grows by 11.3% from 2005 to 2030 as the 65+ population more than doubles, to 2.3 million. Then, from 2030 to 2050, the elderly population increases at a much slower pace, which is reflected in the slower growth of health care use in the total population.

Since this policy shows health care supply adapting to population size, it results in much less variability caused by demography than our first policy scenario, both in terms of the share of population that uses care and of the portion of GDP allocated to health care. We see this in the closeness of the three demographic scenarios in charts 2.4.ii and 2.4.iv.

Unsurprisingly, these simulations reveal that costs would need to increase substantially in comparison with current spending levels to keep congestion constant as Québec's population ages. By 2025, costs would need to increase by at least $ 17B for all demographic scenarios, and

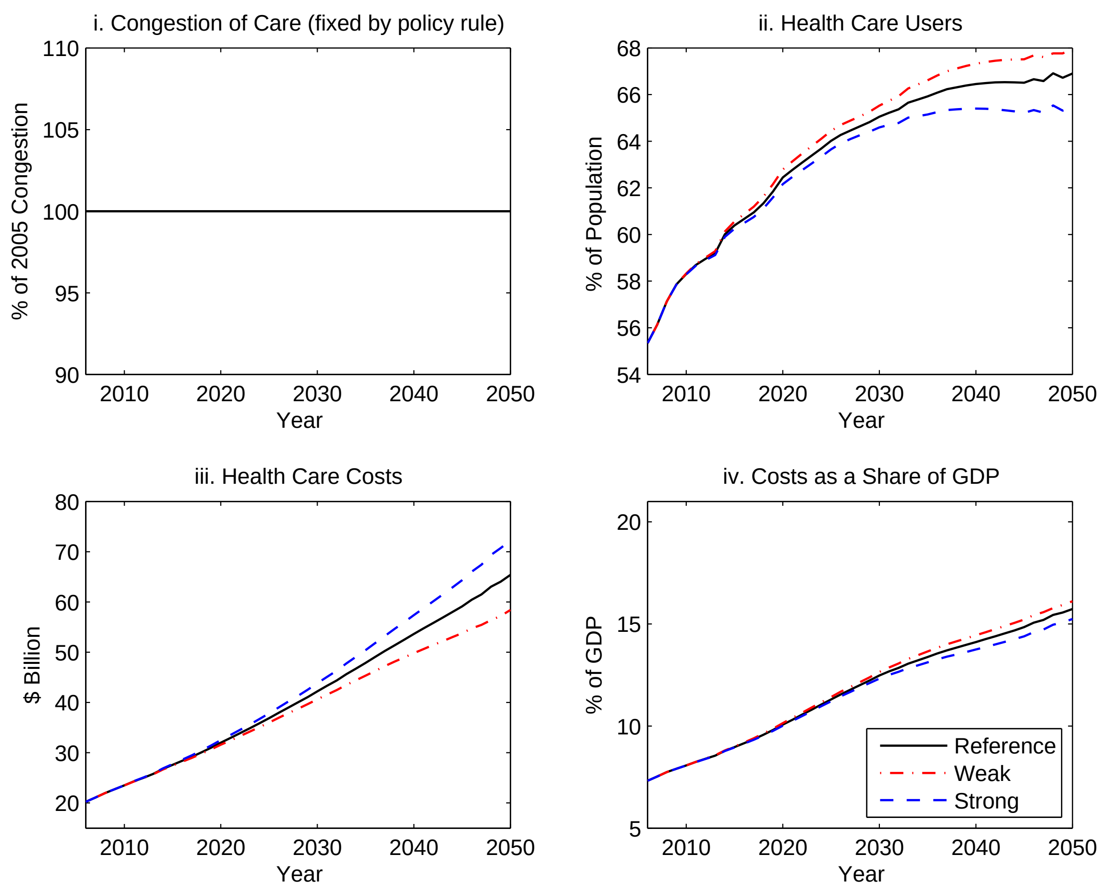


Figure 2.4 Predicted Outcome of Maintaining a Steady Congestion Level

their share in the economy by 4%. For all demographic scenarios, costs would be superior to those of Policy 1 until 2030, after which maintaining a steady access to care will require a slower budget growth. By 2050, health care costs would reach 15.7% of GDP, more than doubling from their 2005 level. This leads us to our third policy stance, in which the government maintains the share of the economy allocated to its health care system constant.

Policy 3: Maintaining Health Spending Constant as a Share of GDP

Our third policy scenario keeps the share of public medical spending in the economy, denoted λ, constant at 6.98% of GDP.[9] This scenario reflects the reticence of policymakers and taxpayers to allow an ever larger share of the economy to be spent on publicly funded health care. In our Policy module, this policy is obtained, quite simply, by choosing $\tau_t^{HC} = \lambda$ for each year.

In Québec, eight public inquiries, special committees and working groups were formed to address increasing health care spending since 1988. While these efforts have yet been largely unsuccessful at stabilizing costs, it remains important to assess the impact such a policy would have, were it to succeed, in an aging population context.

The simulated results, shown in Figure 2.5, are catastrophic. Policymakers implicitly rely on waiting times as a non-monetary price to clear the health care market. However, the demand for care is inelastic with regard to waiting times. In the severe population aging context of Québec, this means that waiting times would need to literally explode in order to maintain currents shares of GDP; according to chart 2.5.i, they would increase by a factor of 15 compared to their 2005 length. Only older agents in very bad health would remain in the health care system, which would be used by less than 35% of the population by 2025. These extreme outcomes are obtained with all demographic scenarios.

In our view, these results attest to the impractical nature of aiming to maintain public health care costs low in an aging population context while providing the newest medical technology. Because of Québec's demographic context, the relative size of the working age population shrinks (Figure 2.1.ii). Meanwhile, as discussed in Section 2.3, technological progress implies

[9] In 2005, the total cost of health assumed by the public sector was of \$ 21.4B, as reported by the Canadian Institute for Health Information. Since we focus on medical spending, we withdraw from that amount the cost of public medication insurance of \$ 2.4B, as reported by the Régie de l'assurance maladie du Québec. The resulting public medical spending is of \$ 19.0B, 6.98% of the \$ 272.0B GDP for that year.

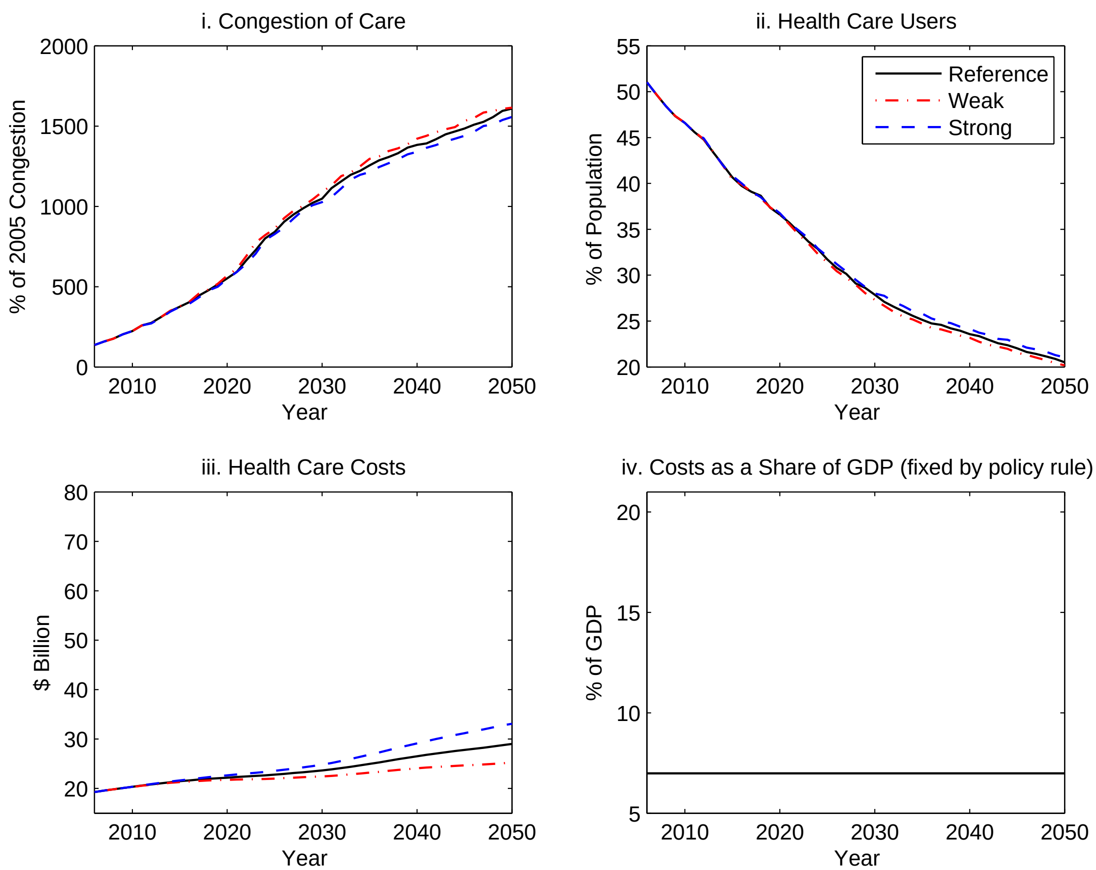


Figure 2.5 Predicted Outcome of Maintaining Health Spending Constant as a Share of GDP

annual increases of the unit cost of treatment of 1.8%. These combined factors restrict Québec's ability to pay for the newest health care technology.

In practice, to maintain τ_t^{HC} low without an explosion of waiting times, policymakers would either need to dramatically increase the productivity of medical expenditures, increase the productivity of workers and GDP, or ration the adoption of new technology. The latter strategy is the basis for our fourth policy scenario.

Policy 4: Rationing the Adoption of Technological Developments

Our last policy scenario also holds the ratio of health care to GDP is held constant at its 2005 level, and has the government choosing to introduce new medical technology only if it reduces the costs of treatments.

In our Policy module, this policy is obtained by choosing $\tau_t^{HC} = \lambda$ for each year, as in the previous scenario, and by replacing our baseline technological trends values by $z_b = 1$ and $z_d = 0.99$, respectively. This value of z_d is chosen based on the estimation by Jones (2002) that the quality-adjusted cost of existing treatments decreases annually by between 1% and 2%. This could be interpreted as aggressively rationing technological progress with the purpose of containing health care costs. Alternatively, the same results would be obtained with a fixed technology if productivity gains of 1% were obtained in the production of medical care.

The results, displayed in Figure 2.6, differ drastically from those of Policy 3. They show the lowest congestion of care and highest use of the health care system of all our policy scenarios. For all demographic projections, waiting times would be eliminated by 2040. With zero waiting times, the cost of using health care is nil and the model predicts that 100% of the population would consume health care. In reality, however, transportation costs, factors such as the duration of treatments and the risk of contagion in health care facilities would prevent such an outcome.[10]

Such a policy stance would certainly prove difficult to implement in practice. In 2050, the newer technology of the first three scenario is 2.2 times more effective than that of this scenario. Understandably so, patients would quickly demand that the more effective (and costly) treatments be available to them. Nonetheless, these simulated results show how productivity gains and discriminating against the adoption of pricey, "grey area" technology may contribute

[10] This figure does not include the Health Care Costs and Costs as a Share of GDP charts that were presented for the other three policies, since they are the same as those of Policy 3, already shown in Figure 2.5.

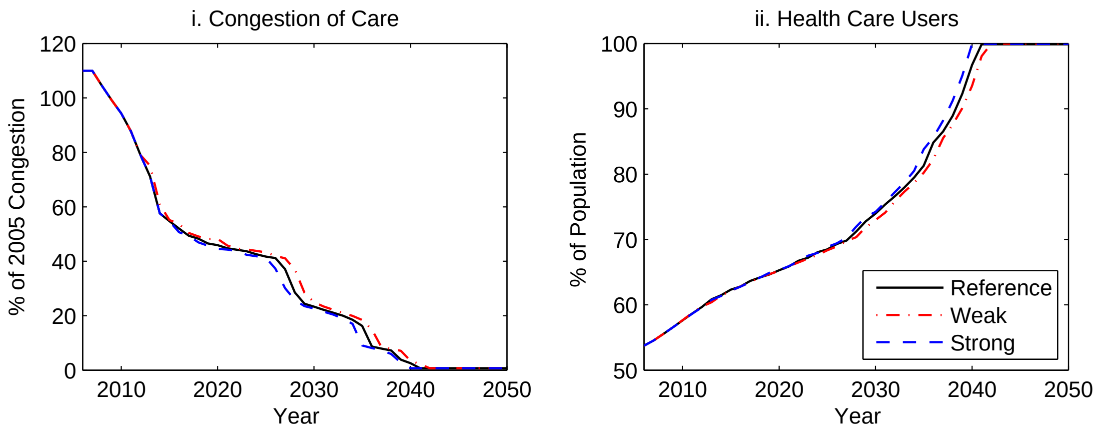


Figure 2.6 Predicted Outcome of Rationing the Adoption of Technological Developments

to alleviate the challenge of aging.

2.5 Discussion

In this article, we study the impact of population aging on Québec's public health care system. Our analysis focuses on waiting times, demand for care and health care costs. To that effect, we build a health care model that predicts Québec's aggregate health care outcomes for the years 2005 to 2050. We then simulate the impact of four different policy stances according to the three reference demographic scenarios produced by the Institut de la Statistique du Québec (ISQ).

While Québec's current budgeting approach of increasing health spending by 5% each year is easy to implement, it fails to adapt to the needs of the population. Our simulations show that these needs will be greatly amplified by Québec's aging demography: an institutional *status quo* would result in twice longer waiting times by 2030. Despite a more difficult access to care, utilization would increase steadily, as the baby-boom generation grows older and its health depreciates. According to our simulations, if the government aimed to maintain waiting times at current levels, the share of health spending in the economy would need to double by 2050. An amelioration of the health status of the elderly would moderate these results, but we find no clear evidence of such a trend in the Canadian Community Health Survey. In this context, we find that maintaining a steady share of GDP allocated to health care would be catastrophic. Since the demand for care is inelastic with regard to waiting times, such an approach in the face of Québec's demography would lead to a 15-fold expansion of waiting times.

These results assume that Québec's health care will remain publicly financed with zero costs to users in the decades to come, and that medical innovation will be automatically introduced. Reforms, such as the inclusion of moderating fees or the privatization of a portion of the services currently paid for by the province, could indeed lighten the increase of waiting times with no additional taxation. Such policies would however transgress the Comprehensiveness and Accessibility criteria of the Canada Health Act, and would expose the province to severe financial sanctions.[11] Another strategy could be to agressively ration the adoption of new technology; we find that it may lead to the elimination of waiting times by 2040. Such a strategy would however be difficult to implement in practice, since patients would quickly demand for the more effective technology to be available to them. Finally, of course, the province could use public debt to pass the expansion of its health costs over to future generations. From our simulation results, we can infer that such a strategy would quickly become unsustainable for Québec: by 2025, maintaining waiting times at their current level without increasing taxes would generate a $ 137B cumulated deficit. This figure is equivalent to 42% of that year's predicted GDP.

This article shows that, as the baby-boom generation gets older and consumes more care, it will be extremely difficult for policymakers in countries with a comprehensive public health care system to meet the population's expectations of a timely delivery of care, an access to the newest medical technology and a moderate fiscal burden. Facing this challenge will constitute a difficult balancing act between these three objectives. It is of note that our results do not reveal which policy stance is *better*. Future research should focus on finding the optimal course of action for countries facing this dilemma. With the demographic transition of the baby-boomers into retirement already underway, time is running out.

[11] In 2005, the federal Canada Health Transfer to Québec was of $ 5.0B, over a fourth of the province's health care budget.

3 FORECASTING POPULATION HEALTH, HEALTH SPENDING AND MORTALITY

Although old-age disability has decreased among the elderly in the US over the last decades, some evidence support the possiblity that it could increase in the future due, for example, to the alarming trend in obesity and its consequences such as diabetes. On the other hand, technological progress in medical care, in particular curative care, may continue to alleviate some of these pressures, allowing a higher quality of life for ill and older individuals, and sustain the rise in life expectancy. However, both an increasing prevalence of chronic conditions and technological progress may raise the demand for health care, resulting in a higher allocation of resources towards health care. Using a dynamic stochastic health production model calibrated to the US experience over the last 25 years, we show that it is possible to simultaneously obtain an increase in longevity and health spending, while at the same time observing a deterioration of population health in terms of chronic conditions. The pace of technological progress in curative care plays a key role in producing this outcome. Looking ahead, our model predicts that chronic conditions prevalence will continue to increase in the years to come, but will not directly contribute to increased medical spending. Unless income or medical innovations were to grow at a slower pace, we find that spending will likely increase to over 22% of GDP.

3.1 Introduction

Needless to say, predicting future health, mortality and health spending is vital for evaluating the fiscal outlook in the US and other countries. Two popular approaches for forecasting these outcomes are extrapolation based on times-series and dynamic microsimulation. For example, demographers use the Lee-Carter model and other projection models to project how mortality will evolve in the future (Lee and Carter, 1992). The Congressional Budget Office (CBO) projects health care costs using exogeneous cost increases applied to an otherwise standard microsimulation model with labour supply, taxes and Social Security (Congressional Budget Office, 2009). One of the main drawback with extrapolation methods is that it often predicts an unconstrained growth of health care spending. If current rates of expenditure growth are maintained, we would be spending all our ressources on health by 2100. To counter this, the CBO imposes an ad hoc limitation to stop the rise of health expenditures as a fraction of GDP.[1]

Dynamic microsimulation models assume reduced-form relationships between health, mortality and health spending (e.g. Goldman et al., 2004). These models cannot foresee behavioral responses, such as a change in the consumption of medical care in response to income shocks or changes in the relative price of health. They are also subject to the problem of unbounded growth. In those models, spending is not productive (in the sense of reducing disease prevalence) and is kept fixed as a function of other characteristics of the household. Further, spending does not respond to relative price changes or changes in lifetime resources of the population.

In this paper, we explore another avenue where the behavior of future agents, including how much to spend on health, evolves optimally over the life-cycle in a way which is consistent with economic theory. This allows to model the behavioral responses to the changing circumstances of households, in particular technological change, in a way which does not lead to unbounded growth in expenditures. We use our model to project the evolution of its chronic condition prevalence, health spending and longevity in the US over the period 2010-2050.

In order to project forward, it is important to look back at what has happened. Over the last 50 years, four stylized facts about the population health have received considerable attention. First, there has been an alarming increase in the prevalence of various chronic and acute health conditions, across cohorts at the same age. In 1980, fewer than 3% of the adult

[1] The Office imposes a gradual decline of the Medicare and Medicaid "excess cost growth", the increase of health spending per capita relative to the growth of GDP per capita, in its projections (Manchester, 2012).

American population had type 2 diabetes while in 2010, 9% had diabetes according to the Center for Disease Control (CDC). A major contributor to the higher prevalence of diabetes is obesity. In 1980, 22.9% of Americans were obese compared to 30.5% in 2000 (Flegal et al., 2002) and 35.7% in 2012 according to the CDC. These trends are potentially the result of more sedentary lifestyles, in part caused by technological innovation both in the economy in general and in particular in the production of food (Lakdawalla and Philipson, 2009; Cutler, Gleaser and Shapiro, 2003).

Second, there has been a steady decline in the prevalence of disability at older ages, at a rate of 1.5% per annum (Manton, Gu and Lamb, 2006). A number of studies have documented a strong decline in the fraction of disabled respondents citing cardio-vascular disease, vision problems and musculoskeletal conditions (Freedman et al., 2007).

Third, we have witnessed a steady, quasi-linear, rise in life expectancy over the last 50 years (Oeppen and Vaupel, 2002). This trend has been driven by a proportional reduction in mortality rates after the age of 40. Murphy and Topel (2005) show that since the 1970s, remaining life expectancy at age 50 has risen much faster than life expectancy at birth.

Finally, health spending has risen as a share of GDP from less than 4% to 16% between 1950 and 2005. Meara et al.(2004) document that this increase accrued disproportionally after age 65. Relative to those aged 35-44, per capita spending was on average five times higher for those age 75+ in 2000 compared to 2.5 times in 1963. The literature aimed at explaining those facts has focused extensively on technology (Newhouse, 1992; Cutler, Deaton and Llearas Muney, 2006; Shoeni et al., 2008; Fonseca et al., 2009). This informs us that a model aimed at forecasting those outcomes into the future should have as one of its main driving force advances in medical technology.

Advances in medical technology may in fact affect differentially different facets of the health of individuals, since a common view among epidemiologists and gerontologists is that population health is multidimensional (Crimmins, 1996). Indeed, medical technology can prevent the onset of health conditions, increase quality of life of individuals with a condition and lenghten remaining lifespan. When progress is "biased" towards treating conditions which emerge with age, such as cancer, resources may be optimally allocated towards older rather than younger ages from a cross-sectional perspective. In turn, this bias may lead to technological change improving health at older ages while the health at younger ages deteriorates. Added to this, the expectation of technological progress may also give the incentives, at the individual level, to delay health investments and thus exacerbate the trends observed.

A good example of the role of technology in raising expenditures but also delivering large benefits in terms of reduced mortality and disability at older ages can be found in the evolution of cardio-vascular disease and its treatments. Some of the risk factors for cardio-vascular disease have been on the rise over the last 30 years (in particular hypertension and obesity). But from 1960 to 2000, close to 70% (4.88 years) of the increase in life expectancy of new borns could be traced back to a reduction in the mortality rates from cardio-vascular disease (Cutler et al., 2006). The introduction of angioplasty, beta-blockers and statins are good examples of medical technologies and advances that have reduced significantly mortality and disability rates from cardio-vascular disease (Cutler, Deaton and Llearas Muney, 2006; Schoeni et al., 2008). These technologies may have increased health spending, particularly at older ages, but with considerable value (Cutler and McClellan, 2001). Simultaneously, the prevalence of cardio-vascular disease may have increase as a result of the lower mortality.

We propose a dynamic stochastic model of health production where three underlying forces shape the future environment of the population. First, a deterioration of health, due to an increase in sedentary lifestyles which will then result in a trend in the prevalence of chronic conditions in the population. Second, an increase in the productivity of health care in transforming money into health. This will be the pathway to build in the effect of improvements in medical technology into the forecasts. Finally, a change in real income of households over time. The allocation of resources both within and across cohorts will depend on the relative strength of these forces. The model can generate both positive or negative longevity and old-age disability trends depending on the strength of these trends, and can be used to assess a variety of different scenarios.

We calibrate our model and our trend parameters to match the 1985-2005 U.S. experience. We then show that it is able to replicate the stylized facts about population health we just mentioned. Projecting into the future, we find that spending will continue to increase steadily if current obesity, medical innovation and income trends were to continue at their recent pace. By 2050, we find that medical spending will account for 22% of GDP. Despite our trend of worsening lifestyles, our baseline projection suggests that life expectancy will continue to rise in the decades to come, due to higher available income and improvements in medical technology. We then consider a number of scenarios aimed at understanding the sensitivity of those forecasts to the underlying trends. According to our simulations, spending per capita is mostly amplified by income growth, and would lower considerably in the case of an downward income shock. While income directly affect the level of medical spending, we find that medical innovations are the factor that leads to the increase of spending as a share of the economy. If future medical productivity grew at half of its current pace, the share of GDP allocated to health would stabilize and eventually decrease to 15% by 2050. All our other scenarios find a growth in the share of the economy allocated to health care in the future.

The remainder of the article is organized as follows. Section 3.2 describes the dynamic health care demand model. Section 3.3 presents our main data sources. Section 3.4 presents our calibration. Section 3.5 shows and discusses our results. Finally, Section 3.6 discusses the implications of our results suggests future research avenues.

3.2 Model

Since the pioneering work of Picone et al.(1998), numerical methods have proved invaluable to study questions in which the stock of health-capital[2] and mortality of agents are endogenous and uncertain. Recent examples include French (2005), Jung and Tran (2007), Blau and Gilleskie (2008), Galama et al.(2008), Halliday et al.(2009), Scholz and Seshadri (2010) and Baicker and Skinner (2011).

In a paper closely linked to the objectives of this article, Hall and Jones (2007) introduced an income growth trend to the health care problem of agents and concluded that the rise of the share of health spending in the economy is optimal, given that health spending constitute a superior good with an income elasticity above one. Looking ahead, they predicted that this share will exceed 30% by 2050. Fonseca et al.(2009) pushed the analysis of Hall and Jones further, adding changes in insurance coverage and technological developement as driving forces of health care spending, and found that these forces accounted for over 50% of the increase in spending since 1965.

The model we propose is an extension of Hall and Jones that includes savings, uncertainty in health and dynamics in health. Health is modeled as a stock that decreases through life and can be affected by negative persistent shocks (chronic conditions). Agents spend in medical care in order to increase their expected longevity. Our model makes technology a prominent force of

[2]A concept introduced by Grossman (1972).

the recent U.S. experience.

3.2.1 Timing

Between 1985 and 2005, US health expenditures grew at a brisk pace, and their share in the economy increased by over 5%. In our model, we consider five-year periods starting with 1985, so that $t = 1985, 1990, 1995, ...$ Agents are born at age 25 and may live up to a maximum age of 110. In the remainder of the text, we refer to agents as being of *cohort* t if they were of age 25 in year t. The death of agents is both stochastic and endogenous to their choices: death may occur at the end of any period, but its probability increases as the health of agents decreases.

The timing of the consumer problem is as follows. At the beginning of period t, an agent of age a_t, financial wealth W_t, health H_t and chronic condition state ϵ_t chooses how much to consume, spend in medical care and save during the period. He obtains periodic utility $u(C_t)$, increasing in his consumption. At the end of the period, his health level depreciates and is affected by medical spending and chronic conditions, and becomes H_{t+1}. Depending on his end-of-period health, a lottery determines if he survives and has a chronic condition in the next period. If he survives, he faces the same problem in $t+1$ with his updated state variables.

Three kinds of trends affect this dynamic problem: an income trend, a health depreciation trend (resulting from changing lifestyles) and a technology trend. Because of those period effects, agents face a different problem than their elders: their health depreciates faster but they have access to more productive health care and more income to allocate between consumption, medical spending and savings. We assume that agents have a perfect foresight of the impact of these trends on their optimization problem.

3.2.2 Preferences

An agent born in year t_0 chooses his periodic consumption C_t and medical expenditures M_t (which jointly define next-period financial assets W_{t+1}) in order to maximize his expected flow of life-time utility:

$$\max_{C_t, M_t} E\left[\sum_{t=t_0}^{t_0+\bar{a}-5} \beta^t u(C_t)\right] \tag{3.1}$$

Where β is the discounting factor of the agent and $\bar{a}$ the maximum age he can expect to reach. The utility function is:

$$u\left(C_t\right) = b + \frac{C_t^{1-\gamma}}{1-\gamma} \tag{3.2}$$

This utility function has a constant relative risk aversion of γ and is increasing in contemporary consumption. Parameter b corresponds to the baseline utility of being alive during the period, and is calibrated to ensure life to be desirable over death for all agents. Medical consumption and health do not increase periodic utility directly, but are desirable to increase longevity.[3]

3.2.3 Budget Constraint

We assume that there is no income uncertainty and that all working-age agents work full time. The agent's budget constraint is determined by his wealth W_t and income $Y\left(t, a_t\right)$, where age $a_t = t - t_0 + 5$.[4] Denoting $X\left(t, a_t, W_t\right)$ the cash-on-hand, we have:

$$X\left(t, a_t, W_t\right) = W_t + Y\left(t, a_t\right) \tag{3.3}$$

We obtain an exogenous trend of rising household revenue by increasing the periodic income at rate $z_y > 1$: $Y\left(t, a_t\right) = Y(1985, a_t) \cdot z_y^{t-1985}$. Thus, an agent of age a_t has z_y more revenue than an agent of the same age in $t-1$.

There is a strict borrowing constraint imposed so that the agent is not allowed to borrow against future income. His assets in $t+1$ are the total amount of resources available but not spent in t, increased at exogenous rate $R = \left(1+r\right)^5$:

$$W_{t+1} = R\left(X\left(t, a_t, W_t\right) - C_t - OOP_t\left(M_t\right)\right) \tag{3.4}$$

Where $OOP_t\left(M_t\right)$ is the portion of medical spending M_t spent out-of-pocket by the agent.

---

[3]This function corresponds to the baseline utility function of Hall and Jones (2007), in which the quality-of-life component was set to zero.

[4]While we could rewrite our model in terms of t_0, we introduce a_t for simplicity.

3.2.4 Health Insurance

All agents are insured for their health spending. As in Fonseca et al. (2009), the out-of-pocket payments are determined by the total cost of medical treatment M_t and the deductible and co-insurance rate of the policy (μ_1 and μ_2, respectively):

$$OOP_t\left(M_t\right) = \begin{cases} M_t & \text{if } M_t \leq \mu_1 \\ \mu_1 + \mu_2\left(M_t - \mu_1\right) & \text{otherwise} \end{cases} \tag{3.5}$$

3.2.5 Health Law of Motion

As first modelled by Grossman (1972), health acts as a human-capital in the agent's problem. We consider that health varies from a minimum of zero to a full health level of $\bar{H}$.

We choose a formulation of the health production function similar to Picone et al.(1998). Health depreciates at rate $1-\delta\left(t, a_t\right)$ each period, where $\delta\left(t, a_t\right) = \delta_0\left(t\right)\exp\left(-\delta_1 \cdot a_t\right) \leq 1$ is the share that remains in $t+1$ in the absence of chronic conditions and medical care consumption. The parameter $\delta_0(t) < 1$ imposes a natural depreciation of health for all ages, while $\delta_1 > 0$ accelerates the depreciation as agents grow older. Imposing a faster depreciation with age results in slowly declining health levels and low mortality rates at younger ages, and rapidly declining health and mortality rates after middle-age.

As with real income growth, we aim to replicate the observed trends of faster declining health over time. This trend can be thought of as the toll from the adoption of lifestyles that can be hazardous to health in the long term, such as changes in nutrition and exercise habits, higher rates of sedentary employment, etc. Such lifestyle changes are associated with higher obesity rates and lower health among the youth, which contribute to the prevalence of medical conditions such as diabetes. We thus model the natural depreciation parameter as declining at rate $z_\delta < 1$ in time: $\delta_0\left(t\right) = \delta_0 \cdot z_\delta^{t-1985}$. As we will see in our simulations, this trend will be correlated with the model's predictions of chronic conditions prevalence.

As in Picone et al. (1998), the stock of health can also decrease due to the occurrence of a stochastic health shock $\epsilon_t \in \{-d, 0\}$. The effect of ϵ_t is only realized at the end of the period, which allows the agent to use medical expenditures in order to minimize its impact. An important distinction from Picone et al. is that our shocks are perfectly persistent; agents who had $\epsilon_t = -d$ will also have a negative shock in $t+1, t+2, ...$ This modeling approach replicates the long term cumulative effect of the prevalence of chronic conditions, such as hypertension and diabetes, on the health of agents. The possibility of obtaining a first shock $\epsilon_t = -d$ adds a to the depreciation of health in a stochastic way, and creates an incentive to invest in preventive care.

The full effect of medical expenditures M_t on health is given by health production function $\theta(t, H_t, M_t, \epsilon_t)$, defined as:

$$\theta(t, H_t, M_t, \epsilon_t) = \left(\frac{\bar{H} - H_t}{\bar{H}}\right)(\theta_1 - \theta_2(t)\epsilon_t) M_t^{\theta_3} \tag{3.6}$$

This function is increasing in M_t and decreasing in both health capital level H_t and chronic condition state ϵ_t. Parameter θ_1 dictates the portion of health care productivity available to all agents, while θ_2 adds to the effectiveness of medical care for agents with chronic conditions.[5] By premultiplying health production by$\left(\frac{\bar{H}-H_t}{\bar{H}}\right)$ and imposing $\theta_2(t) > 0$, we enforce the idea that health care is more productive for agents with low health and chronic conditions. Finally, parameter $\theta_3 \in (0,1)$ imposes a decreasing marginal productivity of medical spending.

Combining these three elements, the end-of period health-capital is given by:

$$H_{t+1}(t, a_t, H_t, M_t, \epsilon_t) = \delta(a_t, t) H_t + \theta(t, H_t, M_t, \epsilon_t) + \epsilon_t \tag{3.7}$$

Our reading of the recent history is that most technological innovations in medical care are targeted to curing diseases or minimizing their symptoms rather than preventing them. Notably, as we discussed in the introduction, over half of the increase in life expectancy since the 1960s is linked to progress in the curative treatment of cardio-vascular disease. This observation suggests that we ought to model technological developments as an improvement in the productivity of *curative* care. With this objective, we model $\theta_2(t)$ as increasing at rate $z_\theta > 1$: $\theta_2(t) = \theta_2 \cdot z_\theta^{t-1985}$.

[5] As discusssed by Chang (1995), a patient that is suffering from a severe disease could expect to receive a greater appreciation of his health capital by having access to care than a patient with a mild disease. This is also a feature of Picone et al.(1998).

3.2.6 Uncertainty

Two components of the individual demand model remain to be specified. First, as mentioned when we described the timing of the model, the probability of being alive in $t+1$ is increasing in the end-of-period health of the agent, H_{t+1}. Like Scholz and Seshadri (2010), we formulate the survival function as a cumulative distribution function:

$$p_s\left(a_t, H_{t+1}\right) = \begin{cases} 1 - \exp\left(\nu_1 H_{t+1}^{\nu_2}\right) & \text{if } a_t < \bar{a} \\ 0 & \text{if } a_t = \bar{a} \end{cases} \tag{3.8}$$

Where parameters ν_1 and ν_2 determine the shape of the function and the probability of surviving for a given health. We will chose their value in order to replicate mortality and longevity moments. Thus, by investing in his health in t, the agent can add to his probability of being alive in $t+1$ and his life expectancy as a whole.

Similarly, the probability of contracting a chronic condition is decreasing in the end-of-period health. The chronic conditions are persistent for the remainder of the agent's lifetime, and thus the probability of having $\epsilon_{t+1} = -d$ is

$$p_\epsilon\left(a_t, H_{t+1}, \epsilon_t\right) = \begin{cases} \exp\left(-\psi_1 H_{t+1}^{\psi_2}\right) & \text{if } \epsilon_t = 0 \\ 1 & \text{if } \epsilon_t = -d \end{cases} \tag{3.9}$$

3.2.7 Bellman Equation

With respect to equations (3.1) to (3.7), the Bellman equation summing this dynamic problem is

$$V\left(t, a, W, H, \epsilon\right) = \begin{cases} \max_{C,M} u\left(C\right) + p_s \beta \left(p_\epsilon V\left(t+1, a+1, A', H', \epsilon'\right)\right. \\ \qquad \left. + \left(1 - p_\epsilon\right) V\left(t+1, a+1, A', H', \epsilon'\right)\right) & \text{if } a < \bar{a} \\ \max_{C,M} u\left(C\right) & \text{if } a = \bar{a} \end{cases} \tag{3.10}$$

$$\text{s.t.} \quad \begin{cases} C \geq 0, \, OOP\left(M\right) \geq 0, \, A' \geq 0 \\ A' = R\left(X\left(t, a, W\right) - C - OOP\right) \end{cases}$$

Since maximal age $\bar{a}$ is fixed at 110 years old, we can solve this dynamic problem numerically with standard backward induction.

3.2.8 Aggregation of Results

The objective of this paper is to predict aggregate medical consumption, prevalence of chronic conditions and longevity over the years to come. To obtain these, we need to solve a distinct equation (3.10) for all cohorts of agents alive in a given year. This arises because an agent of age 40 in 1985 faces a quite different problem than an identical agent in 2005, since the environment (in our case: medical technology, income and lifestyle) greatly affects his expectations and behaviour.[6] We predict the future of these cohorts by computing the tree of all their future possible states with probabilities p_s and p_ϵ. We save these intermediate probabilistic results and then draw from them to compute aggregate variables. These calculations are presented in Appendix C.1.

3.3 Data Sources

3.3.1 Distribution of Agents

The main data sources that we draw from to build our distribution of agents are the 1985 and 2005 waves of the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal household survey that originated in 1968 and contains information about income, wealth, expenditures, education, health and numerous other topics. We use it to extract average income, assets and health status of the American population aged 25 and over in 1985 and 2005. The health status of respondents is derived from the self-rated health status question, in which respondents choose between five qualitative choices (“Excellent”, “Very good”, “Good”, “Fair” or “Poor”). We construct a health index by performing principal-component analysis and create a

[6]This assumes that agents have a perfect foresight. In other words, having witnessed the steady evolution of lifestyles, technology and income over their lifetime, they are able to correctly predict that these trends will continue in the future.

score bounded between 0 and 100.[7]

The prevalence of chronic conditions is absent from the 1985 edition of the PSID; hence, we obtain it from the National Health Interview Survey (NHIS). We define a respondant as having a chronic condition ($\epsilon_{1985} = -d$) if he ever had hypertension, diabetes, a heart condition, a stroke in the past, or more than one of these. We then assign the prevalence per age in our distribution of agents based on that found in the NHIS. We note that cancer prevalence was missing in the 1985 NHIS survey. Thus, we underestimate the real prevalence of chronic conditions.

The size of our cohorts is obtained from the Human Mortality Database (HMD) for years 1985 to 2005. For our simulations of the future, we use the 2010-2050 population projections of the 2009 census. A summary of the initial distribution is presented in Table 3.1.

3.3.2 Moments

As we discuss in the next section, our calibration strategy aims to replicate a number of essential moments, including the recent evolution of medical spending, longevity gains, mortality rates and chronic condition prevalence.

Total health spending for 1985 are obtained from Meara et al.(2004), which estimate per person medical spending by age group by combining household surveys and total spending data. Their measure includes hospital care, physician and clinical services, prescription drugs, home health care, nursing home care, dental care, durable medical equipment and other professional services. We estimate 2005 health spending by applying their methodology to more recent data. We use total medical spending from the 2004 Medical Expenditure Panel Survey (MEPS), to which we apply the adjustments of Meara et al. for the institutionalized nursing home population (excluded from the MEPS). This adjustment consists of adding 18% to the spending of the population aged over 75. We then adjust the the MEPS totals to obtain the National Health Expenditure (NHE) per capita spending in 2005, of \$ 6,868. In Figure 3.1, we present average medical spending per age in 1985 and 2005. Both years present a steady increase of spending with age. However, 2005 spending levels were much greater, more than doubling for all age groups. In particular, medical costs increased by over \$ 10,000 (200%) for Americans

---

[7] We choose this scale for simplicity. We later adapt these scores to a different scale in our numerical model.

Table 3.1 1985 Distribution Summary

Age	N[a]	Income[b]	Wealth[b]	Health index[c]	Chronic condition prevalence (%)[d]
25	21,564	32.6	33.7	95	9.2
30	19,772	40.4	66.8	94	11.4
35	17,273	49.2	117.1	95	13.6
40	13,891	57.5	147.2	92	17.1
45	11,523	59.2	186.7	90	21.5
50	10,921	58.5	203.7	83	24.5
55	11,282	57.2	228.5	84	28.6
60	10,875	52.7	195.1	69	35.4
65	9,271	38.3	206.7	70	39.5
70	7,473	34.9	181.1	67	42.2
75	5,447	28.0	162.7	59	46.4
80	3,331	27.2	140.6	52	44.7
85	1,697	27.0	204.3	52	45.2
90	676	24.9	65.3	41	26.3
95+	181	14.9	26.2	41	36.1

[a]Population in thousands. Source: Human Mortality Database (HMD).

[b]Thousands of 2005 dollars. Source: Panel Study of Income Dynamics (PSID).

[c]The scale is from 0 to 100. Source: PSID, authors' calculations.

[d]Percent of respondants claiming to have hypertension, diabetes or a heart condition, or to have had a stroke. Source : NHIS.

over 75 years old.

Longevity gains are approximated by life expectancy, which we obtain from the National Center for Health Statistics (2011). Since we want our calibration to be consistent with our initial distribution in Table 1, we compute mortality rates directly from the PSID database rather than use external data.

Finally, we use the evolution of GDP per capita over the period 1985 to 2005 to calibrate income growth. The average 5-year growth, of 10.3%, is obtained from Bureau of Labor Statistics (BLS) data.

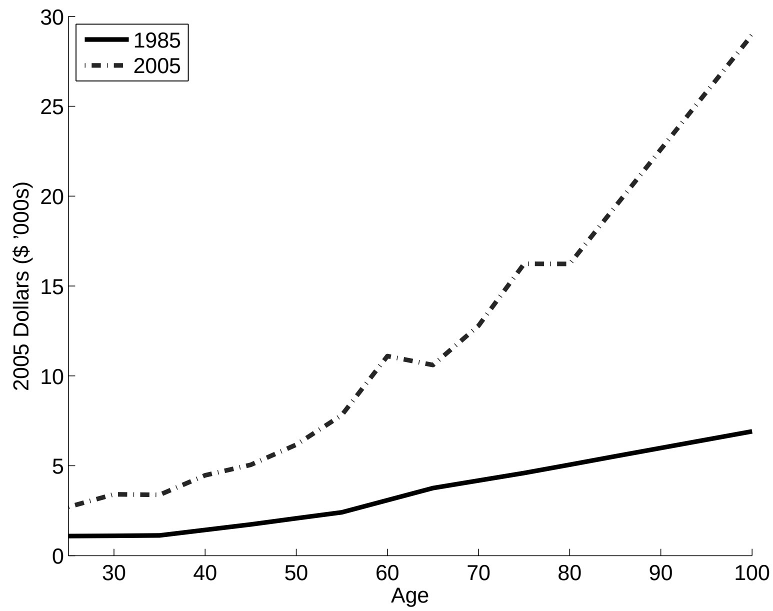


Figure 3.1 Total Medical Expenditures Vs. Age

Sources: Meara et al. (2004) for 1985 totals; authors' calculations based on MEPS and NHE data for 2005 totals.

3.4 Calibration

When possible, we assign values for our parameters according to existing literature or estimate them directly from the data. That said, many of the parameters, such as those that govern health production and depreciation, cannot be obtained through conventional means. Thus, we calibrate their value by manually choosing them so that we fit moments in terms of health, mortality and health spending. We note that the set of parameters presented here may not be the only one that replicates our moments, and that others may provide an equally good fit to our model. Our complete calibration is presented in Table 3.2.

3.4.1 Health Insurance Parameters

As discussed by Fonseca et al.(2009), fixing unique values for insurance parameters (deductible μ_1, co-payment rate μ_2) is complicated, given the variety of plans available and the fundamental differences between individuals younger than 65 (ineligible for Medicare) and the elderly. For instance, some elements of medical spending included in the Meara et al. (2004) medical spending calculation may not be covered even for agents who are insured. Also, while

Table 3.2 Calibration

Periodic Parameters	
Period length (in years)	5
Minimum age ($t = 1$)	25
T: Maximum age	110
r: Annual return rate of savings	1.04
β: Discounting factor	0.951
Health Insurance Parameters	
μ_1: Deductible	0
μ_2: Co-payment rate	0.2
Health Parameters	
$\bar{H}$: Maximum health	191.5
δ_0: Natural depreciation of health	0.9950
δ_1: Health depreciation curve coefficient	0.0230
d: Health depreciation due to chronic condition	7.620
Probability Functions Parameters	
ν_1: Survival probability parameter	0.160
ν_2: Survival probability parameter	0.640
ψ_1: Negative health shock probability parameter	0.235
ψ_2: Negative health shock probability parameter	0.525
Medical technology	
θ_1: Preventive health care productivity coefficient	8.126
θ_2: Curative health care productivity coefficient	1.066
θ_3: Health care productivity curve coefficient	0.0015
Utility Function Parameters	
b: Baseline utility	0
γ: Relative risk aversion	0.0722
Trends Parameters	
z_δ: Health depreciation trend parameter	0.9971
z_θ: Medical productivity trend parameter	1.1070
z_w: Earning trend parameter	1.1039

Medicare part B has a common deductible and co-insurance structure, Medicare part A cost-sharing depends on hospital stay.

In this paper, we choose to greatly simplify this aspect of the problem and only impose a co-payment rate of 20%. This value is both the median rate found by Blau and Gilleskie (2008) among insured individuals under 65 and the rate in Medicare part B. Hence, we believe it successfully conveys the marginal cost of medical care faced by most Americans over their life-cycle. A caveat of this choice is that it implies that everyone is *de facto* covered by an insurance plan.

3.4.2 Health Depreciation and Probability Parameters

We calibrate the parametric values of δ_0, δ_1,ν_1, ν_2, ψ_1, ψ_2 and d to match observed health and mortality measures. These parameters govern health transitions and mortality in the absence of trends and health care consumption. Hence, we assign their value in order to match data for year 1985 (which is our initial year and thus precedes trends) while *witholding* agents to use health care. We conduct this step by finding the mortality, chronic condition prevalence and life expectancy when imposing a zero productivity of care ($\theta_1 = \theta_2 = 0$) in our numerical model.

We choose our parameters in order to replicate the 1960 life expectancy of 69.7 years old. The underlying assumption is that, if agents did not have access to health care in 1985, they would have had the same life expectancy as in 1960. We choose to replicate 1960's statistic because this year predates the medical improvements that played a major role in the rapid increase of life expectancy over the last half century. Simultaneously, the parameters are set to produce very low mortality and rare occurrences of health shocks among younger agents and both high mortality rates and prevalence of chronic conditions as agents reach ages of 80 and over, as observed in the PSID data. We set the upper bound of our health scale by computing $\bar{H} = (-\log(0.01)/\nu_1)^{1/\nu_2}$, in order for an agent with maximum health to have a 99% probability of surviving in the next period.[8]

[8] In our simulations, we adjust the health of our distribution of agents presented in Table 3.1 accordingly.

3.4.3 Medical Technology Parameters

Next, we assign the values of medical technology parameters θ_1, θ_2 and θ_3 to replicate the impact of medical consumption on life expectancy. To do this, we allow agents to consume health care but hold income, medical technology and lifestyle parameters constant at their 1985 level (by imposing $z_y = z_\theta = z_\delta = 1$). The medical technology parameters are then set to reproduce the observed 1985 life expectancy of 74.7 years old with the agents' optimal medical spending. This means that our model's 1985 medical technology enables agents to improve their expected longevity by five years over an environment with no health care. This is in line with our interpretation that the life expectancy gains were driven in large part by medical care developements over the last decades.

Also, we assign these parameters to impose a rapidly decreasing marginal medical productivity, via curve coefficient θ_3 =0.0015. Given that health depreciation due to chronic conditions parameter d is set to -7.62, we set our parameters θ_1 and θ_2 so that preventive and curative care were equal in their impact on the health of agents in 1985. Given the health law of motion (3.6), we thus iterate on these parameters with the restriction $\theta_1 = 7.62 \cdot \theta_2$ until we replicate 1985 life expectancy. With the passage of time and our medical technology process, the curative component becomes gradually more important and dominates preventive care in later years. We present a sensitivity analysis with different relative productivity of curative and preventive care in Appendix C.2.

3.4.4 Utility Parameters and Discount Factor

We set relative risk aversion parameter γ, baseline utility b and discount factor β simultaneously with the medical technology parameters described above to reproduce the average savings per age found in the PSID and lifetime medical expenditure for year 1985. Specifically, the medical expenditure moment we replicate is the average lifetime medical spending of an agent who lives his whole life in a 1985 environment. It is computed as:

$$m_{1985} = \frac{\sum_{a=25}^{110} f_C\left(1985, a\right) \cdot M\left(1985, a\right)}{f_C\left(1985, 25\right)} \tag{3.11}$$

Where $f_C\left(1985, a\right)$ corresponds to the number of Americans of age a in 1985 according to the HMD, and $M\left(1985, a\right)$ their average medical spending according to the results of Meara et al. (2004). Using the model's predictions with a given set of parameters, we find the average spending per age $\tilde{M}(1985, a)$ and compute the corresponding simulated moment:

$$\tilde{m}_{1985} = \frac{\sum_{a=25}^{110} f_C(1985, a) \cdot \tilde{M}(1985, a)}{f_C(1985, 25)}$$

Both wealth and medical spending are directly driven by relative risk aversion parameter γ, which encourages agents to both increase their savings and spend more on medical care. In our calibration exercise, we allow for values comprised between 0 (risk neutrality) and 2 for this parameter. As discussed by Hall and Jones, the latter bound is a reference value in the economic literature due to its success in replicating asset accumulation decisions of agents over their life-cycle. However, this literature usually studies contexts of either infinite or exogenous lifespans for agents. Thus, only the budget constraint of future periods is uncertain. In our context, agents are both subjected to uncertainty regarding future medical spending and an endogenous and uncertain lifespan. Given this fundamental difference, we find that lower values of γ perform better in our setting, while values above unity usually result in extremely high medical spending through the agents' life-cycle. We ultimately assign a value of 0.07, close to risk neutrality. We conduct a thorough senstivity analysis with larger values of this parameter in Appendix C.2.

Baseline utility b is set in order for utility to be positive and thus enforces life in $t+1$ to be desirable for all agents even in the event of a very low consumption level. Considering our specification of the utility function, this can be obtained by:

$$b = \begin{cases} -(\underline{c}^{(1-\gamma)})/(1-\gamma) & \text{if } \gamma > 1 \\ 0 & \text{otherwise} \end{cases}$$

Where $\underline{c}$ is a minimal consumption value.[9] Since our choice of γ is below unity, we fix $b = 0$.

Meanwhile, we assign $\beta = 0.951$, which is consistent with a one-year discount factor of 0.99. This high appreciation for future periods compensates for our low value of γ in terms of replicating observed savings patterns. We note that such high values have previously been found to be plausible by the precautionary savings literature that studied the Euler function

[9] We use $\underline{c} = 10$.

using simulated methods (Cagetti, 2003).

3.4.5 Trend Parameters

Finally, we choose the trend parameters in order to best recreate the evolution of key variables over the period 1985 to 2005. The wage trend is directly computed from the real GDP per capita growth over the period. It produces a real income increase of 10.3% every 5 years, or just under 2% per year. The two remaining parameters, medical productivity growth z_θ and the impact of lifestyle changes on health depreciation z_δ, are set simultaneously to reproduce the evolution of two crucial moments: the growth of aggregate medical spending (as computed by equation (3.11)) and longevity growth. Our five-year medical productivity growth of 10.7% (2.1% in annual terms) is within the range found in the relevant literature: Fonseca et al. and Hall and Jones use a 1.8% and 2.3% annual productivity growth for their simulations, respectively. Finally, parameter z_δ, despite being close to unity, has an important compounding effect on the natural health depreciation factor $\delta_0(t)$. For instance, $\delta_0(1985) = 0.9950$, while $\delta_0(2010) = 0.9834$.

The implication of these parameters is that the health of agents decreases faster over time, but agents have more resources available to spend on curative care for chronic conditions, which also becomes more productive. We will discuss the success of our calibration, in particular with respect to producing increases of the prevalence of chronic conditions, in the next section.

3.5 Forecasting Results

In this section, we use our model to forecast health, longevity and spending oucomes for the decades ahead. First, we present the model's predictions of medical spending, prevalence of chronic condition and life expectancy according to a baseline scenario in which our three trends (medical innovations, income growth and worsening lifestyles) continue at their recent pace in the years to come. Second, we create four alternative scenarios, in which we allow each trend to grow weaker in the future, and see how each scenario affects our baseline projections. Finally, we run a sensitivity analysis to determine how different values of our relative risk aversion and technology parameters would affect our results.

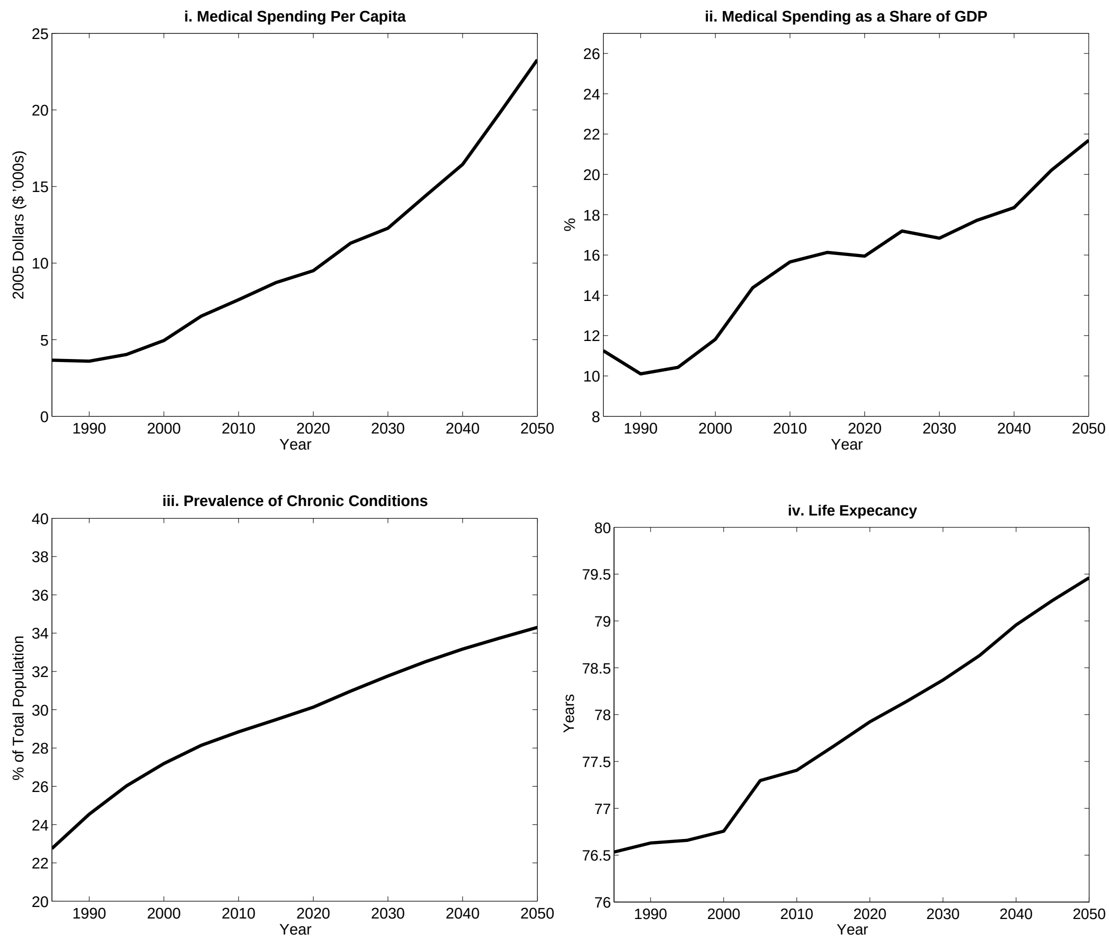


Figure 3.2 Baseline Simulation Results over the Period 1985 to 2050

3.5.1 Baseline Projections

Medical Spending

A main interest of this research project is to anticipate the evolution of medical spending in the future assuming that current trends remain unchanged. In Figure 3.2.i and 3.2.ii, we present our model's prediction of medical spending per capita and as a share of GDP. The variable presented here corresponds to total health expenditure (M_t) and not out of pocket medical expenditure. According to both of these measures, health spending is expected to continue to increase in the foreseeable future. Medical spending per capita, which increased by 79% to \$ 6,500 between 1985 to 2005, would reach \$ 11,000 in 2025. It would then more than double by 2050, to \$ 23,000. The somewhat bumpy spending and longevity results obtained between 1985 and 2000 are attributable to changes in the income per age of our distribution for 1985 to 2005, which are obtained from the PSID. For instance, the decline in spending per GDP in 1990 occurs as the income of young agents increases faster than that of older agents. Similarly, the jump in life expectancy in 2005 reflects a higher health spending path than that of agents born in 2000. Past 2005, the income distribution of 2005 is updated at a steady pace with z_w.

Our model predicts a relatively slow growth of the share of medical spending in the economy, to 21.5% in 2050. This projection contrasts with the results of Hall and Jones, which project health spending to reach over 30% over the same period. This slow increase of medical spending may be caused by our concept of technological growth. We model it as increasing the productivity of dollars spent in medical care over time, to reflect the ability of individuals to reach better health despite chronic conditions. However, Chandra and Skinner (2011) point out that many new treatments increase costs without enabling significant health improvements. According to them, the adoption of such technology may explain in large part the fast increase of medical spending in the United States. Since this facet of innovation is lacking in our model, we project the evolution of future spending if new treatments with poor cost-effectiveness were systematically rejected. Even in this optimistic setting, we find that the share of the economy allocated to health spending will rise by over 5% in the decades to come. In the next sub-section, we will see that decreases of the share of the economy spent on health may occur, depending on the evolution of technological improvements and income.

Since our model presents three main drivers of change over time, it remains unclear at this point which are responsible for the increase in costs, and to what extent. To grasp a better understanding of the underlying dynamics at play, we run three additional simulations. For each simulation, we allow only one of our three trends to take place and find the predicted impact on per capita spending over the period 1985 to 2050. The results are presented in Table 3.3. Unsurprisingly, the rise of available income and new medical technology both lead to higher spending in accordance with the results of Fonseca et al. (2009) and Hall and Jones (2007).

However, the result of our simulation with only the health depreciation trend (z_δ) is quite striking. By itself, this trend, which reproduces the impact of worsening lifestyles, leads to a slight *decrease* of medical spending. This may appear surprising, since a declining health is generally associated with higher medical spending. In this simulation, this is dominated by a concurrent factor. The faster decline of health leads to higher prevalence of chronic conditions and worse health for all age groups, which shorter lifespans and the number of periods during which agents spend on health. As we see in Table 1, this effect is quite important. Thus, our results suggest that worsening lifestyles and the rise in chronic conditions may not be an important factor of medical spending growth over the life-cycle, despite their well established impact on life expectancy and quality of life.

The difference between the total per capita spending increase of \$ 19,610 and those produced with each isolated trend provides information about the amount of complementarity between the forces at play. We find that this difference is of only 0.2%, which suggests that virtually no spending over time is incurred because of the simultaneity of the trends.

Chronic Conditions Prevalence

Our baseline scenario produces a smooth, almost linear increase of chronic condition prevalence over the 1985 to 2050 period (Figure 3.2.iii). We note that this increase, which is consistent with the stylised facts on chronic conditions noted in the introduction, is endogenous to the model and was not directly imposed by our calibration strategy.

Three main factors drive this result, two of which are endogenous to the model. First, prevalence increases most directly because of the health depreciation trend z_δ, which increases the probability of contracting conditions at young ages. Second, prevalence increases because agents suffering from them have access to better care and, as we saw, spend more on medical care. They can thus expect to live longer with their condition. Combining a faster depreciation

Table 3.3 Simulated Impact of Isolated Trends on Medical Spending

	Per Capita Spending
Total annual medical spending increase (1985 to 2050)[a]	**$ 19,610.8**
Increase due to medical productivity trend (z_θ)[b]	25.8%
Increase due to income trend (z_w)	75.9%
Increase due to health depreciation trend (z_δ)	-1.9%
Increase due to interactions between sources[c]	0.2%

[a]Obtained with our baseline projections.

[b]Corresponds to the percent of the baseline 1985-2050 increase of per capita spending obtained when only the medical productivity trend is applied. We conduct this simulation by imposing $z_\theta = 1.070$, $z_w = 1$ and $z_\delta = 1$.

[c]Difference between the total medical spending increase and the sum of the increases with the individual trends.

of health and higher survival rates of agents with chronic conditions, we find that the prevalence of chronic condition at all ages should continually increase over time (see Figure 3.3). Finally, an important cause of the increased chronic conditions prevalence is exogenous to our model. It is the rise of the share of the elderly population due to the aging of the baby-boom generation. Since such conditions are more present in the elderly population, the rapid growth of the elderly population will play a considerable role (see Figure 3.4).

Life Expectancy

Our model predicts that life expectancy should continue to rise despite the noted increase in the prevalence of chronic conditions (Figure 3.2.iii). All other things being equal, an increase in chronic conditions would decrease life expectancy, via a quicker depreciation of health. However, in our baseline scenario, agents suffering from chronic conditions have access to better care, and optimally choose to spend more on medical care. We believe our simulations provide a quite plausible explanation for recent trends, namely the simultaneous decline in the prevalence of disability among the elderly, the increase in longevity and the increase in the prevalence of chronic conditions.

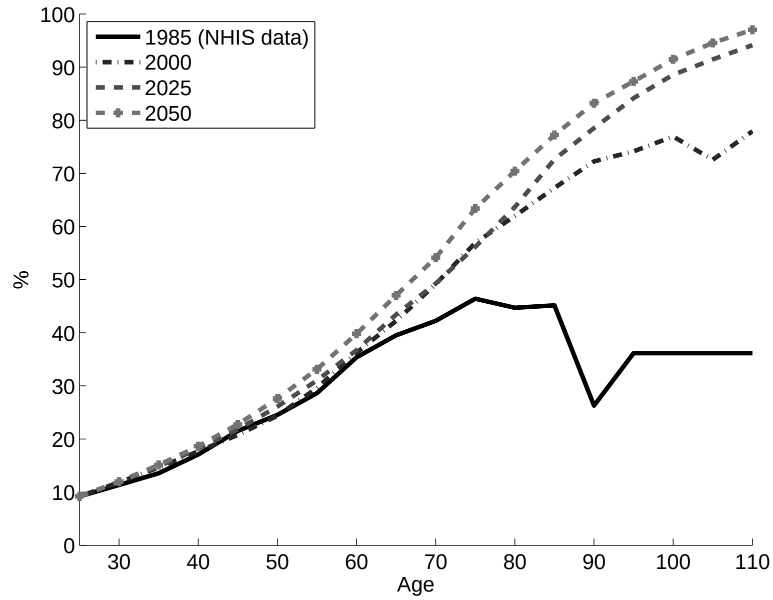


Figure 3.3 Simulated Prevalence of Chronic Conditions per Age

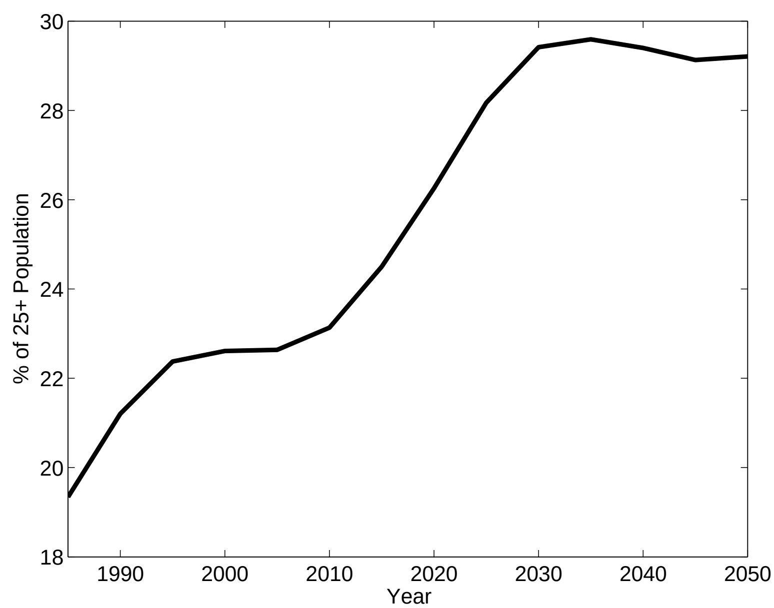


Figure 3.4 Simulated Share of 65+ Population

3.5.2 Alternative Scenarios

Our baseline scenario reveals that, if recent trends were to continue at the current pace, life expectancy, chronic condition prevalence and health spending will all continue to rise in the decades ahead. But what if these trends were to contract? After decades of impressive technological growth that greatly improved our ability to treat acute conditions, perhaps new discoveries will show less improvement and fall more and more in the costly, "grey area" category of Chandra and Skinner (2011). Also, the societal trend that lead to unhealthy lifestyles may flatten as more Americans value better nutrition, exercise, etc. In particular, policy efforts to encourage healthy lifestyles may contribute to such a reversal in the years to come. Finally, our baseline scenario of 2% annual increases of real incomes may prove too optimistic. Over the period 2005-2010, which is not considered in our calibration, the US GDP per capita has declined by 0.22% annually, and it is uncertain if and when past growth rates will be reached again. Even if GDP growth were to revert to its 1985-2005 average and remain at such a level until 2050, it is unclear whether the income of all age groups will rise, which we assume in our baseline scenario. In the PSID data, we find that the annual average income growth of respondents of the same age was of only 0.8% over the 1985-2005 period, manifestedly less than real GDP per capita growth.

To shed some light on the uncertainty of future trends, we run our baseline scenario over years 1985 to 2005, then alter one of our three trend parameters until 2050. We then capture the impact this change would have on the outcomes presented in figure 3.2. We create three alternative scenarios:

1. Slower medical productivity growth: Our first alternative scenario reduces the growth of medical technology to $z_\theta = 1.0535$, half of the baseline value.

2. Stabilization of health depreciation: Our second scenario reduces the health depreciation parameter to $z_\delta = 1$. This means that the recent adoption of sedentary employment and poor nutrition habits halts and that lifestyles will remain unchanged in the future.

3. Slower pace of income growth: This scenario has the real income of agents growing at the rate found in the PSID database, i.e. $z_w = 1.0439$. In comparison with the baseline, this scenario imposes a sharp drop in the anticipated lifetime earnings of agents in year 2010. As we mentioned, this rate is considerably slower than 1985-2005 per capita GDP growth.

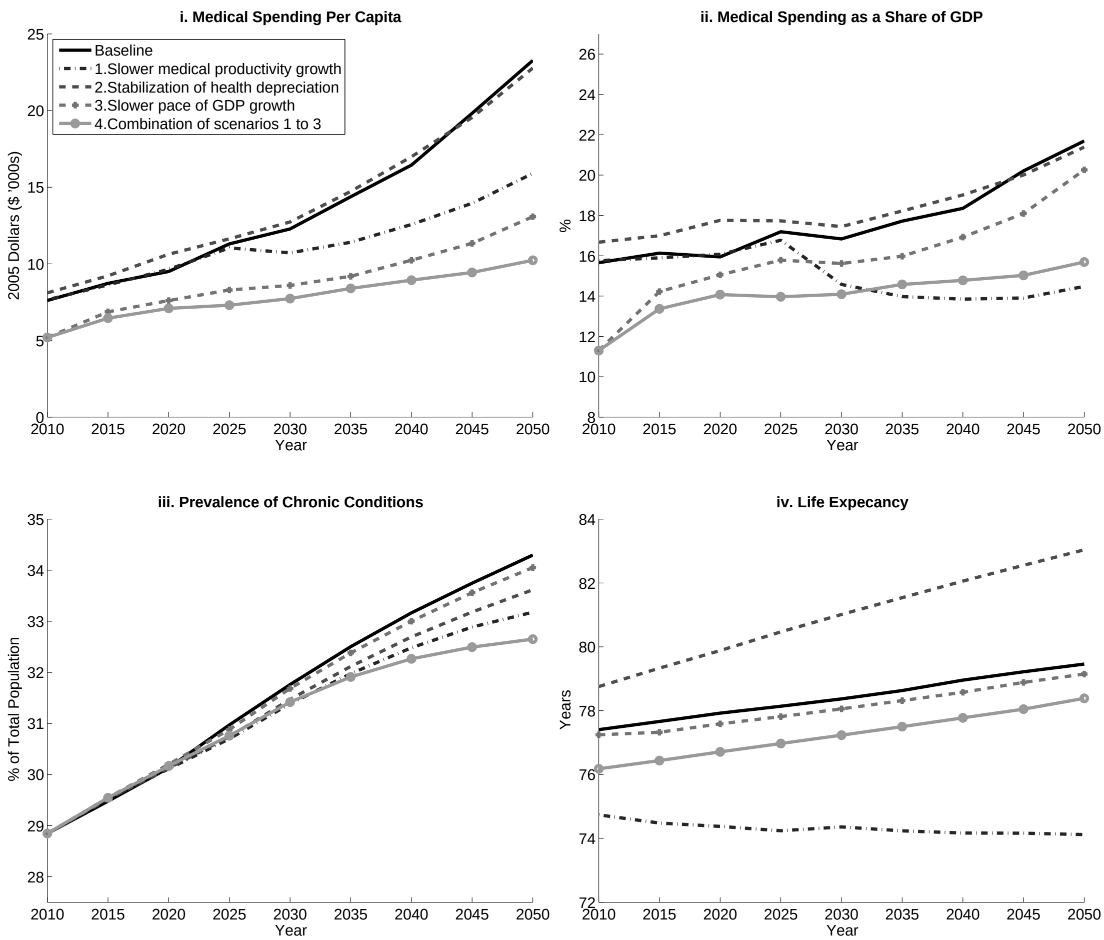


Figure 3.5 Simulations of Alternative Scenarios over the Period 2010 to 2050

4. Combination of scenarios 1 to 3: This scenario combines the slower medical productivity growth, stabilization of health depreciation and slower income growth of scenarios 1, 2 and 3. This scenario aims to capture possible complementary effects between these underlying forces.

The results of these simulations are contrasted with our baseline projections in Figure 3.5.

Medical Spending

Unsurprisingly, we find that a slower income growth (our third scenario) would have the largest impact on medical spending levels. According to this scenario, agents perceive an important drop in their future earning flows, and respond immediately by sharply lowering their medical spending. This is in line with the observed decline in out-of-pocket spending per capita over 2007 to 2010 as real incomes fell (see Figure 3.6). However, the observed decline was of only 5%, whereas our model produces a 21% drop in per capita spending.[10] This reflects the permament nature of the decline in GDP growth in our simulation, in contrast with the temporary nature of the recent economic downturn. Past that initial drop in spending levels, we predict that spending would continue to increase over the remainder of the period, both in dollars per capita and as a share of GDP.

If medical technology was to improve at a slower pace than in the past (our first scenario), we find that medical spending would increase in a fashion similar to our baseline projections until 2025. This corresponds to the period during which the rise of the share of the elderly population will be the fastest (Figure 3.4). Spending per capita would then increase at a slower pace than the baseline until 2050, and the share of GDP would peak at 16% before falling down and stabilizing at 14% of GDP. This scenario reveals that higher incomes may not automatically translate into ever growing shares of GDP being allocated to health care, as Hall and Jones (2007) find to be optimal. Medical productivity also matters. In this scenario, we find that agents will eventually lower the share of their spending allocated to health care if medical productivity fails to increase fast enough.

In Table 3.3, we saw that worsening lifestyles have a rather ambiguous impact on medical spending. On the one hand, this phenomenon leads to higher costs per age over time; on the other hand, all other things being equal, it decrease life expectancy and the number of years during which agents pay for medical care. Thus, when we hold the health depreciation at its 2005 level over years 2005 to 2050 (our second scenario), the resulting projection is very close to the baseline scenario.

Our fourth scenario combines a deceleration of the growth of incomes and medical productivity with a stabilization of the depreciation of health. In this scenario, both the ability to pay for care and the productivity of medical spending are lower than the baseline, resulting in the lowest projected health spending per capita. This would result in the share of health spending oscillating between 14 and 15.5% of GDP between 2020 and 2050, below the current share.

[10]Calculated between the per capita spending in 2005 according to the base scenario, and our projection for 2010.

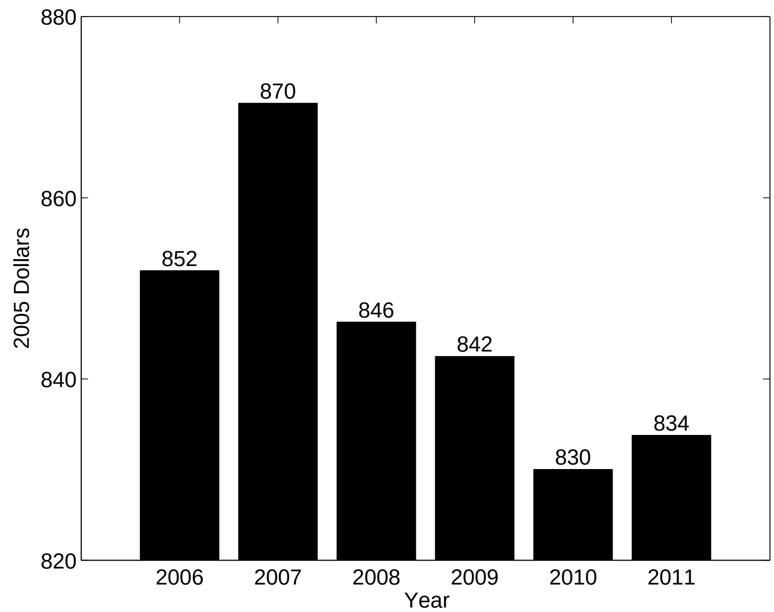


Figure 3.6 Out-of-Pocket Medical Expenditures Per Capita

Sources: Out of pocket spending and population data originates from Centers for Medicare & Medicaid Services. Spending levels are deflated with the All Items, U.S. City Average CPI from the Bureau of Labor Statistics.

Chronic Conditions Prevalence

As seen in Figure 3.5.iii, all scenarios lead to very similar predictions in terms of chronic conditions prevalence in the future. In large part, this is caused by the demographic transition of Figure 3.4. The resulting spread of prevalence rates among our simulations is only 1.65% in 2050.

The baseline scenario combines a faster health depreciation with efficient technology to treat chronic conditions and a high ability to pay. It thus produces the best conditions for high prevalence rates, which would reach 34.3% of the population in 2050. Among our scenarios 1 to 3, we find that a slower increase of medical productivity growth would lead to the lowest rates. A relatively lower curative medical productivity in the future means that Americans with chronic conditions will not live as long as in the baseline scenario, which reduces prevalence.

When we stabilize all trends (scenario 4), the difference in prevalence with the baseline scenario (1.65%) is roughly analogous to the sum of the reduced prevalence observed with the three first scenarios (2.05%).

Life Expectancy

Our life expectancy results, presented in Figure 3.5.iv, show more variation. This variability results from our conception of life expectancy, which is forward-looking and anticipates future mortality rates (see Appendix C.1). Thus, moving from one scenario to another greatly affects current life expectancy.

In terms of longevity, medical technology (scenario 1) and lifestyles (scenario 3) prove to be the most important factors. In our first scenario, medical technology improves at a much slower rate than the baseline scenario, and is insufficient to compensate for the worsening lifestyles imposed by z_δ. This decreases life expectancy over the decades to come. At the opposite, a stabilization of lifestyles would improve the health of agents over their life-cycle, and improve life expectancy.

The reduced medical spending due to a lower economic growth (scenario 2) would lower life expectancy of 2050, but by only 0.3 year. We model dollars spent on medical care as having a rapidly decreasing marginal productivity, via parameter θ_3. Spending levels thus have a lower impact on health than medical technology and lifestyles.

When we combine scenarios 1 to 3, we find an initial reduction in life expectancy, caused by the deceleration of technological progress. However, the simultaneous stabilization of health depreciations means that even a slower technological growth leads to improvements in life expectancy over time, in contrast with scenario 1.

3.5.3 Sensitivity of Results

In Appendix C.2, we conduct sensitivity analyses to evaluate how different choices of relative risk aversion and technology parameters would influence our results.

For higher values of γ, we find that the negative income shock of our third scenario would induce an *increase* in the share of GDP spent on medical spending, the opposite of the prediction of our calibrated model of the impact of an income shock on health spending to GDP ratio. However, Hall and Jones (2007) indicate that health spending should have an income elasticity well above one, invalidating this possibility. This analysis supports our choice of a low value of γ.

With regard to our technology parameters, we test the implications of changing productivity of preventive care θ_1 while adjusting $\theta_2(t)$ to maintain the 2010 productivity of health care $(\theta_1 - \theta_2\,(2010)\,\epsilon_t)$ constant for agents with chronic conditions. This modifies the relative productivity of preventive and curative care, which we arbitrarily set as equal in 1985 in our calibration. We find that both our baseline projections and scenario simulations are qualitatively robust to these changes.

3.6 Conclusion

In this paper, we develop a dynamic stochastic health care demand model that incorporates endogenous chronic conditions and mortality. We use a numerical version of the model calibrated with the U.S. experience over the period 1985 to 2005 to project medical spending, longevity and chronic conditions prevalence in the decades to come.

We find that income growth and technological progress in curative care will be the main drivers of spending in the decades ahead, in line with the findings of Hall and Jones (2007) and Fonseca et al. (2009). In our simulations, income growth are responsible for roughly three quarters of the increase of per capita spending, against one quarter for medical innovations. However, we find that innovations are the largest drivers of the increase of the share of GDP allocated to health care. Unless income growth stagnates or medical productivity does not improve as quickly as in the recent past, we project that medical spending will continue to increase at a brisk pace in the foreseeable future. In our baseline projections, in which only cost-effective technology is adopted, the share of health spending is expected to reach 22% of GDP by 2050.

A main result of our simulations is that chronic conditions prevalence is most likely to continue increasing in the years to come, but will not constitute an important generator of medical spending. Chronic conditions lead to higher medical spending at all ages, but also decrease the number of years during which agents consume medical care, via a lower life expectancy. While curbing sedentary lifestyles and obesity would certainly lead to better quality of life, less chronic conditions and higher life expectancy for Americans, we do not find that it would reduce aggregate medical spending.

Our model can be extended to study a number of essential questions. A question of interest is how the optimal consumption and population health would be affected by a reduction in insurance coverage, as medical spending impose an increasing toll on the fiscal outlook over time. Our model can be used to assess the evolution of the tax needed to finance medical spending not spent out-of-pocket over the years to come, and how imposing such a tax would affect welfare. Also, a facet that remains to be integrated in our model is the causes of technological improvements in curative care, which we assume to be exogenous. Acemoglu and Linn (2004) find that innovation in pharmaceutics is largely determined by future market size, defined as the number of individuals who would potentially consume new drugs. Following this logic, our predicted increase in chronic condition prevalence will spur research efforts in curative care, which will in turn generate more costs. This potentially important link between chronic conditions and medical costs will need to be explored in the future.

Finally, the model could be adapted in a straightforward fashion to study the health and health care costs outcomes of a number of policy scenarios, such as the shift towards Accountable Care Organizations and the adoption of specific components of the Patient Protection and Affordable Care Act of 2010.

A HEALTH CARE DEMAND AND IMPACT OF POLICIES IN A CONGESTED PUBLIC SYSTEM

A.1 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 (1.10) 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:

$$E[V_0'] = \int V(a+1, H \cdot (1 - \delta(a, \epsilon_\delta)), A_o') f(\epsilon_\delta) d\epsilon_\delta \tag{A.1}$$

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 (A.1) as the probability of being alive in $t+1$ times the expected value conditional to being alive plus the probability of dying times $\underline{V}$:

$$E[V_0'] = \left(1 - F\left(\underline{\epsilon_\delta}\right)\right) \cdot E\left[V(a+1, H_0', A_o') \mid \epsilon_\delta > \underline{\epsilon_\delta}\right] + F\left(\underline{\epsilon_\delta}\right) \cdot \underline{V} \tag{A.2}$$

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

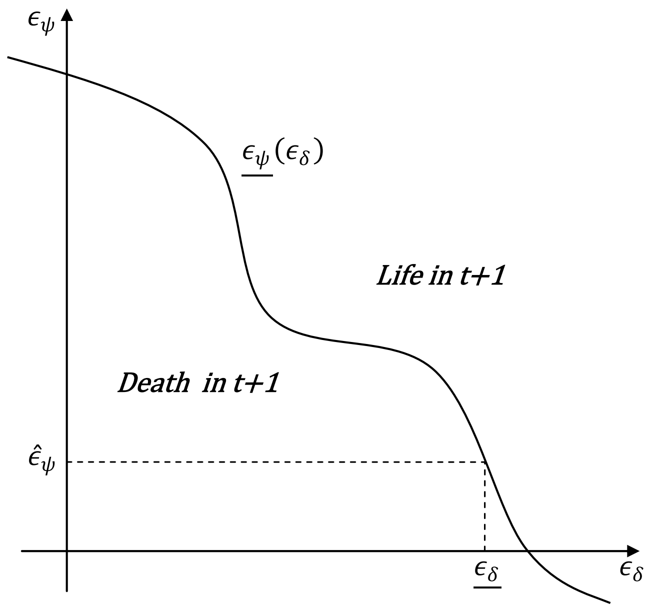


Figure A.1 Life and Death According to Stochastic Realizations

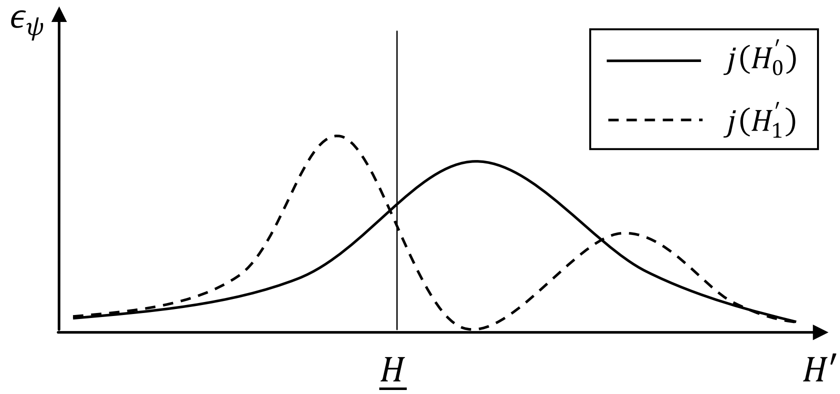
and the second on the success of health care, we cannot find a single critical value of ϵ_ψ such that $H' = \underline{H}$. Instead, $\underline{\epsilon_\psi}$ is a continuum of critical values with respect to the realization of the first shock, ϵ_δ, such that $H \cdot \left(1 - \delta(a, \epsilon_\delta) + \psi\left(H, \underline{\epsilon_\psi}\right)\right) = \underline{H}$. 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 A.1 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 $\underline{\epsilon_\psi}(\epsilon_\delta)$. $\hat{\epsilon}_\psi$ 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 $\underline{\epsilon_\delta}$.

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:

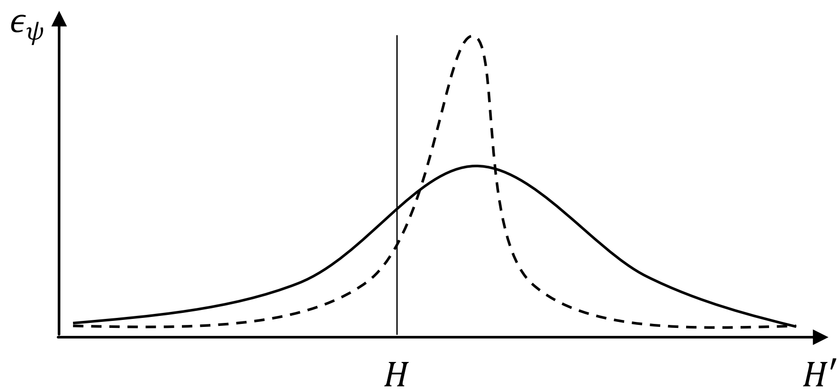
$$E\left[V_1'\right] = \left(1 - \int G\left(\underline{\epsilon_\psi}\left(\epsilon_\delta\right)\right) f\left(\epsilon_\delta\right) d\epsilon_\delta\right) \cdot E\left[V\left(a+1, H_1', A_1'\right) \mid \epsilon_\psi > \underline{\epsilon_\psi}\left(\epsilon_\delta\right) \forall \epsilon_\delta\right] + \int G\left(\underline{\epsilon_\psi}\left(\epsilon_\delta\right)\right) f\left(\epsilon_\delta\right) d\epsilon_\delta \cdot \underline{V} \tag{A.3}$$

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 (A.2) and (A.3) 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 1.3. 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, $f\left(\epsilon_\delta\right)$ and $g\left(\epsilon_\delta\right)$, whose joint distribution result in a $t+1$ health distribution. Denoting this distribution $j\left(H_\alpha'\right)$, we present in Figure A.2 a graphical analysis of three scenarios. The first two present a situation in which one of the component of dynamic gain is positive and the other is negative. In Figure A.2.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 A.2.ii, on the other hand, presents a case in which treatment increases the agent's probability of being alive in the next period, but lowers his expected quality of life. This could be observed in the case of the amputation of an infected limb, for instance. Figure A.2.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 seminal article by Arrow (1963), which emphasizes the importance of uncertainty on individual health care consumption choices.

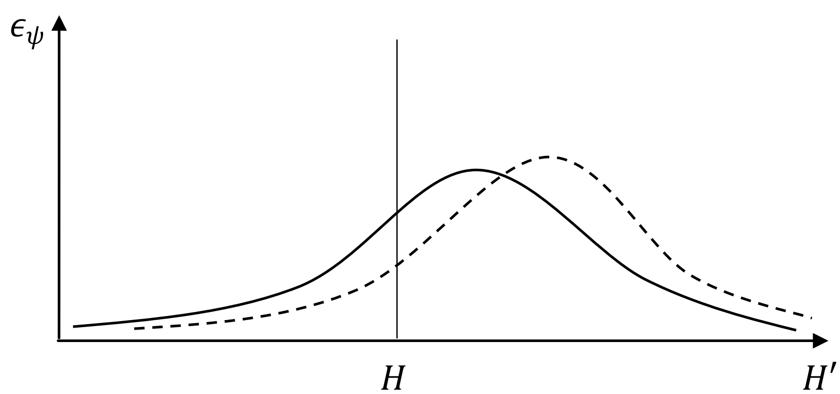
In summary, the left-hand side of (1.10) is mainly determined by the variables affecting the agent's future health. These variables are the depreciation function $\delta\left(a, \epsilon_\delta\right)$, the health production function $\psi\left(H, \epsilon_\psi\right)$, the agent's age via its impact on the depreciation of health, the agent's level of health in the current period and the distribution of health shocks, $f\left(\epsilon_\delta\right)$ and



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

Figure A.2 Dynamic Gain Examples

$g(\epsilon_\psi)$. 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 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. 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 (1.10), 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 $A = A' = 0$.[1] Under this assumption, the only way to impact the inter-temporal aspect of his problem is through the use of health care. Assuming the agent is of working age, 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 (1.2) and consumption constraint (1.3), the level of consumption and leisure time. Interestingly, this problem is both static and very simple:

$$\max_{WT} u\left(w \cdot WT \cdot (1-\tau), P - LT(H, \alpha\gamma) - TW\right) \tag{A.4}$$

$$\text{s.t. } WT \in [0, P - LT(H, \alpha\gamma)]$$

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. For utility functions that present convex indifference curves, the single constraint of (A.4) 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

[1] As will be obvious in Section 1.5, 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.

long waiting times. Unsurprisingly, differentiation shows that this utility loss is increasing in the accrued lost time incurred by using health care. This is directly relevant for policy purposes, LT being increasing in 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 allowing the congestion of the system to grow constitutes a plausible mean for governments to restrain health care consumption.

All other things being equal, the second parameter determined by the government, τ, has a much less direct impact on the cost of using care, as it affects the agent's utility in t for both values of α. Also, its impact on the optimal choice of working time depends on the utility function, since income or substitution 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 doing here. The same applies to the agent's productivity w, the second determinant of real wage in the model. 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.

A.2 Data Manipulations

To obtain a distribution of agents compatible with our numerical model, we adapted the CCHS database in several ways. First the age of adult respondents in the dataset 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 replicating 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 of this question alone is not sufficient to calibrate the health of agents, as we wish to know the health level *before* their health care consumption 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" and "much worse". Using both variables, we infer the health level of the respondents at the beginning of the year, before the health care decision. 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 $\gamma*$ 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 infer the number of consultations with specialists as the total number of consultations minus those with general practitioners. In effect, numerous Québécois visit 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 consumption decision. 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 A.3, 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 1.5.

A.3 Mortality, Life Expectancy and Savings Calculation

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

$$Q\left(a+1,H,A\right) = \sum_{H>\underline{H}}\sum_{A} Q\left(a,H,A\right)\cdot 1_{A'=A} \cdot\left[(1-\alpha*)\sum_{m=1}^{M} p_m \cdot 1_{H_0'(H,a,\epsilon_{\delta m})=H'} + \alpha*\sum_{m=1}^{M}\sum_{n=1}^{N} p_m\cdot p_n\cdot 1_{H_1'(H,a,\epsilon_{\delta m},\epsilon_{\psi n})=H'}\right], \tag{A.5}$$

where A' and $\alpha*$ 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:

$$Q_{\underline{H}}\left(a\right) = \sum_{A} Q\left(a+1,\underline{H},A\right) \tag{A.6}$$

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 for our calibration algorithm (see step 6 of Appendix A.4).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:

$$MR_{model,k} = \frac{\sum_{a=\underline{a_k}}^{\overline{a_k}} Q_{\underline{H}}\left(a\right)}{\sum_{a=\underline{a_k}}^{\overline{a_k}}\sum_{H>\underline{H}}\sum_{A} Q\left(a,H,A\right)}, \tag{A.7}$$

where $\underline{a_k}$ and $\overline{a_k}$ 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:

$$LE_{model} = \frac{\sum_a a \cdot Q_{\underline{H}}(a)}{Q_{18}} \tag{A.8}$$

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

$$\hat{A}(a, H) = \frac{\sum_A Q(a, H, A) \cdot A}{\sum_A Q(a, H, A)} \tag{A.9}$$

Whenever an age-health combinations is 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 $Q(a, H, A) > 0$.

A.4 Calibration Algorithm

We use a recursive approach to calibrate the free 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 GDP of the province, and the mortality rates and health expectancy published by the Institut de la Statistique du Québec (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 (1.6) by iterating on the value function of agents. Since this step is central to our numerical analysis, we describe it in Appendix A.5. 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 $\gamma*$ 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 5 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 A.3 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 rates, 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.[2]

[2] The reasons why we fix the optimal choices at this stage are twofold. First, steps 1 to 5 are extremely

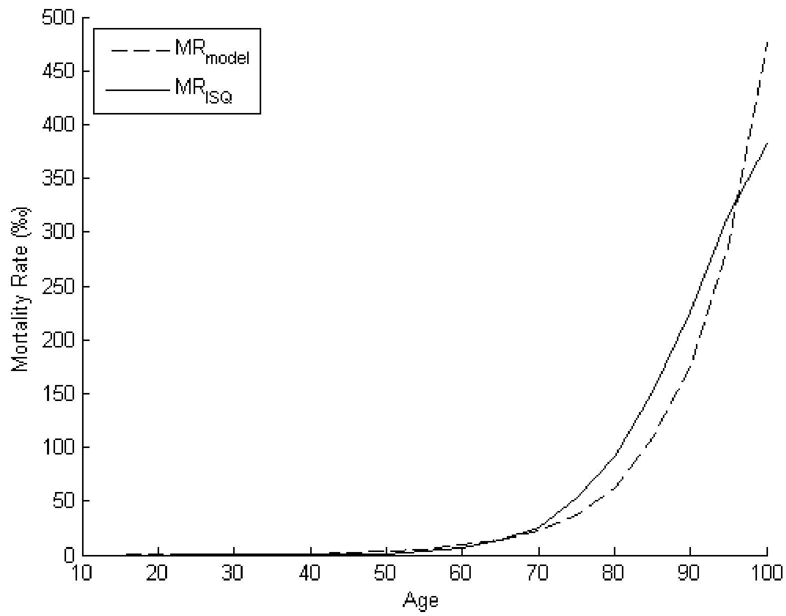


Figure A.3 Comparison of Model and Data Mortality Rates

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 life expectancy error is of 0.05 years, the SSE_{uM} is of 15 155 and the SSE_{wM} is of 8.4, resulting in a mortality criterion of 36.2. As can be seen in Figure A.3, 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.

---

calculation-heavy and necessitate a prohibitive amount of time 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 1.5.1. 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.

A.5 Bellman Equation Solving Algorithm

To solve the dynamic problem represented by Bellman equation (1.6), we iterate on the value function following Banach's fixed-point theorem. First, we choose an *ad hoc* initial value function, set to $V(a, H, A) = .0$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:

$$E\left[V_o'(a,H) \mid A'\right] = \begin{cases} \sum_{m=1}^{M} p_m \cdot V\left(a+1, H_0'\left(H, a, \epsilon_{\delta m}\right), A'\right) & \text{if } a < \overline{a} \\ \sum_{m=1}^{M} p_m \cdot V\left(\overline{a}, H_0'\left(H, a, \epsilon_{\delta m}\right), A'\right) & \text{if } a = \overline{a} \end{cases} \tag{A.10}$$

$$E\left[V_1'(a,H) \mid A'\right] = \begin{cases} \sum_{m=1}^{M} \sum_{n=1}^{N} p_m \cdot p_n \cdot V\left(a+1, H_1'\left(H, a, \epsilon_{\delta m}, \epsilon_{\psi n}\right), A'\right) & \text{if } a < \overline{a} \\ \sum_{m=1}^{M} \sum_{n=1}^{N} p_m \cdot p_n \cdot V\left(\overline{a}, H_1'\left(H, a, \epsilon_{\delta m}, \epsilon_{\psi n}\right), A'\right) & \text{if } a = \overline{a} \end{cases} \tag{A.11}$$

We then find the optimal savings and working time in $t+1$ for *both* choices of α using expected values (A.10) and (A.11), as follows:

$$V_\alpha(a,H,A) = \max_{WT,A'} u(c,L) + E\left[V_\alpha'(a,H,A) \mid A'\right] \tag{A.12}$$

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

Assuming that the expected values computed previously were right, this process explicits the value functions for both possibilities of α. We can then find the value function and optimal choice of α:

$$V(a,H,A) = \max_{\alpha} V_\alpha(a,H,A) \tag{A.13}$$

Replacing our initial *ad hoc* function with this new one, we repeat steps (A.10) to (A.13) until the supremum of the point-by-point distance between two successive value functions is nil.

A.6 Relative Risk Aversion Sensitivity Analysis

Table A.1 presents how the main results of our paper would be 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$\in (0,1)$, keeping all the other calibration parameters of Table 2 intact. The first sub-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 1.5.3.

Throughout the table, column 7, with $\rho = 0.9$, presents the largest differences with our base results (column 4, in bold characters). With a 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 and cannot work as much. 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 note quantitative variations but find largely similar qualitative results. In Table A.1.ii, the elasticity of demand is small and negative at the proximity of γ^* for all values of ρ , and table A.1.iii indicates that the expected impact of modifying the tax rate on the number of users, waiting times and welfare are quite similar. The general impact of different risk aversion values may be summarized as follows: when characterized by 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 support our choice of $\rho = 0.5$ by indicating that the results presented in the remainder of the paper would only be affected quantitatively by other values of γ.

Table A.1 Sensitivity of Results to Relative Risk Aversion Values (ρ)

i. Sensitivity of global outcome

		1	2	3	**4**	5	6	7
	$\rho =$	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
GDP		285.7	273.6	272.4	**271.9**	274.0	271.6	242.1
HC^S ($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
	H 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 $\gamma*$

		1	2	3	**4**	5	6	7
	$\rho =$	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

		1	2	3	**4**	5	6	7
	$\rho =$	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
	ΔConsumption	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
	ΔConsumption	105.2	105.5	109.4	**112.2**	117.5	131.9	115.4

B WAITING TIMES FOR MEDICAL CARE IN AN AGING POPULATION CONTEXT: THE CASE OF QUÉBEC

B.1 Evolution of Self-Perceived Health in the Canadian Community Health Survey (CCHS)

As described in section 2.2.2, we update our distribution by changing the weight of each age group according to demographic projections by the ISQ. By doing so, we keep the health distribution among a given age group constant, which constitutes a strong assumption when projecting health decisions in the long run. To validate this choice, we examine if the self-perceived health variable of the CCHS presents a noticeable evolution over its four bi-annual cycles. These cycles range from 2000/2001 to 2007/2008.

The wording of the self-reported health question is "In general, would you say your health is (excellent, very good, good, fair or poor)?" In Table B.1, we present the distribution of Québec respondents among the five categorical answers using the CCHS weighting variable to compute provincially representative percentages. Other than the first cycle presenting a higher number of "excellent" responses, we notice little change among the cycles.

This could, however, mask a trend among different age groups. To investigate this further, we map the categorical responses along the health-capital scale of our numerical health care demand model. We assign scores of 0.95 to "excellent", 0.85 to "very good", 0.75 to "good", 0.65 to "fair" and 0.55 to "poor" responses. We then calculate the average health score of Québécois for four different age groups for each CCHS cycle. The results are presented in the Figure B.1. On a year-to-year basis, we find statistically significant ($p<0.05$) decreases of self-reported health between cycles 1 and 2 for the population aged 18 to 49, and increases between cycles 2 and 3 for the population aged 35 and over. All other year-to-year variations are not significant. Since we find no evidence of a clear trend in self-perceived health over the period 2000-2008 in the CCHS, our population update procedure appears to be justified. We however discuss how a

Table B.1 Categorical Self-Perceived Health Responses of Québécois

	CCHS Cycle[a]			
	1	**2**	**3**	**4**
Excellent	28.3%	23.2%	23.4%	24.5%
Very good	32.9%	33.7%	36.0%	34.7%
Good	27.7%	32.3%	30.2%	30.8%
Fair	8.9%	8.6%	8.3%	7.8%
Poor	2.2%	2.1%	2.0%	2.0%
Missing	0.0%	0.1%	0.1%	0.1%
N	**22,012**	**27,599**	**29,165**	**23,545**

Source : Statistics Canada.

[a]Cycle 1 data was collected in 2000/2001, Cycle 2 in 2003, Cycle 3 in 2005 and Cycle 4 in 2007/2008.

trend of improving health for the elderly population would affect our results in Section 2.4.

B.2 Population Smoothing Procedure

The 2005 initial distribution is taken from our previous effort. It was built by randomly allocating a weight to each CCHS respondant from Québec to mimic 2005 population estimates. Because these estimates predate the 2006-2050 projections and this random allocation was imperfect in replicating each age, we observe some unrealistic discrepancies between the initial distribution and the 2006 projections. Figure B.2 presents these discrepancies graphically. While the general shape is the same, there are two types of inconsistencies that are problematic. First, in some instances, the population aged $a+1$ in year 2006 is *larger* than the population aged a in 2005. This is the case, for instance, for ages 45 to 49 and 80 to 90, where the 2006 population is clearly larger than its 2005 counterpart. Second, in some cases, the difference between the population of age a in 2005 and that of age $a+1$ in 2006 would require a mortality rate significantly above those produced by the ISQ for these age groups in 2005. For instance, over 15‰ Québécois of ages 20 to 24 would need to have died to meet the 2006 projection, while the ISQ's mortality rates for this age group was of only 0.4‰in 2005.

To avoid major differences between 2005 and 2006 results to be caused by these discrep-

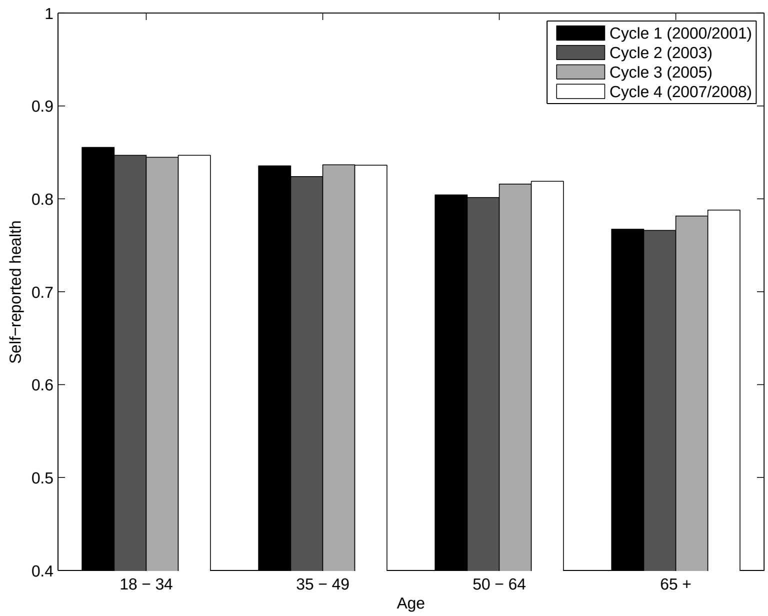


Figure B.1 Average Health per Age and CCHS Cycle

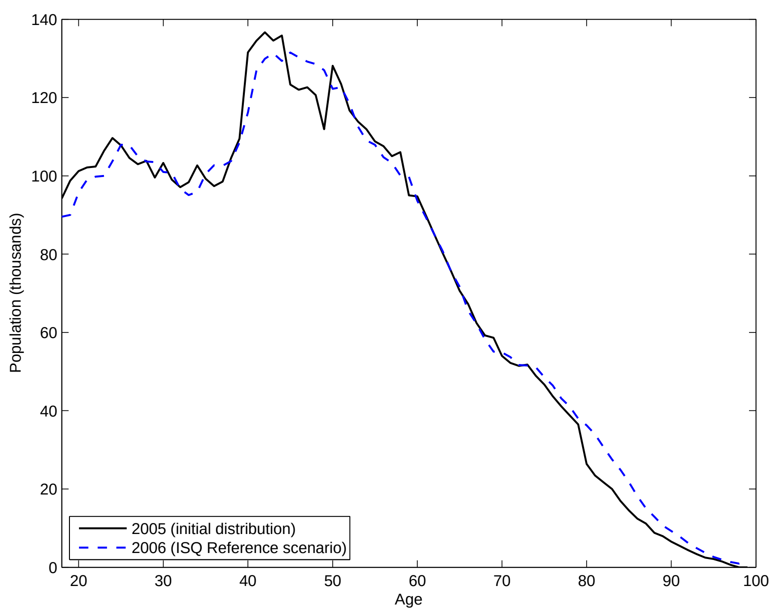


Figure B.2 2005 Initial Distribution vs. 2006 ISQ Projection

ancies, we use a simple smoothing procedure. Let us denote

$$MR_t(a) \equiv \frac{(N_t(a) - N_{ISQ,t+1}(a+1))}{N_t(a)}$$

the mortality rates of age a implied by the difference between the distribution in t and the ISQ's projection for age $a+1$ in $t+1$, and $MR_{ISQ}(a)$ the mortality rates produced by the ISQ for the same age in 2005. We smooth aberrations between $N_t(a)$ and $N_{ISQ,t+1}(a+1)$ by imposing:

$$N_{t+1}(a+1) = \begin{cases} N_t(a) \cdot (1 - 0.5 \cdot MR_t(a)) & \text{if } MR_t(a) < 0 \\ N_t(a) \cdot (1 - 1.5 \cdot MR_{ISQ}(a)) & \text{if } MR_t(a) > 1.5 \cdot MR_{ISQ}(a) \\ N_{ISQ,t+1}(a+1) & \text{otherwise} \end{cases}$$

This procedure allows for a gradual catching up towards ISQ's projection scenarios rather than an abrupt jump between the 2005 and 2006 populations. For all three projection scenarios, we find that the distributions are identical by 2025.

B.3 Dynamic Demand Model

This appendix aims to provide an overview of the life-cycle demand model at the heart of the Health Care module. A more detailed description of this model and analytic developments are presented in the first chapter.The model's main concepts and notation draw from Grossman (1972) and Ehrlich and Chuma (1990). Its numerical version is inspired by Picone, Uribe and Wilson (1998), who first applied numerical methods to study health care demand.

Overview

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:

$$\max \sum_{t=0}^{\infty} \beta^t E\left[\frac{\left(c_t^{1-\omega} L_t^{\omega}\right)^{1-\rho} - 1}{1-\rho}\right], \tag{B.1}$$

where β is the discounting factor and c_t and L_t are the consumption and leisure *in good health* during the period t. As mentioned in section 2.2.4, the fundamental basis of this 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. In public health care system with zero monetary prices such as Québec's, the quantity of health care received by a patient is not restricted by the costs of treatment. However, 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.

Constraints

We assume that the agent is unable to predict future fluctuations of his net wage, which depends on his productivity w_t and the government's income taxation rate τ_t. Thus, the agent's maximizes while holding productivity w and rate τ constant in the future. In addition to his work income, the agent can consume some of his savings A in any period. The amount of savings is not permitted to be negative at any time. For agents in the workforce, it follows that the constraint on consumption c is:

$$c \leq A + w \cdot WT\,(1 - \tau) \tag{B.2}$$

What is left of the agent's savings A at the time of their retirement can be used for consumption, in addition to a fixed exogenous public pension. Accordingly, we model retirement as the following consumption constraint when agents reach retirement age a_θ:

$$c \leq A + \theta \text{ if } a \geq a_\theta, \tag{B.3}$$

where θ is the annual public transfer received by retired agents.

During a given year, the agent faces a certain amount $LT\,(H, \alpha\gamma)$ of lost time, defined as:

$$LT\,(H, \alpha\gamma) \equiv 365 \cdot \left(1 - H + \alpha\gamma \left(\frac{H - \underline{H}}{\overline{H} - \underline{H}}\right)\right) \tag{B.4}$$

where H is his health, $\overline{H}$ and $\underline{H}$ are the health capital levels corresponding to full and lowest health, respectively, and γ the congestion level of the health care system. 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. When using the health care system, 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 Québec system, patients with severe health problems are prioritized, which means that the waiting time portion of LT increases with health. What is left of the period, the agent's available time can be freely divided between working time WT and leisure L. Thus, a one-year period is divided as follows:$LT(H, \alpha\gamma) + WT + L = 365$. French (2005) uses a similar interpretation of intra-period time distribution.

Health

The agent's stock of health-capital decreases at an uncertain depreciation rate each year, and may also be affected by the healthcare impact of treatment if he chooses to seek care. We formulate that the depreciation rate is age-dependent and affected by a stochastic shock, ϵ_δ. If the agent chooses to use health care, the expected impact of treatment is dependent on the health level of the patient, though the sign of this relationship is debatable. 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. The movement of the agent's health-capital stock is given by:

$$H'(a, H, \epsilon_\delta, \epsilon_\psi) = \begin{cases} H \cdot \left(1 - \left(\frac{a - \underline{a} - \epsilon_\delta}{\overline{a} - \underline{a}}\right) + \alpha\left(b \cdot \frac{\left(\overline{H} - H\right)(1 + \epsilon_\psi)}{\left(\overline{H} - \underline{H}\right)}\right)\right) & \text{if } H > \underline{H} \\ \underline{H} & \text{otherwise} \end{cases} \tag{B.5}$$

where $\underline{a}$ and $\overline{a}$ are the minimum and maximum ages considered in the calibration. The health capital depreciation rate is increasing in the age of the agent. We assume that agents in bad health - for whom $\left(\overline{H} - H\right)$ is high - can expect a higher gain from using health care. The parameter b dictates the capacity of current technology to improve the health of the agent towards $\overline{H}$. We assume both shocks to be distributed normally and chose their distribution, as well as the value of b (see table 2.3), to replicate essential 2005 moments.

We consider that death occurs when health level H falls below $\underline{H}$, after which the agent receives a certain *value of death* $\underline{V}$. To enforce life to be desirable over death, we assign

$$\underline{V} = {}^{-1}/_{(1-\rho)(1-\beta)} \tag{B.6}$$

Since our calibrated value of $\rho = 0.5$, this value of death is equal to an infinite sequence of periods in which the agent neither consumes nor enjoys leisure.

Bellman Equation

With respect to equations (B.1) to (B.6), the agent's maximization problem can be summarized by the following Bellman equation:

$$V\left(a, H, A\right) = \begin{cases} \max_{\alpha, WT, c} \frac{\left(c_t^{1-\omega} L_t^{\omega}\right)^{1-\rho} - 1}{1-\rho} & \text{if } H > \underline{H} \\ +\beta E\left\{V\left(a + 1', H'\left(a, H, \epsilon_{\delta}, \epsilon_{\psi}\right), A'\right)\right\} & \\ \underline{V} & \text{otherwise} \end{cases} \tag{B.7}$$

$$\text{s.t.} \begin{cases} \alpha \in \{0, 1\} \\ WT \in \left[0, P - LT\left(H, \alpha\gamma\right)\right] \\ c \in \begin{cases} \left[0, A + w \cdot WT\left(1 - \tau\right)\right] & \text{if } a < a_{\theta} \\ \left[0, A + \theta\right] & \text{if } a \geq a_{\theta} \end{cases} \\ A' = \begin{cases} \left(1 + r\right)\left(A + w \cdot WT\left(1 - \tau\right) - c\right) & \text{if } a < a_{\theta} \\ \left(1 + r\right)\left(A + \theta - c\right) & \text{if } a \geq a_{\theta} \end{cases} \end{cases}$$

This constitutes the basic health care demand model used in this paper. Since health care usage is a binary choice, 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 (B.7),

$$\beta \cdot \left(E\left[V_1'\right] - E\left[V_0'\right]\right) > u\left(c_0, l_0\right) - u\left(c_1, l_1\right) \tag{B.8}$$

This inequality can be interpreted quite easily. 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 care is uncertain and set in the future, while the cost is immediate and certain.

B.4 Bellman Equation Solving Algorithm

In our previous effort, we solved the dynamic problem represented by Bellman equation (B.7) by iterating on the value function, following the Banach fixed-point theorem. This approach enabled us to find a value for the ages over $\bar{a}$ without enforcing death at any age but, given the complexity of the dynamic problem, proved to be quite slow to solve. In this paper, we iterate on γ and τ to solve the equilibrium for 45 different years (2006 to 2050), four policy stances and three demographic scenarios. Thus, we choose a simpler algorithm that remains coherent the indeterminate lifespan of our previous effort. To do this, we assume that the expected value at ages over $\bar{a}$, $E\left[V'_\alpha\left(\bar{a},H,A\right)\right]$, will remain unchanged until 2050 at the value found in our previous effort. Then, for all ages $a \in [\underline{a},\overline{a}]$, we solve the dynamic problem using regular backward induction.

Given this assumption, the value at age $\bar{a}$ for both choices of α and for all possible savings in the next period is:

$$
\begin{aligned}
V_\alpha(\bar{a},H,A) \quad &= \quad \max_{WT,A'} u\left(c,L\right) + E\left[V'_\alpha\left(\bar{a},H,A\right) \mid A'\right] \qquad \text{(B.9)} \\
\text{s.t.} \quad & \begin{cases} c = \begin{cases} A + w \cdot WT\left(1-\tau\right) - \left(1+r\right)A' & \text{if } a < a_\theta \\ A + \theta - A' & \text{if } a \geq a_\theta \end{cases} \\ L = P - LT\left(H,\alpha\gamma\right) - WT \\ u\left(c,L\right) = -\inf \qquad\qquad \text{if } c < 0 \end{cases}
\end{aligned}
$$

This process defines the value at age $\bar{a}$ for both possibilities of α. We then find the value function and the optimal choice of α:

$$V\left(\bar{a},,H,A\right) = \max_\alpha V_\alpha\left(\bar{a},,H,A\right) \qquad \text{(B.10)}$$

Using this value, we calculate the expected $t+1$ value at age $\bar{a}-1$, for both choices of α

and for all possible choices of savings in the next period:

$$E\left[V_0'\left(\bar{a}-1,H\right) \mid A'\right] = \sum_{m=1}^{M} p_m \cdot V\left(\bar{a}, H_0'\left(H,\bar{a}-1,\epsilon_{\delta m}\right), A'\right) \tag{B.11}$$

$$E\left[V_1'\left(\bar{a}-1,H\right) \mid A'\right] = \sum_{m=1}^{M}\sum_{n=1}^{N} p_m \cdot p_n \cdot V\left(\bar{a}, H_1'\left(H,\bar{a}-1,\epsilon_{\delta m},\epsilon_{\psi n}\right), A'\right) \tag{B.12}$$

We then use these expected values to maximize the consumer problem at age $\bar{a}-1$ by repeating steps B.9 and B.10. We solve this problem until reaching minimal age $\underline{a}$.

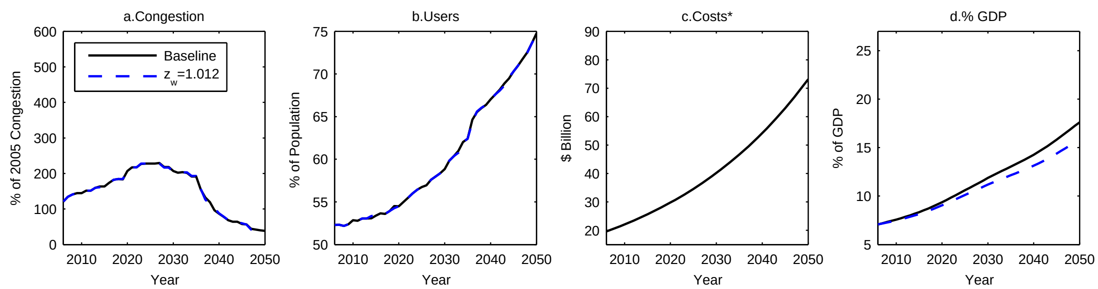
B.5 Sensitivity Analyses

This appendix aims to evaluate how our choices of trend parameters z_w and z_c influence our results.

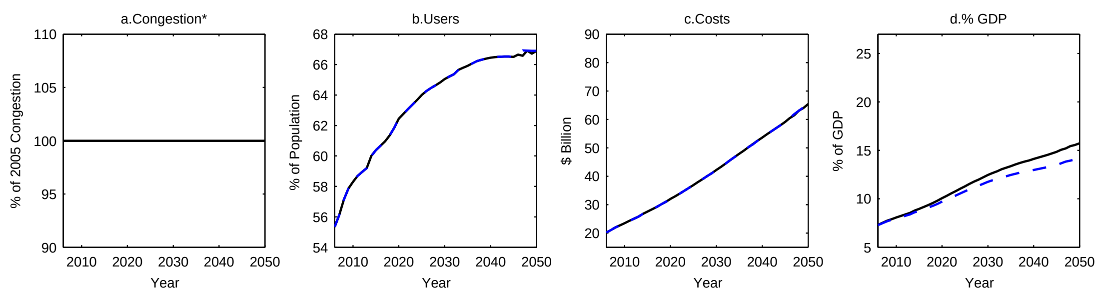
Labor Productivity Trend

On the one hand, the main implication of labor productivity growth z_w in our model is to increase Québec's collective ability to pay for care over time. As we mention in Section 2.3, we increase the productivity of labor annually by 0.96%, its observed annualized rate during years 1981 to 2010. However, Québec's Finances department anticipates a faster pace of 1.2% over the next decades. We thus simulate our three main policy scenarios with this 50% larger value and compare the results with our baseline calibration. In the Population module, we use the ISQ's "A-Reference" population scenario for this analysis.

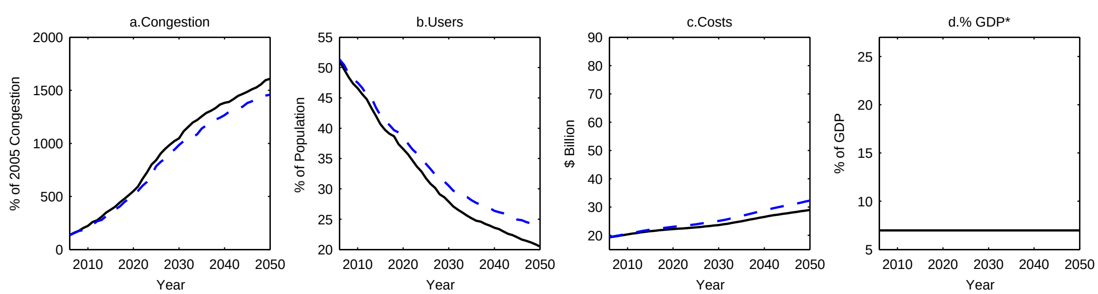
The results of this analysis for our three main policy scenarios are presented in Figure B.3. We notice that a higher productivity growth would only affect the share of GDP spent on medical care in our first two policy scenarios (Figure B.3.i and B.3.iii), since the amount of spending in the dollars is identical to the baseline. In spite of a 50% increase in the rate of productivity growth, we find a relatively small impact on the share of the economy that would be spent on health in these scenarios. In comparison with the base scenario, that share would decrease by 0.7% of GDP in 2025 and 1.6% in 2050. For the last policy scenario (Figure B.3.iii), we find that a stronger productivity trend would not do much to offset the large congestion increases predicted by our baseline scenario when maintaining the share of medical spending at 6.98% of GDP in the decades ahead.



i. Increasing Health Care Costs at Steady Pace



ii. Maintaining a Steady Congestion Level



iii. Maintaining Health Spending Constant as a Share of GDP

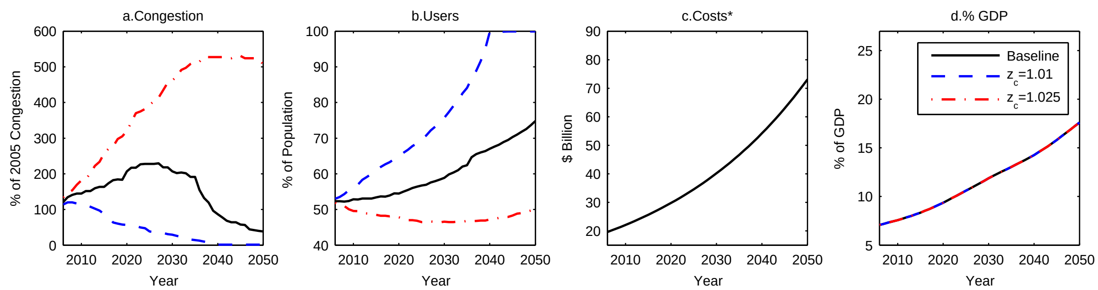
*: Path or level fixed by policy rule.

Figure B.3 Income Growth Trend Sensitivity Analysis

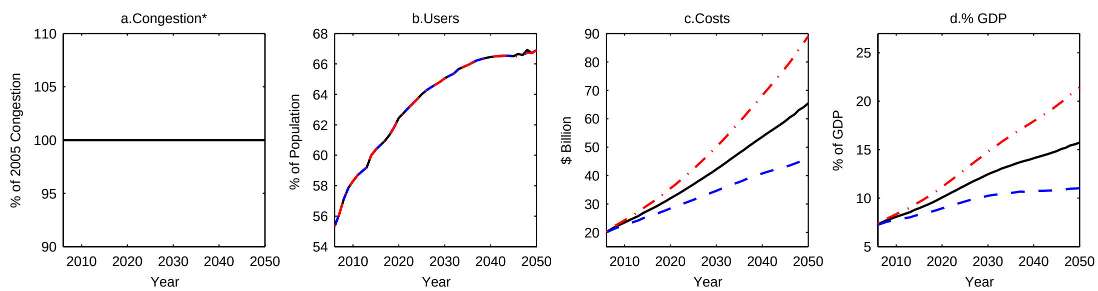
Medical Cost Trend

On the other hand, the impact of technological improvements on costs, z_c, increases the unit cost of treatment over time, which has important ramifications in our simulations. With regards to our policy scenarios, this parameter directly affects the amount of treatments supplied for a given spending and the taxes needed to maintain waiting times constant in the future. While we posed that the annual increase in the unit cost of treatment due to technological progress would be of 1.8%, analoguous to the increase in their productivity, the literature suggests that it could go both ways. We thus conduct our policy simulations with both slower (1%) and faster (2.5%) growths of unit costs, assuming that population evolves according to ISQ's "A-Reference" scenario. This exercise also sheds light on the implication of the health care system's stance towards the application of new medical technology. If all new technologies that produce health gains are adopted with little regard to their cost effectiveness, the unit cost of treatments will tend to grow faster than their productivity.

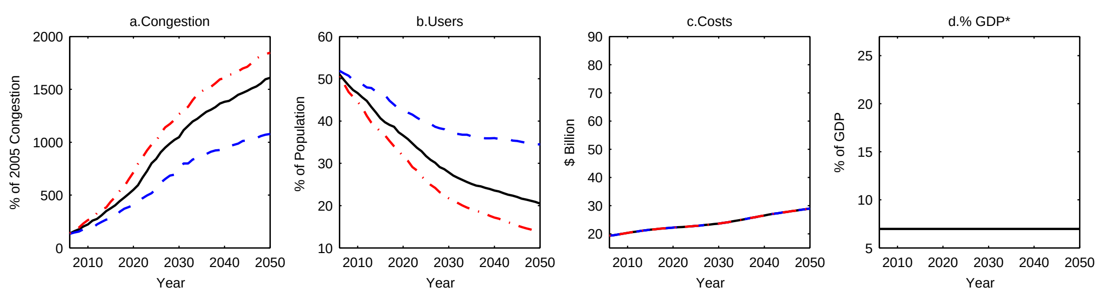
The results, presented in Figure B.4, show the far-reaching implications of the unit cost of treatments. Notably, if the cost increase is close to that of labor productivity in the other sectors of the economy ($z_c = 1.01$), the increase in the share of GDP required to maintain at their current levels (sub-figure B.4.ii.d) appears sustainable, stabilizing at 10% of GDP after 2030. Even in that optimistic scenario, however, maintaining the current share of GDP spend on health (Figure B.4.iii) would result in an explosion of congestion as the elderly population grows older. At the opposite, we find that up to 21% of GDP may be needed to be spent on health by 2050 if costs increase at a 2.5% rate, a threefold expansion from 2005.



i. Increasing Health Care Costs at Steady Pace



ii. Maintaining a Steady Congestion Level



iii. Maintaining Health Spending Constant as a Share of GDP

*: Path or level fixed by policy rule.

Figure B.4 Medical Innovation Cost Trend Sensitivity Analysis

C FORECASTING POPULATION HEALTH, HEALTH SPENDING AND MORTALITY

C.1 Results Calculations

Aggregate Medical Expeditures

Denoting $f(t, a, W, H, \epsilon)$ the number of agents with state variables (a, W, H, ϵ) in year t and $M^*(t, H, W, \epsilon)$ their optimal medical consu mption as predicted by the model in that year, we obtain aggregate medical expenditure for 2005, M_{2005}, by summing individual medical spending over cohorts and state variables:

$$M_{2005} = \frac{\sum_{a=25}^{110} \sum_H \sum_W \sum_\epsilon f(2005, a, W, H, \epsilon) \cdot M^*(2005, a, H, W, \epsilon)}{\sum_{a=25}^{110} \sum_H \sum_W \sum_\epsilon f(2005, a, W, H, \epsilon)}$$

Thus, calculating aggregate medical spending in 2005 requires solving the problem of 18 different cohorts (of ages 25 to 110 in that year), which are all facing a different dynamic problem looking forward.

Prevalence of Chronic Conditions

We proceed likewise for the prevalence of chronic conditions in a given year. In 2005, it is computed as:

$$CC_{2005} = \frac{\sum_{a=25}^{110} \sum_H \sum_W f(2005, a, W, H, -d)}{\sum_{a=25}^{110} \sum_H \sum_W \sum_\epsilon f(2005, a, W, H, \epsilon)}$$

Life Expectancy

We compute the life expectancy of a cohort with the model's predicted size of that cohort in the future. For instance, the life expectancy of the 2005 cohort, denoted LE_{2005}, is:

$$LE_{2005} = \sum_{a=25}^{110} \sum_{H} \sum_{W} \sum_{\epsilon} \frac{f\left(2005 - 25 + a, a, W, H, \epsilon\right)}{f\left(2005, 25, W, H, \epsilon\right)}$$

This concept of life expectancy reflects future trends that will affect the cohort's health and longevity, and thus differs from the usual concept of life expectancy that only takes into account mortality rates observed in a given year. In section 3.5, we study different scenarios in which we modify our baseline trend parameters. Because of the compounded impact of these changes, these scenarios have a large impact on current life expectancy as well as future life expectancy.

C.2 Sensitivity of Results

This appendix aims to evaluate how different choices of relative risk aversion and technology parameter would influence our results.

Relative Risk Aversion Parameter

As discussed in Section 3.4, our calibration effort assigns $\gamma = 0.0722$ simultaneously with technology parameters in order to be consistent with observed 1985-2005 spending patterns. As we noted, this risk aversion value is low in comparison with the literature that studies asset accumulation in the life-cycle, and means that the preferences of our agents are close to risk neutrality. In Figure C.1, we show how the baseline results of Figure 3.2 would be affected if we chose larger values of γ for the period 2010 to 2050.

As expected, higher risk aversion parameters translate into higher medical spending in figures C.1.i and C.1.ii. With $\gamma = 0.3$, spending in 2010 would be of almost 25% of GDP; similar tests with $\gamma = 1.001$, not presented here, produce a 70% share of GDP allocated to health for that year. It is not surprising to find higher spending levels when we increase the risk aversion of our agents. However, we also find that spending would tend to converge with the predictions of our calibrated model over time. In 2050, both spending levels and spending as a share of

GDP are similar for all values of γ. Such a trend towards similar spending values reflects the limitations of health care spending productivity imposed by our low value of θ_3 and the severity of illness factor in equation (3.7). Despite the observed differences in spending levels and their trajectory, projected chronic condition prevalence and life expectancy are quite similar for all values of γ.

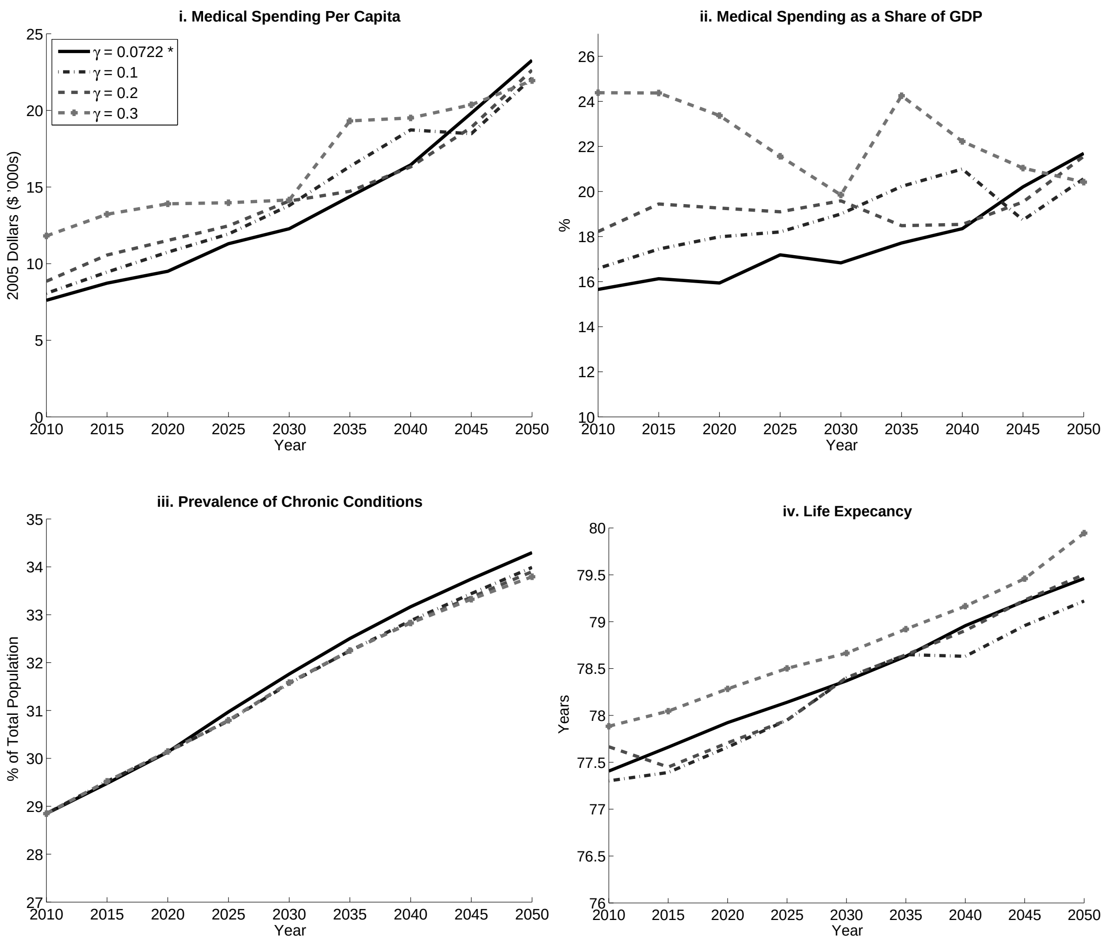
To further investigate the implications of this parameter, we conduct simulations of our four alternative scenarios of Section 3.5.2 with $\gamma = 0.1$ and $\gamma = 0.2$, and see how they affect our results. In Figure C.2, we present the projections of spending to GDP ratios and life expectancy obtained with our scenarios with these parameters. These compare to Figures 3.5.ii and 3.5.iv of Section 3.5.2, respectively.

Like in Section 3.5.2, we find that a decline in the rate of medical innovations (scenario 1) would produce the lowest health spending to GDP ratios among the first three scenarios. Also, we find that the stabilization of health depreciation (scenario 2) would have only a moderate effect on spending in comparison with the baseline scenario, though this effect seems to increase with γ.

A main difference with the results of Figure 3.5.ii concerns the reaction of users to a decline in GDP growth (scenario 3). With $\gamma = 0.2$, this scenario produces the highest health spending to GDP ratios. Agents with such preferences thus increase the share of their income allocated to health when confronted with a negative income shock; hence, health care consititutes a necessity good for them. Similarly, when combining our three scenarios (scenario 4), we obtain the lowest shares of GDP spent on health with $\gamma = 0.1$ (as we found in Figure 3.5.i with $\gamma = 0.07$) and the second largest shares when $\gamma = 0.2$, reflecting the difference in the reaction of agents to income shocks depending on their aversion to risk.

Finally, all life expectancy projections of Figures C.2.i and C.2.ii are quite similar to those found in Section 3.5.2.

This sensitivity analysis reveals that, with suffiently low values of γ, our simulations are quite similar to those of Section 3.5.2. Higher relative aversion parameter values eventually produce an income elasticity of demand of care smaller than one, and invert our calibrated model's prediction that a negative income shock would lower health spending to GDP ratio. Nevertheless, Hall and Jones (2007) stress that standard preferences imply that health spending should have an income elasticity well above one, invalidating this possibility. This analysis thus



*: Calibrated value.

Figure C.1 Sensitivity of Spending Projections to Relative Risk Aversion Parameter over the Period 2010 to 2050

strengthens our choice of a low γ.

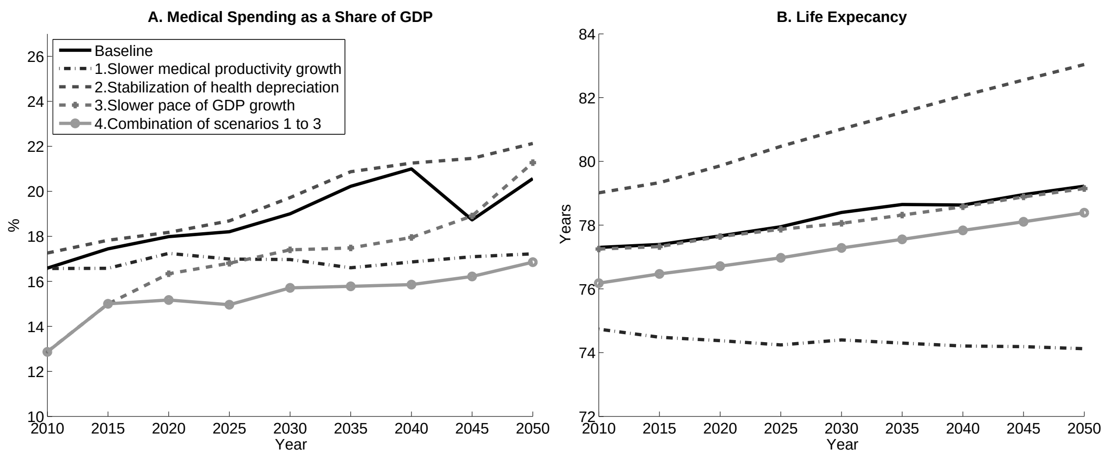
Health Care Productivity

As detailed in Section 3.4, we assumed that the impact of preventive care $(\theta_1 \cdot M)$ was equal to the impact of curative care $(-\theta_2\,(1985)\,d \cdot M^{\theta_3})$ in 1985. To assess the impact of this *ad hoc* choice, we modify the effectiveness ratio of θ_1 and θ_2 in 2010 and conduct a similar sensitivity analysis. For our sensitivity tests, we assign $\tilde{\theta}_1 = \lambda \cdot \theta_1$ with $\lambda \neq 1$. We then keep 2010 productivity constant for individuals with chronic conditions by choosing $\tilde{\theta}_2\,(2010) = {}^{(\theta_2(2010)d - \theta_1(1-\lambda))}/_{d}$. In our first sensitivity test, we ask how our results would differ with relatively more productive curative care, and choose $\lambda = 1.2$; this results in values of $\tilde{\theta}_1$ and $\tilde{\theta}_2(2010)$ of 9.751 and 1.388, respectively. In our second test, we choose $\lambda = 0.8$, which translates in technology parameters of $\tilde{\theta}_1 = 6.501$ and $\tilde{\theta}_2(2010) = 1.815$. In Figure C.3, we show how the baseline results of Figure 3.2 are affected by these different technology ratios.

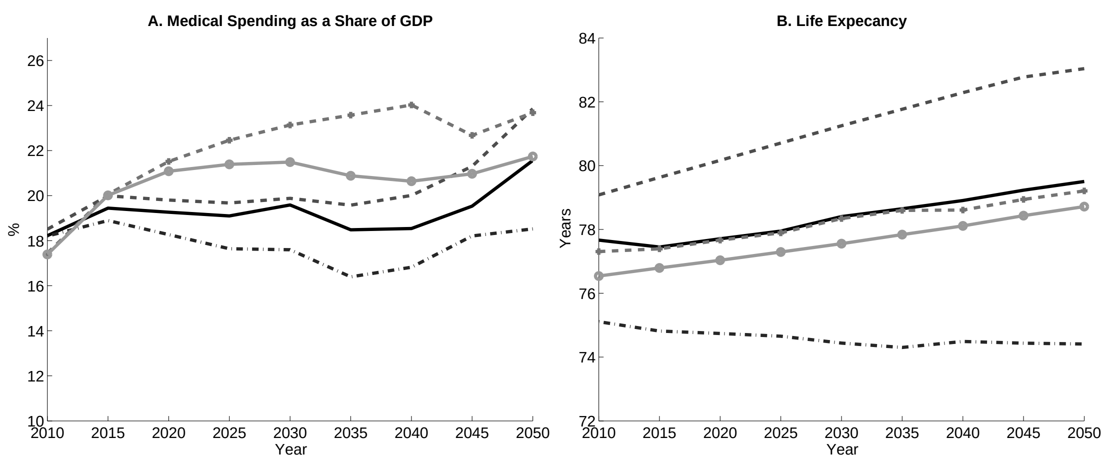
The evolution of spending in the decades to come (Figures C.3.i and C.3.ii) are robust these different technology parameters. With $\lambda = 1.2$ $(\theta_1 = 9.75; \theta_2(2010) = 1.39)$, the 2010 medical productivity is the same for agents with chronic conditions, and 20% higher for agents without conditions. Since preventive care, which is available to all agents, is more productive, this sensitivity scenario yields higher spending throughout the horizon and 0.6% less chronic conditions than the baseline by 2050 (Figure C.3.iii). However, the long-term path of medical productivity is lower than with our calibrated values, since the compounded improvements of medical productivity gains are applied to a lower $\theta_2(2010)$. This yields lower life expectancies than the baseline in Figure C.3.iv, despite higher medical spending. For $\lambda = 0.8$, we find the opposite.

In Figure C.4, we simulate the four alternative scenarios of Section 3.5.2 with these modified technology parameters. The simulated spending with more productive preventive care (Figure C.4.i.A) are quite similar to those of our calibrated model (Figure C.2.ii). In both figures, Scenarios 1, 3 and 4 produce lower health spending to GDP ratios than the baseline. Also, the stabilization of health depreciation scenario produces spending close to the baseline, which rise to about 22 % in 2050, in both figures.

With a higher relative curative care (Figure C.4.ii.A), we notice the same general trends, but with some notable differences. First, the 2010 health to GDP ratios are lower, which is

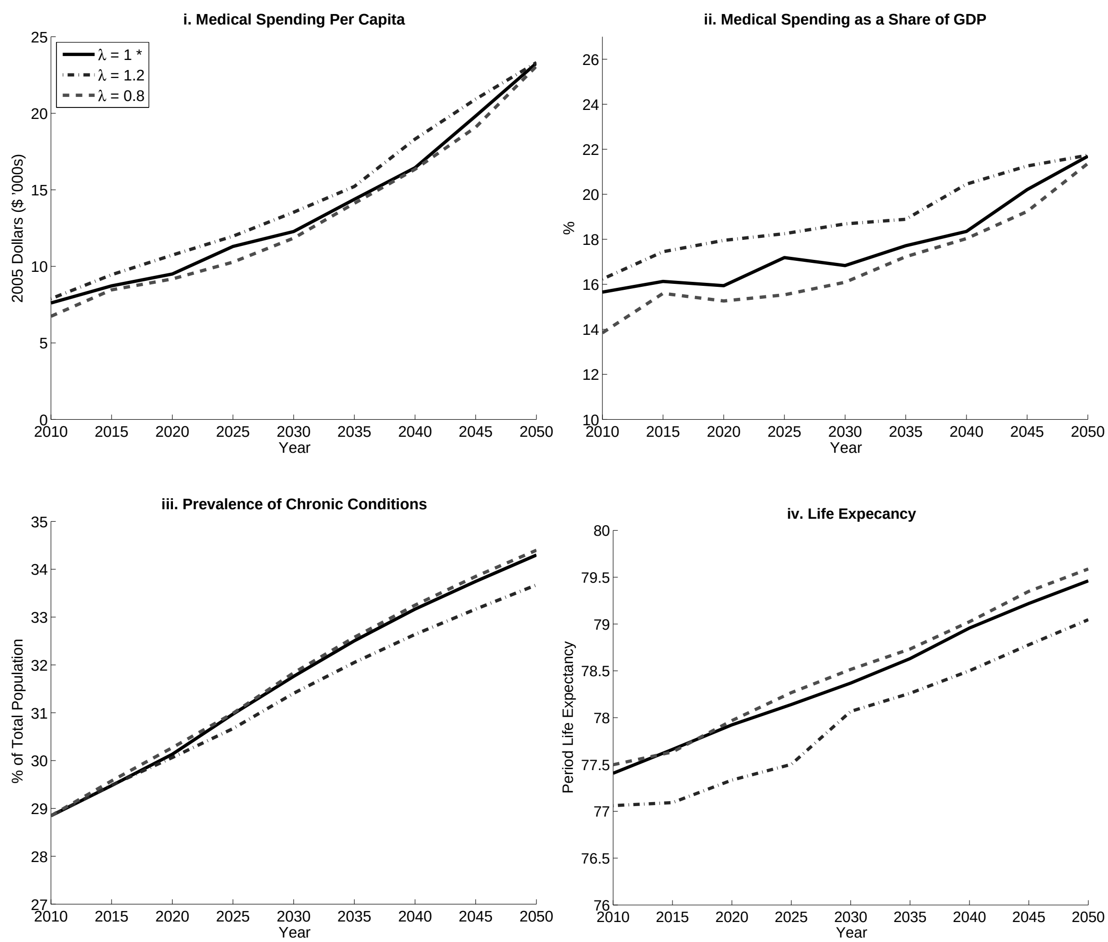


i. Scenario Simulations with $\gamma = 0.1$



ii. Scenario Simulations with $\gamma = 0.2$

Figure C.2 Sensitivity of Alternative Scenarios to Relative Risk Aversion Parameter over the Period 2010 to 2050



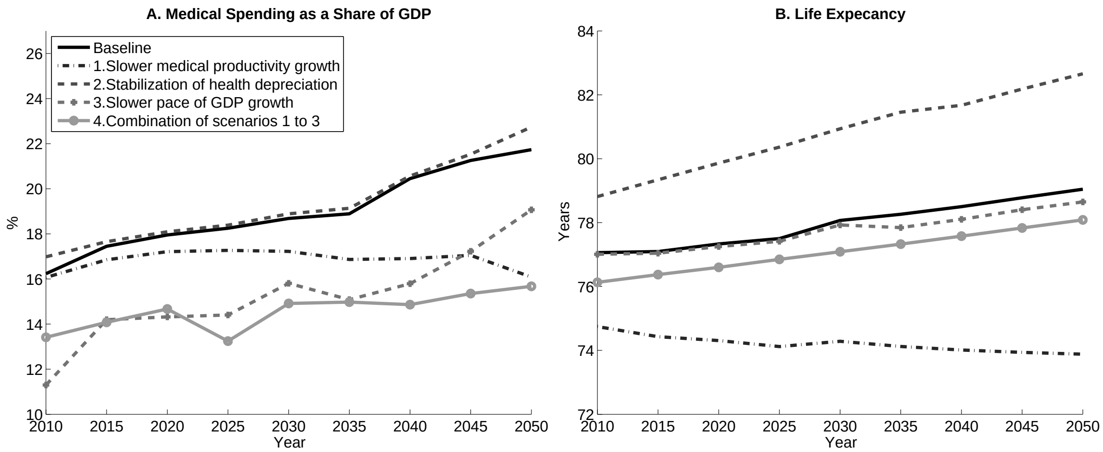
*: Calibrated values.

Figure C.3 Sensitivity of Projections to Technology Parameters over the Period 2010 to 2050

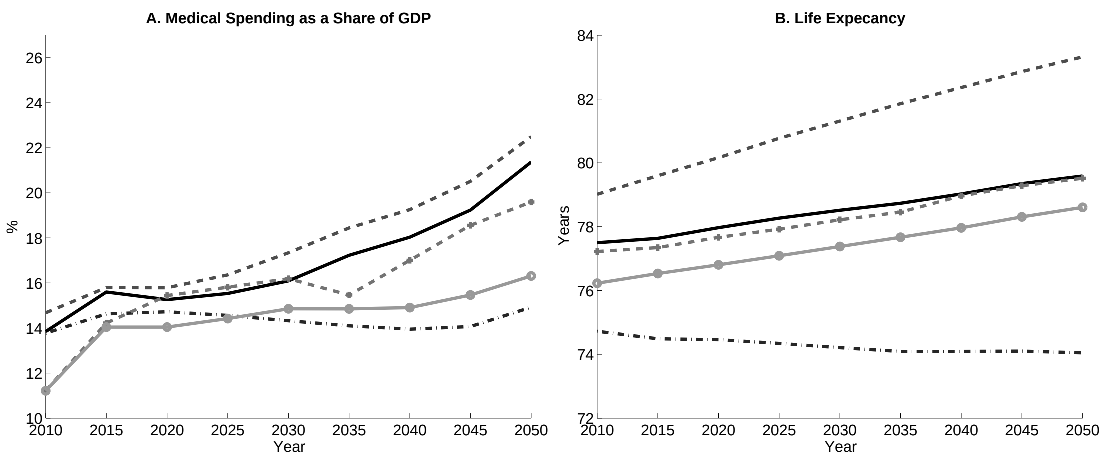
consistent with our reading of Figure C.1.ii. Second, the impact of scenarios 2 and 4, which both contain a slower pace of GDP growth, are more moderate in comparison with the baseline scenario then we found in Figures C.2.ii and C.4.i.A. This lesser impact of income on spending is caused in part by the higher chronic condition prevalence noted in Figure C.3.iii, and in part by the higher productivity path stemming from the compounded impact of medical innovation being applied to a greater $\theta_2(2010)$. This second factor also explains why the share of GDP allocated to care increases faster past 2035 for the Baseline scenario and scenarios 2 and 3 in Figure C.4.ii.A; these three scenarios allow for a fast pace of productivity growth.

Finally, the impact of the modified technology parameters on life expectancy projections, presented in Figures C.4.i.B and C.4.ii.B, are very close to those of our baseline.

For both of these senstivity tests on our key medical technology parameters, we find that our scenario simulations are qualitatively similar to those reported in Section 3.5.2.



i. Sensitivity of Scenario Simulations with $\lambda = 1.2$ ($\theta_1 = 9.751$ and $\theta_2 = 1.388$)



ii. Sensitivity of Scenario Simulations with $\lambda = 0.8$ ($\theta_1 = 6.501$ and $\theta_2 = 1.815$)

Figure C.4 Sensitivity of Alternative Scenarios to Technology Parameters over the Period 2010 to 2050

BIBLIOGRAPHY

Acemoglu, D., and Linn, J. (2004). "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," *Quarterly Journal of Economics*, 119(1), 1-37.

Anderson, G. (1997). "Willingness to Pay to Shorten Waiting Time for Cataract Surgery," *Health Affairs*, 16(5). 181-190.

Arrow, K.J. (1963). "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53(5), 941-973.

Bagust, A. Hopkinson, P. K., Maslove, L. and Currie, C. J. (2002). "The projected health care burden of Type 2 diabetes in the UK from 2000 to 2060," *Diabetic Medicine*, 19(suppl.4), 1-5.

Baicker, K., and Skinner, J.S. (2011). "Health Care Spending Growth and the Future of U.S. Tax Rates," *NBER Working Paper*, No. 16772.

Blau, D.M. and D. Gilleskie (2008). "The Role of Retiree Health Insurance in the Employment Behavior of Older Men," *International Economic Review*, 49(2), 475-514.

Cagetti, M. (2003). "Wealth Accumulation Over the Life Cycle and Precautionary Savings," *Journal of Business and Economic Statistics* (21(3), 339–353.

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.

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.

Castonguay, C. and Laberge, M (2010). "La longévité : une richesse," *CIRANO Research Paper* 2010RP-01.

Chang, F-R (1995). "Uncertainty and Investment in Health," *Journal of Health Economics*, 15(3), 369-376.

Crimmins, E. M. (1996). "Mixed Trends in Population Health among Older Adults," *Journal of Gerontology: Social Sciences*, 51B(5), S223-S225.

Chandra A. and Skinner J (2011). "Technology Growth and Expenditure Growth in Health Care," *Journal of Economic Literature*, 50(3), 645-80.

Chang, F-R (1995). "Uncertainty and Investment in Health", *Journal of Health Economics*, 15(3), 369-376.

Comité consultatif sur l'économie et les finances publiques (CCEFP) (2009). *Le Québec face à ses défis* (Booklet 1), ISBN 978-2-550-57837-6.

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.

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.

Cutler, D. M., Glaeser, E.L, Shapiro, J.M. (2003). "Why Have Americans Become More Obese?," *Journal of Economic Perspectives*, 17(3), 93-118.

Cutler, D.M, Deaton, A, and Lleras-Muney, A. (2006). "The Determinants of Mortality," *Journal of Economic Perspectives*, 20(3), 97-120.

Cutler, D.M. and McClellan, M. (2001): "Is Technological Change in Medicine Worth It?," *Health Affairs*, 20(5), 11-29.

Cutler, D.M., Rosen, A.B. and Vijan, S. (2006). "The Value of Medical Spending in the United States, 1960-2000," *New England Journal of Medicine*, 355, 920-927.

Cutler, D. M., and Richardson E. (1997), "Measuring the Health of the U.S. Population," *Brookings Papers on Economic Activity, Microeconomics*, 217-282.

Department of Finance Canada (2012). *Transfer Tables* (website), http://www.pourunquebeclucide.info/documents/manifeste.pdf.

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.

Finances Québec (2010). *Vers un système de santé plus performant et mieux financé*, ISBN 978-2-550-58433-9.

Flegal, K.M., Carroll, M.D., Odgen, C. L., and Johnson, C. L. (2002). "Prevalence and Trends in Obesity Among US Adults, 1999-2000," *Journal of American Medical Association*, 288(14), 1723-1727.

Fonseca R., Michaud, P.-C., Galama, T., and Kapteyn, A. (2009). "On the Rise of Health Spending and Longevity," *RAND Working Paper*.

Freedman, V.A., Schoeni, R.F., Martin, L.G. and Cornman, J.C. (2007). "Chronic Conditions and the Decline in Late-Life Disability," *Demography*, 44(3):459–77.

French, E. (2005). "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, 72(2), 395-427.

Galama, T., Kapteyn, A., Fonseca, R., and Michaud, P.-C. (2008). "Minimally Productive Health and Retirement," *RAND Working Paper.*

Gravelle, H. and Siciliani, L. (2007). "Is Waiting-Time Prioritisation Welfare Improving?," *Health Economics*, 17(2), 167-184.

Goldman, D., et al. (2004). *Health Status and Medical Treatment of the Future Elderly: Final Report*. Santa Monica, CA, RAND Corporation.

Grossman, M. (1972). "On the concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223-255.

Häkkinen, U., Martikainen, P., Noro, A., Nihtilä, E., Peltola, M. (2008). "Aging, Health Expenditure, Proximity to Death, and Income in Finland," *Health Economics, Policy and Law*, 3, 165-195.

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.

Halliday, T.J., Hui, H. and Zhang, H. (2009). "Health Investment over the Life Cycle," *IZA Discussion Paper 4482*.

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.

Institut de la statistique du Québec (ISQ) (2009). *Perspectives démographiques du Québec et des régions, 2006-2056.* ISBN 978-2-550-56456-0.

Jones, C.I. (2002). "Why Have Health Expenditures as a Share of GDP Risen so Much?," *NBER Working Paper 9325.*

Jung, J. and Tran, C. (2007). "The Macroeconomics of Health Savings Accounts," *CAEPR Working Paper 2007-023.*

Keehan, S., Sisko, A., Truffer, C. Smith, S. Cowan, C., Poisal, J. and Clemens, M.K. (2008). "Health Spending Projections Through 2017: The Baby-Boom Generation is Coming to Medicare," *Health Affairs*, 27(2), w145-w155.

Kydland, F.E., and Prescott, E.C. (1982). "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345-137.

Lakdawalla, D. Tomas, P. (2009). "The growth of obesity and technological change," *Economics and Human Biology*, 7(3), 283-293.

Lee, R. and Carter, L.R. (1992). "Modelling and Forecasting U.S. Mortality," *Journal of the American Statistical Association*, 87, 659–71.

Lindsay, C.M., and Feigenbaum, B. (1984). "Rationing by Waiting Lists," *The American Economic Review*, 74(3), 404-417.

Madore, O. (2005). "The Canada Health Act: Review and Options," *Current Issue Review 94-4E.*

Manchester, J. (2012). "CBO's Long-Term Projections for Medicare and Medicaid Spending in the United States," OECD Expert Workshop on Improving Health Expenditure Forecasting Methods.

Manton, KG, Gu, X., and Lamb, V.L. (2006). "Change in Chronic Disability from 1982 to 2004/2005 as Measured by Long-Term Changes in Function and Health in the U.S. Elderly Population," *Proceedings of the National Academy of Sciences of the United States of America*, 103(48), 18374-18379.

Martin, S. and Smith, P.C. (1999). "Rationing by Waiting Lists: an Empirical Investigation," *Journal of Public Economics*, 71(1), 141-64.

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.

Meara, E., White, C. and Cutler, D.M. (2004). "Trends in Medical Spending By Age, 1963-2000," *Health Affairs*, 23:4, pp. 176-183.

Ministère de la santé et des services sociaux du Québec (MSSS) (2008). *In Brief: The Québec Health and Social Services System*, ISBN 978-2-550-51671-2.

Murphy, K. and Topel, R. (2006). "The Value of Health and Longevity," *Journal of Political Economy*, 114:5, 871-904.

National Center for Health Statistics (2011). *Health, United States, 2011: With Special Feature on Socioeconomic Status and Health*. Hyattsville, MD.

Newhouse, J.P. (1992). "Medical Care Costs: How much Welfare Loss?," *Journal of Economic Perspectives*, 6:3, pp. 3-21.

Organisation for Economic Co-operation and Development (OECD) (2013). "Life expectancy," in *OECD Factbook 2013: Economic, Environmental and Social Statistics*, OECD Publishing. http://www.oecd-ilibrary.org/economics/oecd-factbook-2013/life-expectancy_factbook-2013-95-en.

OECD (2010). *OECD Health Data 2010* (database), http://www.oecd.org/health/healthdata.

OECD (2012). *OECD Health Data 2012* (database), http://www.oecd.org/health/healthdata.

Oeppen, J. and Vaupel, J. W. (2002). "Broken Limits to Life Expectancy," *Science*, 296, 1029-1031.

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.

Régie de l'assurance-maladie du Québec (RAMQ) (2013a). *Health Insurance*, Pamphlet K-0506-1.

Régie de l'assurance-maladie du Québec (RAMQ) (2013b). *Health Insurance: Medical Services*, http://www.ramq.gouv.qc.ca/en/citizens/health-insurance/healthcare/Pages/medical-services.aspx.

Régie de l'assurance-maladie du Québec (RAMQ) (2013c). *Prescription Drug Insurance: Description*, http://www.ramq.gouv.qc.ca/en/citizens/prescription-drug-insurance/Pages/description.aspx.

Reinhardt, U.E. (2000). "Health Care for the Aging Baby Boom: Lessons from Abroad," *The Journal of Economic Perspectives*, 14(2), 71-83.

Seshamani, M. and Gray, A. (2004). "Ageing and Health-Care Expenditure: The Red Herring Argument Revisited," *Health Economics*, 13, 303-314.

Schoeni, R. F., Freedman, V. and Martin., L.G. (2008). "Why is Late-Life Disability Declining?," *Milbank Quarterly*, 86 : 47-89.

Scholz, J.K., and Seshadri, A. (2010). "Health and Wealth in a Life-Cycle Model," *Michigan Retirement Research Center Working Paper 2010-224*.

Shoven, J.B., Topper, M.D., Wise, D. (1994). "The Impact of the Demographic Transition on Government Spending", in *Studies in the Economics of Aging*. David Wise, ed. Chicago: University of Chicago Press.

Siciliani, L. and Hurst, J. (2004). "Explaining Waiting-Time Variations for Elective Surgery Across OECD Countries," *OECD Economic Studies*, No. 38, 95-123.

Stock, L.M., Bradley, G.E., Lewis, R.J., Baker, D.W., Sipsey, 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.

Wolfe, J.R. (1985). "A Model of Declining Health and Retirement," *Journal of Political Economy*, 93(6), 1258-1267.

Zveifel, P., Felder, S., Meiers, M. (1999). "Ageing of Population and Health Care Expenditure: A Red Herring?," *Health Economics*, 8, 485-496.