Chaire de recherche
sur les enjeux économiques
intergénérationnels

COMPAS

COMPAS: A Health Microsimulation Model for Quebec and Canada

David Boisclair, Yann Décarie, François Laliberté-Auger,
Pierre-Carl Michaud

Document technique
Technical Document
Décembre / December 2019

HEC MONTRÉAL ESG UQÀM

Chaire de recherche
sur les enjeux économiques
intergénérationnels

est une chaire multi-institutionnelle qui s'appuie
sur un partenariat avec les organisations suivantes :

COMPAS: A Health Microsimulation Model for Quebec and Canada

Technical Appendix

PROJECT TEAM:[1]

David Boisclair, HEC Montréal
Yann Décarie, HEC Montréal
François Laliberté-Auger, HEC Montréal
Pierre-Carl Michaud, HEC Montréal[2]

December 2019

[1]We wish to thank the following former collaborators of the team for their contribution to previous versions of COMPAS: Aurélie Côté-Sergent, Jean-Yves Duclos, Alexandre Lekina and Steeve Marchand.

[2]Contact: Département déconomie appliquée, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherine, Montréal (Québec), H3T 2A7, Canada.

Contents

List of Figures

List of Tables

Chapter 1

COMPAS: Context and model overview

1.1 Context

The share of national income dedicated to health care globally has increased steadily over time, except for short periods. In Canada, this share has increased from 7.1% in 1981 to 11.0% in 2014 (Canadian Institute for Health Information, 2014). The same observation can be made with respect to health care spending as a share of the state's budget. This is particularly true in Canada, where around 70% of health expenditures are public. The share of total government expenditures dedicated to health care has risen from 33% in 1993 to nearly 36% in 2013, with a peak at around 39% in 2004 (Canadian Institute for Health Information, 2014).

In most industrialized countries, the increase in spending has been accompanied by an increase in life expectancy. Thus, a Canadian born in 1960 had a life expectancy at birth of 71.0 years, compared to 75.3 years if born in 1980 and 82.3 years if born in 2017 (World Bank, 2019). A major share of the gains have been achieved for those over the age of 50: among men, between 1991 and 2008, life expectancy at 50 years old rose from 27.6 to 31.0 years, while women also saw appreciable gains, going from 32.8 to 34.7 years (Bohnert, 2013).

Some important trends are underway. To start with, long-term demographic trends such as an aging population and longer life expectancy are exerting pressure on public finances and the health system as a whole. Also, medical technological progress and the capacity to provide home care to elderly persons are key factors in the future evolution of the amount of resources devoted to health.

Moreover, we observe a progression of chronic illness in Canada, a progression which has increased pressures on the health system and the budget of the state. For instance, the prevalence of hypertension rose from 14.4% to 17.6% between 2003 and 2014 (Statistics Canada, 2016b). Among the population aged 65 years and over, the prevalence of type 2 diabetes increased by nearly one-half between 2003 and 2014, from 12.5% to 18.2% (Statistics Canada, 2016b). It is thus imperative to be able to adequately model and analyse these trends to obtain reliable projections of the future situation.

Some studies, such as Clavet et al. (2013), use a microsimulation model and project that health spending in Quebec will amount to at least 56.5% of public revenues in 2030 – and could even exceed 68% according to a plausible scenario. Using a methodology which makes use of aggregate data, Godbout et al. (2014) obtain results of the same magnitude (58.6%), as have many previous similar works going back to the early 2000s. However, evaluation of policies which aim to attenuate these pressures while continuing to improve the population's health requires models capable of accounting for the aforementioned factors in a detailed manner.

1.1.1 Long-term microsimulation

Since the observed trends in disease prevalence are complex, modelling at the aggregate level does not allow to treat these questions while accounting simultaneously for interactions between diseases. Among other aspects, it cannot account for increasing heterogeneity in life trajectories. This element was brought up during the work of a panel aiming to study the future evolution of health expenditures in the United States (National Research Council (US) Committee on National Statistics, 2010). The *Future Elderly Model* (FEM), originally developed by a team of researchers at the RAND Corporation, is one of the only long-term models in the United States and in the world to incorporate complex interactions between the diverse trends noted above. Initially built using an administrative database on 100,000 Medicare beneficiaries (thus Americans aged 65 years and over), the model now covers the entire population of the United States aged 50 years and over. It allows users to project health status, use of medical resources, as well as labour income, labour supply and individual retirement decisions; the impact of future changes in health status, longevity and medical technologies can also be forecast. The model has been used in various studies analyzing different questions (see for example Goldman et al., 2013; Michaud et al., 2012). An adapted version of the FEM for European countries, using the Survey of Health, Ageing and Retirement in Europe (SHARE) data, has been developed (Atella et al., nd).

This type of model always begins with an initial scenario where there is no future change in government policy.[1] These models are nevertheless very useful, both to consider medium- and long-term scenarios, and to evaluate the potential benefits of policies which aim to improve the population's health using "alternative" scenarios constructed by users. These benefits go well beyond reducing health expenditures since they also cover the economic benefits flowing from improved health and longevity and from the effects on revenues and expenditures of different public programs. It is in this perspective that we develop a multidimensional health microsimulation model for Canada and its provinces/regions.

1.1.2 Microsimulation linked to aging and health in Canada

A certain number of microsimulation models have been developed and applied to aging or health in Canada over the years. Some have been general in nature, such as Statistics Canada's Lifepaths, while others were always oriented towards health, such as POHEM, also developed by Statistics Canada. To

[1] The idea underlying the base (or "reference") scenario is not that policies will actually not change, but rather that the long-term projections should be carried out on the basis of what is known of government policy at the time of performing the simulations.

our knowledge, however, none of the existing models has been developed specifically for provinces/regions, and none of these model simultaneous transitions between different states of health defined by the presence of various illnesses and risk factors. These existing models are briefly described below, based on the inventory performed by Décarie et al. (2012) as well as the summary included in Clavet et al. (2012).

The oldest microsimulation model found in Canada is DYNACAN, set up in the 1970s in what is today the federal department of Employment and Social Development Canada. This dynamic model, based on the 1971 Census and on the American CORSIM model, is oriented towards public pension programs (Old Age Security, Canada Pension Plan) and on the effect of policies relating to these. DYNACAN was abandoned at the end of the 2000s.

The model most recently used by the Canadian government — and used by a number of researchers — is the Lifepaths model, developed by Statistics Canada over the past 25 years. As opposed to others mentioned here, this model is based on the concept of a dominant person in a household, and simulates this person's situation as a whole. The model covers dimensions such as retirement and institutionalization, but has very little on health status and nothing on health care use (only one measure based on self-reported health status is included, in addition to one measure of disability). A large number of articles and analyses have been produced using Lifepaths (Clavet et al., 2012).

To our knowledge, the only health-oriented microsimulation model existing in Canada is named the POpulation HEalth Model (POHEM). Since the 1990s, Statistics Canada has built many modules of POHEM in order to model different illnesses and risk factors. To date, however, the illnesses and risk factors are modelled separately and only permit an analysis "by illness". POHEM uses many sources of data, including several major surveys, the Census, medical and hospitalization records, etc.

We have developed a dynamic health microsimulation model using existing administrative and survey data sources. This model is named COMPAS; it was created in 2013 and its development is ongoing. The model has been used for several purposes already, such as in Boisclair et al. (2018) to estimate the economic benefits of reducing cardiovasular diseases in Canada. The following section presents an overview of COMPAS.

1.2 Dynamics of COMPAS

COMPAS has two important components: 1) a dynamic component that allows users to simulate individuals progressing through their life cycle, and thus also to incorporate the evolution of their health status, captured by various indicators such as the presence of illnesses, the presence of activity limitations, etc.; and 2) a cross-sectional component aiming to quantify the cost of medical resource use associated with the health status of persons alive in a given year. The second component benefits from the infrastructure of existing data, including – for Quebec – administrative data from the Régie de l'assurance maladie du Québec (RAMQ) as well as from the Ministry of Health and Social Services.

The model begins in 2010, even if we only output results from 2018 on. These four cycles of 2 years allow the model to stabilize, and allow to recalibrate mortality against the 2016 projections by Statistics Canada (2015). For the first year, an initial population is created using an existing survey.

This initialization phase gives a statistically representative sample of Canada – as well as Quebec, Ontario and other provinces grouped into regions – in 2010. Then begins the process of demographic evolution of the population up to the target year: 2050 by default. Depending on the desired output variables, the model can keep estimating its variables up to the year 2130 if any remaining agents are still alive. This could be useful, for example, to analyze the evolution of life expectancy of each cohort entering the model. In effect, cohorts entering the model between 2010 and 2050 will see a major proportion of agents who remain living in 2050 (the oldest will only be 70 years old). Figure 1.1 summarizes how the model works and the dynamics of COMPAS.

The microsimulation model includes several components. The structure of the model is almost entirely based on the population dynamics. The model tracks agents, characterized by socioeconomic and health attributes, which transition from one period to another. After the initial year, new agents enter the simulation at the beginning of each simulation cycle[2] at the default starting age(s), which we set at 30-31 years old, or at any age if they immigrate.[3] Their life in the model ends at their time of death or when they reach the maximum age permitted in the model, which is set at 120 years old.

Each period, the living population is characterized by a set of demographic, economic and health variables. The health status variables make it possible to attribute to individuals a level of health care use and related expenses in a given period. This population then passes through a transition phase where health status may change. For example, an individual aged 80 years with hypertension would have a set of probabilities of suffering from a stroke. These probabilities depend on his current health status. During this transition, each individual faces a risk of dying or of developing one or more disabilities which, for instance, increase their probability of entering a long-term care facility in subsequent cycles.

In each simulation cycle, we can observe not only individuals' health status, but also their use of health care. For the purpose of running this cross-sectional component, the population includes agents who are alive at the end of a given period. In each simulation cycle, the population moves forward in time. It loses some agents due to mortality, and gains some because there is renewal – i.e., a new cohort of agents aged 30-31 years old enters the model. Finally, the population gains and loses members due to migration. Each simulated agent has a given demographic weight; thus, statistics of interest can be computed at the population level.

Since the simulations include a stochastic element, the model enables the user to make several replications in order to ensure that the result obtained is not purely a matter of chance. In order to take into account the uncertainty from the transitions model, we have also estimated 100 sets of parameters using a bootstraping approach. For each set of parameters, one can make several replications in order to take into account both parameter distributions and stochastic uncertainty.

The next chapter introduces the two main sources of data for COMPAS, namely the National Population Health Survey (NPHS) and the Canadian Community Health Survey (CCHS). Chapter 3 explains in detail how the data is formatted for the first year of simulation. In chapter 4 we present transitions models, such as the mortality model and disease models. Chapter 5 explains how the renewal of the population is implemented. Chapter 6 sets out how trends in mortality and immigration are integrated

[2] In COMPAS, a simulation cycle corresponds by default to a period of two years.

[3] Age of entry is set at 30-31 years old in order to avoid having to model individual education trajectories and choices, and because illnesses and disabilities considered in the model are uncommon before this age.

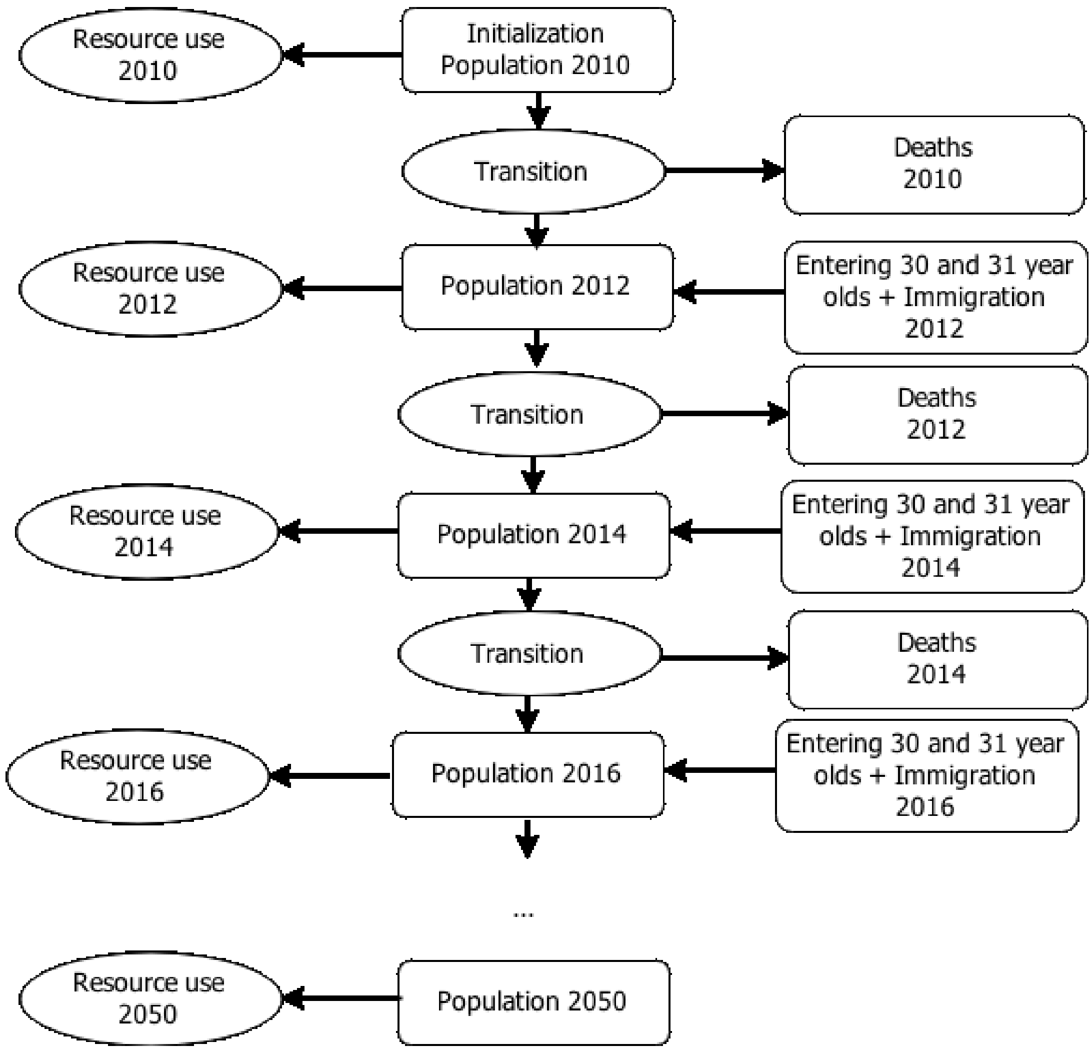

Figure 1.1: Dynamics of the COMPAS model with 2-year simulation cycles

in COMPAS. Chapters 7 and 8 present the cross-sectional part of COMPAS, namely health care use and health care costs. Finaly chapter 9 exposes how COMPAS deals with the inherent uncertainty of microsimulation models.

Chapter 2

Data

This chapter explains how the data used to build the microsimulation model is prepared. We use a number of different databases to construct the model. By far the most important survey is the National Population Health Survey (NPHS), which is the only longitudinal survey on health in Canada. We use it to build the transition module and the health care use module. The 2008-2009 and the 2010 Canadian Community Health Survey (CCHS) are used to construct the model's initial population. The latter two surveys are representative of the Canadian population around 2010. For the operation of the model, we minimally need these three surveys, the use of which is described in the following sections. For convenience and because some of them are otherwise difficult to locate, the questionnaires of these surveys are made available in our GitHub archive (https://github.com/CEDIA-models/compas2019).

In addition, we use the General Social Survey (GSS) to estimate health care use. We also use the Labour Force Survey (LFS) to estimate trends and to impute certain levels of education. We also use more data provided by Statistics Canada regarding population size by age to mimic its distribution in the model's initial population. Finally, we use some data provided by the Régie de l'assurance maladie du Québec (RAMQ) to calculate the costs associated with certain variables of health care use. More details regarding these complementary sources of data are provided in the next chapters.

2.1 NPHS Description

The NPHS is a longitudinal survey carried out on a sample of the Canadian population. The questionnaire was administered on the same sample of respondents every two years from 1994-1995 onwards. There are nine cycles in this survey, the last one being 2010-2011. While there are three components to the NPHS (the Households component, the Health Institutions component, and the North component), only the Households component is used in this model. The North component has not been implemented by the NPHS since 2000-2001, while the Health Institutions component was terminated in 2002-2003.

In the Households component, all respondents interviewed necessarily live in a private household in the first cycle of the survey. The survey thus excludes residents of Indian reserves, on Crown land and in remote regions of Quebec and Ontario, as well as full-time members of the Canadian Forces and persons living in institutions (Statistics Canada, 2012b). It should be noted that if a respondent from the Households component transfers to a long-term care institution between two cycles of the survey, she is still followed in the Households component.

In 1994, there were 12,797 respondents aged 30 and over in Canada, including 3,143 aged 65 years old or over. In comparison, there were 1,866 (3,838) respondents in Quebec (in Ontario), including 365 (984) aged 65 years or over. This comparison suggests that the number of observations in each region is small, especially for the population aged 65 and over. Given that certain health problems are relatively rare, it seems prudent to use the entire Canadian sample and to allow deviations for each region considered in the model. These regions, five in total, are the Atlantic Provinces (Nova Scotia, New Brunswick, Prince Edward Island and Newfoundland), Quebec, Ontario, the Prairies provinces (Manitoba, Saskatchewan and Alberta), and British Columbia. Table 2.1 presents the number of observations in the NPHS by age group for Quebec and the rest of Canada.

	Atlantic Provinces	Quebec	Ontario	Prairies	British Columbia	Canada
30 to 34 years old	396	283	594	410	251	1,934
35 to 39 years old	360	285	515	369	241	1,770
40 to 44 years old	314	252	435	288	210	1,499
45 to 49 years old	315	214	395	284	184	1,392
50 to 54 years old	243	180	318	218	150	1,109
55 to 59 years old	221	141	311	202	132	1,007
60 to 64 years old	202	147	286	192	116	943
65 to 69 years old	210	131	304	201	115	961
70 to 74 years old	204	116	307	199	83	909
75 to 79 years old	147	65	190	163	73	638
80 years old and older	154	52	183	165	81	635
Total (30 years and older)	2,766	1,866	3,838	2,691	1,636	12,797

Table 2.1: Number of observations for each Canadian region and for Canada in the NPHS (1994-2011)

2.1.1 Weighting

In order to proceed with estimations using NPHS data, one must use sampling weights. This allows the selected sample to be representative of the entire Canadian population. Each person in the sample represents a certain number of individuals in the population. An initial weight is calculated and then adjusted to account for certain survey particularities. This first weight is the inverse of the selection probability. Thus, the greater the chance of an individual of being selected into the sample, the lower its weight. This weight is calculated in the first cycle of the survey (1994-1995) and does not vary in time. Since there is no resampling in the NPHS, this weight enables us to represent the population of interest as it was in 1994, when initial observations were presented (Statistics Canada, 2012b).

Statistics Canada (2012b) then computes other weights which allow a subset of the sample to be representative of the Canadian population. The second weight only applies to individuals having responded to all nine cycles of the survey (or who died or were in an institution during some cycles). Indeed, some individuals did not respond to all survey cycles.

A longitudinal survey on health requires that respondents who could not be interviewed be thoroughly followed up. In particular, it is important to be able to clearly identify those who are deceased. It is also important to verify the frequency at which people leave NPHS and check if their characteristics are different from those who remain. We therefore address this last point before discussing mortality.

2.1.2 Attrition

Table 2.2 shows cumulative response rates, attrition rates (the share of those having responded in the preceding cycle who did not respond in the current cycle) and the share of those entering (respondents who left between two previous cycles but who returned in the current cycle). For those who entered, the software used by Statistics Canada to record responses ensures that there are no inconsistencies in the responses of individuals between the cycle preceding their absence from the survey and the cycle following it (Statistics Canada, 2012b).

	Response rate	Attrition rate	Share of entrants
1994-1995	100.00%	.	.
1996-1997	92.86%	7.14%	0.00%
1998-1999	87.55%	6.75%	1.44%
2000-2001	83.61%	7.24%	3.29%
2002-2003	79.62%	7.57%	3.59%
2004-2005	74.66%	8.53%	3.57%
2006-2007	73.13%	6.50%	4.97%
2008-2009	65.62%	10.05%	2.54%
2010-2011	64.79%	6.50%	5.66%

Table 2.2: Response and attrition rates as well as share of entrants by cycle in the NPHS

Attrition rates in the NPHS are generally low and are comparable to those found in other frequently used longitudinal surveys, such as the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID) in the U.S. The attrition rate in the HRS was 35.7% between 1992 and 2002 (Kapteyn et al., 2006). In 2002, 64.3% of the initial sample had responded to all cycles. As for the PSID, observed attrition rates were between 2.5% and 3% per year between 1968 and 1989 (Fitzgerald et al., 1998). In comparison, between 1994 and 2004, the attrition rate in the NPHS was only 22.4%. Finally, in the NPHS, a large share of respondents who leave the panel eventually return.

In Table 2.3, we analyze whether the health status of individuals remaining in the panel continuously from 1994-1995 to 2010-2011 is significantly different from that of non-deceased individuals who did not respond to one or more cycles. The table presents Student's t-statistics for the difference in averages, calculated relative to 1994 data. For example, a value of more than 1.96 or less than -1.96 for a variable means that we reject at the 5% confidence level the null hypothesis that the mean of this variable for individuals having responded throughout the entire surveying period is identical to that of individuals who will eventually stop responding. Otherwise stated, a value of more than 1.96 or less than -1.96 indicates a difference between the two groups, at a 5% threshold. Significant differences could cause problems, because in that case nonresponses would reduce the sample's representativeness.

The tests prove significant at the 5% threshold for some variables. For some age groups, the presence of diabetes (65 years and over) and heart diseases (40 to 64 years) is higher in 1994 for those who will later stop responding to the survey. A few other groups are also significantly larger among future non-responders: former smokers under 65 years old; individuals under 40 years old or 65 or more with at least one limitation of activities of daily living, or ADL; and individuals aged 65 or more with at least one limitation of instrumental activities of daily living, or IADL. Conversely, prevalence is higher among individuals who respond to all survey cycles for current smoking among those under 65; for the absence of disability among individuals 65 or older; and the presence of cognitives impairments among respondents aged 40 to 64.

	Under 40 years	40 to 64 years	65 years and over
Diabetes	-0.7853	-0.3678	-2.1647
Hypertension	0.0033	-1.9597	-0.7893
Cancer	0.1757	-1.5085	-1.5145
Heart diseases	1.2394	-2.2566	-1.6630
Stroke	0.5612	-0.3785	-0.3279
Lung diseases	1.3966	0.0352	-0.9286
Dementias	.	0.3679	-0.4029
Current smokers	4.2614	2.3738	-1.2198
Former smokers	-3.2797	-4.1568	-1.5910
No disability	-1.0161	-1.3617	5.1766
Cognitive impairments	1.4869	2.0129	-1.7816
One IADL or more	0.0348	0.3048	-5.2798
One ADL or more	-2.1155	1.1364	-2.8001

Table 2.3: Statistical test on selection, ENSP (1994-2011): Difference in proportions between individuals responding to all cycles and those for whom there are missing cycles. T-statistics are presented. Note: a negative value indicates a higher prevalence among individuals who eventually leave.

Having presented the bias potentially introduced by the selection, we need to asses whether this affects the results. To do so, we calculate the averages for the variables in Table 2.3 by weighting with weights that take the probability of nonresponse into account. In doing so, we use the same methodology as Michaud et al. (2011). When comparing averages for respondents in 2010 obtained with standard weights provided by Statistics Canada to those obtained using weights that control for nonresponse, we get very similar results. We conclude that the use of the weights provided by Statistics Canada in COMPAS is justified.

2.1.3 Mortality

The NPHS provides two ways of detecting mortality. The first is for an interviewer to obtain the information when attempting to contact the respondent. The second is by validation with the national registry of deaths. Such a validation is only possible when respondents had given permission for their survey data to be linked to administrative data. About 90% of respondents had given permission for data linkage.

Table 2.4 presents the annual mortality rates of the NPHS by cycle for the population aged 30 years and over. In earlier cycles, the rates increase over time. In the later cycles, the annuel mortality rate decreases. We believe this is mainly due to problems with the validation from the national death registry. Indeed, the validation with the registry is performed retrospectively for the previous cycle. As we have seen, there are nine cycles in the NPHS; but no deaths have been validated with the registry in the last cycle (2010-2011), which looked at deaths that occurred among individuals surveyed in the previous cycle.[1] We therefore exclude the eighth wave (2008-2009) for mortality estimation purposes, since death records are not complete for these individuals. Furthermore, no information on deaths is available after the last cycle of the survey (i.e. for the individuals surveyed in 2010-2011).

	Mortality rates
1994-1995	0.98%
1996-1997	1.13%
1998-1999	1.13%
2000-2001	1.15%
2002-2003	1.28%
2004-2005	1.25%
2006-2007	1.05%
2008-2009	0.85%
2010-2011	-
1994-2010	1.10%

Table 2.4: Annual mortality rate by NPHS cycle, population aged 30 and over

Figure 2.1 presents the mortality rates by age from the NPHS from 1994 to 2006, as well as the 95% confidence interval around these rates (dotted lines). We compare these rates with the periodic mortality rates from the Human Mortality Database for a similar period, i.e. from 1995 to 2004. All Canadian data comes from Statistics Canada.

The fit is satisfactory up to about 75 years of age. However, there is a divergence after 75 years. There is two main reasons for that divergence. First, that is the moment where the number of individuals in institutions strongly increases and the follow-up is minimal. Second, if someone stops responding at one cycle and dies after that, even if his death is known from the mortality linkage, his characteristics at the time of death are not known. In this situation we do not count this person in the mortality presented in figure 2.1 because he will not be included in the mortality model. However, we discuss in the next chapter how mortality rates are calibrated in the first year of the simulation to take these differences into account.

[1] This is shown in a chart in the NPHS's User Guide, only available in Statistics Canada's Research Data Centres (RDCs).

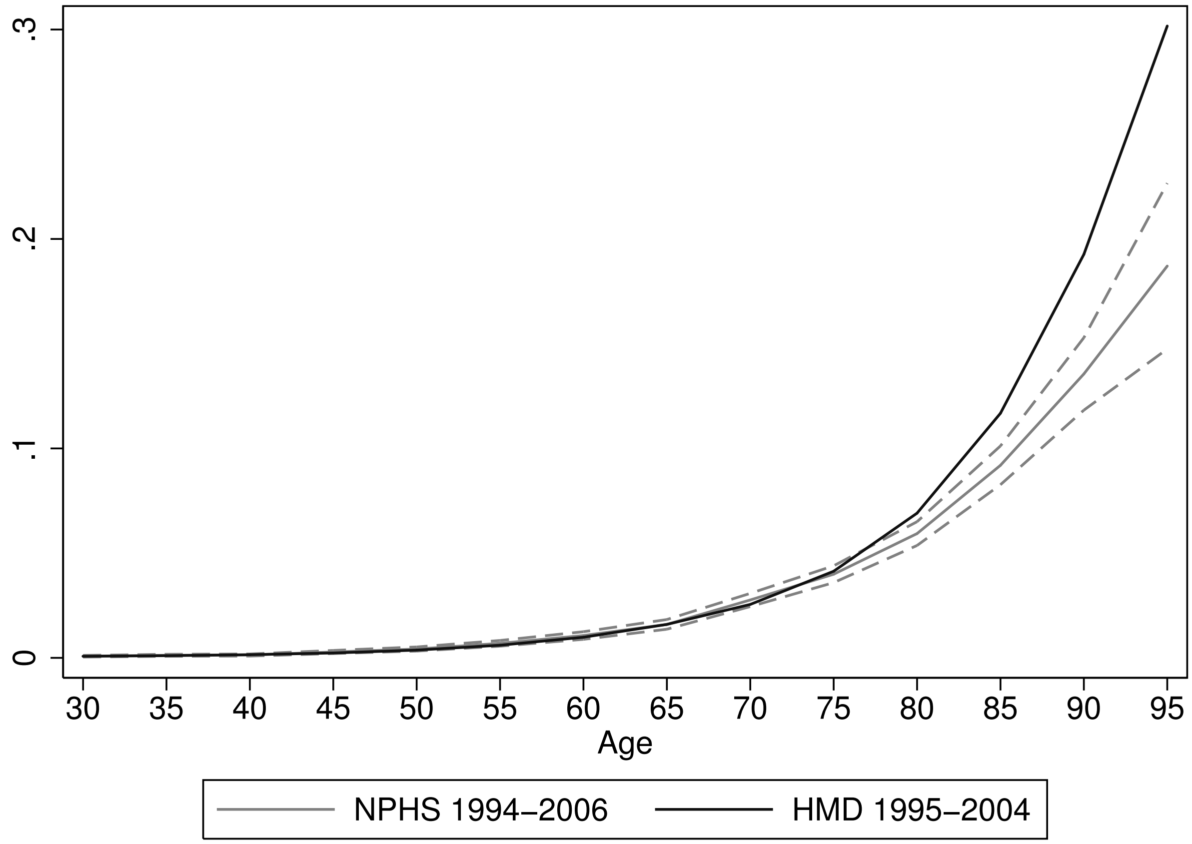

Figure 2.1: Comparison of mortality rates: NPHS vs. Human Mortality Database (HMD).
Note: dotted lines show the 95% confidence interval.

2.2 Constructing the variables

The NPHS questionnaire is rich, and the questions are for the most part identical from one cycle to the next. We divide this section into three parts. First, we present how variables related to self-reported health problems are constructed. These variables deal with measures of incidence (new cases of a disease arising between two points in time) and prevalence (all people suffering from a disease at a given moment). Then, we describe the measures of disability. We also present the measures of risk factors. Finally, we describe the health care use variables. A person's health status includes many dimensions in the model; the dimensions considered are presented in Table 2.5. For health care use variables, we use the dimensions shown in Table 2.6.

We use the sampling weights to produce the following statistics. In this version, we have not stratified by province nor by gender. However, in order to check the existence of differences between certain

Diabetes
Hypertension
Cancer
Stroke
Heart diseases
Lung diseases
Alzheimer's and other dementias
Obesity (BMI)
Tobacco use
Presence of disabilities

Table 2.5: Dimensions of health status accounted for in COMPAS

Number of visits to a generalist
Number of visits to a specialist
Number of nights of short-term hospitalization
Use of medicines (yes / no)
Use of home care services (yes / no)
Use of home care services (formal / informal / both)

Table 2.6: Dimensions of health care use accounted for in COMPAS

regions and the country as a whole, we present for all results Student's comparison tests, which indicate whether the proportion for Canada is significantly different from the proportion for each region.[2]

2.2.1 Self-reported health conditions

The model covers seven self-reported health conditions: presence of diabetes, of hypertension, of cancer, of heart diseases, of strokes, of lung diseases, and of dementias (Alzheimer's and others). These diseases were selected primarily for their high prevalence in the population and their presence in others micro-simulation models like the *Future Elderly Model* (Goldman et al., 2005), allowing for comparisons. The aforementioned model includes the same diseases with the exception of dementias, which are absent.

The questions are asked as follows in the NPHS: *"We're interested in conditions diagnosed by a health professional [...]. Do you have diabetes?, ... hypertension?, etc."* (Statistics Canada, 2012c). We consider the presence of diabetes, hypertension, lung disease or dementia to be an absorbing state. We therefore recode these variables such that after a first positive response, individuals are considered to have the condition in all the following cycles. For the three others health conditions (cancer, heart diseases and stroke), we model remissions, and as such we keep the data as is. Once aggregated, the answers of all respondents lead directly to measures of prevalence. To develop the measures of incidence, we use positive responses in the current cycle when the response was negative in the preceding cycle – accounting for the absorbing states defined as mentioned above.

[2]Throughout the tables in this chapter, a negative (-) t-statistic indicates a lower proportion in the designated region compared to Canada as a whole, while a positive t-statistic indicates a higher proportion in the region. The significance of these differences depends on the size of the statistics; e.g., an absolute value of more than 1.96 indicates a significant difference at a 5% threshold. The direction of the difference depends on the sign of the t-statistic, as explained.

Interpreting the prevalence by age is difficult for two important reasons: a) cohort effects (intergenerational differences for a given age) in disease incidence; and b) increased mortality among individuals suffering from a disease. Thus, we could very well observe that incidence increases with age while prevalence decreases. We present both prevalence and incidence for each disease.

Table 2.7 gives the prevalence rates by 10-year age group for each disease. Unsurprisingly, we find that prevalence increases with age, except for certain diseases (diabetes, cancer) for which prevalence decreases after 90 years. Higher mortality among individuals with those conditions may explain this pattern. The most common disease among individuals aged 40 years and older is hypertension, by a fairly wide margin – though heart disease prevalence increases substantially at older ages.

Age	Diabetes	Hypertension	Cancer	Heart diseases	Stroke	Lung diseases	Dementias
30 to 39 years	1.4%	5.6%	0.4%	0.9%	0.2%	6.1%	0.3%
40 to 49 years	3.6%	11.0%	0.8%	1.9%	0.3%	6.5%	0.3%
50 to 59 years	7.4%	26.3%	2.1%	5.0%	0.9%	7.2%	0.4%
60 to 69 years	12.7%	42.4%	3.6%	10.5%	2.2%	9.7%	1.1%
70 to 79 years	17.9%	52.0%	4.5%	17.6%	4.41%	13.1%	3.3%
80 to 89 years	19.6%	58.0%	6.4%	24.9%	7.5%	16.2%	9.9%
90 years and older	18.0%	57.3%	7.0%	22.8%	10.6%	17.2%	17.7%
Total (30 yrs +)	7.6%	24.9%	2.1%	6.5%	1.5%	8.3%	1.3%
t-statistics of difference between rates in a given region and those in Canada							
Atlantic Provinces	-5.17	-9.45	-1.79	-0.79	1.10	-2.89	0.92
Quebec	1.33	2.14	4.61	2.80	0.81	0.31	0.15
Ontario	-1.59	-2.30	-2.27	-5.05	-1.44	-2.48	2.59
Prairies	2.67	4.08	-0.71	7.04	0.04	5.45	-4.51
British Columbia	3.22	3.80	-0.55	3.26	3.32	2.05	1.41

Table 2.7: Self-reported disease prevalence in the NPHS (1994-2011)

In Table 2.8, we present the incidence rates by age group and for each disease. Among those aged 40-49 years, incidence is highest for hypertension, with 2.3%. Indeed, hypertension exhibits the highest incidence up until 70-79 years. We also note that incidence generally increases with age, although it declines for some diseases at advanced ages. A number of reasons could explain that, including selection and the biological processes leading to these diseases.

In Table 2.9, we present the remission rates by age group for cancer, heart disease and stroke, the three conditions for which remissions are allowed. The percentage of remissions for cancer decreases sharply with age, from 70.6% for individuals between 30 and 39 years old to 37.6% for people 90 years and over. We do not observe the same pattern for heart diseases, where remissions vary between 25.6% and 30.9% – except for the 30-39 years old and the 90 years and over, who experience a higher remission rate. Remission from suffering from the effect of a stroke varies between 16.7% and 24.8%, with the exception of the oldest age group in which we observe the highest rate of remissions at 33.3%.

Age	Diabetes	Hypertension	Cancer	Heart diseases	Stroke	Lung diseases	Dementias
30 to 39 years	0.35%	1.15%	0.31%	0.75%	0.03%	0.59%	0.02%
40 to 49 years	0.64%	2.30%	0.65%	1.22%	0.20%	0.54%	0.06%
50 to 59 years	1.36%	4.97%	1.42%	2.55%	0.40%	0.83%	0.10%
60 to 69 years	1.76%	5.69%	2.32%	4.87%	1.01%	1.28%	0.34%
70 to 79 years	2.16%	5.58%	3.01%	8.31%	2.12%	1.54%	1.53%
80 to 89 years	1.77%	4.89%	4.08%	12.15%	3.36%	2.10%	5.19%
90 years and older	1.20%	5.28%	4.55%	13.26%	7.91%	4.80%	8.67%
Total (30 yrs +)	1.05%	3.32%	1.35%	2.93%	0.75%	0.88%	0.53%
t-statistics of difference between rates in a given region and those in Canada							
Atlantic Provinces	-2.47	-0.76	-0.93	-0.76	-0.20	-1.07	1.96
Quebec	0.07	-0.87	0.59	-0.64	0.51	-1.59	-0.08
Ontario	-0.31	-0.18	-0.27	-1.00	-0.20	0.73	0.32
Prairies	1.55	0.74	-0.46	2.24	-0.11	0.04	-1.57
British Columbia	0.91	1.86	1.29	1.97	0.44	2.00	1.06

Table 2.8: Self-reported disease incidence in the NPHS (1994-2011)

	Cancer	Heart diseases	Stroke
30 to 39 years	70.62%	45.07%	19.68%
40 to 49 years	57.87%	30.51%	20.36%
50 to 59 years	56.38%	25.61%	16.73%
60 to 69 year	54.12%	30.94%	19.34%
70 to 79 years	53.82%	26.75%	23.16%
80 to 89 years	48.61%	29.97%	24.77%
90 years and older	37.59%	49.21%	33.29%
Total (30 yrs +)	54.86%	29.56%	21.77%

Table 2.9: Self-reported disease remission in NPHS (1994-2011)

2.2.2 Disability

We wanted to build variables that could take into account cognitive and physical disabilities. Thus, we use the variables from the questionnaire to build a measure of cognitive impairment and two measures of physical disability. In this section, we present each of these measures.

Cognitive impairment

We use a self-reported variable indicating whether individuals have frequently had memory problems that negatively affect their activities.

Presence of instrumental activities of daily living (IADL) limitations

We have created a variable that indicates the presence of at least one limitation in instrumental activities of daily living (IADL) declared by the respondent. The questions are asked as follows in the NPHS: *"Because of any physical condition or mental condition or health problem, do you need the help of another person: ... in preparing meals?... with getting to appointments and running errands such as shopping for groceries?"* The activities considered are preparing meals; doing everyday housework; and shopping.

Presence of activities of daily living (ADL) limitations

We have created a variable that captures the presence of at least one limitation in activities of daily living (ADL). The questions asked regarding ADL in the NPHS are phrased like those for IADL. The activities included are moving about inside the house as well as three categories of personal care: washing, dressing, and eating.

Disability variables constructed for the model

Table 2.10 shows the distribution of these three variables by age group alongside the absence of disabilities. Among individuals aged 40-49 years, few respondents have disabilities: 6.5%. This rate increases to 11.1% for the 60-69 years age group. Among individuals aged 90 and over, more than three-quarters suffer from at least one disability.

	No disability	Cognitive impairment	At least 1 IADL	At least 1 ADL
30 to 39 years	94.7%	1.5%	2.9%	0.7%
40 to 49 years	93.5%	1.3%	4.2%	1.3%
50 to 59 years	91.3%	1.4%	6.5%	1.8%
60 to 69 years	88.9%	1.6%	8.9%	2.6%
70 to 79 years	77.8%	3.3%	18.8%	6.8%
80 to 89 years	55.2%	9.8%	39.1%	19.4%
90 years and older	23.8%	18.8%	67.5%	44.9%
Total (30 yrs +)	88.6%	2.2%	3.2%	8.8%
t-statistics of difference between rates in a given region and those in Canada				
Atlantic Provinces	-1.65	-1.49	1.65	2.10
Quebec	3.33	0.51	1.74	-2.01
Ontario	-0.91	1.17	-0.49	0.40
Prairies	-1.37	-1.16	-1.55	0.83
British Columbia	0.45	-0.30	0.17	-0.65

Table 2.10: Prevalence of disability according to the NPHS (1994-2011)

2.2.3 Risk factors

COMPAS considers two risk factors for the time being, i.e. obesity and smoking. Regarding the latter, we have defined three smoking statuses: 1) never smoked; 2) current smoker; and 3) former smoker. Table 2.11 shows the distribution of respondents by age according to their tobacco use. We see that the rate of tobacco use reaches its peak at 27.0% between 30 and 39 years old, and then declines with age. The share of former smokers rises up to the age of 80 and declines thereafter, which probably reflects cohort effects in tobacco-related habits, and possibly the higher mortality rate among current and former smokers.

In Table 2.12, we show, by age group, the initiation and quit rates, or transitions between smoking statuses. Transition rates are generally higher among younger individuals.

Age	Current smokers	Former smokers
30 to 39 years	27.0%	26.4%
40 to 49 years	24.9%	32.9%
50 to 59 years	21.2%	39.3%
60 to 69 years	16.0%	44.6%
70 to 79 years	11.6%	46.6%
80 to 89 years	6.5%	41.2%
90 years and older	3.8%	34.0%
Total (30 yrs +)	20.9%	36.3%
t-statistics of difference between rates in a given region and those in Canada		
Atlantic Provinces	-3.57	-4.59
Quebec	-6.11	-2.99
Ontario	3.99	5.53
Prairies	-0.83	0.23
British Columbia	7.90	-3.56

Table 2.11: Tobacco use in the NPHS (1994-2011)

As for obesity, we have constructed a variable based on the body mass index (BMI), with three categories (under 30, 30-35, and 35+). According to World Health Organization (WHO) terminology, a person with a BMI of 30 or more is considered obese. Since we need a finer categorization, we distinguish between classes of obesity. As in the WHO classification, class I obesity in our model includes individuals with a BMI between 30 and 35, while the "classes II-III" category brings together all respondents with a BMI of 35 or over (World Health Organization, 2006). Table 2.13 gives the distribution of respondents by age group. We see that class I obesity is at its highest in the 60s, while classes II-III is highest in the 50s.

Table 2.14 presents the individual transitions between levels of obesity for two groups: those aged under 65 years and those 65 years and older. The rows indicate the level of obesity in the current cycle, while columns complete the crosstab with respect to the following cycle. On average, a little more than 5% (3%) of non-obese individuals under 65 years (65 and older) become obese between each 2-year cycle. On the other hand, 19.8% (23.7%) of individuals with class I obesity become non-obese in the next cycle, while 2.6% (4.2%) progress to classes II-III obesity.

Age	Initiation	Cessation	Reuptake
30 to 39 years	1.00%	3.77%	2.02%
40 to 49 years	0.85%	4.08%	2.32%
50 to 59 years	0.77%	4.30%	2.10%
60 to 69 years	0.99%	3.51%	1.60%
70 to 79 years	1.09%	3.64%	1.77%
80 to 89 years	0.78%	2.98%	1.41%
90 years and older	1.59%	3.62%	1.35%
Total (30 yrs +)	0.91%	3.86%	1.99%
t-statistics of difference between rates in a given region and those in Canada			
Atlantic Provinces	1.32	-2.20	-2.29
Quebec	0.00	-0.09	-2.31
Ontario	-0.25	1.96	2.67
Prairies	-0.90	-1.61	-1.12
British Columbia	-0.29	0.36	1.52

Table 2.12: Smoking initiation and cessation rates over 2 years in the NPHS (1994-2011)

2.2.4 Health care use

The NPHS is rich in data on health care use (Statistics Canada, 2012c). Six variables were created for modelling purposes. These were chosen first because the information was available in the NPHS. Secondly, they represent an important part of health expenditures. The first two variables are the number of visits or telephone consultations over the previous 12 months, on the one hand with a family doctor, paediatrician or general practitioner; and on the other hand with another doctor or specialist (e.g. a surgeon, an allergist, an orthopedist, a gynecologist or a psychiatrist). These two variables were censored at the 99$^{\text{th}}$ percentile because the last observations presented a high number of consultations, which might skew the means.

Since doctor consultations do not include stays in a care facility, a third variable was created: the number of nights spent in a hospital, a non-residential nursing home or a convalescent home. As was the case for medical consultations variables and for the same reason, this variable was censored at the 99$^{\text{th}}$ percentile. This third category does not include same-day surgeries and emergency room visits, since these are excluded from the NPHS. They are therefore excluded from COMPAS for the time being.

The fourth variable is a medicines use indicator, consisting of a binary variable indicating whether the individual took at least one medication (prescription or over-the-counter) in the previous month.

The fifth variable takes three possible values, indicating whether the individual 1) is not receiving long-term care; 2) is receiving home care; or 3) resides in a long-term care facility. We consider that individuals have received home care in the previous 12 months if they received nursing care, other health services (e.g. physiotherapy or nutritional counselling), personal hygiene care (e.g. help with bathing) or help with housework, meal preparation, meal delivery, or running errands. Individuals are considered institutionalized if, in the NPHS, they reside in one of the following types of facilities:

Age	Absence of obesity BMI < 30	Class I obesity $30 \leq$ BMI < 35	Classes II and III obesity BMI ≥ 35
30 to 39 years	84.0%	12.0%	4.0%
40 to 49 years	81.4%	13.7%	4.9%
50 to 59 years	77.9%	15.8%	6.2%
60 to 69 years	77.9%	16.8%	5.3%
70 to 79 years	82.6%	14.3%	3.1%
80 to 89 years	90.4%	8.1%	1.4%
90 years and older	92.3%	6.6%	1.0%
Total (30 yrs +)	81.4%	14.0%	4.7%
	t-statistics of difference between rates in a given region and those in Canada		
Atlantic Provinces	10.29	-7.17	-6.86
Quebec	-7.78	5.24	6.23
Ontario	1.82	-0.83	-2.01
Prairies	7.54	-5.60	-4.66
British Columbia	-6.59	4.29	5.58

Table 2.13: Body mass index (BMI) distribution in the NPHS (1994-2011)

	Under 65 years of age		
	Obesity in following cycle		
Obesity in current cycle	Absence of obesity	Class I obesity	Classes II and III obesity
Absence of obesity (BMI < 30)	94.67%	5.16%	0.17%
Class I obesity ($30 \leq$ BMI < 35)	19.76%	70.46%	9.78%
Classes II and III obesity (BMI ≥ 35)	2.63%	18.94%	78.43%
	65 years and older		
	Obesity in following cycle		
Obesity in current cycle	Absence of obesity	Class I obesity	Classes II and III obesity
Absence of obesity (BMI < 30)	96.35%	3.49%	0.16%
Class I obesity ($30 \leq$ BMI < 35)	23.70%	69.96%	6.34%
Classes II and III obesity (BMI ≥ 35)	4.19%	28.69%	67.12%

Table 2.14: Transitions by 2-year period between levels of obesity using the NPHS (1994-2011)

1. institutions for the aged, including residential care facilities for the aged and extended/chronic care hospitals;

2. cognitive institutions, including residential care facilities for emotionally disturbed children, psychiatrically disabled and developmentally delayed people, and psychiatric hospitals;

3. other rehabilitative institutions, including rehabilitation, paediatric and other speciality hospitals, general hospitals with long-term units as well as residential care facilities for people with physical disabilities.

The sixth variable also takes three possibles values, indicating whether the home care received was 1) formal (provided by a professional); 2) informal (provided by a family member, friend or neighbour); or 3) both formal and informal. To do so, two questions help determine if these are services *with the cost being entirely or partially covered by government* or *with the cost not covered by government (for example: care provided by a private agency or by a spouse or friends)* (Statistics Canada, 2012c). In the first case, we identify the person as having received formal care. In the second case, a further question determines whether the person has received help from a nurse from a private agency, a homemaker from a private agency, a family member, a friend or a neighbour. For the first two sources of care we identify the individual as having received formal care. For the other sources we identify the person as having received informal care. Formal and informal home care are not mutually exclusive: a person can receive both.

Table 2.15 gives the distribution by age for the first five variables, the *Home care services (yes)* and *Institutionalization* columns being two of the three categories of long-term care use. The number of consultations with a general practitioner increases fairly constantly with age, especially beyond 50 years (one may suppose that, between the ages of 30 and 39 years, pregnancy involves a higher number of visits). Over the course of the 12-month reference period, individuals consult a general practitioner three times on average. Consultations with specialists in the previous 12 months, however, drop to under one. The number of nights of hospitalization increases faster after 70 years of age. Individuals aged 80 to 89 years spend on average more than three nights in a short-term facility over the previous 12 months. Around 85% of the entire population responded that they had taken at least one medication during the reference period. Above 90 years of age, about 37% of the population receives home care services. Finally, at this age, nearly one-third of individuals reside in a long-term care institution.

Table 2.16 shows the type of home care received among those who received home care (i.e. the columns add to 100% and are a breakdown of the "Home care services (yes)" column of the the previous table). The proportion of individuals receiving formal home care (i.e. formal and both, combined) varies between 58.2% for the 40 to 49 years old and 78.4% for the 90 years and older. The proportion of individuals receiving informal home care varies between 39.4% for the 80 to 89 years old and 53% for the 60 to 69 years old. The proportion of individuals receiving formal home care mostly increases with age.

Table 2.17 indicates that the average number of consultations with a generalist is 1.8 to 3.5 times higher among individuals suffering from one of the health conditions considered. The difference is statistically significant in every case. Individuals who suffer from diabetes, cancer, heart diseases, stroke or dementia consult with a specialist more than once a year on average. This difference is statistically significant. Those who suffer from any form of dementia, a heart disease, or who have had a stroke spend more than five times as many nights in hospital as those who do not. These differences are also statistically significant.

Among people having diabetes, hypertension, a heart disease or a stroke, the prevalence of medication use is highest at over 97%. It seems that medication use does not increase by much with age. However, due to the definition of this variable, a respondent who is in overall good health but takes a single medication in the entire year cannot be distinguished from a respondent suffering from many diseases and taking several medications. Thus, 80% of individuals aged 30 to 39 years had taken at least one medication in the reference period. The more than 10-percentage-point increase in the proportion of medication use between individuals aged 30-39 years and those aged over 90 is fairly important given the definition of the variable.

Over 30% of people having had a stroke or suffering from a form of dementia receive home care services, a proportion about five times higher than that of individuals who never suffered from these — once again, the difference is statistically significant. We also notice that about 32% of individuals suffering from a form of dementia are institutionalized.

Age	Nb consultations w/ generalist	Nb consultations w/ specialist	Nb hospitalization nights
30 to 39 years	2.69	0.70	0.46
40 to 49 years	2.57	0.66	0.45
50 to 59 years	3.01	0.78	0.70
60 to 69 years	3.44	0.83	0.96
70 to 79 years	4.26	0.88	1.90
80 to 89 years	4.94	0.80	3.09
90 years and older	5.12	0.56	3.93
Total (30 yrs +)	3.12	0.75	0.86
t-statistics of difference between rates in a given region and those in Canada			
Atlantic Provinces	-10.39	3.48	-3.50
Quebec	21.85	-1.98	-2.80
Ontario	-5.59	-3.77	4.40
Prairies	-1.51	6.14	-1.20
British Columbia	-8.05	2.43	4.25
Age	1+ medication	Home care services (yes)	Institutionalization
30 to 39 years	80.3%	3.6%	0.0%
40 to 49 years	82.2%	2.6%	0.2%
50 to 59 years	84.3%	3.1%	0.2%
60 to 69 years	88.5%	5.7%	0.3%
70 to 79 years	91.1%	14.5%	2.9%
80 to 89 years	92.8%	26.6%	12.7%
90 years and older	91.7%	36.6%	32.9%
Total (30 yrs +)	84.6%	6.5%	1.4%
t-statistics of difference between rates in a given region and those in Canada			
Atlantic Provinces	-5.14	-0.54	1.21
Quebec	3.84	-2.18	0.51
Ontario	-3.01	1.70	0.43
Prairies	-2.79	0.93	-2.76
British Columbia	3.14	1.72	2.06

Table 2.15: Health care use by age according to the NPHS (1994-2011)

	Formal Home care	Informal Home care	Form. & Infor. Home care
30 to 39 years	58.9%	30.1%	11.0%
40 to 49 years	49.2%	41.8%	8.9%
50 to 59 years	57.4%	34.8%	7.8%
60 to 69 years	47.0%	36.5%	16.5%
70 to 79 years	57.0%	28.6%	14.4%
80 to 89 years	60.6%	23.5%	15.9%
90 years and older	56.6%	21.6%	21.8%
Total (30 yrs +)	55.9%	30.1%	14.0%
t-statistics of difference between rates in a given region and those in Canada			
Atlantic Provinces	4.56	-5.03	0.53
Quebec	-5.10	4.75	1.42
Ontario	1.80	-1.33	-0.75
Prairies	1.36	-0.81	-0.82
British Columbia	0.79	-0.95	0.15

Table 2.16: Type of home care received according to the NPHS (1994-2011)

2.2.5 Other socio-economic variables

We have constructed socio-economic variables that will be useful in the analysis.

- Immigrant: dummy variable equal to 1 if the individual is an immigrant and 0 otherwise.
- Woman: dummy variable equal to 1 if the individual is a woman and 0 otherwise.
- Education: three dummy variables that combine to indicate the highest level attained among high school, college and university.

We have chosen to model only these variables because we can reasonably assume that they are fairly stable after the age of 30. If we were to include other variables, such as marital status or income, we would have to model individual behaviours over time. Future versions of the model might include such variables.

Table 2.18 presents descriptive statistics drawn from the NPHS regarding these variables. We note that the share of immigrants increases slightly with age, as does the proportion of women — in particular over the age of 60 for women. Due to important cohort effects, the proportion of university graduates decreases with age.

2.3 CCHS : description and construction of variables

We also use the Public Use Microdata Files (PUMFs), of the CCHS (Statistics Canada, 2008, 2010). Two waves of this survey are used to create the initial population in 2010, aged 30 to 110 years. We cannot use the NPHS to this end because its population is only representative of the Canadian population in 1994. The statistics in this section are from the CCHS 2008-2009 and 2010 combined.

Diseases	Nb consultations w/ a generalist	Nb consultations w/ a specialist	Nb hospitalization nights
Diabetes			
No	2.96	0.73	0.75
Yes	5.45	1.06	2.53
Hypertension			
No	2.68	0.68	0.61
Yes	4.68	0.98	1.77
Cancer			
No	3.07	0.71	0.80
Yes	5.55	2.72	3.59
Heart diseases			
No	2.94	0.70	0.67
Yes	5.81	1.45	3.64
Stroke			
No	3.07	0.74	0.79
Yes	6.53	1.09	5.99
Lung diseases			
No	3.00	0.72	0.79
Yes	4.81	1.10	1.83
Dementias			
No	3.09	0.74	0.83
Yes	5.69	1.17	4.66
Diseases	1+ medication	Home care services (yes)	Institutionalization
Diabetes			
No	83.6%	5.8%	0.7%
Yes	97.9%	15.7%	2.4%
Hypertension			
No	81.1%	4.5%	0.4%
Yes	96.7%	12.2%	1.8%
Cancer			
No	84.4%	6.2%	0.8%
Yes	94.0%	25.5%	2.9%
Heart diseases			
No	83.6%	5.5%	0.6%
Yes	98.9%	21.1%	3.9%
Stroke			
No	84.4%	6.2%	0.7%
Yes	98.8%	33.3%	10.7%
Lung diseases			
No	83.9%	5.9%	0.7%
Yes	93.8%	14.9%	1.7%
Dementias			
No	84.5%	6.2%	0.4%
Yes	93.5%	37.3%	32.3%

Table 2.17: Use of health care services by presence of disease, NPHS (1994-2011). Note: all these differences are statistically significant at the 5% threshold.

Age	Immigrant	Woman	High school	College	University
30 to 39 years	16.2%	50.2%	43.5%	24.5%	20.8%
40 to 49 years	19.8%	50.0%	42.8%	23.8%	20.0%
50 to 59 years	22.8%	49.1%	40.1%	19.9%	19.6%
60 to 69 years	24.7%	51.7%	36.1%	15.4%	15.0%
70 to 79 years	25.4%	57.8%	31.5%	12.7%	9.7%
80 to 89 years	26.7%	63.9%	29.8%	9.1%	8.4%
90 years and older	28.7%	69.8%	31.5%	7.5%	7.7%
Total (30 yrs +)	21.2%	51.7%	39.7%	20.1%	17.8%
t-statistics of difference between rates in a given region and those in Canada					
Atlantic Provinces	57.62	-1.70	6.68	-0.92	10.75
Quebec	24.17	-0.91	9.71	2.74	0.67
Ontario	-27.03	0.70	-1.58	0.11	-7.01
Prairies	16.57	0.26	-4.75	0.59	4.33
British Columbia	-11.43	0.85	-8.32	-4.05	-3.19

Table 2.18: Socio-economic characteristics drawn from the NPHS (1994-2011)

The CCHS 2008-2009 and 2010 share many variables with the NPHS, which we have been able to use directly. We thus have access to all the variables regarding the presence of the diseases considered in the model, with the exception of Alzheimer's and other dementias. We will expand on the problems caused by the absence of those questions in the next chapter. Table 2.19 presents disease prevalence in the CCHS (2008-2009 and 2010 combined). With the exceptions of hypertension and cancer, the questions asked are limited to the presence of the health condition at the time of the interview. The question used for hypertension is retrospective, which means that we can build lifetime prevalence for this health condition. The difference between prevalence in the NPHS and in the CCHS can therefore be explained by two elements. First, as stated in the previous section, in the NPHS we recode the variables to treat three health conditions as absorbing states, something we cannot do with the CCHS for diabetes and lung diseases. This explains the higher prevalence found in NPHS data for those two health conditions. Second, we use lifetime prevalence variables in the CCHS for hypertension, which explains its higher measured level when using CCHS data. Prevalence for cancer, heart diseases and stroke are similar in both surveys.

As we observed in the NPHS, the prevalence of most diseases increases with age. The prevalence of diabetes increases up to 79 years and then decreases slightly afterwards. The most common disease is hypertension, with a prevalence of 44.2% among persons aged 60 to 69 years.

The risk factors (smoking and obesity) are also present in both the CCHS 2008-2009 and 2010. As shown in Table 2.20, the proportion of current smokers is 24% among individuals aged 30 to 49. It is within this age group that the proportion of smokers is highest. On the contrary, the highest proportion of former smokers is among older individuals, peakingat 55% for the 70 to 79 years old.

The Table 2.21 shows the prevalence of obesity in the CCHS 2008-2009 and 2010. The prevalence of obesity classes II-III increases with age among those aged 30 to 69 years and then gradually decreases. The prevalence of class I obesity is particularly high among those aged 50 and 69, at about 16%.

We can also calculate the number of IADL and ADL limitations, although they are not entirely consistent with those in the NPHS, which means that we cannot have disability measures directly comparable

Age	Diabetes	Hypertension	Cancer	Heart diseases	Stroke	Lung diseases
30 to 39 years	1.9%	4.3%	0.3%	0.7%	0.15%	0.8%
40 to 49 years	3.9%	11.9%	0.9%	1.9%	0.47%	2.3%
50 to 59 years	8.6%	23.7%	1.9%	5.7%	0.78%	4.3%
60 to 69 years	14.2%	38.4%	3.8%	11.8%	2.19%	6.3%
70 to 79 years	18.7%	50.9%	6.5%	20.6%	3.61%	8.4%
80 years and older	15.4%	53.2%	6.5%	29.6%	5.36%	9.2%
Total (30 yrs +)	8.5%	28.4%	8.3%	7.5%	1.5%	4.1%
t-statistics of difference between rates in a given region and those in Canada						
Atlantic Provinces	-6.23	-9.73	-4.15	-5.98	-2.72	-4.45
Quebec	2.10	3.84	2.70	-2.66	1.88	-1.11
Ontario	-2.70	-1.93	1.09	0.05	0.46	1.91
Prairies	1.77	0.41	-0.43	2.30	-2.12	0.66
British Columbia	4.66	4.48	-2.61	5.38	0.69	0.27

Table 2.19: Self-reported disease prevalence in the CCHS (2008-2009 and 2010 combined)

Age	Current smokers	Former smokers
30 to 39 years	24.5%	32.5%
40 to 49 years	24.4%	37.8%
50 to 59 years	23.6%	46.1%
60 to 69 years	16.4%	53.0%
70 to 79 years	10.2%	55.0%
80 years and older	5.4%	52.2%
Total (30 yrs +)	20.5%	43.2%
t-statistics of difference between rates in a given region and those in Canada		
Atlantic Provinces	-4.59	-7.09
Quebec	-7.42	-7.30
Ontario	4.43	9.51
Prairies	-3.43	-0.22
British Columbia	9.22	-2.77

Table 2.20: Tobacco use in the CCHS (2008-2009 and 2010 combined)

to those of the NPHS. Moreover, the individuals in institution are not included in the CCHS 2008-2009 and 2010. We present the methodology to overcome these problems in the next chapter.

Age	Absence of obesity BMI $<$ 30	Class I obesity 30 $\geq$ BMI $<$ 35	Classes II and III obesity BMI $\geq$ 35
30 to 39 years	82.6%	12.8%	4.6%
40 to 49 years	79.6%	14.3%	6.1%
50 to 59 years	77.4%	16.0%	6.5%
60 to 69 years	76.9%	16.3%	6.8%
70 to 79 years	81.4%	14.3%	4.3%
80 years and older	87.9%	10.0%	2.1%
Total (30 yrs +)	79.8%	14.6%	5.7%
t-statistics of difference between rates in a given region and those in Canada			
Atlantic Provinces	13.96	-9.55	-9.60
Quebec	-6.93	3.67	6.50
Ontario	1.82	-0.91	-1.77
Prairies	8.09	-5.41	-5.76
British Columbia	-12.42	9.21	7.56

Table 2.21: Body mass index (BMI) in the CCHS (2008-2009 and 2010 combined)

Finally, we know the level of education of individuals, but unlike in the NPHS, no distinction is made between college and university degrees. Table 2.22 details some socio-economic characteristics of individuals within the CCHS 2008-2009 and 2010. It shows that the proportion of individuals with a college or university degree decrease age.

Chapter 3 explains in more details how we solve problems related to CCHS 2008-2009 and 2010 variables, and how we use these surveys as a starting point for the simulation.

Age	Immigrant	Woman	Secondary	College and university
30 to 39 years	26.2%	48.6%	17.4%	76.6%
40 to 49 years	27.3%	50.4%	23.3%	68.4%
50 to 59 years	23.1%	49.3%	26.3%	61.0%
60 to 69 years	25.0%	51.0%	23.3%	54.7%
70 to 79 years	26.4%	53.6%	21.5%	40.1%
80 years and older	26.2%	61.7%	20.3%	31.4%
Total (30 yrs +)	25.8%	51.4%	22.6%	61.0%
	t-statistics of difference between rates in a given region and those in Canada			
Atlantic Provinces	36.86	-0.84	6.22	3.10
Quebec	31.03	0.38	- 11.90	1.58
Ontario	-32.93	-1.45	-4.90	-0.71
Prairies	14.59	2.30	-3.74	0.37
British Columbia	-17.37	0.29	-7.31	-3.41

Table 2.22: Socio-economic characteristics in the CCHS (2008-2009 and 2010 combined)

Chapter 3

Initialization

Initialization involves creating a database for the initial simulation year (e.g. 2010). We cannot use the NPHS as the initial database because it was designed to be representative of the Canadian population at the time of sampling (1994). Therefore, it is not representative of the population's state of health in 2010.

We therefore use a combination of the CCHS 2008-2009 and 2010 as the database to construct the initial population. These surveys have the advantage of giving an up-to-date picture (in 2010) of the state of health of the Canadian population. We use two surveys rather than one to have a greater number of observations. The disadvantage of the CCHS (2008-2009 and 2010) compared to the NPHS is that, as opposed to the latter, the former are not longitudinal in nature. This feature, however, is not required to create an initial database.

Using a combination of surveys introduces a difficulty: the initial database must contain all the variables needed for the transitions and health care use models. For most of the variables that we use, this is not a problem since the NPHS and CCHS variables are identical. However, in a few cases, we need to impute values for certain variables that are missing from the CCHS. Moreover, we rely on the public-use version of the surveys, which means that the data is limited for certain variables.[1] This chapter describes the assumptions which allow us to create the initial database — and thus population — using the CCHS 2008-2009 and 2010. For the remainder of this section, unless otherwise stated, "CCHS" refers to the combination of CCHS 2008-2009 and 2010.

3.1 Imputation and calibration

This section describes imputation procedures in detail. Imputation involves a random process, so we create 100 initial datasets as described below to account for the randomness of the processes. We then use a different initial database for each replication of the model. This procedure is linked to the global treatment of uncertainty in the model, in section 9.1.

[1] We rely on public-use CCHS data to construct the initial database to ensure that we can initialize the model at will outside the walls of a Statistics Canada Research Data Centre (RDC). Using restricted-access data would mean that the user would be forced to run the model within a RDC.

3.1.1 Imputation of age

The public-use CCHS (the PUMFs) classifies individuals into five-year age groups until 79 years of age and provides an open age category for individuals aged 80 years and over.

To overcome this problem, we attribute a precise age to each individual within his/her age group. In order to recover the distribution of observations at each age we estimate a multinomial logit model (presented in chapter 4,section 4.1.5) for each five-year age group until 79 years old. These models include dummy variables for sex, education, region, presence of diseases, smoking status, obesity status, and disability. All the models for the imputation of age are estimated on the restricted-access CCHS version, available in Statistics Canada's Research Data Centres (RDCs).

For the 80 years and older, we have to deal with more single years of age and fewer observations. First, we use a multinomial logit model to impute each individual into smaller categories (80-82,83-85,86-88,89+). Second, we use multinomial logit models to impute single years of age for the 80-82, 83-85 and 86-88 years old categories. For the 89+, we first use a logit model (presented in section 3.1.4) to split between 89-90 and 91+ years old. Then we use a second logit model to split between 89 and 90 years old. Finally, we use the 2011 Census to calculate a probability for each precise age conditional on being 91 or older, and we draw a random value from a uniform distribution to assign a specific age. The oldest age imputed is 100 years old.

This burdensome procedure was necessary to recover the distribution of characteristics by age, even at old ages where the limited possibilities for estimating a multinomial logit model with too many categories led us to use chain imputation. The estimated models — codes and parameter estimates — are available upon request.

3.1.2 Imputation and calibration of education level

The education levels available in the public-use CCHS (the PUMFs) are "less than secondary school graduation", "secondary school graduation, no post-secondary education", and "post-secondary education'". "Post-secondary" levels include "apprenticeship or other trades certificate" as well as college and university degrees. In order to have similar levels to those available in the NPHS, we have to impute "apprenticeship or other trades certificate", to group them with "secondary school graduation, no post-secondary education"; and to impute university degrees among individuals in the "post-secondary education" category. While it would have been possible to use the NPHS to impute the education level, we were dissuaded by the smaller number of respondents in the survey.

To carry out the imputation, we use the restricted-access CCHS version in the same way as for the imputation of age. We use a multinomial logit model to predict in which of the three categories an individual with "post-secondary" levels of education falls ("apprenticeship or other trades certificate", university degree, or other post-secondary education). The table 3.1 shows the results of the imputation model :

	Apprenticeship or other trades certificate	College degree	University degree
Aged between 35 and 39	0.0144	−0.0094	−0.0050
Aged between 40 and 44	−0.0092	0.0047	0.0045
Aged between 45 and 49	0.0636**	0.0428	−0.1064***
Aged between 50 and 54	0.0617**	0.0051	−0.0668**
Aged between 55 and 59	0.0654**	0.0155	−0.0809**
Aged between 60 and 64	0.0556**	−0.0310	−0.0246
Aged between 65 and 69	0.0941***	−0.0655**	−0.0286
Aged between 70 and 74	0.1256***	−0.0689**	−0.0567*
Aged between 75 and 80	0.1388***	−0.0632*	−0.0756*
Aged 80 and older	0.1871***	−0.0319	−0.1552***
Woman	−0.1224***	0.1409***	−0.0186
Reside in Quebec	0.0190	−0.0561**	0.0371*
Reside in Ontario	−0.1222***	0.0556***	0.0666***
Reside in Prairies	0.0010	−0.0017	0.0007
Reside in British Columbia	−0.0291	−0.0204	0.0494*
Presence of diabetes	0.0400*	0.0282	−0.0682**
Presence of hypertension	0.0170	0.0328*	−0.0499**
Presence of cancer	−0.0197	0.0111	0.0086
Presence of heart disease	0.0120	−0.0383	0.0263
Presence of Stroke	0.0332	0.0186	−0.0518
Presence of lung diseases	0.0609**	0.0676	−0.1285***
Current smoker	0.1679***	0.0738***	−0.2417***
Former smoker	0.0708***	0.0159	−0.0867***
Class I obesity	0.0314*	0.0691***	−0.1005***
Classes II-III obesity	0.0778***	0.0730**	−0.1508***
Presence of IADL limitation	−0.0059	0.0245	−0.0186
Presence of ADL limitation	0.0152	−0.0093	−0.0059
Presence of cognitive impairment	0.0275	−0.0013	−0.0262
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01			

Table 3.1: Average marginal effects of variables on the probability of being in one of 3 levels of education — CCHS 2010-2011

3.1.3 Imputation of disability

The CCHS allows us to count the number of activities of daily living (ADL) limitations and instrumental activities of daily living (IADL) among the following:

- (ADL) needing help preparing meals;
- (ADL) needing help for personal care (washing, dressing, eating and taking medication);
- (IADL) needing help moving about inside the house;
- (IADL) needing help doing everyday housework.
- (IADL) needing help running errands;

Compared to the ADL definition we use from the NPHS to compute the transitions (see following chapter), the only difference is that "needing help for personal care" includes help in "taking medication" in the CCHS, but not in the NPHS. Nonetheless, the three other components (washing, dressing, eating) are identical in both surveys. We cannot separate the question about the "needing help taking medication" from the three other components of personal care since the question is not addressed elsewhere in the NPHS nor, to our knowledge, by any other survey. We therefore use the question as is.

The difference in how the question is phrased in both surveys should have a marginal impact since the only individuals that would be assigned a positive value to the variable "one ADL or more" in the CCHS but not in the NPHS are those — rare — cases in which the person requires assistance to take medication, but not to bathe, to get dressed, to eat or to move about inside the house.

The other questions used to obtain the other disability variables (presence of at least one cognitive impairment or at least one IADL limitation) are identical in both surveys.

3.1.4 Imputation of institutionalization status

The CCHS does not include individuals in institutions. We must therefore impute the institutionalization status. We estimated a logit model which provides the effect of individual characteristics on the probability of being institutionalized. The structure of the model is as follows:

$$y_i^* = \boldsymbol{\beta}\mathbf{x}_i + \epsilon_i \tag{3.1}$$

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* \geq 0 \\ 0 & \text{if} \quad y_i^* < 0 \end{cases}, \tag{3.2}$$

where y_i is equal to 1 if individual i is institutionalized and 0 otherwise; y_i^* is the latent variable; $\mathbf{x}_i$ is the vector of explanatory variables for individual i; and ϵ_i is a random term that follows a logistic distribution.

This is estimated with the 2010-2011 NPHS data. The NPHS households component did not include people in institutions in its first wave in 1994. However, people who transfer into institutions continue to be followed in the survey. We are confident that, 16 years after the beginning of the survey, the prevalence should be fairly representative considering the rather rapid renewal of the population of individuals living in institutions. We include a spline at 75 years in order to allow the age effect to change after 75 years; a gender effect; an immigrant status effect; and three binary variables for the three types of disability cognitive impairment; ADL; IADL). Table 3.2 shows the results.

	Marginal effect
Age (if under 75 years)	0.0006^{***}
Age (if 75 years or over)	0.0006^{***}
Immigrant	-0.0058^{**}
Women	0.0003
Presence of at least one ADL limitation	0.0229^{***}
Presence of at least one IADL limitation	−0.0034
Presence of cognitive impairment	0.0059^{***}
Number of observations	7,757
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01	

Table 3.2: Average marginal effects of variables on the probability of being institutionalized — NPHS 2010-2011

We see that the age effects are statistically significant. The effect for women is not significant, but that for immigrants is negative. The dummy variables for the presence of one ADL limitation and for that of cognitive impairment have large coefficients, because individuals who suffer from important physical or cognitive limitations are more likely to be in an institution.

The model coefficients are then used to impute the status of institutionalization in the CCHS. For each observation i, the probability of being institutionalized is estimated with the following formula:

$$P(institution_i = 1|\mathbf{x_i}) = \frac{\exp\left(\mathbf{x_i'}\hat{\boldsymbol{\beta}}\right)}{1 + \exp\left(\mathbf{x_i'}\hat{\boldsymbol{\beta}}\right)} \tag{3.3}$$

We then draw a random value from a uniform distribution (0,1) for each observation and impute the status of institutionalization if it is smaller than the probability calculated above.

3.1.5 Imputation of help received for home care

The help received for home care is present in the public-use data of the CCHS 2008-2009, in a survey on aging, but not in the CCHS 2010. The information is nonetheless available in the restricted-access files for CCHS 2010, so we impute using them. The model is identical to the one used for institutionalization. The estimation results are presented in Table 3.3.

We find that the probability of receiving home care increases with age, mostly after 75 years of age. The most important effect is the presence of at least one ADL limitation, which increases the probability of needing LTC by 4%. Other disability variables (presence of cognitive impairment and presence of at least one IADL limitation) also have significant positive effects.

	Marginal effect
Age (if under 75 years)	0.0009***
Age (if 75 years or more)	0.0038***
Immigrant	0.0020
Woman	0.0053
Presence of at least one ADL	0.0418***
Presence of at least one IADL	0.1140***
Presence of cognitive impairment	0.0255**
Number of observations	15,765
Legend	* p <0.10 ; ** p <0.05 ; *** p <0.001

Table 3.3: Average marginal effects of variables on the probability of receiving home care — CCHS 2010

3.1.6 Imputation of Alzheimer's and other dementias

Alzheimer's and other dementias are the only health condition in COMPAS that cannot be observed in the public-use CCHS. We therefore impute the variable using a logistic regression estimated with 2010-2011 NPHS data. The model is similar to that used for institutionalization, but it includes a binary variable for institutionalization status as an explanatory variable. The estimation results are presented in Table 3.4.

We find that the probability of suffering from dementia increases with age and stops increasing significantly after 75 years. Gender and immigration status do not have a significant effect. As we might expect, the disability variables are strongly associated with the presence of dementia.

	Marginal Effect
Age (if under 75 years)	0.0009***
Age (if 75 years or over)	0.0003
Immigrant	−0.0035
Woman	−0.0041
Presence of at least one ADL	0.0118***
Presence of at least one IADL	0.0148***
Presence of cognitive impairment	0.0362***
Living in an institution	0.0197***
Number of observations	7,749
Legend	* p <0.10 ; ** p <0.05 ; *** p <0.001

Table 3.4: Average marginal effects of variables on the probability of having Alzheimer's or another dementia — NPHS 2010-2011

The model coefficients are then used to impute the presence of dementia in the CCHS. To do so, we use the same method used for the previous variables.

3.1.7 Calibration of the population

A final adjustment is required to have an initial population that is representative of the 2010 population. Indeed, the size of the institutionalized population obtained through imputation differs from that provided in the 2011 Census (Statistics Canada, 2011). The main reason is that the population surveyed in the 2010 CCHS excluded individuals in institutions, who are on average more ill than the population as a whole.

When we impute the institutionalization status, we apply it to individuals who are in fact not living in institutions. The predicted average probability of being institutionalized in the CCHS population is lower than the probability of being institutionalized within the total population. This leads to the institutionalized population being under-represented.

In order to align the number of individuals in institutions in the initial population with the number of institutionalized individuals counted in the 2011 Census[2], we calibrate the CCHS weight with aggregated Census data. We calibrate by region (Atlantic Provinces, Quebec, Ontario, Prairies, British Columbia), by gender and by age group ([30-75[, [75-80[, [80-85[, [85+). Finally, the weights are adjusted so that the total population by region, gender and age is consistent with the Canadian population aged 30 and over in 2010 (Statistics Canada, nda).

3.2 Characteristics of the initial population

After the imputations performed on 2008-2009 and 2010 CCHS data, the initial population of the model contains all the variables needed for the transitions and health care use models. Table 3.5 presents certain statistics regarding the variables of COMPAS.

As the Table shows, the average age in the initial population is about 53 years (as a reminder, individuals under 30 years are excluded) and the proportion of men and women is similar. After the imputation, we find that 25% of individuals hold a university degree, which is similar to the proportion of individuals with a university degree in the NPHS. The prevalence of diseases is below 10%, excluding hypertension which affects more than 28% of individuals. Only 1.4% of the initial population suffers from dementia.

As for risk factors, a little more than 20% of the initial population smokes and 20% of individuals suffer from obesity. The proportion of individuals who have a cognitive impairment, at least one ADL limitation or at least an IADL limitation is 9.6%. The proportion of individuals receiving home care is 7.8% while about 1% of individuals are institutionalized.

[2] The category of accommodation we use is that of health facilities and related institution to which we withdraw the retirement homes.

	Average
Age (30 to 110 years old)	53.2
Immigrant status	25.1%
Women	51.3%
Less than high school	18.7%
High school graduate	37.3%
College degree	18.7%
University degree	25.2%
Presence of diabetes	8.6%
Presence of hypertension	28.8%
Presence of heart disease	7.7%
Presence of stroke	1.4%
Presence of cancer	2.4%
Presence of lung disease	4.2%
Presence of dementia	1.2%
Never smoked	35.6%
Current smoker	21.1%
Former smoker	43.4%
No obesity	79.6%
Class I obesity	14.6%
Classes II and III obesity	5.7%
Disability: no disability	90.2%
Disability: presence of cognitive impairment	1.6%
Disability: presence of at least one IADL limitation	8.5%
Disability: presence of at least one ADL limitation	2.9%
Long-term care: none	90.8%
Long-term care: home care	7.8%
Long-term care: living in an institution	1.4%

Table 3.5: Description of the initial population (100 replications mean)

Chapter 4

Transitions

Using the transitions models, we can calculate the probability that the state of health or behaviour of individuals changes as a function of their individual characteristics. Such models require estimates of the parameters of discrete choice econometric models, with the transitions that we want to model serving as dependent variables.

In COMPAS, these parameters are estimated using data from the NPHS because its longitudinal nature makes it possible to observe transitions from one state to the other. The estimated parameters are then used to calculate the transition probabilities of the simulated individuals as a function of their individual characteristics and thereafter to simulate the future evolution of their health conditions.

In section 4.1, we present the econometric models used to estimate transitions in COMPAS. They help estimate the effects of individual characteristics on transitions occurring between cycles of the NPHS. By default, these are two-year cycles, so the individual probabilities of changing state of health are calculated for two-year intervals.

4.1 Econometric models

4.1.1 Models for diseases

We use a transitions model to estimate incidence probabilities for each disease under consideration. The probabilities are a function of age, of risk factors, of socio-economic characteristics and, in certain cases, of the presence of other diseases. Restrictions imposed on the latter include the fact that certain diseases do not impact the incidence of other diseases, which prevents the creation of unwarranted links between them. The restrictions were evaluated by a panel of experts in the context of the Future Elderly Model (FEM) and are based on medical research on the relationship between various diseases (Goldman et al., 2005).

Table 4.1 shows the various effects permitted in the model. For example, an " $\times$ " at the intersection of the row "Diabetes" and the column "Hypertension" means that we allow the presence of diabetes to have an effect on the probability of getting hypertension. Moreover, diabetes, hypertension, lung diseases and dementias are considered "absorbing" states. This means that once an individual suffers from one of these diseases, he or she has it until the end of his or her life (see explanations on prevalence provided in chapter 2). For the three others conditions (cancer, heart diseases and stroke), we model remission transitions.

	Diabetes	Hypertension	Cancer	Lung diseases	Heart diseases	Stroke	Dementias
Diabetes	n.a.	x			x	x	
Hypertension		n.a.			x	x	
Cancer			n.a.			x	
Lung diseases				n.a.			
Heart diseases					n.a.	x	
Stroke						n.a.	
Dementias							n.a.

Table 4.1: Permitted effects of diseases (rows) on other diseases' incidence (columns)

The specification is as follows:

$$inc^*_{i,j,t+1} = \boldsymbol{\lambda}_j \mathbf{y}_{i,t} + \boldsymbol{\beta}_j \mathbf{x}_{i,t} + \epsilon_{i,j,t} \tag{4.1}$$

with

$$inc_{i,j,t+1} = \begin{cases} 1 & \text{if} \quad inc^*_{i,j,t+1} > 0 \\ 0 & \text{if} \quad inc^*_{i,j,t+1} \leq 0 \end{cases} \tag{4.2}$$

where:

- $inc^*_{i,j,t+1}$ is a latent variable for the incidence of disease j for the individual i in period $t+1$;
- $inc_{i,j,t+1}$ is a binary variable indicating the prence of disease j for the individual i in period $t+1$;
- $\boldsymbol{\lambda}_j$ is a vector that includes the effects of the various diseases on the probability of incidence of disease j, accounting for the permitted links described in table 4.1;
- $\mathbf{y}_{i,t}$ is a vector of binary variables indicating the presence or absence of each disease ($j = 1, ..., 7$) in the individual i in period t;
- $\mathbf{x}_{i,t}$ includes all explanatory variables accounted for;
- $\boldsymbol{\beta}_j$ includes the effects of these variables on disease j;
- $\epsilon_{i,j,t}$ is a random term specific to the individual i and disease j in period t.

We estimate the parameters only on the population that, in period t, does not yet have disease j. We assume that the distribution function of $\epsilon_{i,j,t}$ follows an extreme value distribution so we can use a complementary log-log model (Sueyoshi, 1995). This model differs from the discrete choice logit and probit models, which are more commonly used, in that it relaxes the assumption that $\epsilon_{i,j,t}$ is symmetric around 0. It is therefore a better specification for cases in which the occurrence of the dependent variable is rare. The estimated parameters are used to compute the individual transition probabilities in the simulation using the following formula:

$$\hat{P}(inc_{i,j,t+1} = 1 | \mathbf{y_i}, \mathbf{t}, \mathbf{x_i}) = 1 - \exp(-\exp(\boldsymbol{\lambda}_j \mathbf{y}_{i,t} + \boldsymbol{\beta}_j \mathbf{x}_{i,t})). \tag{4.3}$$

	Diabetes	Hypertension	Cancer	Heart diseases
Age (if 60 years or under)	0.070***	0.071***	0.083***	0.063***
Age (if over 60 years)	0.008*	0.001	0.026***	0.048***
Current smoker	0.093	0.191***	0.182*	0.405***
Former smoker	0.119	0.095*	0.082	0.272***
Class I obesity	1.103***	0.544***	−0.039	0.208**
Classes II-III obesity	1.973***	0.904***	0.331***	0.349***
Woman	−0.292***	−0.061	0.045	−0.325***
Immigrant	0.027	0.142**	−0.144	−0.274***
High school graduate	0.030	−0.122**	0.124	−0.118*
College degree	−0.143	−0.014	0.349**	−0.087
University degree	−0.440***	−0.211**	−0.022	−0.220**
Resides in Quebec	−0.101	−0.033	−0.266***	0.128
Resides in Ontario	−0.060	−0.039	0.058	0.291***
Resides in the Prairies	−0.351***	−0.125**	−0.039	−0.067
Resides in British Columbia	−0.166	−0.174**	−0.146	−0.045
Presence of diabetes	—	0.543***	—	0.383***
Presence of hypertension	—	—	—	0.637***
Presence of heart disease	—	—	—	—
Presence of cancer	—	—	—	—
Year 2002 and after	—	—	—	−0.203***
Constant	−8.134***	−6.642***	−8.910***	−7.302***
Number of observations	65,000	53,700	67,900	64,600
Average incidence	0.0105	0.0332	0.0087	0.0160

	Stroke	Lung diseases	Dementias
Age (if 85 years or under)	0.060***	—	0.126***
Age (if over 85 years)	0.067**	—	0.036
Age (if 45 years or under)	—	−0.013	—
Age (if over 45 years)	—	0.038***	—
Current smoker	0.577***	0.942***	0.184
Former smoker	0.087	0.421***	0.184*
Class I obesity	−0.170	0.111	−0.473**
Classes II-III obesity	−1.033***	0.596***	−0.500
Woman	−0.105	0.301***	−0.146
Immigrant	0.246*	−0.320***	0.226*
High school graduate	−0.132	−0.266***	−0.197
College degree	−0.739**	−0.550***	−0.914**
University degree	−1.126***	−0.465**	−0.629***
Resides in Quebec	−0.156	0.125	0.353*
Resides in Ontario	0.105	0.072	0.330*
Resides in the Prairies	0.132	−0.038	0.577***
Resides in British Columbia	−0.082	−0.122	0.176
Presence of diabetes	0.764***	—	—
Presence of hypertension	0.295**	—	—
Presence of heart disease	0.688***	—	—
Presence of cancer	−0.033	—	—
Year 2002 and after	−0.455***	−0.541***	—
Constant	−8.624***	−4.562***	−13.600***
Number of observations	52,100	64,600	28,800
Average incidence	0.0054	0.0088	0.0053

Legend: * p <0.10 ; ** p <0.05 ; *** p <0.01

Table 4.2: Coefficients of variables on the probabilities of disease incidence over two years

Table 4.2 presents the coefficients for each disease. The first two effects are age effects, constructed like the in the following example:

- up to 60 years: $min(age, 60)$;
- over 60 years: $max([age - 60], 0)$.

The above example applies to diabetes, hypertension, cancer and heart diseases, for which a spline at 60 years old is imposed — meaning that the effect of age in the model is allowed to change after 60 years. We use splines at different ages for the different conditions, either at 45, 60 or 85 years old. The ages were chosen for each disease by testing which one had the highest explanatory power.

Smoking positively and significantly affects the probability of incidence of certain diseases, such as hypertension, heart disease, stroke and lung disease. Being a former smoker also positively impacts the probabilities of incidence, but less so than currently smoking. Smoking does significantly affect the likelihood of developing cancer, but only at at the 90% confidence level, which may seem counter-intuitive as 13.5% of all diagnosed cancers are lung cancers (Canadian Cancer Society's Advisory Committee on Cancer Statistics, 2015). One possible explanation is the very short life expectancy of lung cancer patients: only 32% survive one year after their diagnosis (Cancer Research UK, nd). Individuals are surveyed only at two-year intervals, so many might be diagnosed and die between two survey cycles.

Obesity has a generally positive effect on the probabilities of disease incidence. Women seem less affected by diabetes and heart disease and more affected by lung disease. As for immigrants, they have less of a chance of being affected by heart and lung diseases and more of a chance of being affected by hypertension, stroke and Alzheimer's and other dementias.

A higher educational attainment is generally associated with lower diseasw incidence, with the exception of cancer. The estimated effects of education may capture two effects that might counter one another. Better educated individuals are, on the one hand, generally healthier but, on the other hand, more likely to consult with a health professional, which might increase their disease incidence.

The dummy variables for province of residence are significant in some cases. However, they are statistically significant when tested all together (F-test).

The effects of the presence of other diseases on disease incidence are generally positive and significant where we allow for non-null effects. The exceptions are that the presence of cancer does not seem to significantly affect the probability of stroke incidence.

A dummy variable was included indicating years 2002 and after. Upon statistical inspection, year 2002 was found to be the most likely one to allow for series breaks.

The same variables as in incidence models are used in remission models. Age has a negative effect on the probability of remission before 60 years old for both heart diseases and cancer, and a positive effect after that for heart diseases only. Smoking significantly increases the probability of remission from cancer, which may seem counter-intuitive; but it must be reminded that this is the remission rate among survivors for at least 2 years following diagnosis, and that many patients likely die without having been re-surveyed. However smoking has a negative effect on the probability of remission from stroke. Women have better chances of remission. People with a university degree have a better chance of cancer remission but a lower chance of remission from stroke. Residing in any region West of the Atlantic provinces generally increases the probability of remission from cancer and heart diseases. However, residing in Ontario reduces the probability of remission from stroke. The presence of hypertension reduces the probability of remission from heart diseases.

	Cancer	Heart diseases	Stroke
Age (if 50 years or under)	—	—	−0.086
Age (if over 50 years)	—	—	0.011
Age (if 60 years or under)	−0.019*	−0.025***	—
Age (if over 60 years)	0.007	0.010*	—
Current smoker	0.523***	0.280**	−0.424*
Former smoker	0.221	0.071	0.078
Class I obesity	0.335*	−0.211*	0.377
Classes II-III obesity	−0.286	−0.008	−0.150
Woman	0.379**	0.281***	0.845***
Immigrant	−0.170	0.267**	−0.036
High school graduate	0.121	0.126	−0.045
College degree	−0.114	0.240	−0.029
University degree	0.640***	−0.006	−1.252***
Resides in Quebec	0.786***	0.521***	−0.315
Resides in Ontario	0.446**	0.314**	−0.434*
Resides in the Prairies	0.295	0.283**	−0.456
Resides in British Columbia	0.793***	0.281*	−0.334
Presence of diabetes	—	0.051	−0.011
Presence of hypertension	—	−0.286***	0.259
Presence of heart disease	—	—	−0.192
Presence of cancer	—	—	−0.039
Constant	0.269	−0.010	2.524
Number of observations	67,900	64,600	52,100
Legend: * p <0.10 ; ** p <0.05 ; *** p <0.01			

Table 4.3: Coefficients of variables on the probabilities of disease remission

4.1.2 Model for mortality

The mortality model is similar to that for disease incidence, but we include additional explanatory variables: disability statuses and whether the person is living in a long-term care facility. Moreover, we do not impose any restrictions on the effects of the various diseases' presence.

Table 4.4 presents the coefficients of the variables on the probability of dying in the two subsequent years. Unsurprisingly, the probability of dying increases significantly with age, both before and after 50 years of age. It may seem surprising, however, that the increase slows down after 50 years, but this means that death at older ages is better predicted by other variables in the model that are themselves correlated with age. Indeed, the presence of diseases positively affects the probability of death, and this relationship is statistically significant in many cases (diabetes, cancer, heart disease and stroke).

Being a current or a former smoker also significantly affects the probability of death. The link between obesity and death is negative: this can be explained by the weight loss that often occurs in the last stages of life.

Age (if 50 years or under)	0.104***
Age (if over 50 years)	0.058***
Presence of diabetes	0.315***
Presence of hypertension	0.061
Presence of cancer	1.060***
Presence of heart disease	0.364***
Presence of stroke	0.220**
Presence of lung disease	0.078
Presence of dementia	−0.013
Current smoker	0.831***
Former smoker	0.370***
Class I obesity	−0.274***
Classes II-III obesity	−0.146
Woman	−0.446***
Immigrant	−0.046
High school graduate	−0.254***
College degree	−0.577***
University degree	−0.468***
Resides in Quebec	0.005
Resides in Ontario	0.065
Resides in the Prairies	−0.117
Resides in British Columbia	0.105
Presence of cognitive impairment only	0.573**
Presence of IADL limitations only	0.792***
Presence of ADL limitations only	0.425**
Presence of cognitive impairment and IADL limitations	0.973***
Presence of ADL and IADL limitations	1.159***
Presence of cognitive impairment and of ADL and IADL limitations	1.566***
Living in a long-term care facility	0.421***
Constant	−9.993***
Number of observations	83,261
Average mortality	1.98%
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01	

Table 4.4: Coefficients of variables on two-year probability of death

Being a woman decreases the probability of death, which is consistent with the higher life expectancy of this gender. Higher educational attainment also leads to a significant decrease in the probability of death.

As discussed in chapter 2, mortality rates observed in the NPHS are slightly different from official mortality rates, especially at higher ages. As a result, the number of deaths per year in our model could be biased. We calibrate the estimated probabilities to fit the mortality rates provided by Statistics Canada (Statistics Canada, 2018).

We calibrate mortality at the first year of simulation. To do so, we first estimate the average probability of death by age and gender from the 100 starting populations. We subsequently compute an adjustment factor by age and sex, equal to the rate estimated using the moving average divided by the mortality rates derived from Statistics Canada. Finally, the relevant adjustment factor is applied to the calculated probability of death in each simulation year.

Furthermore we dynamically re-calibrate during simulation, at year 2016, to take in account the stabilization that occurs during the first few cycles of the simulation. This calibration is mostly the same that takes place in the first years, except that we use mortality rates projected by Statistics Canada (Statistics Canada, 2015) and that each run (or replication) of the simulation is calibrated individually. Because we calibrate each run individually, such that the number of observations by age is relatively small, we use a moving average over three years of age.

4.1.3 Models for smoking

We estimate three transitions models for smoking:

1. Smoking initiation (estimated for individuals who have never smoked);
2. Smoking cessation (estimated for current smokers);
3. return to smoking (estimated for former smokers).

Once again, we use a complementary log-log model for all three transitions. Even if the initial population is 30 years and older, we model smoking initiation as we observe some such transitions in the NPHS data. The presence of diseases is, however, excluded from the explanatory variables since smoking is considered a risk factor. Indeed, smoking increases the probability that individuals develop certain diseases (hypertension, lung and heart diseases, stroke) and then die as a consequence. Table 4.5 presents the coefficients of the three models.

Up to 50 years, age has a negative or null effect on all transition probabilities. Above 50 years, age increases the probability of quitting smoking and decreases that of restarting. We observe a negative effect of obesity on initiation and reuptake and a positive effect on cessation. Being a woman reduces the probability of initiation and cessation, and being an immigrant increases the chances of quitting. Finally, higher educational attainments are associated with a higher chance of quitting smoking and a lower probability of initiating or restarting.

	Initiation	Cessation	Reuptake
Age (if 50 years or under)	-0.014^{**}	-0.003	-0.043^{***}
Age (if over 50 years)	-0.003	0.030^{***}	-0.027^{***}
Class I obesity	-0.212^{*}	0.106^{*}	-0.121^{*}
Classes II-III obesity	0.080	0.059	-0.077
Woman	-0.802^{***}	-0.103^{**}	0.026
Immigrant	0.064	0.275^{***}	-0.059
High school graduate	-0.355^{***}	0.177^{***}	-0.201^{***}
College degree	-0.786^{***}	0.476^{***}	-0.620^{***}
University degree	-0.968^{***}	0.680^{***}	-0.483^{***}
Resides in Quebec	0.236^{**}	-0.076	-0.160^{*}
Resides in Ontario	0.009	-0.177^{***}	-0.147^{*}
Resides in the Prairies	0.132	-0.067	-0.014
Resides in British Columbia	0.031	-0.008	-0.192^{*}
Constant	-2.386^{***}	-1.697^{***}	-0.463^{**}
Number of observations	33,899	18,428	31,471
Average incidence	0.0091	0.0386	0.0199
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01			

Table 4.5: Coefficients of variables on two-year probability of transition between smoking statuses

4.1.4 Model for obesity

We model the transitions between the various possible states of obesity (BMI < 30, $30 \leq$ BMI < 35, and BMI ≥ 35) using a multinomial logit model. The model is presented in section 4.1.5. The model does not include the presence of disease and states of disability as explanatory factors, since obesity is considered a risk factor that affects these variables as opposed to a consequence of them. Table 4.6 presents the average marginal effects.

We find that each additional year after the age of 50 reduces by 0.07% the probability of beginning to suffer from class I obesity and that of suffering from classes II-III obesity. Being a former smoker has a positive and significant effect on the incidence of classes II-III obesity. This last relationship could be observed because of weight gain after quitting smoking (Dare et al., 2015).

	Obesity in the next period*	
	Class I obesity $30 \leq$ BMI < 35	Class II-III obesity BMI ≥ 35
Age (if 50 years or under)	0.008^{**}	0.016^{***}
Age (if over 50 years)	-0.021^{***}	-0.051^{***}
Resides in British Columbia	-0.109^{*}	-0.192^{*}
College degree	-0.213^{***}	-0.245^{**}
Constant	-2.952^{***}	-6.726^{***}
Former Smoker	0.123^{***}	0.317^{***}
High school graduate	-0.169^{***}	-0.191^{***}
Immigrant	-0.100^{**}	-0.210^{**}
Class I obesity	4.198^{***}	5.485^{***}
Resides in Ontario	-0.041	0.017
Resides in the Prairies	0.053	0.076
Resides in Quebec	-0.226^{***}	-0.340^{***}
Woman	-0.093^{**}	0.280^{***}
smoker	0.037	0.064
University degree	-0.329^{***}	-0.460^{***}
Classes II-III obesity	4.939^{***}	9.567^{***}
Number of observations	69,538	
Average incidence	0.1453	0.0489
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01		

*No obesity (BMI < 30) is the refererence category.

Table 4.6: Coefficients of variables on two-year probability of transition between obesity classes

The effects of the obesity dummy variables are of course very significant, since individuals tend to remain in the same state from one period to the next. Women have a slightly lower likelihood than men of suffering from class I obesity, but have a greater risk of suffering from classes II-III obesity. Higher educational attainments generally have a negative effect on the probability of becoming obese.

4.1.5 Model for disability

We model the transition probabilities between all the states of disability in the model. The three states in which we are interested are the presence of cognitive impairment, the presence of at least one limitation in activities of daily living (ADL) and the presence of at least one limitation in instrumental activities of life daily (IADL). We then create all possible permutations from these three states. Therefore, each individual can be in one of eight combinations of states ($2 \times 2 \times 2$, since each state involves two possible values — presence or absence).

However, we have eliminated one of these eight combinations: cognitive impairment and at least one ADL limitation without IADL limitations. Very few people end up in this category, and it seemed highly unlikely that an individual with a cognitive disorder needing help to eat (or to get dressed) would still be able to go shopping (or to cook) on his or her own. Instead, we placed these individuals in the cognitive impairment with ADL and IADL limitations combination.

The estimate is obtained using a multinomial logit model, which estimates the probability for an individual to end up in each of the seven possible combinations of states of disability in the next cycle.

We assume that individual characteristics affect the probability of being in each combination through seven latent variables (one for each combination of states):

$$disab^*_{i,j,t+1} = \boldsymbol{\beta}_j \mathbf{x}_{i,t} + \epsilon_{i,j,t},\ \forall j = 1,2,3,4,5,6,7 \tag{4.4}$$

with

$$Pr(disab_{i,t+1} = j|\mathbf{x}_{i,t}) = Pr(disab^*_{i,j,t+1} > disab^*_{i,k,t+1}),\ \forall k \neq j \tag{4.5}$$

where:

- $disab^*_{i,j,t+1}$ is the latent variable for the state of disability j for individual i in period $t+1$;
- $disab_{i,t+1}$ is the categorical variable for the combination of states of disability of individual i in period $t+1$;
- $\mathbf{x}_{i,t}$ is the vector of individual characteristics for individual i at period t, including the presence of diseases;
- $\boldsymbol{\beta}_j$ is the vector of parameters of the effects of the $\mathbf{x}_{i,t}$ variables on the latent variable for the combination j of states of disability;
- $\epsilon_{i,j,t}$ is a random term specific to the individual i and the disability combination j in period t.

Vector $\mathbf{x}_{i,t}$ includes the presence of all diseases, all risk factors, socioeconomic characteristics as well as a dummy variable for each combination of states of disability in the preceding period. We assume that $\epsilon_{i,j,t}$ follows the standard type 1 extreme value distribution, i.e. that the difference between the error terms follows a logistic distribution. Since we use the difference between the latent variable values to calculate each probability, we have to normalize to 0 the parameters vector for one of the combinations of states. We can nevertheless obtain the marginal effects for all combinations.

Table 4.7 presents the estimated average marginal effects for each combination of states of disability. In general the increase in age is associated with an increases in the probability of becoming disabled. The presence of diseases generally has a positive and significant effect on the probability of suffering from the different combinations of disability.

Smoking increases the risk of suffering from almost all combinations of states of disability, with the exception of ADL limitation only. Being a former smoker only has an impact on combinations that include cognitive impairment. Classes II and III obesity only has a positive and significant effect on the risk of having one or more ADL limitation (by themselves) and ADL and IADL limitations combined.

Being a woman significantly affects the probability of having most disability combinations. Higher educational attainment is associated with a greater probability of having no disability at all. The effects of disability dummy variables are obviously strong because individuals tend to remain in the same state of disability from one period to the next.

4.1.6 Model for long-term care

We model the probability of transitions between different states of long-term care (LTC) use. There are three possible states: no LTC use, receiving home care, or residing in a LTC facility. Institutionalization thus characterizes both disability and use of LTC in the model. As with the model for disability, the estimate is made using a multinomial logit model, which estimates the probability for an individual to end up in each of the possible states of LTC use in the next cycle. Living in a LTC facility is an absorbing state, thus when an individual enters it no exit is modelled.

Table 4.8 presents estimates of average marginal effects for each state of LTC use. Being a woman increases the likelihood of receiving home care. After 50 years, age increases the likelihood of receiving LTC. The presence of diabetes, cancer, heart disease and lung disease have a positive and significant effect on the likelihood of receiving home care. Except for diabetes and dementia, diseases do not have a significant effect on institutionalization.

Having a limitation in ADL or IADL strongly increases the probability of receiving home care, and all limitations increase the probability of living in a LTC facility. The dummy variable for current home care use is a strong predictor of LTC use in the next period.

The distinction between home care received from formal and informal sources is covered in section 7. We do not use the distinction in the transitions model because varying sources of help are not necessarily related to different states of health or of LTC use. The goal of the transitions model is to determine the *level* of LTC use of an individual, not to characterize the services used.

	Disability in the next period					
	Cognitive impairment only	IADL limitations only	ADL limitations only	Cognitive impairment and IADL lim.	ADL and IADL limitations	Cognitive impairment & ADL & IADL lim.
Aged between 60 and 64	−0.337*	0.430***	0.378*	−0.101	0.272*	0.436*
Aged between 65 and 69	−0.078	0.430***	0.022	−0.780**	0.616***	0.107
Aged between 70 and 74	0.089	0.836***	1.305***	0.261	0.850***	1.803***
Aged between 75 and 79	0.038	1.280***	1.147***	0.901***	1.457***	2.249***
Aged between 80 and 110	0.242	1.846***	2.064***	1.287***	2.379***	3.760***
Presence of diabetes	0.267**	0.366***	0.386**	0.549***	0.592***	0.569***
Presence of hypertension	0.072	0.311***	0.082	0.297*	0.142**	−0.070
Presence of cancer	0.128	0.264**	−0.428	0.697*	0.166	−0.087
Presence of heart disease	0.578***	0.394***	−0.176	0.820***	0.414***	0.260*
Presence of stroke	0.871***	0.815***	0.891**	1.254***	1.424***	0.973***
Presence of lung disease	0.159	0.550***	0.768***	0.677***	0.571***	0.681***
Presence of dementia	1.322***	0.849***	0.430	2.272***	1.063***	2.449***
Current smoker	0.397***	0.380***	−0.030	0.506***	0.250**	0.501**
Former smoker	0.355***	0.126**	0.070	0.381**	−0.013	0.316**
Class I obesity	−0.101	0.077	0.224	−0.147	0.174**	−0.257
Classes II-III obesity	−0.021	0.098	0.770***	−0.190	0.228**	−0.186
Woman	0.141	0.746***	−0.036	0.271**	0.165**	0.258**
Immigrant	0.009	0.105*	0.401**	0.291	0.418***	0.258*
High school graduate	−0.281**	−0.150**	−0.193	−0.574***	−0.228***	−0.558***
College degree	−0.920***	−0.213*	−0.327	−0.656***	−0.566***	−0.918***
University degree	−0.455***	−0.414***	−0.673***	−1.174***	−0.936***	−0.892***
Resides in Quebec	−0.383**	−0.066	0.182	−0.184	0.084	−0.090
Resides in Ontario	−0.026	−0.124*	−0.180	−0.057	0.041	0.117
Resides in the Prairies	−0.118	−0.055	0.054	0.027	0.075	0.175
Resides in British Columbia	−0.184	0.043	−0.049	0.060	0.179	0.064
Presence of cognitive impairment only	3.672***	0.400*	0.109	3.305***	1.121**	3.135***
Presence of IADL limitations only	0.859***	2.683***	1.040***	2.722***	2.851***	2.465***
Presence of ADL limitations only	1.475*	1.529***	2.422***	0.874*	2.446***	1.551**
Presence of cog. imp. & IADL lim.	3.553***	2.639***	3.138***	5.519***	3.373***	4.847***
Presence of ADL and IADL limitations	−0.191	2.889***	2.512***	3.374***	4.906***	4.466***
Presence of cog. imp. & of ADL & IADL lim.	4.096***	3.513***	−12.599	5.779***	5.507***	7.634***
Constant	−4.950***	−4.187***	−6.106***	−6.687***	−5.524***	−7.615***
Number of observations				69,126		
Average mortality	0.0091	0.0609	0.0042	0.0049	0.0231	0.0083
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01						

Table 4.7: Coefficients of variables on two-year probability of transition between the combinations of states of disability

	LTC use in the next period	
	Receives home care	Living in a LTC facility
Age (if 50 years or under)	-0.024^{***}	0.120
Age (if over 50 years)	0.073^{***}	0.164^{***}
Presence of diabetes	0.393^{***}	0.503^{*}
Presence of hypertension	0.070	-0.003
Presence of cancer	0.371^{**}	0.142
Presence of heart disease	0.208^{**}	-0.243
Presence of stroke	0.251	0.275
Presence of lung disease	0.248^{***}	0.071
Presence of dementia	0.401^{*}	1.245^{***}
Current smoker	0.228^{**}	0.306
Former smoker	0.063	0.186
Class I obesity	0.207^{**}	0.428^{*}
Classes II-III obesity	0.341^{***}	-2.478^{**}
Woman	0.373^{***}	0.379^{*}
Immigrant	-0.262^{***}	-0.472^{*}
High school graduate	-0.006	-0.413^{**}
College degree	0.048	-0.283
University degree	0.092	-2.188^{***}
Resides in Quebec	0.166^{**}	0.370
Resides in Ontario	-0.126	0.360
Resides in the Prairies	-0.124	0.385
Resides in British Columbia	-0.160^{*}	-0.071
Presence of ADL limitations	0.814^{***}	1.026^{***}
Presence of IADL limitations	1.103^{***}	1.641^{***}
Presence of cognitive impairment	0.038	0.801^{**}
Receives home care services	1.383^{***}	0.999^{***}
Constant	-3.157^{***}	-15.627^{***}
Number of observations	29,659	
Average mortality	0.0660	0.0166
Legend : * p <0.10 ; ** p <0.05 ; *** p <0.01		

Table 4.8: Coefficients of variables on two-year probability of transition between states of use of LTC

4.1.7 Transitions validation

To validate the quality of our estimates, we simulated the evolution of the population from 1994 to 2010 using the previously estimated parameters, and compared the characteristics of that population with those obtained in the NPHS population. We started with the population of the 1994 wave, and simulated the first transitions. We used the simulated population of 1996 and simulated the transitions, and so on, until 2010. With incorrect dynamics, we would end up with very different prevalences in 2010. The results are shown in Table 4.9. Although there are differences, the results are generally satisfactory.

	Diabetes	Hypertension	Cancer	Heart disease	Stroke	Lung disease	Dementias
1994	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1996	0.0005	-0.0013	0.0019	0.0003	0.0008	-0.0018	-0.0002
1998	0.0043	0.0022	-0.0005	0.0011	0.0040	0.0020	0.0006
2000	0.0050	0.0047	-0.0011	0.0021	0.0042	0.0066	0.0020
2002	0.0049	0.0019	0.0013	0.0046	0.0032	0.0113	0.0031
2004	0.0034	0.0027	-0.0014	0.0016	0.0051	0.0100	0.0045
2006	0.0042	0.0031	0.0025	0.0069	0.0074	0.0171	0.0082
2008	0.0065	0.0054	-0.0009	0.0053	0.0128	0.0202	0.0086
2010	0.0089	0.0133	0.0049	0.0092	0.0165	0.0278	0.0076
1994-2010	0.0042	0.0036	0.0007	0.0035	0.0060	0.0104	0.0038

Table 4.9: Difference in prevalences between simulated population and NPHS population — 1994-2010

A second validation is done by comparing the projections of period life expectancy obtained in COMPAS with the ones produced by Statistics Canada (Bohnert et al., 2015).[1] The results from COMPAS are close to official projections. In 2050, the 90% confidence interval for COMPAS projections lies between Statistics Canada's Medium (M1) and High projection scenarios.

	Life expectancy at 30 years old			
	StatsCan-Low	StatsCan-M1	StatsCan-High	COMPAS
2030	54.1	54.9	56.2	55.6 (55.3, 56.0)
2050	55.9	57.3	59.4	58.4 (58.0, 58.9)
	Life expectancy at 65 years old			
	StatsCan-Low	StatsCan-M1	StatsCan-High	COMPAS
2030	21.5	22.0	23.0	22.7 (22.3, 23.0)
2050	22.7	23.7	25.4	24.9 (24.7, 25.3)

Table 4.10: Comparison of projected life expectancies for the province of Quebec
Source: COMPAS, Statistics Canada (ndd) and authors' calculations.
Note: 5^{th} and 95^{th} percentiles of estimates shown in brackets.

[1] The projections by province published by Statistics Canada stop in 2038, so we extend the projection for Quebec until 2050. Because they are not publicly available, we have obtained the detailed projections of mortality by age, sex, province and year used in preparing Population Projections for Canada (2013-2063) directly from Statistics Canada to reconstruct the period life expectancies.

Chapter 5

Renewal

In COMPAS, entry of a new cohort of individuals aged 30 and 31 years old is implemented in each simulation cycle, that is every 2 years. Each entering individual has many characteristics. In proportion, these should at all times reflect the joint distribution of the initial conditions (at $t = 0$) of the target population aged 30 and 31 years, as well as potential trend changes that we will have applied. In this chapter, we start by describing our methodology in section 5.1, then in section 5.2 we present some results of our renewal procedure.

5.1 Modeling

In each simulation cycle, we introduce individuals of a fixed age into COMPAS. Denote by $y = (y_1, \dots, y_M)$ their M individual characteristics. Table 5.1 gives the list of characteristics of an individual entering the model. Each of these characteristics corresponds to only one type of variable: binary, ordinal, or integer.

Each year, the average characteristics of the cohort entering the model change. These characteristics are correlated with each other. Let $F_t(y)$ be the joint probability distribution of these characteristics in year t and let $F_m(y_m)$ be the marginal distribution of the variable y_m.

We wish to model the evolution of entering cohorts over time. To do this, using information from the public-use microdata files (PUMFs) discussed in chapter 2, we look at the evolution of $\{F_{m,t}(y_m),\ m = 1, \dots, M\}$. We arrive at this by modeling the evolution of the expected value of y_m. The correlation between three variables is accounted for: education level, tobacco use, and obesity. These are the three variables for which we apply an exogenous trend for the entering cohorts.

Let $\{\tau_{m,t},\ m = 1, \dots, M\}$ be a sequence such that:

$$E_{m,t}[y_m] = \tau_{m,t} E_{m,0}[y_m], \tag{5.1}$$

where $E_{m,0}[y_m]$ is the average of y_m in the year of initialization of the model ($t = 0$), and $E_{m,t}[y_m]$ is the expected value of y_m at t. In COMPAS, the year of initialization is the year 2010.

Variables	Types
Age	integer
Birth year	integer
Sex	binary
Immigrant	binary
Education level	ordinal
Diabetes	binary
Hypertension	binary
Heart diseases	binary
Stroke	binary
Cancer	binary
Lung diseases	binary
Dementias	binary
Tobacco use	ordinal
Obesity	ordinal
Disability	ordinal
Long-term care	ordinal

Table 5.1: Types of variables associated with entering individuals

To preserve the correlation between education level, tobacco use and obesity, we first determine education level using a multinomial logit with sex and province fully interacted as independent variables. Let $\{p(n), n = 1, \ldots, N\}$ be the probabilities to have education level n where $\{p(n), n = 1, \ldots, N\}$ is determined by a logistic regression detailed in chapter 7. As with transitions, we first obtain the probability of being in each category and then adjust probabilities by $\tau_{m,t}$ to take trends into account. Afterward we draw a random number between 0 and 1 from the uniform distribution to determine the level of education.

In a second step we estimate a multinomial logit with all 9 possible combinations of tobacco use and obesity (3 x 3) as the dependent variable, and sex, province and education level as independent variables. As in the case of education level above, we apply the trend $\tau_{m,t}$ and draw a random number to determine tobacco use and obesity level.

5.2 Implementation

5.2.1 Historical trends

In order to identify historical trends concerning obesity, education and tobacco use and, based on these, construct factors of change to be applied to future cohorts of entrants, we use the public-use microdata files of the CCHS, the Canadian Tobacco Use Monitoring Survey (CTUMS) from 2000 and 2012, and the Census of 2006 and 2016. The sample considered consists of the whole Canadian population aged 25 to 34 years. This means we implicitly assume that the evolution of these three characteristics is the same in each region, although their levels may differ. We retain individuals aged 25 to 34 years for obesity and tobacco use in order to have enough observations while remaining as close as possible to the age of entering cohorts, which is 30 and 31 years old. We use individuals aged between 30 and 34 for education level as we use the Census. We assume that the trends are the same for men and women and by regions for tobacco use and obesity. For education level, we have trends by sex and region.

According to Statistics Canada (2000, 2012a, see table 5.2), between 2000 and 2012 the obesity rate increased for the population aged 25 to 34 years. The annual average growth rate (AAGR) of classes II-III obesity was just under 2% per year while class I obesity only increased 0.25% per year.

Statistics Canada (2006, 2016a, see table 5.3) indicates an important decrease in the proportion of Canadians aged 30 to 34 years who do not have any diploma, with the exception of men in Ontario. Conversely, it is not surprising to find that the share of individuals with a university diploma rose, by 2.10% per year on average between 2000 and 2012.

While the trend is upward for obesity, the inverse is observed for tobacco use (Statistics Canada, 2000, 2012a, see table 5.2). The share of smokers declined by 2.15% per year. The proportion of former smokers increasd by 0.64% per year and the share of Canadians who never smoked increased during the period under study, at an AAGR of 0.75%.

Table 5.2 presents a summary of the AAGRs calculated using the survey data.

Data sources	Characteristics	AAGR
	Absence of obesity	-0.12%
Canadian Community Health Survey	Class I obesity	0.25%
	Classes II-III obesity	1.99%
	Never smoked	0.75%
Canadian Tobacco Use Monitoring Survey	Current smoker	-2.15%
	Former smoker	0.64%

Table 5.2: Annual average growth rate (AAGR) of the shares of certain individual characteristics, Canadian population aged 25 to 34 years old, 2000 to 2012

Regions	Sex	No diploma	High school graduate	College degree	University degree
Resides in Atlantic	Men	-4.07%	0.46%	-0,61%	2.46%
	Women	-5.51%	-1.41%	-0.90%	3.52%
Resides in Quebec	Men	-1.16%	1.00%	-1.97%	0.72%
	Women	-2.72%	-0.23%	-1.15%	1.73%
Resides in Ontario	Men	0.12%	-0.48%	-0.45%	0.83%
	Women	-1.32%	-1.86%	-0.73%	1.84%
Resides in Prairies	Men	-3.43%	0.88%	-1.66%	1.85%
	Women	-3.73%	-1.07%	-0.93%	2.85%
Resides in British Columbia	Men	-2.77%	0.33%	-1.34%	1.45%
	Women	-2.92%	-1.43%	-1.25%	2.61%

Table 5.3: Annual average growth rate (AAGR) of the shares of population by educational achievement, Canadian population aged 30 to 34 years old, 2006 to 2016

5.2.2 Projections

The recent historical AAGRs obtained for obesity, educational achievement and tobacco use are not used "as is" in each simulation cycle. Instead, the different growth rates are adjusted in order to reflect some uncertainty about the future. Thus, for the obesity and education variables, the 2000-2012 AAGR presented in table 5.2 is only assumed to hold between 2010 and 2020. It is then reduced by half over 2020-2030 and then again over 2030-2040. The AAGR is then set to zero between 2040 and 2050.

Tobacco use in Canada has been relatively stable among young to middle-age adults since 2007. The prevalence of tobacco use fluctuates in the 25-34 years old group, around $21-24\%$ for the period 2009-2012 (Health Canada, 2012). This seems to indicate that the decline in prevalence recorded between 2000 and 2012 is mainly the result of an important decline in the number of smokers in the early 2000s. There is therefore no reason to believe that the share of smokers will decline in future cohorts. Due to this, each new cohort entering the model has, at 30 and 31 years old, approximately the same share of smokers as the preceding cohort. Otherwise stated, the trends are stable for tobacco use in the model : the AAGR is zero. This is precisely what is shown in table 5.4, where the proportions are shown for entering cohorts. However, it is possible to use non-zero trends if deemed necessary.

In table 5.4, we present the distribution of certain individual characteristics among the entering cohorts. These are adjusted using equation **??**. As is the case for imputation in chapter 2, we draw from the multivariate distribution multiple times. For each replication of a simulation we re-draw the characteristics of the entering cohorts. Table 5.4 shows the results of 100 replications. More information on randomness and uncertainty can be found in chapter 9.

Cycle Year	0 2010	1 2012	5 2020	10 2030	15 2040	20 2050
No diploma	8.3%	7.6%	6.4%	5.8%	5.1%	5.5%
High school graduate	34.2%	34.7%	33.8%	32.4%	32.5%	32.5%
College degree	21.1%	23.2%	21.3%	20.6%	19.5%	19.9%
University degree	36.4%	34.5%	38.5%	41.3%	43.0%	42.1%
Never smoked	42.9%	39.7%	40.2%	40.4%	40.4%	40.2%
Current smoker	26.9%	26.3%	26.4%	26.1%	26.2%	26.1%
Former smoker	30.2%	34.0%	33.4%	33.5%	33.4%	33.7%
Absence of obesity	81.3%	80.3%	80.0%	80.0%	80.4%	80.4%
Classe I obesity	13.0%	13.0%	13.3%	13.3%	13.1%	13.2%
Classes II-III obesity	5.7%	6.7%	6.7%	6.7%	6.6%	6.5%

Table 5.4: Distribution of certain individual characteristics in entering cohorts (30-31 y.o.)

Chapter 6

Demographics

The present chapter describes the assumptions with regard to mortality and immigration, as well as the demographic models which serve as the basis for COMPAS. It also describes the modelling of mortality improvements.

6.1 Mortality

6.1.1 Definitions and statistics

The (gross) mortality rate is the ratio of the number of deaths in a year to the average total population in the year. Otherwise stated, it is the probability of dying in the course of a year. Among other things, this varies depending on the age structure of the population: a person's sex; and year. According to Statistics Canada (Martel, 2013), aggregate mortality rates have fluctuated between 7.0 and 7.2 per thousand in Canada between 2001 and 2011, and sat at 7.0 per thousand in 2011.

While this probability is of interest to analyze the evolution of mortality with precision, life expectancy at birth is better at capturing the evolution of mortality as a whole. Life expectancy at birth is defined as "the average number of years that a newborn could expect to live if he or she were to pass through life exposed to the sex- and age-specific death rates prevailing at the time of his or her birth[...]."(World Health Organization, 2015, p.159). For the three-year period of 2009 to 2011, life expectancy at birth in Canada was 79.3 years for men and 83.6 years for women according to Statistics Canada (Martel, 2013).

Historical data shows a clear improvement in mortality rates between 1960 and 2011. Life expectancy rose from 68.2 to 79.5 years for men and from 74.2 to 83.7 years for women during this period (Human Mortality Database, 2005). In COMPAS, given that young cohorts aged 30 and 31 years enter the modelling process up until 2050, the very efficiency of our model requires forecasts of mortality rates over a long period. Moreover, given that an individual may reach 110 years old in COMPAS, the forecasts should ideally include age groups for the oldest among the elderly, or those aged 90 and over.

Improved overall mortality depends mainly on two factors: *(i)* trends linked to diseases and risk factors; and *(ii)* technological progress, particularly in medicine. The contribution of diseases and risk factors is largely modelled in COMPAS even though, as stated in chapter 4, the transition models are not a function of time. The overall combined effect of diseases and risk factors is nearly nil in COMPAS as a result of the various opposing effects. Thus, in the long term, the significant improvement in mortality mostly comes from technological progress. As technological progress is not modelled in COMPAS, we rely on exogenous assumptions with respect to mortality rate improvement.

In the following section, we present the sources of our mortality rate forecasts.

6.1.2 Estimation methods

As an input in population projections, prospective mortality tables have been built around the world (for a description see Macdonald (1997); McDonald et al. (1998)). Numerous models exist, but the development of existing forecasts is largely based on the *Lee-Carter* model introduced in Lee and Carter (1992), as well as on certain variants, such as the Poisson log-bilinear model (Brouhns et al., 2002; Renshaw et al., 1996). In its population projection exercise, Statistics Canada uses a modified *Lee-Carter* model to forecast mortality rates, the Li and Lee (2005) method with some improvements – among others the "implementation of the 'Extended Lee-Carter' model for modelling the evolution of the age patterns of mortality decline (Li et al., 2013)" (Bohnert et al., 2015, p.47).

We use the latest Statistics Canada projections of mortality rates from Population Projections for Canada (2013-2063) (Statistics Canada, 2015). These projections include all the characteristics needed for our simulations. They go to 2063 and they are finely defined by age, sex and province (they are available by province up to 2038).[1] Projections are provided for old ages, up to 110 years old.[2]

We use the mortality table "as is", with three exceptions. First, we group provinces in five regions (Atlantic, Quebec, Ontario, Prairies, British Columbia) by simply calculating the weighted mean of the provincial mortality rates by age, sex and year. Second, we apply to all regions the Canadian evolution of mortality for the period after 2038. Third, we only use Statistics Canada projections until 2050; after this point we keep mortality rates constant at 2050 rates. Table 6.1 shows the annual reduction of mortality rates based on Statistics Canada's forecasts. Hereafter, we present the integration of the prospective mortality table into the COMPAS model.

6.1.3 Integration into COMPAS

Let $s \in S = \{h, f\}$ with S being a set representing the human genus, where the letters h and f respectively represent men ("hommes" in French) and women ("femmes"). Let $\{ga_i,\ i = 1,\ldots,n\}$ be the age groups observed in the population, $\{pp_j,\ j = 1,\ldots,m\}$ be the projection periods and $\{r_k,\ k = 1,\ldots,p\}$ be the Canadian regions. The corresponding samples ranked in increasing age and chronological order are denoted as $ga_{1,n} < \ldots < ga_{n,n}$ and $pp_{1,m} < \ldots < pp_{m,m}$.

[1] Because they are not publicly available, we have obtained the detailed projections of mortality by age, sex, province and year used in preparing Population Projections for Canada (2013-2063) directly from Statistics Canada.

[2] For more information on the methodology used by Statistics Canada, see Bohnert et al. (2015).

Age group	2013-2019	2020-2029	2030-2039	2040-2050
		Men		
30 to 39	2.1%	2.1%	2.1%	2.2%
40 to 49	2.1%	2.2%	2.2%	2.2%
50 to 59	2.6%	2.5%	2.4%	2.4%
60 to 69	2.7%	2.6%	2.5%	2.4%
70 to 79	2.3%	2.2%	2.1%	2.1%
80 to 89	1.1%	1.1%	1.0%	1.0%
90 to 99	0.3%	0.3%	0.2%	0.2%
100 to 110	0.2%	0.1%	0.1%	0.1%
		Women		
30 to 39	1.9%	1.7%	1.7%	1.7%
40 to 49	1.8%	1.7%	1.7%	1.7%
50 to 59	1.8%	1.8%	1.8%	1.8%
60 to 69	1.8%	1.8%	1.8%	1.8%
70 to 79	1.7%	1.6%	1.6%	1.6%
80 to 89	1.0%	0.9%	0.9%	0.9%
90 to 99	0.3%	0.2%	0.2%	0.2%
100 to 110	0.2%	0.1%	0.1%	0.1%

Table 6.1: Annual rates (in %) of mortality rates reduction.
Source: Statistics Canada (2015) and authors' computation.

The prospective mortality table by sex, region, age group (increasing) and period (also increasing) is denoted by $M_{s,p,n,m}$. Let g be the number of years between two demographic transitions as described in chapter 4 (by default and due to the structure of the survey data we use, $g = 2$). Let us denote by *nages* and *ncycles* the number of age ranges and simulation cycles in the COMPAS model:

$$\begin{aligned} nyears &= 1 + (stopyear - startyear)\,, && (6.1)\\ ncycles &= 1 + (stopyear - startyear)\,/g, && (6.2)\\ nages &= stopage - startage + 1, && (6.3)\\ nregions &= 5 && (6.4) \end{aligned}$$

where *startyear*, *stopyear*, *startage* and *stopage* are respectively the year of initialization of the model, the last simulation year, the minimum age an individual can have, and the maximum age at death permitted for an individual in the model. The current default values for these parameters are 2010, 2130, 30 and 110, respectively. In the current version of COMPAS, the data sources provide information by "age groups in the population" of one year. The "age ranges in the model" are of one year as well. However the notions of "age groups in the population" and "age ranges in the model" need not refer to the same mathematical quantities, e.g. were we to use different data sources.

Under the assumption that there is a gradual displacement of mortality rates reduction from younger ages towards older ages (see below), we define the annual decrease in the probability of death, $q_{\text{x}}(s, r, a, t)$, in the following recursive manner:

$$\begin{aligned} q_{\mathrm{x}}(s,r,a,1) &= 1, &(6.5)\\ q_{\mathrm{x}}(s,r,a,t) &= q_{\mathrm{x}}(s,r,a,t-1)\left(1-M_{s,p,i,j}\right)^{g}, &(6.6)\end{aligned}$$

with $a = 1, \ldots, nages$, $t = 2, \ldots, nyears$ and $r = 1, \ldots, nregions$.

On the one hand, note that the minimum age an individual can have in COMPAS, *startage*, his/her current age range a, and the real age of an individual are linked by the relation:

$$age = startage + a - 1. \tag{6.7}$$

Thus, the matrix index i in (6.6) is linked to the real age of an individual and the population age group in the prospective mortality table $M_{s,p,n,m}$ through the relationship:

$$i = \sum_{l=1}^{n} l \times \mathbb{1}_{\left(age \in ga_{l,n}\right)} = \begin{cases} 1 & \text{if } age \in ga_{1,n} \\ 2 & \text{if } age \in ga_{2,n} \\ \vdots & \\ n & \text{if } age \in ga_{n,n} \end{cases}$$

On the other hand, the year of initialization of the model *startyear*, the number of years between two demographic transitions g, the current cycle $c = 1, \ldots, ncycles$ and the current year *year* in the simulation are linked by:

$$year = startyear + g(c-1). \tag{6.8}$$

Thus, the matrix index j in (6.6) is linked to the current year *year* and the projection period in the prospective mortality table $M_{s,p,n,m}$ by the relation:

$$j = \sum_{l=1}^{m} l \times \mathbb{1}_{\left(year \in pp_{l,m}\right)} = \begin{cases} 1 & \text{if } year \in pp_{1,m} \\ 2 & \text{if } year \in pp_{2,m} \\ \vdots & \\ m & \text{if } year \in pp_{m,m} \end{cases}$$

We then define the probability that an individual remains alive, accounting for his/her individual characteristics and after a demographic transition, as

$$p := p_{s,a,c} = 1 - (1 - p_{i',c}) q_{\mathrm{x}}(s,a,c), \tag{6.9}$$

with $p_{i',c} = 1 - \exp(-\exp(\boldsymbol{\beta}_j \mathbf{x}_{i,c}))$ being the probability of death of individual i at cycle. Where $\mathbf{x}_{i,c}$ is a vector of all explanatory variables accounted for, specific to individual i', and including his/her disability status at time c (for details, the reader can refer to sections 4.1.1 and 4.1.2). This probability thus depends on sex, region, projection period, age and individual health status.

The matrix of mortality rate projection is incoporated in the probability of remaning alive through the term $(1 - M_{s,p,i,j})^g$ where the mortality rate for a given sex, region, age and period is $M_{s,p,i,j}$.

The model used by Statistics Canada does not incorporate opinions on the potential evolution of mortality due to advances in medicine, to the appearance of new illnesses or risk factors or to changing lifestyles. In this respect, it is worth noting that Statistics Canada explicitly recognizes that its projections are based on an extrapolation of past trends: *"More than any other component of demographic growth, mortality lends itself to projections based on the extrapolation of past data"* (Bohnert et al., 2015, p.42). Now, in COMPAS we can build scenarios to model advances in medicine by changing the mortality projections, changes in risk factor exposure or modifications in the incidence of diseases.

6.2 Immigration

In order to correctly reflect demographic reality, we should account for individuals moving from one country to another — *international migration* — or from one province to another — *interprovincial migration*.

We use Statistics Canada's 2013-2063 population projections (Bohnert et al., 2015). The agency forecasts that the net migratory balance, which includes both migration between provinces and international migration, will increase, for Canada, from 194,400 in 2013 to 250,000 in 2038 for the population aged 30 and over.[3] In the absence of projections by region past 2038, net migration is maintained constant at its 2038 level until 2050. The projections are decomposed in three age groups (30-44; 45-64; 65+) and by sex following the 2013 population breakdown (Statistics Canada ndb, ndc). We use those forecasts to re-weight the existing immigrant population to reflect the size of the new immigrant population at each new cycle of the simulation. This method assumes new immigrants' health remains stable over time and is perfectible; but for the time being the absence of data on the health of new international immigrants prevents us from integrating "real" new immigrants at each cycle.

6.3 Conclusion

By integrating the mortality model and the migratory balance forecasts in COMPAS and combining the results with Statistics Canada's forecasts for individuals under 30 years old,[4] we obtain a total population of just over 45 million inhabitants in 2050. As shown in table 6.2, the population forecast by COMPAS lays between Statistics Canada's "low" and "medium" scenarios, with a handful of age groups just under the population of the "low" scenario in 2016 and 2030. The population projected by COMPAS shifts from Statistics Canada's "low" scenario towards its "medium" one as the simulation progresses in time.

[3] We used an earlier version of 2013-2063 population projections which have since been updated. We will use updated figures in a subsequent version of COMPAS.

[4] This operation is required here because COMPAS only simulates the population aged 30 and over.

	COMPAS			Stat-Can Low			Stat-Can M1		
Age group	2016	2030	2050	2016	2030	2050	2016	2030	2050
0 to 30	12.8	13.7	15.3	12.7	12.4	12.2	12.8	13.7	15.3
30 to 39	4.9	5.4	5.5	4.9	4.9	4.9	5.0	5.3	5.8
40 to 49	4.8	5.4	5.7	4.8	5.1	4.8	4.8	5.4	5.7
50 to 59	5.3	4.9	6.0	5.4	4.7	5.0	5.4	4.9	5.7
60 to 69	4.3	5.0	5.5	4.3	4.7	4.9	4.3	5.0	5.4
70 to 79	2.5	4.3	4.6	2.5	4.1	4.1	2.5	4.2	4.4
80 to 89	1.4	2.3	3.7	1.2	2.1	3.1	1.2	2.1	3.4
90+	0.4	0.5	1.4	0.3	0.4	1.0	0.3	0.4	1.1
All ages	36.3	41.5	47.7	36.0	38.6	39.8	36.2	41.1	46.9

Table 6.2: Comparison of Projected population(in millions): COMPAS vs. Statistics Canada.
Source: COMPAS, Statistics Canada (ndd) and author's calculations.
Note: Figures for the 0-30 years old in COMPAS are those from the Statistics Canada M1 forecast.

Chapter 7

Health care use

The health care use module makes it possible to evaluate, for each simulation year, the quantity of medical resources used by the population. In order to obtain these results, however, we must first establish a relationship between health status and resource use. This chapter describes the process; in COMPAS, the parameters presented here are then applied to the projected population at the end of each simulation cycle.

COMPAS includes several measures of health care use. We model separately the number of consultations with a generalist physician and with a specialist physician; the number of nights an individual is hospitalized; the consumption of drugs; the use of formal and informal home care; and the number of hours used of both types of home care.

In the next section we present the data and econometric models used for the estimation. This is followed by a presentation of the resulting estimates.

7.1 Modelling

This section presents the data and variables used in estimating health care use and briefly explains the econometric models used. We rely on the NPHS (see chapter 2 on this topic) for all the health care use estimation, with the exception of the number of hours of formal and informal home care. For those, we use the General Social Survey (GSS) of 2012 (cycle 26). With the NPHS, we only use data from 2000-2001 to 2010-2011 to control for the risk that health care use patterns may have changed over time. While the regression type (i.e. the econometric model) used differs depending on the category of health care considered, explanatory variables remain the same in all regressions that use NPHS data. These regressions are only performed using respondents who live in private households in the current period, because there is not sufficient information on health care use for those who live in an institution. For example, if a respondent is in an institution at the time of the last survey cycle, his/her answers to the preceding cycles are still used (only responses from the last cycle are excluded).

The socio-demographic independent variables used in NPHS-based regressions are sex, age, immigrant status, education, and indicators for respondents living in each Canadian region (Atlantic, Quebec, Ontario, Prairies, and British Columbia). As elsewhere in this report, Atlantic is the reference. The effect of age on resource use does not appear to be linear. Specifically, it appears to change after 50 years of age. In order to capture this effect, a spline was created (Goldman et al., 2005). There are therefore two age variables, one for individuals under the age of 50 and another for those aged 50 years and over. This last variable represents the number of years since the individual turned 50 (e.g., it takes a value of 5 for a 55-year-old person). The estimated coefficient of the first age variable thus captures the effect of age up to 50 years on the dependent variable; similarly, the coefficient of the second variable is the effect of age after 50 years on the use of health care resources. We use three variables for disability (1+ cognitive disability, 1+ IADL, and 1+ ADL) and dummies for the presence of the seven diseases included in COMPAS. The risk factors (tobacco use and obesity) are also included. Tobacco use is represented by two binary variables: one for former smokers and one for current smokers. Two binary variables are included in the regression to capture the effect of class I ($30 \leq$ BMI <35) and classes II-III (BMI ≥ 35) obesity. The reference categories for risk factors are thus individuals who have never smoked and those with a BMI under 30.

Regressions using GSS data rely on fewer variables as we need to use variables that have counterparts in the CCHS. We use sex, age and Canadian regions in the same format as in NPHS-based regressions. We include the seven diseases but heart disease, stroke and hypertension are grouped into cardiovascular diseases because only one question is asked in GSS for these three diseases combined. We do not include disability variables, because questions are relative to help received for a disability. The questions are too far from those used in the CCHS and NPHS to include disability variables in our regression.

Consultations with physicians, number of nights hospitalized and hours of home care services are generally characterized by a large number of observations with missing values (or small values used) and an asymmetric distribution. Stated otherwise, the foregoing means that during a reference period, many individuals do not actually use any of these medical resources while others use much more than the average (Frees et al., 2011). In order to account for such a distribution of data, we use a negative binomial regression.

For drugs consumption, which in the current version of the model can take two values (”yes” or “no”), a logistic regression is used. Use of home care services can take three values (informal, formal or both) conditional on using any home care, so we use a multinomial logistic model. The framework of the multinomial logistic model is presented in section 4.1.5; we present the logistic regression and the negative binomial regression in next two subsections. The econometric theory of the models is based on Cameron and Trivedi (2005).

7.1.1 Negative binomial regression

The negative binomial regression is generally used to analyze discrete and countable data, i.e. data that only take whole and non-negative values, which fits well with health care use data. This type of regression, by allowing for overdispersion of data (variance higher than the mean), can capture the asymmetry in the distribution observed in lots of health care use data. Using the Poisson distribution, which assumes equality between mean and variance, would not have matched the distribution of the data as effectively.

The negative binomial regression assumes that the observations are generated from a negative binomial distribution, the first two moments of which are:

$$\mathbb{E}(y_i|\mu_i, \alpha) = \mu_i = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \tag{7.1}$$

$$\mathbb{V}ar(y_i|\mu_i, \alpha) = \mu_i(1 + \alpha\mu_i) \tag{7.2}$$

where:

- $\mathbf{x}_i$ is a vector of explanatory variables for individual i;
- $\boldsymbol{\beta}$ is a vector of coefficients;
- α is the overdispersion parameter.

If $\alpha > 0$, it necessarily follows from the definition of variance that there is overdispersion, given that the variance is greater than the mean (μ_i being positive). This definition of variance is the one used by the *nbreg* command in Stata to perform the negative binominal regression. Estimation of coefficients is done using the maximum likelihood method.

However, the estimated parameters do not directly yield the effects of a variable x on the conditional expected value of y. Thus, it is not the parameters themselves which are interesting when analyzing the explanatory variables' impact, but rather the marginal effects, namely, the change in the conditional expected value of y_i when the value of a variable x_i changes by one unit. As elsewere in this report, we present the average marginal effect (AME). As a reminder, the AME is an average change, over all individuals, in the expected value of y_i when x_i changes by one unit.

7.1.2 Logistic regression

The logistic regression, as mentioned above, is used with variables that can only take two values, such as consumption (or not) of at least one drug. The specification is as follows:

$$cons^*_{i,t+1} = \boldsymbol{\beta}\mathbf{x}_{i,t} + \epsilon_{i,t}, \tag{7.3}$$

with

$$cons_{i,t+1} = \begin{cases} 1 & \text{if} \quad inc^*_{i,t+1} > 0 \\ 0 & \text{if} \quad inc^*_{i,t+1} \leq 0 \end{cases} \tag{7.4}$$

where:

- $cons^*_{i,t+1}$ is a latent variable for the drug consumption of person i at period $t+1$;
- $cons_{i,t+1}$ is a binary variable indicating whether an individual i is taking medication at period $t+1$;
- $\mathbf{x}_{i,t}$ is a vector of all explanatory variables;
- $\boldsymbol{\beta}$ is a vector of the effects of explanatory variables on the latent variable of drug consumption;

- $\epsilon_{i,t}$ is a random term specific to individual i at period t.

We assume that the distribution function of $\epsilon_{i,t}$ follows a logistic law. The dependent variable y thus has a probability p of taking a value of 1 and of $1-p$ of taking a value of 0. This probability is modelled as:

$$p \equiv \mathbb{P}\left(y_i = 1 | \boldsymbol{x}_i\right) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}$$

where:

- $\mathbf{x}_i$ is a vector of explanatory variables for individual i;
- $\boldsymbol{\beta}$ is a vector of coefficients.

Estimation of the parameters is again done by maximum likelihood. As in the case of the negative binominal regression, the estimated parameters are not amenable to direct interpretation. It is the AME that is used to show the impact of the explanatory variables on the probability that y_i will take a value of 1.

The two types of regressions described in this section allow us to model the relation between use of medical resources and the presence of the seven illnesses, two risk factors, and disabilities.

7.2 Results

7.2.1 Physician consultations and hospitalizations

The effect of the different illnesses on use of medical resources is presented in table 7.1. Use is computed over a 12-month period. As such, if an individual is said to have consulted with a specialist 3 times, this indicates that he consulted with a specialist 3 times over the course of a 12-month period. As expected, illnesses have a large effect on the number of consultations with a generalist. The presence of diabetes, hypertension, cancer, heart disease or lung disease increases the number of consultations with a generalist by 0.7 to 1.2 visits. The presence of dementia has no significant effect, but the presence of cognitive impairment does. Classes I and II-III obesity also increase the number of consultations, by 0.3 and 0.6 respectively. However, it is disabilities that lead to the largest increase in consultations. An individual with at least one IADL had 1.5 more consultations on average than an individual without any IADL. On average, Quebecers consult with physicians less often than individuals in the rest of Canada; the gap is between 0.85 and 1.3 consultations over 12 months. Women also consult a generalist more often.

The effect of illnesses on the number of consultations with a specialist is also positive and significant for some illnesses. The presence of hypertension, cancer, heart disease or lung disease increases the number of consultations. Classes II-III obesity increases the number of specialist consultations by 0.1 per year on average, while tobacco use variables have little effect.

Variable	Nb. of consultations: generalist	Nb. of consultations: specialist	Nb. of nights hospitalized
Age (if under 50)	−0.001	0.001	−0.008
Age (if 50 years or over)	0.007*	−0.007***	0.023**
Woman	0.823***	0.244***	0.223
Immigrant	0.210*	−0.046	−0.389*
High school graduate	−0.051	0.211***	−0.144
College degree	−0.089	0.188***	−0.105
University degree	−0.069	0.366***	−0.182
Resides in Quebec	−1.106***	0.164***	0.186
Resides in Ontario	−0.246**	0.151***	−0.511***
Resides in the Prairies	−0.229**	−0.022	−0.100
Resides in British Columbia	0.193	0.086	−0.100
Presence of diabetes	0.887***	0.085*	0.478**
Presence of hypertension	0.993***	0.199***	0.434***
Presence of cancer	1.168***	1.003***	1.593***
Presence of heart disease	1.053***	0.470***	1.251***
Presence of stroke	0.693***	−0.098	1.356**
Presence of lung disease	0.665***	0.107**	−0.001
Presence of dementia	0.021	0.122	−0.175
Class I obesity	0.327***	0.051	−0.016
Classes II-III obesity	0.626***	0.148**	0.333
Current smoker	0.218*	−0.022	0.262
Former smoker	0.224***	0.078*	0.134
Presence of ADL	0.281	0.124	0.854***
Presence of IADL	1.523***	0.524***	1.726***
Presence of cognitive impairment	0.922***	0.265***	0.962**
Number of observations	49,079	49,147	43,050
Average number of consultations or nights	3.12	0.75	0.86
Legend	* p<0.05 ; **p<0.01; *** p<0.001		

Table 7.1: Average marginal effect of different variables on health care use.
Note: use is computed on an annual basis. Binomial negative regression used.

The presence of dementia or lung disease does not have a significant impact on the number of nights spent in short-term hospitalization, but all other types of illnesses have considerable effects. The presence of a cancer or of stroke increases the number of nights hospitalized by just over 1 on average. Individuals suffering from disabilities spend an average of 1.2 more nights in hospital: 0.9 nights for those with ADL, 1.7 for those with IADL, and 1.0 for those with cognitive impairment. Individuals suffering from two or more disabilities spend an average of 4.2 more nights in hospital than those with no disabilities. The coefficient of class I obesity is not significant and is near zero, but individuals with a BMI of 35 or over (classes II-III) spend an average of 0.3 more nights in hospital than those who are not overweight.

7.2.2 Medications

As stated in chapter 2 and shown in table 7.2, a high proportion of the population is taking medication (84.6%). Women are more likely to have taken medication in the last month, as well as individuals with higher education. Without surprise, those with a diagnosed disease are more likely to take medication. Individuals with diabetes, hypertension and heart disease exhibit the highest increase in the probability of taking at least one medication, compared to those without the disease (a probability that is higher by 17% to 23%). The presence of IADL and cognitive impairment also increases significantly the probability of taking medication.

7.2.3 Home care services

Chapter 4 presented the estimation method to determine whether an individual receives home care services. In this section we estimate the types of services used among individuals who receive home care as we prevously determine if an individual reveive or not home care (see section 4.1.5). An individual can receive formal or informal home care services, or both. Table 7.3 indicates that women are less likely to receive formal home care. In the province of Quebec the probability of receiving informal home care is lower and the probability of receiving formal home care is higher. The presence of ADL or IADL increases the probability of receiving both types of care.

Table 7.4 shows the effect of several variables on the number of hours received of each type of care. It appears that the older individuals get over age 50, the more hours of informal and formal care they receive. Individuals with cardiovascular disease receive more informal care but less formal care. The presence of dementia increases the number of hours by as much as 19.6 hours of informal help and 9.9 hours of formal home care services. Home care users residing in Quebec receive significantly fewer hours of formal and informal help than those in any another region, with the exception of BC in the case of informal home care services.

7.3 Conclusion

The main take-away from the analyses presented in this chapter is that the illnesses considered in COMPAS are important factors in explaining the use of health care resources. Disability also has sizeable effects. Finally, there are significant differences between Canadian regions.

Variable	1+ medication
Age (if under 50)	−0.000
Age (if 50 years or over)	0.000
Woman	0.090***
Immigrant	−0.040***
High school graduate	0.026**
College degree	0.033***
University degree	0.039***
Resides in Quebec	−0.004
Resides in Ontario	0.012
Resides in the Prairies	0.001
Resides in British Columbia	−0.025**
Presence of diabetes	0.186***
Presence of hypertension	0.167***
Presence of cancer	0.032*
Presence of heart disease	0.229***
Presence of stroke	0.163***
Presence of lung disease	0.055***
Presence of dementia	0.039
Class I obesity	0.030***
Classes II-III obesity	0.023
Current smoker	0.008
Former smoker	0.028***
Presence of ADL	−0.005
Presence of IADL	0.096***
Presence of cognitive impairment	0.045*
Number of observations	49,279
Average proportion of users	84.6%
Legend	* p<0.05 ; **p<0.01; *** p<0.001

Table 7.2: Average marginal effect of different variables on the probability of taking medication. Note: use is calculated over one month. Logistic regression used.

Variable	Formal home care	Informal home care	Formal and informal home care
Age (if under 50)	-0.001	0.004	-0.003
Age (if 50 years or over)	0.003	-0.005^{***}	0.003^{**}
Woman	-0.060^{*}	0.022	0.038
Immigrant	-0.068	0.069	-0.001
High school graduate	-0.039	0.020	0.019
College degree	-0.031	-0.029	0.060^{*}
University degree	0.002	0.034	-0.036
Resides in Quebec	0.208^{***}	-0.228^{***}	0.020
Resides in Ontario	0.087^{*}	-0.122^{***}	0.035
Resides in the Prairies	0.104^{**}	-0.122^{***}	0.018
Resides in British Columbia	0.058	-0.093	0.035
Presence of diabetes	0.034	-0.020	-0.014
Presence of hypertension	-0.043	0.057	-0.014
Presence of cancer	-0.042	-0.045	0.087^{**}
Presence of heart disease	0.010	-0.005	-0.014
Presence of stroke	0.071	-0.076	-0.005
Presence of lung disease	-0.002	0.025	-0.023
Presence of dementia	-0.043	0.022	0.021
Class I obesity	0.031	-0.028	-0.003
Classes II-III obesity	0.011	-0.025	0.014
Current smoker	-0.031	0.054	-0.023
Former smoker	-0.004	-0.002	0.006
Presence of ADL	0.077^{*}	-0.140^{***}	0.063^{**}
Presence of IADL	-0.089^{**}	-0.002	0.091^{***}
Presence of cognitive impairment	-0.150^{**}	0.176^{***}	-0.027
Number of observations		2,501	
Average proportion of users	47.1	36.2	16.8
Legend	* p<0.05 ; **p<0.01; *** p<0.001		

Table 7.3: Average marginal effect of different variables on home care use.
Note: use is computed on an annual basis. Binomial negative regression used.

Variable	Nb hours informal home care	Nb hours formal home care
Age (if under 50)	−0.156	-0.329^{**}
Age (if 50 years or over)	0.166^{***}	0.180^{***}
Woman	1.261	0.022
Presence of diabetes	3.409	0.293
Presence of cardiovascular disease	5.491^{***}	-2.802^{*}
Presence of cancer	1.771	−0.614
Presence of lung disease	−0.078	−2.593
Presence of dementia	19.608^{***}	9.876^{**}
Resides in Quebec	-5.436^{***}	-8.330^{***}
Resides in Ontario	−2.139	-4.089^{***}
Resides in the Prairies	-4.916^{**}	0.199
Resides in British Columbia	-5.515^{**}	−2.606
Number of observations	2,136	1,442
Average number of hours	16.8	9.8
Legend	* p<0.05 ; **p<0.01; *** p<0.001	

Table 7.4: Average marginal effect of different variables on the number of hours of formal and informal home care services used.
Note: use is computed on an annual basis. Binomial negative regression used.

Chapter 8

Health care costs

This chapter discusses the methodology used to include the costs of health care in COMPAS. At this stage of the model's development, the only costs that we can model are those for Quebec. Thus, we model the costs of hospitalization and consultations with general practitioners and specialists. To do this, we use 2012 data from the Régie de l'assurance maladie du Québec (RAMQ) and from the *Maintenance et exploitation des données pour l'étude de la clientèle hospitalière* (MED-ECHO) database, maintained by the Ministère de la Santé et des Services sociaux (MSSS). Furthermore, we can estimate an average hourly cost for home care services using data collected by the insurance companies. Finally, we can calculate an approximate average annual cost of institutionalization with some publicly available data provided by the Canadian Institute for Health Information (CIHI) and Statistics Canada.

For each health care use variable, we try to estimate the cost of the resource based on a number of characteristics such as age, gender and the presence of diseases (when the data allows). We use these estimates to calculate the cost of the health care received by each simulated individual according to his or her characteristics. We begin by describing in detail the data used to attribute costs to the health care used. We then present the models to calculate these costs.

8.1 MED-ECHO

To estimate the cost of hospitalization, we match the MED-ECHO data to the RAMQ data. MED-ECHO is an administrative database that includes all available information on hospitalizations. However, it excludes stays in psychiatric hospitals, rehabilitation hospitals and long-term care facilities as well as physician salary costs. To capture physician salary costs, we must also use the RAMQ data. These are described in detail in section 8.2. It should be noted that the MED-ECHO data does not include hospital costs related to emergency services. Furthermore, the MED-ECHO data that we have for the moment does not include the cost of hospital day surgeries.

For each hospitalization in the MED-ECHO database, we are given the admission and discharge dates (and therefore the length of stay), the number of medical interventions received by the patient during his or her stay, the primary and secondary diagnoses (according to the International Statistical Classification of Diseases and Related Health Problems, 10^{th} Revision, or ICD-10), and an indicator variable for in-hospital death. Classifying diagnoses based on the ICD-10 allows us to observe the presence (or absence) of various diseases. In addition, each hospitalization is assigned an All Patient Refined-Diagnosis Related Group (APR-DRG) code from the MSSS database of hospital performance.

The APR-DRG classifies hospitalizations according to three criteria: severity of illness, risk of mortality and resource consumption. Severity of illness captures the extent of physiological system or organ function loss. The risk of mortality is the probability that the patient dies during his or her stay, and resource consumption corresponds to the relative use of resources to treat a given disease. The APR-DRG therefore includes, in addition to a diagnosis related group (DRG), two variables with four categories each to capture the severity of illness and the risk of mortality.

MED-ECHO data does not directly provide the cost of hospitalization. The MSSS combines a relative intensity of resource use (*niveau d'intensité relative des ressources utilisées* or NIRRU in Quebec) for each DRG and severity of illness, which yields a particular NIRRU per APR-DRG. The NIRRU is an index to calculate the total cost of hospitalization (excluding physician salary costs) by measuring the amount of resources used during the hospitalization. It includes treatment and intervention costs, medicine costs, the cost of transportation, of hospital bed, of depreciation, etc. The database provided by the MSSS includes a NIRRU per hospitalization in addition to a "cost per unit" of NIRRU (or the cost associated with a NIRRU of 1). Combining both data points gives the cost of hospitalization, except for the medical treatment provided by physicians.

8.1.1 NIRRU

In this section, we briefly present the methodology behind the NIRRU. First, we need to calculate the unit cost of a NIRRU, based on cost of hospitalizations in the United States in 1994-1995 and adjusted for Quebec. It presupposes that the cost of hospitalization for a given APR-DRG may be different in the U.S. and in Quebec, but that the cost of all cases of a given APR-DRG should be similar. To calculate the cost of a NIRRU, we first calculate the cost of hospitalization for an APR-DRG i in Quebec according to U.S. costs as well as Quebec and U.S. lengths of hospital stays. The cost of hospitalization for an APR-DRG i in Quebec can be calculated as follows(Ministère de la Santé et des Services sociaux, 2014)[1] :

$$Cost_{Q(i)} = Cost_{US(i)} \left(1 - Ratio_{Q(i)} + \frac{ALS_{Q(i)}}{ALS_{US(i)}}\right) \tag{8.1}$$

where:

- $Cost_{Q(i)}$ is the average cost per hospitalization for the APR-DRG i in the United States adjusted for Quebec;
- $Cost_{US(i)}$ is the average cost per hospitalization for the APR-DRG i in the United States;
- $ALS_{Q(i)}$ is the average length of stay per hospitalization for the APR-DRG i in Quebec;

[1] In addition to being provided with the database, the white paper can be found at the following address: http://www.ecosante.fr/QUEBFRA/11200.html (dated July 19, 2016).

- $ALS_{US(i)}$ is the average length of stay per hospitalization for the APR-DRG i in the United States;
- $Ratio_{Q(i)}$ is the ratio of routine and ancillary costs of hospitalization to average cost per day for the APR-DRG i.

Thus, the cost of hospitalization in Quebec is the cost of hospitalization in the United States, but adjusted for differences in length of stay and routine costs. It is essential to include a correction for routine and ancillary costs since lengths of stay may vary. However, it is reasonable to assume that the vast majority of hospitalization costs incurred during the first days (operating room, radiology, intensive care unit, etc.) and this number of days are the same in Quebec and in the United States. The relative costs for a given APR-DRG should therefore be the same in Quebec and in the United States for the first days of hospitalization. However, there may be important differences between the two areas in the cost of the last days of a hospitalization.

The cost of these last days is calculated by finding the cost per day of hospitalization (the average cost per hospitalization over the average length of stay) for a given APR-DRG in the United States multiplied by the ratio of routine and ancillary costs of hospitalization to average cost per day in Quebec. These routine and ancillary costs are those "normally distributed throughout the stay (e.g. nursing, pharmacy, laboratory, radiology, etc.)" (Ministère de la Santé et des Services sociaux, 2014, p. 178). The adjustment for the number of days is based on the marginal cost of an extra day of hospitalization, not on the average cost of stay. The marginal cost of the last days is exclusively comprised of routine and ancillary costs.

Once we have the average cost of hospitalization for each APR-DRG, we can calculate the unit cost of a NIRRU as follows:

$$Cost_{NIRRU} = \frac{\sum\limits_{i=APR-DRG} (Cost_{Q(i)} \times Case_{Q(i)})}{Case_Q} \tag{8.2}$$

where:

- $Cost_{NIRRU}$ is the cost of a unit of relative intensity of resource use in Quebec;
- $Cost_{Q(i)}$ is the average cost per hospitalization for the APR-DRG i in the United States adjusted for Quebec;
- $Case_{Q(i)}$ is the number of hospitalizations for the APR-DRG i in Quebec;
- $Case_Q$ is the total number of hospitalizations in Quebec.

The cost of a NIRRU depends on the cost of hospitalization in the U.S. adjusted for Quebec and the number of hospitalizations observed in Quebec for each APR-DRG. We can then find the NIRRU of a typical hospitalization for each APR-DRG.

A hospitalization is deemed typical if, during the hospital stay, the patient did not receive long-term care, did not die, did not leave without permission, etc. When it is indeed typical, the NIRRU for the APR-DRG i in Quebec:

$$NIRRU_i = \frac{Cost_{Q(i)}}{Cost_{NIRRU}} \tag{8.3}$$

where:

- $NIRRU_i$ is the relative intensity of resource use for a hospitalization for the APR-DRG i in Quebec;
- $Cost_{Q(i)}$ is the average cost per hospitalization for the APR-DRG i in the United States adjusted for Quebec;
- $Cost_{NIRRU}$ is the cost of a unit of relative intensity of resource use in Quebec.

The NIRRU for atypical hospitalizations is the NIRRU for a typical hospitalization to which is added or subtracted units of NIRRU to consider the length of stay of typical and atypical cases, the maximum length of stay and the ratio of routine and ancillary costs of hospitalization. A detailed description of how to calculate the NIRRU for atypical cases is available in the white paper for the hospital performance database (Ministère de la Santé et des Services sociaux, 2014).

Finally, the cost of hospitalization is the NIRRU for an APR-DRG i times the cost of a NIRRU. For each hospital stay, the total cost is the cost of a NIRRU to which physician-related costs are added. Spending is censored at the 99$^{\text{th}}$ percentile as some observations have outlying values that could skew the means.

Even if this procedure yields a cost for the NIRRU and although the MSSS provides a cost for a NIRRU by year, the reader should be advised that these costs are in fact relative costs. The cost of the NIRRU is adjusted for the relative intensity of ressource use in Quebec, but remains in US dollars and refers to the total costs for Maryland.

8.1.2 Length and cost of hospital stay

Table 8.1 shows the annual average length of hospital stays for men and women separately by age group. The number of days spent in hospital generally increases with age. A slight decrease compared to previous age groups can be seen in the length of stay for men aged between 90 and 94 years compared to previous age groups and for women above 95 years. Men spend more time in the hospital until the age of 65. Thereafter, the total lenght of annual hospital stays is higher for women.

Table 8.2 shows the average annual cost of hospitalizations. This includes all costs associated with the NIRRU and all physician salary costs. On average, the average annual cost of hospital stays is respectively \$13,256 and \$10,609 for men and women 30 years and older. Similarly to the length of hospitalization, cost increases relatively steadily with age. For women, the average annual cost of hospitalization increases until 89 years, when it reaches \$15,367. It then decreases slightly, which can partly be explained by a decrease in annual hospital stays at advanced ages. For men, the average

Age	Men	Women
30 to 34 years	7.4	3.2
35 to 39 years	6.3	3.8
40 to 44 years	8.3	5.6
45 to 49 years	8.7	6.9
50 to 54 years	10.2	9.2
55 to 59 years	10.7	11.7
60 to 64 years	12.3	11.6
65 to 69 years	12.5	13.3
70 to 74 years	15.6	15.9
75 to 79 years	17.1	18.5
80 to 84 years	19.8	21.8
85 to 89 years	24.1	25.1
90 to 94 years	20.8	25.5
95 years and older	26.3	22.9
Total (30 yrs +)	14.0	12.5

Table 8.1: Cumulative annual length of hospital stays — MED-ECHO (2012)

annual cost increases gradually up to age 79 and fluctuates between $14,122 and $15,404 thereafter. Men aged 95 and older have the highest hospitalization costs, i.e. $15,404 on average per year.

8.2 RAMQ

Thanks to data concerning the medical services billed to the RAMQ, we can capture the cost of the procedures performed by physicians working on fee-for-service schedules or mixed remuneration arrangements.[2] According to Boulenger and Castonguay (2012), this data covers the vast majority of Quebec physicians, with 84% of their total compensation in 2009 taking one of these forms (73% being fee-for-service and 11% mixed).

In addition to the fee charged and the procedure code of the service performed, RAMQ data tells us the date of the act, the identity of the patient and whether the service was performed by a general practitioner or a specialist. Although it can also give the physicians' diagnosis during consultation, it is not mandatory for them to provide the information in their claims. It is therefore not verified by the RAMQ and thus not reliable enough for us to use in the model.

As explained in section 2.2, we model consultations with general practitioners and specialists separately from the number of hospitalization nights. Therefore, we must tell apart the consultations that are part of a hospitalization and those that are not, since the cost of services performed on a hospitalized patient should be included in the cost of hospitalization and not that of a consultation. Those costs for hospitalized patients are added to the cost of the NIRRU as described in section 8.1.

[2] A doctor is paid on a fee-for-service basis if he or she bills the RAMQ for each service rendered. A doctor receives a mixed remuneration if he or she receives a fixed amount per day as well as a supplement for each service rendered.

Age	Men	Women
30 to 34 years	8,503	4,908
35 to 39 years	8,356	5,593
40 to 44 years	9,248	7,072
45 to 49 years	10,407	8,729
50 to 54 years	11,107	9,782
55 to 59 years	12,322	11,849
60 to 64 years	13,572	11,809
65 to 69 years	13,979	12,943
70 to 74 years	14,782	13,155
75 to 79 years	15,272	14,273
80 to 84 years	14,839	14,624
85 to 89 years	15,315	15,367
90 to 94 years	14,122	14,427
95 years and older	15,404	13,823
Total (30 yrs +)	13,256	10,609

Table 8.2: Average annual cost of hospitalization ($) — MED-ECHO (2012) and RAMQ (2012)

We have selected all the services that correspond to a procedure code from the billing manual for general practitioners and the one for specialists (Régie de l'assurance maladie du Québec, 2015a,b). To reconcile RAMQ data with the data gathered from the NPHS question on physician visits, we consider all services performed on the same day and billed by the same doctor for the same patient to make up a single consultation, since the patient saw the doctor only once during that day.[3] Finally, costs are censored at the 99th percentile since some observations have outlying values that could skew the means.

Table 8.3 shows the number of consultations (not associated with a hospitalization) with general practitioners and specialists for men and women separately, by age group. On average, men and women consult a general practitioner around four times over the course of 12 months. The number of consultations with a specialist is slightly higher at around five. It is worth pointing out that the number of consultations (with general practitioners and specialists) increases with age but varies in older ages. This is true for both men and women.

Table 8.4 presents the annual cost of consultations with general practitioners and specialists, for men and women separately and by age group, for services provided to patients who were not hospitalized. The average cumulative annual cost of visits to a general practitioner is $218 for men and $228 for women. The total of annual consultations with a specialist cost on average $451 for men and $399 for women.

The annual cost of consultations with general practioners increases almost constantly with age. The trend by age is not as obvious when looking at the cost of consultations with specialists: the cost

[3] We may be slightly underestimating the number of visits, since some patients could be seeing the same doctor twice in a day and then report having consulted twice. Nonetheless, we believe that this situation only applies to a very small number of individuals.

Age	GP consultations		Specialist consultations	
	Men	Women	Men	Women
30 to 34 years	2.9	3.9	3.5	4.7
35 to 39 years	3.3	3.9	3.8	4.5
40 to 44 years	3.1	3.2	3.6	3.7
45 to 49 years	3.5	3.7	4.2	4.2
50 to 54 years	3.2	3.5	4.0	4.3
55 to 59 years	3.6	4.0	4.8	5.2
60 to 64 years	3.4	3.7	4.8	5.1
65 to 69 years	3.8	4.2	6.0	6.2
70 to 74 years	3.9	4.3	6.2	6.1
75 to 79 years	4.8	5.0	7.3	6.6
80 to 84 years	5.1	5.4	6.9	5.7
85 to 89 years	6.2	6.4	7.0	5.7
90 to 94 years	6.4	6.9	5.4	4.5
95 years and older	8.4	8.2	5.3	4.1
Total (30 yrs +)	3.8	4.1	5.2	5.0

Table 8.3: Annual number of consultations — RAMQ (2012)

Age	GP consultations		Specialist consultations	
	Men	Women	Men	Women
30 to 34 years	156	205	315	336
35 to 39 years	177	202	337	351
40 to 44 years	166	169	315	308
45 to 49 years	189	193	378	327
50 to 54 years	177	187	355	335
55 to 59 years	200	210	428	413
60 to 64 years	191	198	418	400
65 to 69 years	212	223	511	481
70 to 74 years	251	266	525	477
75 to 79 years	300	306	607	534
80 to 84 years	322	336	583	482
85 to 89 years	385	395	599	505
90 to 94 years	400	417	454	402
95 years and older	465	463	475	419
Total (30 yrs +)	218	228	451	399

Table 8.4: Annual cost of consultations ($) — RAMQ (2012)

generally increases between 30 and 79 years, but declines slightly after 89 years. For both men and women, the decrease might come from the drop in the number of consultations at older ages (see Table 8.3).

8.3 Models

We use the models presented in this section to calculate the annual cost of hospitalizations, physician consultations, home care and institutionalization. These models are estimated using data from the RAMQ, MED-ECHO, insurance companies, CIHI and Statistics Canada. The estimated parameters are used to calculate the individual annual cost of health care use.

8.3.1 Hospitalizations

To calculate the cost of hospitalization, we estimate two models: one for men and one for women. Both models are linear, but include a polynomial of degree four for the annual length of hospital stays (LS). The LS for an individual is the sum of all nights (or days) spent in hospital during the year. The specification is as follows:

$$cost\_hospitalization_i = \lambda_1 LS_i + \lambda_2 LS_i^2 + \lambda_3 LS_i^3 + \lambda_4 LS_i^4 + \beta_i \mathbf{x}_i + \epsilon_i, \tag{8.4}$$

where:

- $cost\_hospitalization_i$ is the annual cost of hospital stays for the individual i;
- LS_i is the length of stay (in days) for the individual i;
- LS_i^2 is the length of stay (in days) for the individual i raised to the power of 2;
- LS_i^3 is the length of stay (in days) for the individual i raised to the power of 3;
- LS_i^4 is the length of stay (in days) for the individual i raised to the power of 4;
- $\mathbf{x}_i$ includes all other explanatory variables;
- ϵ_i is a random term specific to individual i.

The explanatory variables are limited because MED-ECHO is an administrative database. We therefore include the presence (or absence) of all diseases (excluding dementias, which are very difficult to capture in the data). We also add a dummy variable for ages 70 years and over to capture a possible effect of age.

Table 8.5 presents the estimation for the annual cost of hospitalizations. The model is estimated only among individuals who had at least one hospital stay in 2012.

The dummy variable for age has a negative effect on the annual cost of hospitalization for both men and women. This suggests that the marginal effect captures a residual effect of age on the cost of hospitalization. Indeed, hospital stay variables and the presence of diseases are likely to capture much

	Men	Women
70 years and more	−2,391.16216***	−1,807.18420***
Stroke	−1,172.86396***	−609.38623*
Diabetes	7.78941	271.91647
Hypertension	178.74447	−173.52830
Lung disease	1,939.62067**	221.98001
Heart disease	1,959.85891***	1,736.42148***
Cancer	816.82019***	1,900.19386***
Length of stay	1,116.83700***	906.97822***
(Length of stay)^2	−11.54489***	−8.10794***
(Length of stay)^3	0.04656***	0.02849***
(Length of stay)^4	−0.00006***	−0.00003***
Number of observations	12,886	17,541
Average cost ($)	13,256	10,609
Legend	* p <0.10 ; ** p <0.05 ; *** p <0.001	

Table 8.5: Average marginal effect of the different variables on the annual cost of hospitalization

of this. One possible explanation is that older people might tend to have longer hospital stays because they do not have a suitable place to go to at their discharge from hospital. These additional nights cost less on average than the nights of younger individuals who owe the entire length of their stay to their health condition.

The presence of a majority of the diseases has a positive effect on the annual cost of hospitalization, except for the presence of a past stroke (men and women) and hypertension (for women only), which are associated with a lower cost. The presence of heart disease has a very important effect on the cost: for men (women), it increases the annual cost of the hospitalization by $1,959 ($1,736). Unsurprisingly, the length of stay is also positively related to the cost of hospitalization.

8.3.2 Consultations

We calculate the cost of consultations using four linear models that include a polynomial of degree four for the number of consultations (NC) in the last 12 months. The first estimates the cost of visits to a general practitioner and the second estimates the cost of consultations with a specialist. The model specification is as follows:

$$cost\_consultation_{i,j} = \lambda_1 NC_i + \lambda_2 NC_i^2 + \lambda_3 NC_i^3 + \lambda_4 NC_i^4 + \epsilon_i,\ \forall j = 1, 2 \tag{8.5}$$

where :

- j is an indicator for whether the cost is for consultation with generalists or specialists;
- $cost\_consultation_{i,j}$ is the annual cost of consultations for the individual i;
- NC_i is the number of consultations for the individual i;

- NC_i^2 is the number of consultations for the individual i raised to the power of 2;
- NC_i^3 is the number of consultations for the individual i raised to the power of 3;
- NC_i^4 is the number of consultations for the individual i raised to the power of 4;
- ϵ_i is a random term specific to the individual i.

We estimate these two models for men and for women separately, for a total of four models. They are estimated among individuals who visited a doctor at least once in the last 12 months. Table 8.6 presents the estimate of the determinants of the annual cost of consultations.

	GP consultations	
	Men	Women
Number of consultations	50.48885***	45.83090***
(Number of consultations)^2	0.56933***	1.11615***
(Number of consultations)^3	−0.01681***	−0.04247***
(Number of consultations)^4	0.00006	0.00034***
Number of observations	92,624	143,107
Average cost ($)	218	228
	Specialist consultations	
	Men	Women
Number of consultations	82.95037***	78.75006***
(Number of consultations)^2	−0.10072	0.33086*
(Number of consultations)^3	0.00189	−0.01997***
(Number of consultations)^4	−0.00007	0.00019***
Number of observations	78,634	127,092
Average cost ($)	451	399
Legend	* p <0.10 ; ** p <0.05 ; *** p <0.001	

Table 8.6: Average marginal effect of each variable on the annual cost of consultations

8.3.3 Long-term care facilities

As presented in Laliberté-Auger et al. (2015), we estimate the cost of long-term care housing by combining many data sources. First, we use total spending (public and private) on long-term care facilities calculated by the Canadian Institute for Health Information (2014). Spending in this category amounted to $5.9 billions for 2011. Then, to obtain a cost "per bed-year", we divide this total spending ($5.9 billion) by the number of people who reported living in a health care and related facility in the 2011 Census (Statistics Canada, 2011). We get, for the 138,760 people residing in one such facility, a unit (annual) cost of $42,784.

This figure needs to be taken with caution for two reasons. First, facility-type definitions differ in the Canadian Institute for Health Information (2014) and in the Census (Statistics Canada, 2011).[4] Secondly, COMPAS is calibrated on a smaller set of facilities that excludes "residences for senior citizens". The cost of these residences being lower than that of nursing homes and long-term care hospitals, the cost we obtain is lower than the actual cost per bed-year estimated by the model. In the end, this figure is still the best approximation possible with the data at our disposal.

8.3.4 Home care services

At this stage of COMPAS development, we do not have data that allow us to accurately calculate the cost per hour of formal home care services. We simply use data from the Canadian Life and Health Insurance Association (sd) and Sun Life Financial (2014) to estimate an average hourly cost. Table 8.7 shows the hourly rate for preparing meals at home, personal care, skilled nursing and for company and monitoring. For the average, we simply calculated the mean of the median hourly costs for each type of care or services.

	CLHIA	Sun Life Financial
In-home meal preparation	\$3.00 to \$25.25	\$2.50 to \$26.50
Personal care (bathing and dressing)	\$12.50 to \$25.25	\$3.00 to \$26.50
Skilled nursing	\$15.00 to \$85.00	\$21.50 to \$68.00
Companionship and supervision	—	\$3.00 to \$26.50
Mean	\$27.63	\$22.19

Table 8.7: Compas expanse

The hourly cost for most services (in-home meal preparation, personal care as well as companionship and supervision) is between \$3 and \$26.50. For their part, nurses receive an hourly wage ranging between \$15 and \$85. On average, home care services cost about \$28 according to CLHIA and \$22 according to Sun Life Financial. We thus assign an hourly cost of \$25 for formal home care services in COMPAS.

8.4 Aggregate costs

In order to obtain aggregate costs for hospitalizations and physicians that are in line with actual government expenditure on those items, we need to rescale the total expenses obtain with our cost models, for a few different reasons. First, as explained in section 8.1, the costs we obtain for hospitalizations are relative costs. Second, the question we use from the NPHS does not ensure we obtain a total number of physician consultations that reflects the number actually charged by doctors. For instance, a radiologist analyzing an x-ray will charge the government, but a survey respondent would surely not count this separately when asked "how many times did you see a doctor?". The assumption underlying

[4]See Appendix A in Laliberté-Auger et al. (2015) for details on the different definitions.

the scaling is that even if the number of consultations is not the same in administrative data and in survey data, the relative use is.

For scaling we use CIHI data on public health expenditures in Canadian Institute for Health Information (2014). Table 8.8 shows the aggregate costs obtained in CIHI and from COMPAS and the ratios used to scale up the aggregate costs from COMPAS. Because CIHI does not separately identify the cost for generalists and specialists, we use total billing for doctors as reported by RAMQ (Régie de l'assurance maladie du Québec (2012)) to distribute the "physicians" item from CIHI.

	CIHI		COMPAS		Ratio
Hospital	9 304.5	66,7%	5 831.0	84.1%	1.6
Generalist	1 785.1	12,8%	744.8	10.7%	2.4
Specialist	2 851.4	20,5%	360.2	5.2%	7.9
Total	13 941.0	100.0%	6 936.0	100.0%	2.0

Table 8.8: Aggregate expenses on hospitals and physicians in 2012 according to CIHI/RAMQ and COMPAS, and scaling ratios

Chapter 9

Uncertainty

Uncertainty in COMPAS is managed with a Monte-Carlo setup. Multiple sources of uncertainty are take into consideration. We take into account the stochastic (first-order) nature of microsimulation as well as parameter uncertainty (Briggs et al., 2012). Individual heterogeneity is accounted for by the integration of individual characteristics in every econometric model used in COMPAS.

9.1 Types of uncertainty

We re-estimate each transition model (as described in chapter 4) 100 times using a bootstrap approach. In order to take into account the joint distribution of parameters both intra-model and inter-model, we draw with replacement from our datasets and re-estimate each transition model, completing this procedure 100 times. This provides a good estimation of the parameter distribution.

For each joint set of parameters, we then run the simulation 50 times to take into account stochastic uncertainty. We retain the mean of these 50 runs of the simulation as the result for each set of parameters. We obtain 100 sets of results that therefore incorporate first-order, second-order, and heterogeneity uncertainty.

We also integrate two other sources of uncertainty. First, we generate 100 initial databases to take into account the random process of imputation of some variables, as mentioned in chapter 3. Second, we also recalculate the joint distribution among each draw of the entering cohorts.

9.2 Presentation of the uncertainty

We present the 5^{th} and 95^{th} percentiles of the different sets of parameters as an outcome's 90% uncertainty (or confidence) interval. We chose to present these because of the asymmetrical distribution of the estimates around the mean outcome.

9.3 Summary

The whole process to account for uncertainty can therefore be summarized by these 5 steps for a particular scenario:

- 100 bootstrap sets of parameters are estimated.
- For each set of parameters, a simulation is run 50 times to account for stochastic uncertainty.
- Each run uses a different initial dataset and different entering cohorts. These vary as a function of the inherent randomness of the imputation processes.
- We obtain 50 outcome values for each of the 100 sets of parameters, for a total of 5,000 values.
- The mean is reported for each outcome, along with the 5^{th} and 95^{th} percentiles as uncertainty (or confidence) interval computed for each set of parameters.

Bibliography

Atella, V., Belotti, F., Cricelli, C., Dankova, D., Kopinska, J., Palma, A., and Mortari, A. P. (n.d.). euFEM, Future Elderly Model. http://http://www.eufem.org/. Accessed: 2019-12-20.

Bohnert, N. (2013). Mortality: Overview, 2008 and 2009. Technical report, Statistics Canada.

Bohnert, N., Chagnon, J., Coulombe, S., Dion, P., and Martel, L. (2015). Population Projection for Canada (2013 to 2063), Provinces and Territories (2013 to 2038): Technical Report on Methodology and Assumptions. Technical report, Statistics Canada.

Boisclair, D., Décarie, Y., Laliberté-Auger, F., Michaud, P.-C., and Vincent, C. (2018). The economic benefits of reducing cardiovascular disease mortality in quebec, canada. *PLOS ONE*, 13(1):1–13.

Boulenger, S. and Castonguay, J. (2012). Portrait de la rémunération des médecins de 2000 à 2009. *Scientific Series*, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) - 2012s-12.

Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J., and Paltiel, A. D. (2012). Model parameter estimation and uncertainty: A report of the ispor-smdm modeling good research practices task force-6. *Value in Health*, 15(6):835 – 842.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373 — 393.

Cameron, A. and Trivedi, P. (2005). *Microeconometrics: Methods and Application*. New York: Cambridge University Press.

Canadian Cancer Society's Advisory Committee on Cancer Statistics (2015). Canadian cancer statistics 2015. *Canadian Cancer Society*.

Canadian Institute for Health Information (2014). National Health Expenditure Trends, 1975 to 2014. Technical report.

Canadian Life and Health Insurance Association (s.d.). A guide to long-term care insurance. Technical report, Canadian Life and Health Insurance Association.

Cancer Research UK (n.d.). Lung cancer survival statistics Read more at http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/survivalrVU6y20MxJ3HGRm1.99.

Clavet, N.-J., Duclos, J.-Y., Fortin, B., and Marchand, S. (2012). Le Québec, 2004-2030 : une analyse de micro-simulation. *Project Report*, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) - 2012RP-16.

Clavet, N.-J., Duclos, J.-Y., Fortin, B., Marchand, S., and Michaud, P.-C. (2013). Les dépenses en santé du gouvernement du Québec, 2013-2030: projections et déterminants. *Scientific Series*, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) - 2013s-45.

Dare, S., F.Mackey, D., and P.Pell, J. (2015). Relationship between smoking and obesity: A cross-sectional study of 499,504 middle-aged adults in the uk general population. *PLoS One*, 10(4).

Décarie, Y., Boissonneault, M., and Légaré, J. (2012). An inventory of Canadian microsimulation models. SEDAP Research Paper No. 298.

Fitzgerald, J., Gottschalk, P., and Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan panel study of income dynamics. *The Journal of Human Resources*, 33(2):251–299.

Frees, E., Gao, J., and Rosenberg, M. (2011). Predicting the frequency and amount of health care expenditures. *North American Actuarial Journal*, 15(3):377–382.

Godbout, L., St-Cerny, S., Arseneau, M., Dao, N. H., and Fortin, P. (2014). La soutenabilité budgétaire des finances publiques du gouvernement du Québec. Working document 2014/01 of the Chaire de recherche en fiscalité et en finances publiques de l'Université de Sherbrooke.

Goldman, D., Cutler, D., Rowe, J., Michaud, P.-C., Sullivan, J., Peneva, D., and Olshansky, S. (2013). Substantial health and economic returns from delayed aging may warrant a new focus for medical research. *Health Affairs*, 32(10):pp.1698–1705.

Goldman, D., Shekelle, P., Bhattacharya, J., Hurd, M., Joyce, G., Lakdwalla, D., Matsui, D., Newberry, S., Panis, C., and Shang, B. (2005). Health status and medical treatment of the future elderly. Technical report, RAND Corporation.

Health Canada (2012). Smoking Prevalence 1999 — 2012. Technical report.

Human Mortality Database (2005). Complete data series. Technical report.

Kapteyn, A., Michaud, P., Smith, J., and van Soest, A. (2006). Effects of attrition and non-response in the health and retirement study. *RAND Working Paper*, pages 1–41.

Laliberté-Auger, F., Côté-Sergent, A., Décarie, Y., Duclos, J.-Y., and Michaud, P.-C. (2015). Utilisation et coût de l'hébergement avec soins de longue durée au québec, 2010 à 2050. *Assurances et gestion des risques*, 82(3-4):23 – 41.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87(419):pp. 659–671.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, volume 42:pages 575 to 594.

Li, N., Lee, R., and Gerland, P. (2013). Extending the lee-carter method to model the rotation of age patterns of mortality-decline for long-term projection. *Demography*, Volume 50:pages 2,037 to 2,051.

Macdonald, A. (1997). *The Second Actuarial Study of Mortality in Europe: The Companion Disk.* Groupe consultatif des associations d'actuaires des pays des Communautés Européennes.

Martel, L. (2013). Mortality: Overview, 2010 and 2011. Technical Report Catalogue no. 91-209-X, Statistics Canada.

McDonald, A., Cairns, A., Gwilt, P., and Miller, K. (1998). An international comparison of recent trends in mortality. *British Actuarial Journal*, 4:3–141.

Michaud, P.-C., Kapteyn, A., Smith, J. P., and Van Soest, A. (2011). Temporary and permanent unit non-response in follow-up interviews of the health and retirement study. *Longitudinal and Life Course Studies*, 2:145–169.

Michaud, P.-C., Lakdawalla, D., Goldman, D., Zheng, Y., and Gailey, A. (2012). The value of medical and pharmaceutical interventions for reducing obesity. *Journal of Health Economics*, 21(4):pp.630–640.

Ministère de la Santé et des Services sociaux (2014). Banque de données dérivée : APR-DRG (J57) version 24.0. Technical report, Ministère de la Santé et des Services sociaux.

National Research Council (US) Committee on National Statistics (2010). Improving health care cost projections for the Medicare population: Summary of a workshop. *National Academies Press (US).*

Régie de l'assurance maladie du Québec (2012). Table sm.22. Technical report, Régie de l'assurance maladie du Québec.

Régie de l'assurance maladie du Québec (2015a). Manuel des médecins omnipraticiens (no 100). Technical report, Régie de l'assurance maladie du Québec.

Régie de l'assurance maladie du Québec (2015b). Manuel des médecins spécialistes (no 150). Technical report, Régie de l'assurance maladie du Québec.

Renshaw, A., Haberman, S., and Hatzopoulos, P. (1996). The modelling of recent mortality trends in United Kingdom male assured lives. *British Actuarial Journal*, 2:449–477.

Statistics Canada (2000). Canadian Community Health Survey, 2000, Wave 1.1, Annual component. Data sets.

Statistics Canada (2006). 2006 Census of Population, Statistics Canada catalogue no. 97-560-xcb2006031. Technical report, Statistics Canada.

Statistics Canada (2008). Canadian Community Health Survey - healthy aging. Data files, Statistics Canada.

Statistics Canada (2010). Canadian Community Health Survey, 2010, Wave 5.1, Annual component. Data sets, Statistics Canada.

Statistics Canada (2011). 2011 Census of Population, Statistics Canada catalogue no. 98-313-xcb2011024. Technical report, Statistics Canada.

Statistics Canada (2012a). Canadian Community Health Survey, 2012, Annual component. Data sets.

Statistics Canada (2012b). National Population Health Survey—Household component — longitudinal (NPHS). Technical report.

Statistics Canada (2012c). National Population Health Survey, 1994-2010, waves 1-9, Household component- longitudinal. Data sets.

Statistics Canada (2015). Population projections for Canada (2013 to 2063), provinces and territories (2013 to 2038). Technical report, Statistics Canada, National Population Projections team.

Statistics Canada (2016a). 2016 Census of Population, Statistics Canada catalogue no. 98-400-x2016243. Technical report, Statistics Canada.

Statistics Canada (2016b). Table 105-0501 - Health indicator profile, annual estimates, by age group and sex, Canada, provinces, territories, health regions (2013 boundaries) and peer groups, CANSIM (database). Technical report, Statistics Canada, Canadian Community Health Survey (CCHS).

Statistics Canada (2018). Life tables, Canada, provinces and territories 2009 to 2011. Technical report, Statistics Canada.

Statistics Canada (n.d.a). Estimates of population, by age group and sex for July 1, Canada, provinces and territories annual (persons unless otherwise noted). *CANSIM (database). Last updated September 29, 2015.*

Statistics Canada (n.d.b). Table 17-10-0014-01 : Estimates of the components of international migration, by age and sex, annual. DOI: https://doi.org/10.25318/1710001401-eng (Accessed: 2019-12-20).

Statistics Canada (n.d.c). Table 17-10-0015-01 : Estimates of the components of interprovincial migration, by age and sex, annual. DOI: https://doi.org/10.25318/1710001501-eng (Accessed: 2019-12-20).

Statistics Canada (n.d.d). Table 17-10-0057-01 : Projected population, by projection scenario, age and sex, as of july 1 (x 1,000). DOI: https://doi.org/10.25318/1710005701-eng (Accessed: 2019-12-20).

Sueyoshi, G. T. (1995). A class of binary response models for grouped duration data. *Journal of Applied Econometrics*, Vol. 10, 1995,(No. 4):pp. 411–431.

Sun Life Financial (2014). Long term care in Quebec - 2014. Technical report, Sun Life Financial.

World Bank (2019). Life expectancy at birth, total (years) - Canada. Data sets, World Bank.

World Health Organization (2006). BMI classification. Technical report, United Nations.

World Health Organization (2015). World health statistics 2015 - indicator compendium.