Chaire de recherche
sur les enjeux économiques
intergénérationnels

23-01

Evaluating the relationship between income, survival and loss of autonomy among older Canadians

Marie Connolly, Akakpo Domefa Konou and Marie-Louise Leroux

Cahier de recherche
Working paper
Février / February 2023

ESG UQÀM

Chaire de recherche
sur les enjeux économiques
intergénérationnels

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

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

Retraite Québec

Les opinions et analyses contenues dans les cahiers de recherche de la Chaire ne peuvent en aucun cas être attribuées aux partenaires ni à la Chaire elle-même et elles n'engagent que leurs auteurs.

Opinions and analyses contained in the Chair's working papers cannot be attributed to the Chair or its partners and are the sole responsibility of the authors.

© 2023 Marie Connolly, Akakpo Domefa and Marie-Louise Leroux. Tous droits réservés. All rights reserved. Reproduction partielle permise avec citation du document source, incluant la notice ©. Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.

Dépôt légal : Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada, 2023. ISSN 2563-0008

Evaluating the relationship between income, survival and loss of autonomy among older Canadians*

Marie Connolly[†], Akakpo Domefa Konou[‡], Marie-Louise Leroux[§]

February 9, 2023

Abstract

Evaluating the relationship between health at old age and income is crucial for the design of equitable public policies targeted toward the elderly. Using 2016 Canadian survey data on adults aged between 50 and 70, we estimate the relationships between individual income, longevity and dependency at the old age. We use both subjective and objective measures of the probability to survive to age 85, of the probability to have activities of daily living (ADL) limitations, and of the probability of entering a nursing home. We find that income and the (objective and subjective) probability to live to age 85 and over are positively related while income and the (objective and subjective) probability to suffer from ADL limitations are negatively related. We also find that while the objective probability to enter a nursing home is negatively correlated with income, the subjective probability is positively correlated with income. Most of our results are driven by individuals in the highest tercile of the income distribution. Our results are robust to different sensitivity checks.

Keywords: Long term care, Survival probability, Probability to become dependent, Nursing home, Income.

JEL Codes: C36, I10,I14, I18, J14.

*The authors thank participants at the annual meeting of the Société canadienne de science économique (Montréal, 2022) as well as P-C Michaud for their useful comments. M-L Leroux also acknowledges financial support from Fonds de Recherche du Québec - Société et Culture (Grant number: 2019-SE2-252890), and from Social Sciences and Humanities Research Council (Grant number: 435-2020-0787).

[†]Département des sciences économiques, ESG UQAM (Montréal, Canada) and CIRANO (Montréal). Email: connolly.marie@uqam.ca

[‡]Département des sciences économiques, ESG UQAM. Email: konouakakpo@gmail.com

[§]Département des sciences économiques, ESG UQAM (Montréal, Canada), CESifo (Munich), CORE (Louvain-la-Neuve) and CIRANO (Montréal). E-mail: leroux.marie-louise@uqam.ca

1 Introduction

In all developed countries, population is aging. According to the OECD (2021), between 2019 and 2050, the share of the Organisation for Economic Co-operation and Development (OECD) population aged 80 and over is expected to more than double, from 4.6% to 9.8%. Hence people will live longer, but this does not tell whether they will live longer in *good health*. In OECD countries, 50% of the 65 years and older report some limitations in their daily-life activities. By 2050, the number of people with dementia is expected to double to 42 millions (OECD, 2021). Under that scenario, long term care (LTC) will have to be provided at large and this is certainly going to be costly for both individuals and public administrations. LTC is defined as "the day-to-day help with activities such as washing and dressing, or help with household activities such as cleaning and cooking"(OECD, 2011). LTC often comes with additional types of support such as medical assistance. The Canadian population, which we study in this paper, is not exempt from this trend. Clavet et al. (2021) estimate that the number of individuals needing help with activities of daily living (hereafter ADL) in Québec is likely to almost double between 2020 and 2050, from 315,000 to more than 600,000.

It is now quite well documented (see for instance, Cristia (2009), Cutler et al. (2006), Chetty et al. (2016), Delavande and Rohwedder (2011), Kalwij et al. (2013), Lacroix et al. (2021), Milligan and Schirle (2021)) that even though overall longevity has increased in the last decades in all developed countries, there still remains important heterogeneity in longevity across different socioeconomic groups (for example as split along age, ethnicity, gender, income or education lines). Abstracting from any causal relationship, we also know that there is a positive relationship between income (or wealth) and longevity. For instance, Milligan and Schirle (2021) show that, in Canada, among cohorts born between 1923 and 1955, there exists a clear gradient between income and longevity for men: "men in the top 5% of the earnings distribution lived eight years longer than those in the bottom 5% after age 50, a difference of 11% of a lifespan." For women, this gradient is more modest, with the difference in longevity of only 3.6 years.[1]

[1] Boisclair et al. (2015) also show differences in mortality rates across different education levels in the Québec population aged 25 and above, in 2009.

Chetty et al. (2016) find similar results for the US population.

In contrast, the relationship between income (or wealth) and the risk of becoming dependent and of needing LTC assistance is less clear, and the literature on the subject is still quite scarce. For instance, Cambois et al. (2011) study the relationship between occupational status and life expectancy in good health for individuals aged 50 and above. They show a clear disadvantage for manual workers, who have shorter life expectancy and more years with poor health and disabilities. Liu and Wang (2022) find similar evidence that socioeconomic status (as measured by income and education) is correlated with disability of Chinese individuals over 60 years old. Using data from England and the US, Zaninotto et al. (2020) show that people in the bottom wealth tercile can expect to live seven to nine fewer years without disability (that is in terms of impaired activities and instrumental activities of daily living) than those in the highest wealth tercile at the age of 50. Finally, the paper most closely related to ours is Lefebvre et al. (2018), which studies, using data from the Survey of Health, Ageing and Retirement in Europe (SHARE), the relationship between wealth, the subjective survival probability and the probability to become dependent among Europeans aged 50 and above. They find that individuals from the low end of the wealth distribution have a higher probability of becoming dependent and also, that they remain dependent longer. We come back below on how different our paper is from this one.

More empirical evidence on the directions of these relationships is needed to obtain reliable theoretical results and formulate informed policy recommendations. Indeed, from a public policy perspective, it is crucial to determine which individuals are the most likely to suffer from dependency. Do they come from the top or the bottom of the wealth and income distributions? Since LTC provision is expensive, low-income individuals could suffer from a "double penalty" if they are also more likely to become dependent: they would have both more limited means and a higher dependency risk, which would overstretch their financial resources if the risk did materialize and they had to pay for LTC on their own. In this case in particular, public LTC provision may be desirable so as to avoid such a double penalty. Conversely, if dependency concerns primarily wealthy individuals, should the government step in? Or should it let them pay for their loss of autonomy (either out of private savings or through private insurance)? Obviously, whether the government should or should not intervene very much depends on the signs of the correlations between income (or wealth), survival, and the individual risk to become dependent. In order to further convince the reader, we develop in Section 2, a simple model that shows how these correlations may impact individuals' financial planning and the design of optimal intra-generational redistribution policies.[2]

The objective of this paper is then to estimate the relationship between longevity and income, and between dependency and income. To do so, we use data from a survey conducted by Boyer et al. (2020a) in 2016 among 2,000 Canadian respondents aged between 50 and 69 years old, which aimed at understanding better the LTC insurance puzzle (i.e., the fact that agents do not buy private LTC insurance even though the LTC risk is important and the associated financial losses are substantial) in the Canadian context (Institut sur la retraite et l'épargne / Retirement and Savings Institute, 2020). Using this survey data, Boyer et al. (2020a) also ran a health microsimulation model called COMPAS to predict personalized lifetime exposure to death, disability, and nursing home care.[3] Hence, in our study, we will use both subjective (as reported by the survey respondents) and objective (as obtained from COMPAS) measures of longevity and of dependency. Longevity will be approximated by the probability to be alive at 85 years old. Dependency will be assessed using two different measures. The first one is the probability to have at least two ADL limitations at some point in life. The second one is the probability of ever entering a nursing home. To measure financial resources, we will use both reported income and savings.

Using a simple ordinary least squares (OLS) methodology, we obtain the following results. We find that the relationships between objective probabilities and income (whether expressed in log or split in terciles) are always statistically significant. In particular, a higher income is always associated with a higher (objective) probability to be alive at 85 years old and a lower (objective)

[2]Theoretical works such as Leroux et al. (2021), Leroux and Pestieau (2021), and Leroux and Pestieau (2022) also need to know the signs of the correlations between income, survival and dependency at the old age in order to make policy recommendations regarding optimal LTC policies and the taxation of bequests.

[3]More details about COMPAS can be found in Boisclair et al. (2016) and in Section 3.

probability to have ADL limitations, as well as to ever enter a nursing home. Most of these relationships appear to be driven by people at the top of the income distribution (i.e. in the highest income tercile). Regarding the relationships with subjective probabilities, the evidence is more mixed. Being in the second and third income terciles (as opposed to the first) is associated with a higher subjective probability to be alive at 85. Also, a higher income reduces the subjective probability to have ADL limitations (at a 10% significance level). More surprisingly, income is positively associated with the subjective probability to ever enter a nursing home. As we explain below, this may be because respondents have considered other influencing factors (such as their financial situation and their family situation) when answering that question. In a last section, we show that overall, our results are robust to using an instrumental variable strategy instead, and to using a restricted sample (with only non-imputed income variables). We also show that our estimations are very robust to using individual savings instead of household income as the main explanatory variable. We find the same relationships, yet of smaller magnitude.

The present paper can be related to the vast literature on the relationship between income and longevity (see the articles mentioned above) as well as to the literature studying the relationship between the socioeconomic status (including occupation, education, income) and health at different ages (see among others, Cutler et al. (2011), Ettner (1996) and Michaud and van Soest (2008)). Contrary to this literature, the originality of this paper is to concentrate on health *at older ages* and in particular, on the risk to ever become dependent. Unlike many other health conditions, dependency is an absorbing state, is more likely to happen at older ages, and can exhaust individual wealth. Our paper then complements the earlier studies of Lefebvre et al. (2018), Cambois et al. (2011), Liu and Wang (2022) and Zaninotto et al. (2020) in at least three respects. First, we use original Canadian survey data. Second, we estimate the risk of becoming dependent through two different and complementary measures. Third, we compare the results obtained using either objective or subjective estimations of individual survival and dependency probabilities.

The paper is structured as follows. In the next section, we present a theoretical model justifying the importance of our empirical question for formulating public policy recommendations.

In Section 3, we present our data and some descriptive statistics. In Section 4, we provide descriptive evidence of the relationships between income, survival, and the probabilities of having ADL limitations and of entering a nursing home, and we provide our main empirical results. Section 5 shows some robustness checks. The last section concludes.

2 Theoretical model

The following model aims to show how important it is for both individual decisions and the design of optimal redistributive policies to know how income (or wealth) correlates with survival and the risks to become dependent in old age. We develop a model of intertemporal financial planning in which we assume away any efficiency issue and concentrate exclusively on redistributive ones.[4]

To do so, we consider a two-period economy. Period 1 is adulthood (say between the ages of 50 and 70), during which individuals are autonomous, work, and decide how to allocate their wealth between consumption and saving for their old days. Without loss of generality, we assume fixed labour. Period 2 is old adulthood (70 years old and beyond). That period is reached with a probability $0 < \pi \leq 1$. During the old age, individuals do not work anymore and consume their savings. Moreover, individuals reaching old age become dependent with a probability $0 \leq p \leq 1$, in which case they would bear additional exogenous LTC expenditures.

We assume that individuals, indexed by i, differ on three dimensions: income, y_i; survival probability, π_i; and the probability to become dependent in old age, p_i. For simplicity, we assume two income levels: type-H individuals have a high income, whereas type-L individuals have a low income, i.e. $y_H > y_L$. We assume that types are equally distributed in the society and that the size of each group is normalized to 1.

Following the literature on the health-wealth gradient, we assume that individuals with a higher income have also a higher probability to reach the old age, that is $\pi_H > \pi_L$.[5] For the moment, we make no assumption regarding the ranking between type H and type L, of the unconditional probability of old-age dependency, $\pi_i p_i$.

[4] Assuming inefficiencies would give additional reasons for the government to intervene. Differences in sociodemographic characteristics may then also shape the optimal individualized taxation scheme.

[5] See the papers cited in the Introduction.

Individuals preferences are represented by a standard expected utility function. Lifetime welfare is time-additive, and the utility of being dead is normalized to 0. We also assume that temporal utility is state-dependent, in the sense that the transformation of resources into welfare is not the same when the individual is autonomous and when he is dependent. Assuming no time preferences, the expected lifetime utility for an individual of type $i \in \{L, H\}$ is:

$$EU_i = u(c_i) + \pi_i(1 - p_i)u(d_i) + \pi_i p_i v(m_i)$$

where c_i is consumption in period 1, d_i is period 2 (old-age) consumption in case of autonomy, while m_i denotes old-age consumption in case of dependency.

The utility obtained from consumption under good health is such that $u'(\cdot) > 0$, $u''(\cdot) \leq 0$. Also, we assume that the individual utility under old-age dependency is of the following form:[6]

$$v(x) = u(x - Z)$$

where Z is the amount of LTC expenditures the dependent elderly has to bear. This implies that $v'(x) = u'(x - Z) > u'(x)\,\forall x$: marginal utility of an additional dollar is always higher under bad health than under good health. This ensures that individuals are willing to insure against the LTC risk.

2.1 The laissez-faire equilibrium

Let us first characterize the equilibrium in the absence of governmental intervention. We assume a perfect annuity market, yielding actuarially fair returns. Together with the assumption that the interest rate is zero, the return on savings for an individual of type i in case of survival to period 2 is thus $1/\pi_i$. We also assume an actuarially fair private LTC insurance market (yielding a return $1/\pi_i p_i$).[7]

[6]This assumption is common in models dealing with old-age dependency. See among others Cremer and Pestieau (2014); Donder and Leroux (2014); Donder and Pestieau (2017); Klimavicuite and Pestieau (2018); Leroux and Pestieau (2022).

[7]Assuming away the possibility of annuitization and of LTC insurance would lead to modeling accidental bequests. This would complicate our model above what is needed to show that knowing about the correlation between income and demographic characteristics is crucial for understanding individual decisions and making informed policy recommendations.

In the first period, individuals choose savings $s_i \geq 0$ and private LTC insurance $a_i \geq 0$ in order to maximize their expected lifetime welfare subject to their budget constraint:

$$\max_{s_i, a_i} u\left(y_i - s_i - a_i\right) + \pi_i(1-p_i)u\left(\frac{s_i}{\pi_i}\right) + \pi_i p_i v\left(\frac{s_i}{\pi_i} + \frac{a_i}{\pi_i p_i}\right).$$

Solving this problem, we obtain the laissez-faire trade-offs between consumption across time and health status:

$$u'(c_i) = u'\left(d_i\right) = v'\left(m_i\right)$$

implying that for a given type-i agent, $c_i = d_i = m_i - Z$.

Together with the individual's per period budget constraints, this in turn yields:

$$\begin{aligned} a_i^{LF} &= \pi_i p_i Z \\ s_i^{LF} &= \frac{\pi_i}{1+\pi_i}(y_i - \pi_i p_i Z) \end{aligned}$$

The first condition shows that the individual chooses to fully insure against the LTC risk (which is a direct consequence of assuming actuarially fair insurance contracts) and that the higher the unconditional probability to become dependent ($\pi_i p_i$), the higher the amount of insurance bought. Income y_i does not influence the level of LTC insurance in this simple model.[8]

Regarding the level of savings, the higher the agent's income, the higher the amount they wish to transfer to the second period through higher savings. Also, the higher the survival probability, the higher is s_i^{LF}; this is related to the fact that they have more chances of needing resources in the second period. Finally, the smaller the unconditional probability to become dependent, $\pi_i p_i$, the higher is s_i^{LF}; the underlying mechanism goes through LTC insurance: an individual with lower chances to become dependent needs to buy less insurance ($a_i^{LF} = \pi_i p_i Z$), which leaves him more resources to allocate across periods.

Coming back to our assumptions that $y_H > y_L$ and $\pi_H > \pi_L$, we should then obtain that $s_H^{LF} > s_L^{LF}$. However, whether the savings of type-H agents is greater than those of type-L agents is ambiguous since the correlations between income and the unconditional probability to become

[8]With a more complete modeling (assuming for instance different types of goods consumed in different quantities under good health and bad health), this may however not be the case anymore.

dependent is less documented in the literature (establishing the signs of these relationships will be the purpose of our next sections). Hence, if we assume, as in Lefebvre et al. (2018), that lower-income agents have a higher probability to become dependent, i.e. $\pi_L p_L > \pi_H p_H$, $a_L^{LF} > a_H^{LF}$ and this reinforces the ranking of savings, $s_H^{LF} > s_L^{LF}$. In that situation, differences in income and in demographic characteristics both point in the same direction of lower savings for type-L individuals than for type-H. Instead, if we assume that $\pi_H p_H > \pi_L p_L$, we obtain that $a_H^{LF} > a_L^{LF}$ and this would go in the opposite direction of the effect of higher income and higher survival in the ranking of savings. At the extreme, we could even observe a reverse ranking of savings, i.e. $s_L^{LF} > s_H^{LF}$.[9] In this latter case, the size of the differences in sociodemographic characteristics across types matters in the ranking of savings.

2.2 The utilitarian optimum

We now derive the optimal allocation under perfect information. To do so, we rely on the standard utilitarian social welfare criterion, which consists in summing up the individuals' utility subject to the economy resource constraint. The problem of the utilitarian social planner then writes:

$$\begin{array}{cl} \max\limits_{c_i,d_i,m_i} & \sum n_i \left[u\left(c_i\right) + \pi_i(1-p_i)u\left(d_i\right) + \pi_i p_i v\left(m_i\right)\right] \\ \text{s.t.} & \sum n_i y_i = \sum n_i \left(c_i + \pi_i(1-p_i)d_i + \pi_i p_i m_i\right) \end{array}$$

Rearranging the first-order conditions of this problem, it can easily be shown that consumption should now not only be smoothed across periods (like in the laissez-faire) but also *between types*: $c_i = d_i = m_i - Z = \bar{c} \, \forall i$.

The decentralization of the first-best optimum is standard and requires only lump sum taxes and transfers between individuals with different types so as to ensure equal consumption $\bar{c}$. In a purely redistributive system with only two types, the government budget constraint implies that $T_L = -T_H$ where T_i is the lump sum tax faced by individual i and which can be positive

[9] This could be the case if differences in income and in survival are small relative to differences in the probability to become dependent.

or negative:

$$T_i = y_i - \bar{c}(1+\pi_i) - \pi_i p_i Z.$$

In order to understand the direction of redistribution, we proceed by step and assume that agents differ in only one dimension:

- If agents only differ in income with $y_H > y_L$, we obtain that $T_H > 0 > T_L$ and high-income agents should pay taxes that will be redistributed to low-income agents.
- If agents only differ in survival with $\pi_H > \pi_L$, we obtain the reverse: $T_L > 0 > T_H$. This somewhat counterintuitive result is a consequence of the utilitarian criterion: agents with higher chances to survive should be allocated more resources.
- Finally, if agents differ only in the unconditional probability to become dependent with $\pi_L p_L > \pi_H p_H$ (resp. $\pi_H p_H > \pi_L p_L$), then $T_H > 0 > T_L$ (resp. $T_L > 0 > T_H$). The intuition is similar to our previous point: more resources should be allocated to these agents who have a higher probability to become dependent and face higher risks to incur LTC expenses.

It is then clear that the direction of redistribution depends on the relationship between income, survival and the probability to become dependent. Since the different sources of heterogeneity may push the direction of redistribution in opposite ways, it is also important to look at the magnitude of the differences in these sociodemographic characteristics in order to get a clear ranking of the transfers to be made.

All in all, this simple model showed that it is crucial to have information about the correlation between income and survival, and between income and the probability to become dependent, in order to understand individuals' life-cycle decisions as well as in order to make accurate policy recommendations.[10] While the relationship between income and survival is already quite well documented in the literature, the relationship between income and the risk of becoming dependent has been less studied empirically. This is what we aim to do in the following sections.

[10]On this, see footnote 2.

3 Data

3.1 Data sources

We use data from a survey called "Protection against financial risks in retirement: economic analysis of the risk of dependency" (Institut sur la retraite et l'épargne / Retirement and Savings Institute, 2020) conducted by Boyer et al. (2020a).[11] This survey was deployed by the firm Delvinia, a Canadian online survey organization, and conducted on a sample of 2,000 respondents who were part of the web panel "Asking Canadians/What do you think?". Participants were aged between 50 and 70, and residing in Ontario or Québec (the two most populous provinces in Canada). At the end of the survey, participants were rewarded for their participation with loyalty rewards from major Canadian retailers.

A copy of the questionnaire is provided in Appendix C. Information about the socioeconomic characteristics of the respondents, their health condition, their perceived survival probability, and their perceived probability to become dependant was collected.[12] However, some cases of non-response occurred, creating some under-representation of certain groups, in particular, the low-education and low-income groups. Consequently, in order to make the survey data representative of the population aged 50 to 70 in Québec and Ontario, sampling weights, constructed using the 2014 Canadian Labour Force Survey, are provided with the data and used in all our computations (see Boyer et al. (2020a) for more details). For questions where a significant fraction of missing information was expected, such as questions on savings and income, the survey allowed for unfolding brackets. To deal with missing responses, multiple imputation was done to impute missing values with information from the bracketing, conditional on basic sociodemographic covariates.[13] Boyer et al. (2020a) also fed RSI-01 respondents into a health microsimulation model called COMPAS, which is a dynamic microsimulation model that

[11] Data are available through the Retirement and Savings Institute (RSI) of HEC Montréal at: `https://borealisdata.ca/dataset.xhtml?persistentId=doi:10.5683/SP2/PP5U7Y`.

[12] Additional questions (which are not used in the present paper) were also asked, such as whether respondents had LTC insurance, about their financial literacy, their knowledge of LTC public institutions, etc. The last part of the survey (Section 5, removed from the questionnaire in AppendixC) consisted of a stated-preference experiment, which was used in particular in Boyer et al. (2019b, 2020a).

[13] In Section 5, we show that we obtain very similar results when we restrict the sample to observations for which income was reported and exclude any imputed value.

was developed to estimate the long-term evolution of health and health care use in Canada.[14] COMPAS takes the respondents' sociodemograhic characteristics and health factors to obtain as outputs the individuals' *objective* probabilities of living to age 85 and their *objective* probabilities of ever becoming dependent.[15]

In the following, we will use both the *objective* and the *subjective* probabilities of survival to age 85, of ADL limitations, and of moving to a nursing home as the main explained variables. Objective probabilities refer to the probabilities generated through the COMPAS model we just described. Subjective probabilities are obtained from the answers given by the respondents to survey questions Q18 to Q20 (see Appendix C), which we report below:

- Q18: On a scale of 0 to 100, where 0 is absolutely no chance and 100 is absolutely certain, what do you believe is the percent chance you will live to age 85 or more?

- Q19: On a scale of 0 to 100, where 0 is absolutely no chance and 100 is absolutely certain, what do you believe is the percent chance you will live more than 1 year during your lifetime with two or more limitations in activities of daily living? Activities of daily living include eating, bathing, getting dressed, walking about one's home and getting in and out of bed.

- Q20: Of course nobody wishes to go to a long-term care home, but sometimes this becomes necessary. On a scale of 0 to 100, what do you believe is the percent chance that you will have to move to a long-term care home because of important limitations in your activities of daily living?

Note that because of the framing of the questions, Q19 and Q20 allow to obtain the *unconditional* probability to have two ADL limitations or more, or to move to a nursing home. To make a link with the previous theoretical section, it then provides an estimation of $\pi_i p_i$ for each respondent i.

[14]For more details about COMPAS, see Boisclair et al. (2016) and Boyer et al. (2019a, 2020a).

[15]The RSI-01 and COMPAS data are available through Boyer et al. (2020b) at `https://doi.org/10.3886/E115322V1`.

We use total annual household income (Question 11 in the survey) as our main explanatory variable. It includes imputations for missing values, and we restrict the sample to observations with total annual household income between \$1,000 and \$2,000,000, thus removing outliers that would affect our estimates. This removes 42 observations, most of which (38) come from responses of total income under \$1,000. As a robustness check in the last section, we will also use individual savings (Question 16 in Appendix C) instead of income.

3.2 Descriptive statistics

We start with some descriptive statistics of our main dependent variables. Table 1 reports the summary statistics about both objective and subjective probabilities. Note that there were some non responses in the subjective probabilities, as respondents could answer "Don't know" if they were unable to provide an answer, or did not want to provide one. This resulted in missing values and in a smaller number of responses for the subjective probabilities compared to the full-sample objective ones. We thus also present statistics on the objective probabilities for a restricted sample in which the subjective probabilities are non missing.

Interestingly, it appears that respondents are better able to formulate an answer regarding their survival probability than their probability to become dependent, since 1,624 respondents did provide a probability to survive to the age 85 while only 1,271 and 1,322 respondents could, respectively, give an estimation of the probability of having ADL limitations or of entering a nursing home. On average, respondents are optimistic since the average objective survival probability is below the subjective one (63.3% and 67.8% respectively) and the reverse is true for the probability to have ADL limitations (55.8% -on the restricted sample- and 47.9%, respectively). To the contrary, respondents are, on average, pessimistic regarding their probability to enter a nursing home since the average subjective probability is higher than the objective one (35.4% against 26.3% -on the restricted sample-).[16]

Let us now present the main explanatory variables. Income was asked at the level of the household, including all potential sources of income, before taxes and deductions. Looking at

[16]For a complete analysis of the correlations between objective and subjective probabilities, and between survival probabilities and probabilities to become dependent, see Boyer et al. (2019a).

Table 1: Descriptive statistics, dependent variables

Variable	Mean	SD	Min	Max	N
Probability of survival					
Objective (full sample)	0.633	0.136	0.063	0.875	1,958
Objective (restricted sample)	0.633	0.137	0.063	0.875	1,624
Subjective	0.678	0.286	0.000	1.000	1,624
Probability of ADL limitations					
Objective (full sample)	0.565	0.072	0.34	0.767	1,958
Objective (restricted sample)	0.558	0.070	0.34	0.767	1,271
Subjective	0.479	0.337	0.000	1.000	1,271
Probability of nursing home					
Objective (full sample)	0.271	0.114	0.055	0.522	1,958
Objective (restricted sample)	0.263	0.113	0.055	0.522	1,322
Subjective	0.354	0.301	0.000	1.000	1,322

Source: Authors' calculations based on the RSI-01 data.
Note: Includes only observations with total annual household income between 1,000$ and 2,000,000$. Objective probabilities are taken from COMPAS model; subjective probabilities from survey responses. "Restricted sample" means the observations are restricted to those for which the subjective probability is not missing.

Table 2, we find that the median income in our sample is around 75,000$, with a median higher in Ontario (80,000$) than in Québec (65,000$). Savings are measured at the individual level, and include Registered Retirements Saving Plans (RRSPs), which are privately-owned individual pension plans, Tax-Free Savings Accounts (TFSAs), and any other type of savings. The median savings level is of 75,000$ in our sample, with a higher median in Ontario (100,000$) than in Québec (53,000$).

Finally, Table 3 presents summary statistics of individual sociodemographic characteristics, which are used as controls in our regressions. None of the statistics presented in that table are surprising. Almost two thirds of respondents (62.11%) are from Ontario, and 37.89% live in Québec. A majority (72.01%) of individuals are either married (with a higher proportion in Ontario than in Québec) or common-law partners (with a higher proportion in Québec than in Ontario), and 73.53% of our sample has children. The last two lines of Table 3 present information on additional variables that we include in our regressions. First, a majority of respondents (54.72%) declare that they do not contribute to an employer pension plan nor receive benefits from one. The last question relates to the individual degree of risk aversion.

Table 2: Descriptive statistics, income and savings (CAD)

Income	Tercile	Mean	SD	Median	Min	Max
Québec	First	30,904	13,809	30,000	1,800	50,435
	Second	77,007	15,609	78,670	50,442	100,000
	Third	172,638	104,207	150,000	100,231	1,000,000
	All	80,870	74,750	65,000	1,800	1,000,000
Ontario	First	31,761	12,954	34,000	1,509	50,355
	Second	77,160	14,506	75,078	50,573	100,000
	Third	190,206	144,110	160,000	100,539	1,800,000
	All	102,424	107,641	80,000	1,509	1,800,000
Both provinces	First	31,378	13,339	31,672	1,509	50,435
	Second	77,101	14,935	76,439	50,442	100,000
	Third	185,100	133,888	150,000	100,231	1,800,000
	All	94,258	97,048	75,000	1,509	1,800,000
Savings	Tercile	Mean	SD	Median	Min	Max
Québec	First	7,077	9,794	25	0	30,000
	Second	80,660	35,142	75,000	33,000	150,000
	Third	650,302	4,229,901	300,000	160,000	90,000,000
	All	201,566	2,173,217	53,000	0	90,000,000
Ontario	First	7,587	9,806	500	0	31,000
	Second	81,558	35,623	75,000	33,000	150,000
	Third	534,823	831,716	350,000	160,000	15,000,000
	All	225,666	558,126	100,000	0	15,000,000
Both provinces	First	7,370	9,795	300	0	31,000
	Second	81,192	35,402	75,000	33,000	150,000
	Third	569,635	2,422,391	350,000	160,000	90,000,000
	All	216,535	1,407,833	75,000	0	90,000,000

Source: Authors' calculations based on the RSI-01 data.

Note: Includes only observations with total annual household income between 1,000$ and 2,000,000$. Terciles are computed for the income and savings variables separately, for the overall sample (both provinces together).

Most respondents (57.81%) declare to be willing to take average risk when making financial investments. This is associated to moderate risk aversion.

Table 3: Distribution of control variables (in percentage)

		Province of residence		
Variables		Ontario	Québec	Both
Province of residence	Ontario	100	0	62.11
	Québec	0	100	37.89
Marital status	Married	66.44	44.1	57.97
	Common-law	9.03	22.25	14.04
	Widowed	3.74	3.66	3.71
	Separated	3.32	3.27	3.3
	Divorced	8.93	11.66	9.96
	Never married	8.55	15.07	11.02
Number of children	Zero	25.47	28.1	26.47
	One	14.76	21.44	17.29
	Two	39.36	34.51	37.52
	Three or more	20.41	15.96	18.72
Contributes or receives benefits from an employer pension plan	Yes	47.3	41.97	45.28
	No	52.7	58.03	54.72
Risk willing to take when saving or investing	Substantial risk	4.58	3.88	4.32
	Above average risk	18.47	14.03	16.79
	Average risk	61.36	51.98	57.81
	Under average risk	15.58	30.12	21.09

Source: Authors' calculations based on the RSI-01 data.
Note: Includes only observations with total annual household income between 1,000$ and 2,000,000$. The number of respondents is almost equal between Ontario and Québec (978 and 980, respectively), but the use of sampling weights adjusts the percentages in the table to reflect the demographic weight of each province. For each variable, the percentages sum to 100 in each column.

4 Findings

4.1 Descriptive evidence

Before moving to the econometric analyses, we present in this section correlations between our variables of interest, that is between income and the objective and subjective survival probabilities, and between income and the probabilities to become dependent (either estimated by the number of ADL limitations or by the probability to enter a nursing home). In addition, in Appendix A, we run the same analysis using savings instead of income. Table B1 in Appendix

B also confirms the correlations one may expect between these variables. Below, we come back on these correlation values when studying in detail each relationship.

4.1.1 The relationship between survival and income

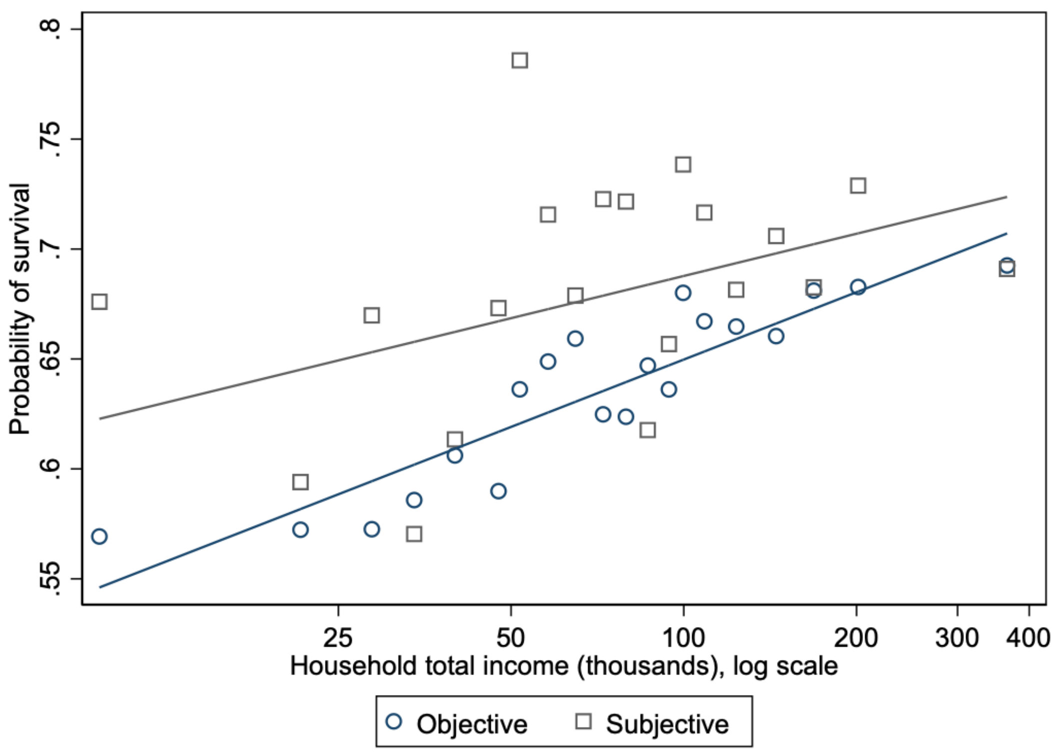
Figure 1 shows the binned scatter plots of income and of (objective and subjective) probabilities of survival to age 85, where each point corresponds to a ventile of household total income, as well as the fitted lines between the probabilities of survival and household income. Note that the X-axis is in logarithmic scale. First, Figure 1 confirms that, on average, respondents overestimate their survival probability (the objective probability line is below the subjective probability one). Second, for both objective and subjective probabilities of survival at 85, there is, as expected, a positive association with log income. Yet, the relationship between income and the objective probability of survival to 85 appears stronger than the relationship between income and the subjective probability (the regression line for the objective probability is steeper than the regression line for the subjective probability). This is also confirmed in Table B1, which shows that the correlation between the subjective survival probability and household log income is 0.067, while it is 0.256 between the objective probability and log income.

In order to confirm our results, in Figure A1 (top left graph) in the Appendix, we plot average survival probabilities by income tercile. We find similar patterns. We also find similar results when we use savings instead of income (see Figure A2 and Table B1).

4.1.2 The relationship between ADL limitations and income

Figure 2 presents the binned scatter plot of log income and the (subjective or objective) probability of ADL limitations. First, objective probabilities are almost always higher than subjective ones, meaning that respondents tend to underestimate their probability to have ADL limitations. Second, this figure shows a negative (but small) relationship between the probability of ADL limitations and income, independently from whether we use subjective or objective probabilities: The slopes of the fitted lines are -0.008 and -0.023 for the objective and the subjective

Figure 1: Probability of survival and household total income



Source: Authors' calculations based on the RSI-01 data.
Note: Each point corresponds to a ventile of household total income (incomes under 1,000$ and above 2,000,000$ excluded), with average probability of survival for the bin on the Y-axis. X-axis is in log scale. Slopes of fitted lines are 0.044 (SE 0.005) and 0.028 (SE 0.013) for objective and subjective probabilities of survival, respectively.

probabilities, respectively.[17] Therefore, it seems that the higher the income, the smaller the (true and perceived) probability to become dependent, and that the relationship is stronger with subjective probabilities than with objective ones. These are results that we will need to confirm with our regression analysis below.

We further graph the average probabilities of ADL limitations by income tercile (see Figure A1 in the Appendix). We find a small decreasing relationship between income terciles and the subjective probabilities to have ADL limitations. The relationship between objective probabilities and income is almost flat. This confirms our earlier results of small negative relationship between income and the probabilities to have ADL limitations.

Finally, Figure A3 in the Appendix presents the relationship between the probability of ADL limitations and savings. Interestingly, the relationship between objective probability to become dependent and savings is decreasing (as with income), while the relationship between subjective probabilities and savings is almost flat (the slope is positive but very small and not statistically different from zero).

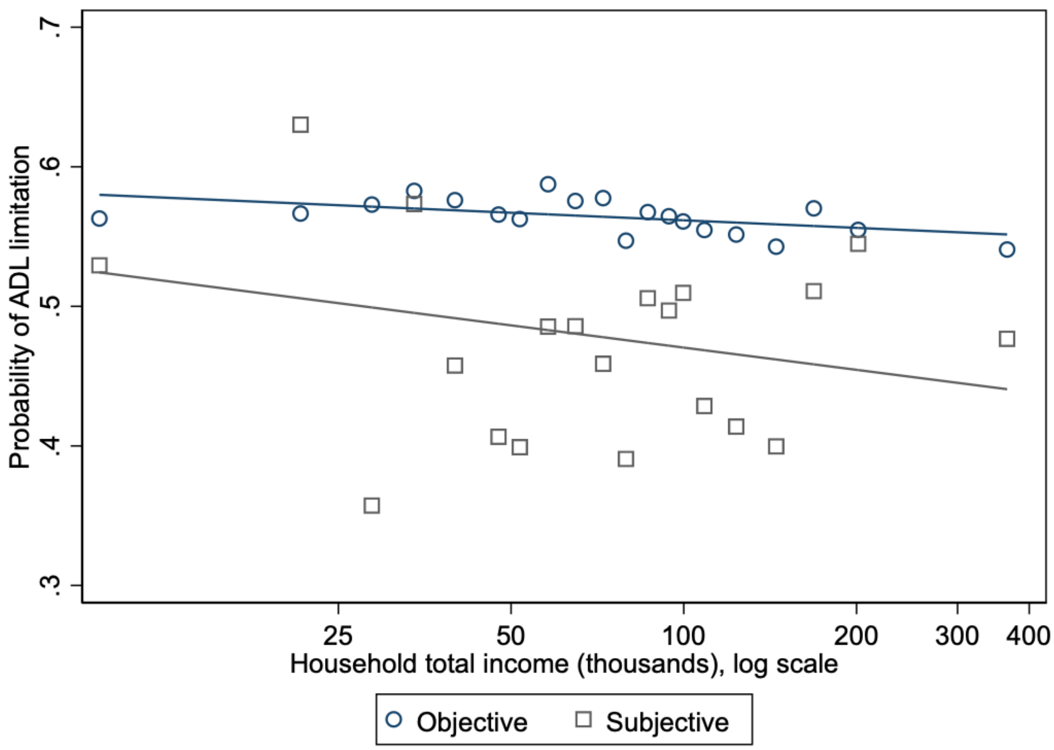
4.1.3 The relationship between moving to a nursing home and income

In addition to the previous measure of dependency–ADL limitations–, we now use a second measure, which is the probability of moving to a nursing home. We believe this measure is particularly relevant as moving to a nursing home concerns primarily individuals needing daily assistance because of an important loss of autonomy. Unlike the previous measure, it is likely to correspond to more intensive care needs.

Figure 3 shows a strong negative relationship between the objective probability of being in nursing home and income (the slope of the fitted line between the objective probability and log income is equal to -0.027, with a standard error of 0.007). This negative relation is consistent with the previous result of a negative relationship between income and the probability to have ADL limitations. More surprisingly, the relationship between income and the subjective probability of entering a nursing home is strongly positive, with a slope equal to 0.051 (standard error

[17] Table B1 in the Appendix reports correlation coefficients of -0.087 and -0.062 between log income and the objective and subjective probabilities to have ADL limitations, respectively.

Figure 2: Probabilities of ADL limitation and household total income



Source: Authors' calculations based on the RSI-01 data.
Note: Each point corresponds to a ventile of household total income (incomes under \$1,000 and above \$2,000,000 excluded), with average probability of ADL limitation for the bin on the Y-axis. X-axis is in log scale. Slopes of fitted lines are -0.008 (SE 0.004) and -0.023 (SE 0.021) for objective and subjective probabilities of ADL limitation, respectively.

of 0.011). This is confirmed by the correlations (see Table B1) between log income and the objective probability to enter a nursing home (equal to -0.188) and between log income and the subjective probability to enter a nursing home (equal to 0.141).

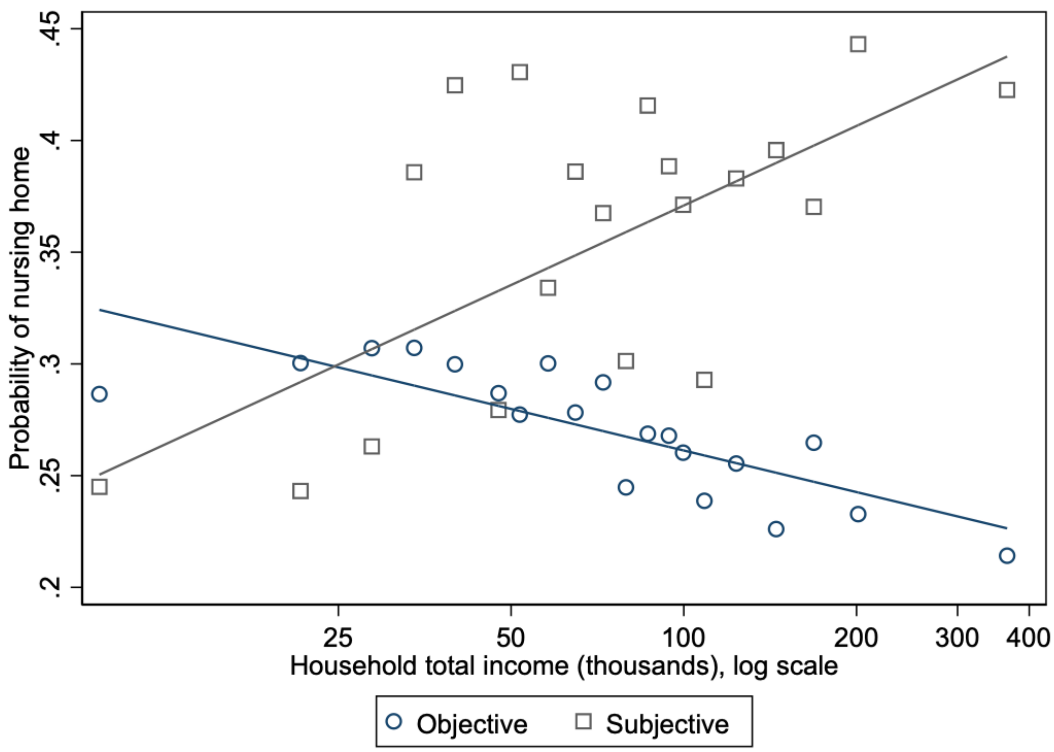
This difference in signs could be explained by the different factors that may influence an individual's decision of moving to a nursing home. Unlike the subjective probability of ADL limitations and of survival, when reporting their subjective probability of being in nursing home, respondents may have taken into account many other factors that could impact this decision, rather than just their perceived future health condition. For instance, their financial situation and their financial ability to pay for the nursing home fees could be strong determinants. In Québec and Ontario, two nursing home systems coexist, the public and the private ones, and unless the elderly has very low resources, they are generally not free to the user. In Québec for instance, private nursing homes are not covered by the public health system and their fees range between $5,000$\$ and $8,000$\$ a month.[18] On the other hand, public nursing homes are means-tested and the maximum fee for a private room is around $2,000$\$ a month.[19] To confirm our intuition, we observe in Figure 3 that, at the bottom of the income distribution, respondents underestimate their probability of entering a nursing home, while above some threshold income, they would overestimate it. Also, whether respondents have relatives or family members to take care of them at home could impact their perception of the probability to ever enter a nursing home. This is something that we will be able to test in our regression analysis. Hence, when answering the survey question on the probability of entering a nursing home, it is possible that the respondents have considered several factors (financial and family situations, etc.) that could have an impact on their move to a nursing home, rather than just the probability that they experience a sharp loss of autonomy.

Figure A1, which shows the relationship between income and the probabilities of moving to a nursing home, using income terciles instead, confirms our earlier results of a negative (resp. positive) relationship between the objective (resp. subjective) probability of entering a nursing

[18] `https://bonjourresidences.com/blogue/couts-hebergement-chsld/`

[19] For the rates in Ontario, see `https://www.ontario.ca/page/paying-long-term-care`, and those in Québec, see `https://www.ramq.gouv.qc.ca/fr/citoyens/programmes-aide/hebergement-etablissement-public`

Figure 3: Probabilities of being in nursing home and household total income



Source: Authors' calculations based on the RSI-01 data.

Note: Each point corresponds to a ventile of household total income (incomes under 1,000$ and above 2,000,000$ excluded), with average probability of being in nursing home for the bin on the Y-axis. X-axis is in log scale. Slopes of fitted lines are -0.027 (SE 0.007) and 0.051 (SE 0.011) for objective and subjective probabilities of being in nursing home, respectively.

home and income terciles.

We also observe the same relationships between subjective and objective probabilities of entering a nursing home and savings. Yet, the strength of the correlations appear smaller in comparison to income ones (see Table B1).

4.2 Regression Results

4.2.1 Methodology

We present here the empirical strategy we use to analyze the relationship between economic status and demographic characteristics at old age. The main equation to be estimated using OLS is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 Z_i + \epsilon_i \tag{1}$$

where the dependent variable y_i is the (objective or subjective) probability of survival, of ADL limitations, or of being in nursing home of individual i. x_i represents total household income (or individual savings), and Z_i is a vector of sociodemographic characteristics that will be used as controls. ϵ_i is the individual error term. Note that we first use the log of income as the independent variable, and then we also use income in terciles, taking the lowest income tercile as the reference category. Also, all probabilities are expressed as proportions, on a scale of 0 to 1.

4.2.2 Estimation of the survival probabilities

In this section, we start by estimating the relationship between household total income and the probability of survival. Our results are summarized in Table 4 (Panel A) for both the objective and the subjective survival probabilities. In this table, we only report the coefficients associated to income (either measured by the log of income or using income terciles). In the Appendix, Tables B2 and B3 report the full set of estimated coefficients.

Income is found to have a positive relationship with the objective survival probability, independently from whether we use income measured in log or income terciles. An increase of 1% in income is associated with an increase in the survival probability of almost five percentage points.

Table 4: Relationship between objective and subjective probabilities and household total income

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Panel A: Probability of survival				
Log(income)	0.0461***	0.0437***	0.0239	0.0244
	(0.0067)	(0.0073)	(0.0123)	(0.0143)
R-squared	0.074	0.091	0.004	0.029
Income by tercile (reference: lowest tercile)				
Second tercile	0.0792***	0.0742***	0.0663**	0.0686*
	(0.0144)	(0.0148)	(0.0253)	(0.0279)
Highest tercile	0.0967***	0.0945***	0.0688**	0.0756*
	(0.0155)	(0.0175)	(0.0256)	(0.0303)
R-squared	0.090	0.107	0.011	0.036
N	1,624	1,624	1,624	1,624
Panel B: Probability of ADL limitations				
Log(income)	-0.0101**	-0.0082*	-0.0252	-0.0425*
	(0.0031)	(0.0035)	(0.0199)	(0.0204)
R-squared	0.014	0.026	0.003	0.033
Income by tercile (reference: lowest tercile)				
Second tercile	-0.0078	-0.0066	-0.0305	-0.0624
	(0.0062)	(0.0065)	(0.0367)	(0.0371)
Highest tercile	-0.0239***	-0.0197**	-0.0468	-0.0795*
	(0.0064)	(0.0070)	(0.0383)	(0.0391)
R-squared	0.017	0.029	0.002	0.032
N	1,271	1,271	1,271	1,271
Panel C: Probability of being in a nursing home				
Log(income)	-0.0243***	-0.0240***	0.0517***	0.0388***
	(0.0042)	(0.0047)	(0.0132)	(0.0134)
R-squared	0.031	0.056	0.019	0.035
Income by tercile (reference: lowest tercile)				
Second tercile	-0.0151	-0.0151	0.0706*	0.0460
	(0.0091)	(0.0091)	(0.0287)	(0.0277)
Highest tercile	-0.0505***	-0.0482***	0.0851**	0.0542
	(0.0092)	0.0099)	(0.0286)	(0.0287)
R-squared	0.032	0.057	0.013	0.030
N	1,322	1,322	1,322	1,322
Sociodemographic controls	NO	YES	NO	YES

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sociodemographic control variables are dummy variables for the province of residence, marital status, number of children, whether the respondent contributes or receives benefits from an employer pension plan, and for the risk preferences. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

We also find that the relationship between the third income tercile and survival is stronger than between the second income tercile and survival (p-values of a test of equality of the coefficients on second and highest terciles are 0.075 and 0.044 in columns (1) and (2), respectively), with the top-tercile respondents having on average an objective probability of survival close to ten percentage points higher than those in the bottom third of the income distribution. The inclusion of sociodemographic controls (column (2)) does not change our results substantially.

The relationship between income and subjective survival probabilities is however more nuanced. While we find no significant relationship between income and the subjective survival probability, we find a statistically significant positive relationship between being either in the second or third income tercile (in comparison to being in the lowest tercile) and the subjective probability of surviving to age 85. This relationship appears slightly stronger for the highest income tercile, but the difference between the second and top terciles are not statistically significant (p-values of 0.904 and 0.728 in columns (3) and (4), respectively). Again, the inclusion of controls does not seem to influence our findings.

In Tables B2 and B3, we also find that some controls have a statistically significant impact on the subjective probability to survive to age 85. For instance, Québec residents in comparison to Ontarians report higher survival probabilities. Also, contributing to an employer pension plan (or receiving benefits from one) is statistically significant at a 5% or 10% level (whether we use the log of income or income terciles) and increases the likelihood to report a higher survival probability at age 85. In the evaluation of the objective survival probabilities, no covariates are found to be statistically significant. These results are also independent of whether we use income terciles or the log of income as the main explanatory variable.

4.2.3 Estimation of the ADL limitations probabilities

Our results regarding the relationships between household income and the objective and the subjective probabilities of having ADL limitations are summarized in Panel B of Table 4. We find a clear negative relationship between the objective probability to have ADL limitations and income, independently from whether we use as the explanatory variable income measured in terciles or in log. A one-percent increase in household total income is associated with a reduction of the objective probability to have ADL limitations by around 1 percentage point. This relationship is smaller when including controls. Looking at the regression using dummies for income terciles, only the third income tercile is statistically significant at a 1% and a 5% level, suggesting that the negative relationship between income and the objective probability to have ADL limitations is driven by the higher incomes in the distribution.

When explaining the subjective probability to have ADL limitations using household income, we find that the relation with log income is significant only at a 10% significance level, and only when we include controls. An increase of 1% in income is associated with a decrease in the subjective probability of around four percentage points. When using income measured in terciles, we also find a negative relationship between income and the subjective probability to face ADL limitations, but only the variable "being in the highest tercile" has a coefficient that is statiscally different from zero, at a 10% significance level. Again, this suggests that the link between income and the subjective probability is again mainly driven by the higher incomes in the distribution.

Finally, being from Québec and being divorced appear to be significant at a 1% and a 10% level respectively, but only for the subjective probability (see Tables B4 and B5). Being from Québec (in comparison to Ontario) and being divorced (in comparison to being married) increase the chance that the respondent reports a lower subjective probability to face ADL limitations.

4.2.4 Estimation of the probabilities of moving to a nursing home

Table 4 (Panel C) confirms the negative relationship between the objective probability of being in a nursing home and income, and the positive relationship between the subjective probability and income, which we had found in Section 4.1.3. When using the log of income, we find that, when including covariates, an increase of 1% in income is associated with a *decrease* of the objective probability to enter a nursing home of more than two percentage points (statistically different from zero at a 1% significance level) but with an *increase* of the subjective probability of 3.9 percentage points (also significant at a 1% significance level). This confirms our earlier interpretation that respondents may have interpreted the survey question in a way such that they would consider entering a nursing home only when they have enough financial means.

When using income measured in terciles, we find that the relationship between the objective probability and income remains statistically significant at a 1% level only for the highest income tercile. Being in the highest income tercile is associated with a reduction of the probability to enter a nursing home of around five percentage points compared to being in the lowest income tercile. On the other hand, for the subjective probability to enter a nursing home, the sign of the coefficient is positive, meaning that higher incomes are associated with a greater probability to enter a nursing home, a result that is statistically significant only when no controls are included.

As shown in Tables B6 and B7, Québec respondents (in comparison to those from Ontario) have, all else equal, lower subjective probabilities of entering a nursing home, at a 5% significance level. Interestingly, it appears that being widowed is positively associated with the subjective probability of entering a nursing home. Again, this confirms our earlier hypothesis that respondents may have considered many other factors than just their declining health when formulating expectations about the probability to ever enter a nursing home. One of these factors is surely whether the partner is still alive or not.

5 Robustness checks

5.1 IV regression

The existence of a positive correlation between economic resources and health is well documented in the economic literature. However, the direction of the causality between these two variables is still an open question (see for instance, Ettner (1996); Carrieri and Jones (2017); Lefebvre et al. (2018); Meer et al. (2003); Michaud and van Soest (2008)). Some authors support the idea that economic resources affect health, since richer agents can afford better health care, healthier food, and healthier environments. Some other studies argue that the causality runs the other way: healthier people often work more and are more able to accumulate economic resources. Thereby, in our main estimated equation (eq. 1), one may suspect that simultaneity between health and income could lead to biased estimates.

In order to solve this problem, we propose to proceed as in Lefebvre et al. (2018) and to estimate our model using an instrumental variable approach, where we choose as the instrument the dichotomous variable about ownership of the primary home. This question was asked in the RSI-01 survey (see Q13 in the questionnaire, Appendix C). For this variable to be considered a valid instrument, it would need to be correlated with income, but not with the error term in our main equation, meaning that its effect on (objective or subjective) probabilities of survival, of ADL limitations, and of being in nursing home can only be through its correlation with income.

Table 5 below presents estimates from the second stage of a two-stage least squares (2SLS) estimation, and Table B8 in Appendix B presents those from the first stage. The instrument appears sufficiently relevant, with F-statistics ranging from 26.12 to 101.01, depending on the specification.[20] Owning a primary residence is positively and strongly correlated with log income. Given that we only have one instrument, we cannot test for its endogeneity. We defer to the existing literature to argue for its endogeneity, but we also focus in our main analysis (in the preceding section) on the OLS approach given the potential doubts about the validity of the instrument.

The 2SLS estimates presented in Table 5 are generally similar to those found using OLS (in Table 4). The point estimates are larger when explaining the probability of survival (0.1007 compared to 0.0461 and 0.1585 compared to 0.0437 for columns 1 and 2, respectively, and similarly for columns 3 and 4), but the signs (positive) and the statistical significance (significant for the objective probabilities, not significant for the subjective ones) are the same. In Panel B, for the probabilities of ADL limitations, the signs obtained using 2SLS are the same as with OLS (i.e., negative), but no relationship is found to be statistically different from zero. Finally, in Panel C, we can see that the different signs for different types of probabilities (negative for objective probabilities, positive for subjective ones) are maintained, but that the coefficients on log income when explaining the subjective probabilities are not statistically significant. By and

[20]While all models are essentially the same, with log income being explained by the instrument, the sample size varies slightly across the different probabilities due to non response, and some specifications include sociodemographic controls.

Table 5: Instrumental variable estimations

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Panel A: Probability of survival				
Log(income)	0.1007***	0.1585**	0.0594	0.1071
	(0.0256)	(0.0537)	(0.0372)	(0.0804)
N	1,624	1,624	1,624	1,624
Panel B: Probability of ADL limitations				
Log(income)	-0.0104	-0.0069	-0.0170	-0.0948
	(0.0087)	(0.0165)	(0.0506)	(0.0850)
N	1,271	1,271	1,271	1,271
Panel C: Probability of being in a nursing home				
Log(income)	-0.0309*	-0.0366	0.0617	0.0396
	(0.0137)	(0.0247)	(0.0400)	(0.0747)
N	1,322	1,322	1,322	1,322
Sociodemographic controls	NO	YES	NO	YES

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Estimations using two-stage least squares, with the indicator variable of owning a home used as an instrument for log income. Sociodemographic control variables are dummy variables for the province of residence, marital status, number of children, whether the respondent contributes or receives benefits from an employer pension plan, and for the risk preferences. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

large, the IV results confirm our previous OLS findings and are consistent with others in the literature (Meer et al., 2003; Kim and Gukbin, 2016; Lefebvre et al., 2018).

5.2 Using only non-imputed values of income

As another robustness check, we estimate the models as we did in Section 4.2, but using only the observations for which income was not imputed. Indeed, as we explained in Section 3, when respondents did not provide their household income, income was imputed using some of their sociodemographic characteristics (education level, home ownership, age, gender, being married and whether the respondent has kids). To alleviate potential colinearity issues between income and our control variables, we thus provide, in Table 6, the same estimates as in Table 4, but from a subsample using only non imputed values. As Table 6 shows, the sample size decreases–as expected–but the signs, the sizes, and the significance levels of the coefficients remain quite close to what we obtained in our baseline regressions (Table 4). We conclude that it is unlikely that the imputation procedure could have impacted our results.

5.3 Individual savings

In this paper, we have used total household income as the main explanatory variable (either in log or in terciles). As a last robustness check, in this section, we now use instead individual savings as the main explanatory variable. Savings may give a better approximation of permanent income than total household income in year 2016, although accumulated savings also reflect consumption patterns and investment behavior.

Table 7 indeed shows that our estimations are robust to using savings instead of income as the main explanatory variable. Comparing with the estimations from Table 4, we find the same relationships. Objective survival probabilities increase with reported savings, but objective probabilities to become dependent (either evaluated by ADL limitations or by the probability of entering a nursing home) decrease with savings. Yet, the magnitude of these relations seems to be smaller. In addition, we find no significant relationships between savings and subjective probabilities, as well as positive signs (again) for the coefficients on savings when explaining the

Table 6: OLS estimations using only non-imputed income

Explanatory variables	Objective probability (1)	(2)	Subjective probability (3)	(4)
Panel A: Probability of survival				
Log(income)	0.0532***	0.0483***	0.0367*	0.0385*
	(0.0098)	(0.0099)	(0.0151)	(0.0165)
R-squared	0.080	0.103	0.009	0.041
Income by tercile (reference: lowest tercile)				
Second tercile	0.0879***	0.0784***	0.0633*	0.0616*
	(0.0177)	(0.0170)	(0.0271)	(0.0282)
Highest tercile	0.0969***	0.0901***	0.0625*	0.0618*
	(0.0198)	(0.0207)	(0.0288)	(0.0306)
R-squared	0.093	0.114	0.010	0.041
N	1,157	1,157	1,157	1,157
Panel B: Probability of ADL limitations				
Log(income)	-0.0120**	-0.0116**	-0.0208	-0.0369
	(0.0037)	(0.0039)	(0.0223)	(0.0225)
R-squared	0.016	0.016	0.001	0.020
Income by tercile (reference: lowest tercile)				
Second tercile	-0.0083	-0.0084	-0.0431	-0.0731*
	(0.0071)	(0.0074)	(0.0381)	(0.0372)
Highest tercile	-0.0265***	-0.0256**	-0.0471	-0.0779
	(0.0074)	(0.0081)	(0.0410)	(0.0410)
R-squared	0.021	0.020	0.002	0.022
N	951	951	951	951
Panel C: Probability of being in a nursing home				
Log(income)	-0.0311***	-0.0336***	0.0554**	0.0353
	(0.0054)	(0.0060)	(0.0189)	(0.0186)
R-squared	0.044	0.055	0.019	0.028
Income by tercile (reference: lowest tercile)				
Second tercile	-0.0210*	-0.0224*	0.0529	0.0262
	(0.0105)	(0.0107)	(0.0343)	(0.0323)
Highest tercile	-0.0612***	-0.0633***	0.0941**	0.0565
	(0.0112)	(0.0121)	(0.0357)	(0.0350)
R-squared	0.045	0.056	0.013	0.026
N	960	960	960	960
Sociodemographic controls	NO	YES	NO	YES

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sociodemographic control variables are dummy variables for the province of residence, marital status, number of children, whether the respondent contributes or receives benefits from an employer pension plan, and for the risk preferences. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, with total annual household income not imputed, and with non-missing values of the subjective probabilities.

Table 7: Relationship between objective and subjective probabilities and individual savings

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Panel A: Probability of survival				
Log(savings)	0.0110**	0.0103**	0.0053	0.0070
	(0.0034)	(0.0034)	(0.0056)	(0.0055)
R-squared	0.021	0.032	0.000	0.028
N	1,427	1,427	1,427	1,427
Savings by tercile (reference: lowest tercile)				
Second tercile	0.0531***	0.0406**	0.0398	0.0324
	(0.0154)	(0.0142)	(0.0271)	(0.0271)
Highest tercile	0.0854***	0.0725***	0.0583*	0.0572*
	(0.0148)	(0.0140)	(0.0239)	(0.0252)
R-squared	0.064	0.082	0.006	0.031
N	1,624	1,624	1,624	1,624
Panel B: Probability of ADL limitations				
Log(savings)	-0.0060***	-0.0050***	-0.0007	-0.0048
	(0.0016)	(0.0015)	(0.0072)	(0.0074)
R-squared	0.020	0.035	-0.001	0.022
N	1,105	1,105	1,105	1,105
Savings by tercile (reference: lowest tercile)				
Second tercile	-0.0047	-0.0026	-0.0139	-0.0263
	(0.0063)	(0.0063)	(0.0366)	(0.0349)
Highest tercile	-0.0298***	-0.0255***	-0.0311	-0.0502
	(0.0060)	(0.0061)	(0.0352)	(0.0332)
R-squared	0.034	0.043	-0.000	0.028
N	1,271	1,271	1,271	1,271
Panel C: Probability of being in a nursing home				
Log(savings)	-0.0084***	-0.0062**	0.0122	0.0095
	(0.0025)	(0.0024)	(0.0069)	(0.0070)
R-squared	0.015	0.039	0.004	0.011
N	1,171	1,171	1,171	1,171
Savings by tercile (reference: lowest tercile)				
Second tercile	-0.0133	-0.0070	0.0549	0.0334
	(0.0096)	(0.0094)	(0.0302)	(0.0291)
Highest tercile	-0.0492***	-0.0392***	0.0796**	0.0496
	(0.0091)	(0.0094)	(0.0287)	(0.0281)
R-squared	0.033	0.054	0.011	0.030
N	1,322	1,322	1,322	1,322
Sociodemographic controls	NO	YES	NO	YES

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sociodemographic control variables are dummy variables for the province of residence, marital status, number of children, whether the respondent contributes or receives benefits from an employer pension plan, and for the risk preferences. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities. Sample size is smaller when using log savings due to savings values of $0.

probability of entering a nursing home (see Panel C).

6 Conclusion

This paper has quantified the relationships between household income, survival at old age, and the probability to become dependent (as measured by the probability of having ADL limitations, and by the probability to ever enter a nursing home). Using survey data representative of the population of Ontario and Québec, we find, as already documented in the literature, that there is a positive relation between income and both the objective and subjective probabilities to survive to age 85. More interestingly, because less documented in the economic literature, we find a negative relationship between income and the objective probability to become dependent (either measured by the probability to have ADL limitations or of entering a nursing home). This relation is mostly driven by individuals in the highest income tercile. To the contrary, the relationship between income and subjective probabilities to become dependent is more nuanced. While income and the subjective probability to have ADL limitations are negatively correlated, income and the subjective probability to enter a nursing home are positively related. This latter relationship may be explained by other factors, such as the financial means necessary to enter a long-term care home. Our results are robust to different specifications.

Our analysis sheds light on important sociodemographic relationships. Higher-income agents have higher survival chances and are less often dependent while lower-income agents have lower survival chances and are more often dependent. These empirical findings should certainly be taken into account by policy makers when elaborating redistributive public policies, and, in particular, those targeted toward the elderly. Indeed, our findings recommend that a redistributive policy not only favours low-income individuals for reasons of lower financial resources, but also because they would be more often confronted to dependency. In the context of aging populations, these are certainly important dimensions to account for; omitting such correlations would lead to less redistribution than otherwise optimal.

References

D. Boisclair, J-Y. Duclos, S. Marchand, and P-C. Michaud. Une analyse économique de propositions visant à bonifier la couverture du risque de longévité. *L'Actualité économique*, 91(4): 499–530, 2015.

D. Boisclair, A. Côté-Sergent, F. Laliberté-Auger, S. Marchand, and P-C. Michaud. A health microsimulation model for Quebec and Canada. *CEDIA, HEC Montréal*, 2016.

M. M. Boyer, P. De Donder, C. Fluet, M-L. Leroux, and P-C. Michaud. A Canadian parlor room-type approach to the long-term care insurance puzzle. *Canadian Public Policy*, 45(2): 262–282, 2019a.

M. M. Boyer, P. De Donder, C. Fluet, M-L. Leroux, and P-C. Michaud. Long-term care risk misperceptions. *Geneva Papers on Risk and Insurance - Issues and Practices*, 44:183–215, 2019b.

M. M. Boyer, P. De Donder, C. Fluet, M-L. Leroux, and P-C. Michaud. Long-term care insurance: Information frictions and selection. *American Economic Journal: Economic Policy*, 12 (3):134–169, 2020a.

M. M. Boyer, P. De Donder, C. Fluet, M-L. Leroux, and P-C. Michaud. Data and code for: Long-term care insurance: Information frictions and selection, 2020b. URL `https://doi.org/10.3886/E115322V1`.

E. Cambois, C. Laborde, I. Romieu, and J-M. Robine. Occupational inequalities in health expectancies in France in the early 2000s: Unequal chances of reaching and living retirement in good health. *Demographic Research*, 25(12):407–436, 2011.

V. Carrieri and A. M. Jones. The income-health relationship 'beyond the mean': New evidence from biomarkers. *Health Economics*, 26:937–956, 2017.

R. Chetty, M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler. The association between income and life expectancy in the United States, 2001–2014: Association between income and life expectancy in the United States. *Journal of the American Medical Association*, 315(16):1750–1766, 2016.

N-J. Clavet, Y. Décarie, R. Hébert, P-C. Michaud, and J. Navaux. Le financement du soutien à l'autonomie des personnes âgées à la croisée des chemins. *CREEI Working paper*, (21-01), 2021.

H. Cremer and P. Pestieau. Social long-term care insurance and redistribution. *International Tax and Public Finance*, 21:955–974, 2014.

J. Cristia. Rising mortality and life expectancy differentials by lifetime earnings in the United States. *B.E Journal of Economic Analysis and Policy*, 28(984–995):551–575, 2009.

D. Cutler, A. Deaton, and A. Lleras-Muney. The determinants of mortality. *Journal of Economic Perspectives*, 20(3):97–120, 2006.

D. Cutler, A. Lleras-Muney, and T. Vogl. Socioeconomic status and health: Dimensions and mechanisms. In *The Oxford Handbook of Health Economics, S. Glied et P. Smith (Eds.)*. Oxford University Press, New York, 2011.

A. Delavande and S. Rohwedder. Differential survival in Europe and the United States: Estimates based on subjective probabilities of survival. *Demography*, 48(4):1377–1400, 2011.

P. De Donder and M-L. Leroux. Behavioral biases and long-term care insurance: A political approach. *B.E Journal of Economic Analysis and Policy*, 14(2):551–575, 2014.

P. De Donder and P. Pestieau. Private, social, and self-insurance for long-term care in the presence of family help. *Journal of Public Economic Theory*, 19(1):18–37, 2017.

S. Ettner. New evidence on the relationship between income and health. *Journal of Health Economics*, 15(1):67–85, 1996.

Institut sur la retraite et l'épargne / Retirement and Savings Institute. RSI-01: Protection against financial risks in retirement: An economic analysis of the risk of dependency / Protection contre les risques financiers à la retraite: analyse économique du risque de dépendance, 2020. URL `https://doi.org/10.5683/SP2/PP5U7Y`.

A. Kalwij, R. Alessie, and M. Knoef. The association between individual income and remaining life expectancy at the age of 65 in the Netherlands. *Demography*, 50(1):181–206, 2013.

J. Kim and K. Gukbin. Relationship between the remaining years of healthy life expectancy in older age and national income level, educational attainment, and improved water quality. *The International Journal of Aging and Human Development*, 83(4):02–417, 2016.

J. Klimaviciute and P. Pestieau. Long-term care social insurance: How to avoid big losses? *International Tax and Public Finance*, 25(1):99–139, 2018.

G. Lacroix, F. Laliberté-Auger, P-C. Michaud, and D. Parent. The effect of college education on health and mortality: Evidence from canada. *Health Economics*, 30(S1):105–118, 2021.

M. Lefebvre, S. Perelman, and J. Schoenmaeckers. Inégalités face à la mort et au risque de dépendance. *Revue française d'économie*, XXXIII(2):75–111, 2018.

M-L. Leroux and P. Pestieau. Age and health related inheritance. *CESifo Working Papers*, (9004), 2021.

M-L. Leroux and P. Pestieau. Age-related taxation of bequests in the presence of a dependency risk. *Journal of Public Economic Theory*, 24(21):92–119, 2022.

M-L. Leroux, P. Pestieau, and G. Ponthiere. Fair long-term care insurance. *Social Choice and Welfare*, 57:503–533, 2021.

H. Liu and M. Wang. Socioeconomic status and adl disability of the older adults: Cumulative health effects, social outcomes and impact mechanisms. *Plos ONE*, 17(2):e0262808, 2022.

J. Meer, D. Miller, and H. Rosen. Exploring the health–wealth nexus. *Journal of Health Economics*, 22(2003):713–730, 2003.

P-C. Michaud and A. van Soest. Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics*, 27:1312–1325, 2008.

K. Milligan and T. Schirle. The evolution of longevity: Evidence from Canada. *Canadian Journal of Economics*, 54(1):164–192, 2021.

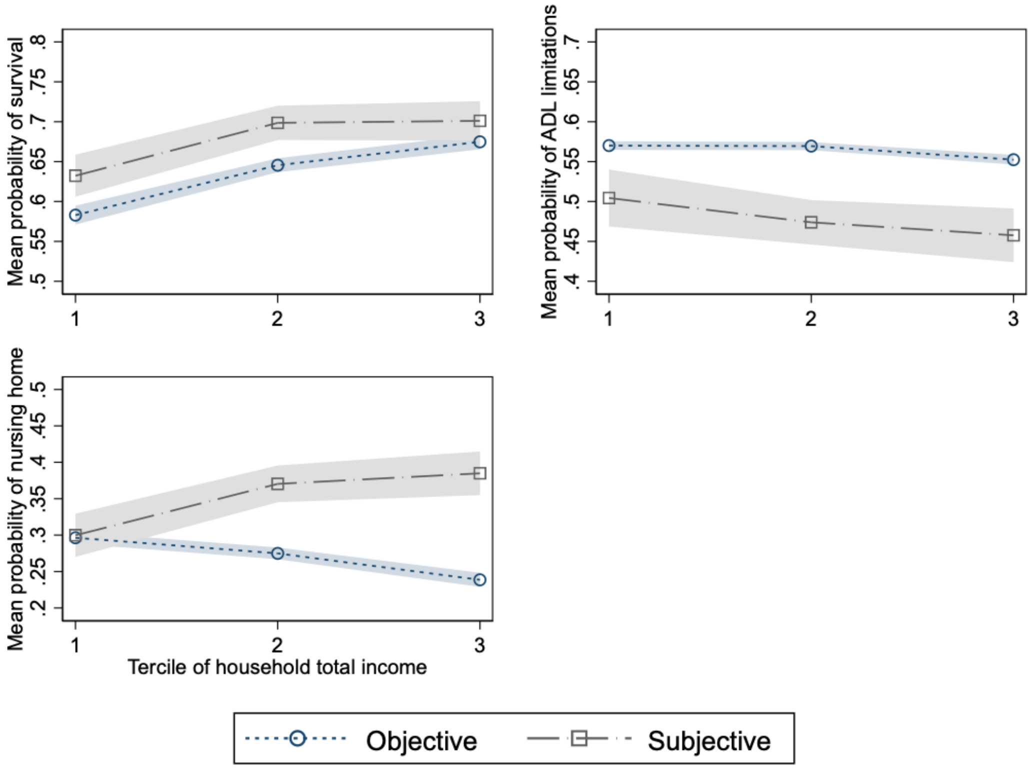
OECD. Help wanted? - providing and paying for long-term care. *OECD Press*, 2011.

OECD. Health at a glance 2021: OECD indicators. *OECD Publishing, Paris*, 2021.

P. Zaninotto, G.D. Batty, S. Stenholm, I. Kawachi, M. Hyde, M. Goldberg, H. Westerlund, J. Vahtera, and J. Head. Socioeconomic inequalities in disability-free life expectancy in older people from England and the United States: A cross-national population-based study. *Journals of Gerontology: Medical Sciences*, 75(5):906–913, 2020.

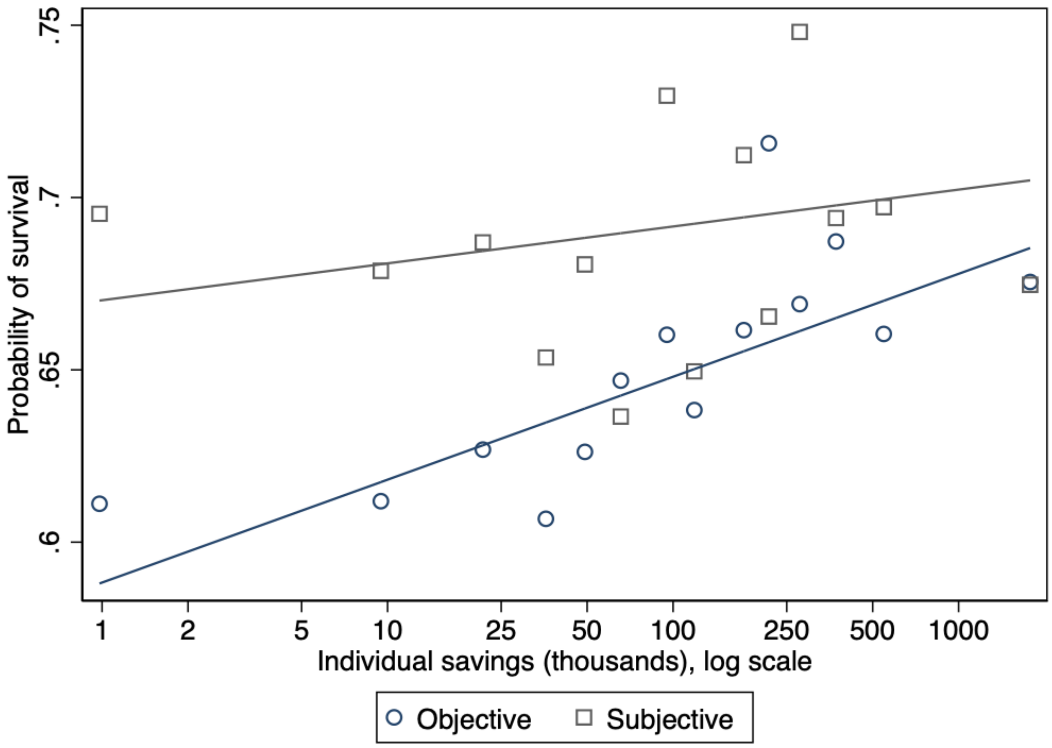
A Appendix Figures

Figure A1: Mean of the objective and subjective probabilities by income tercile



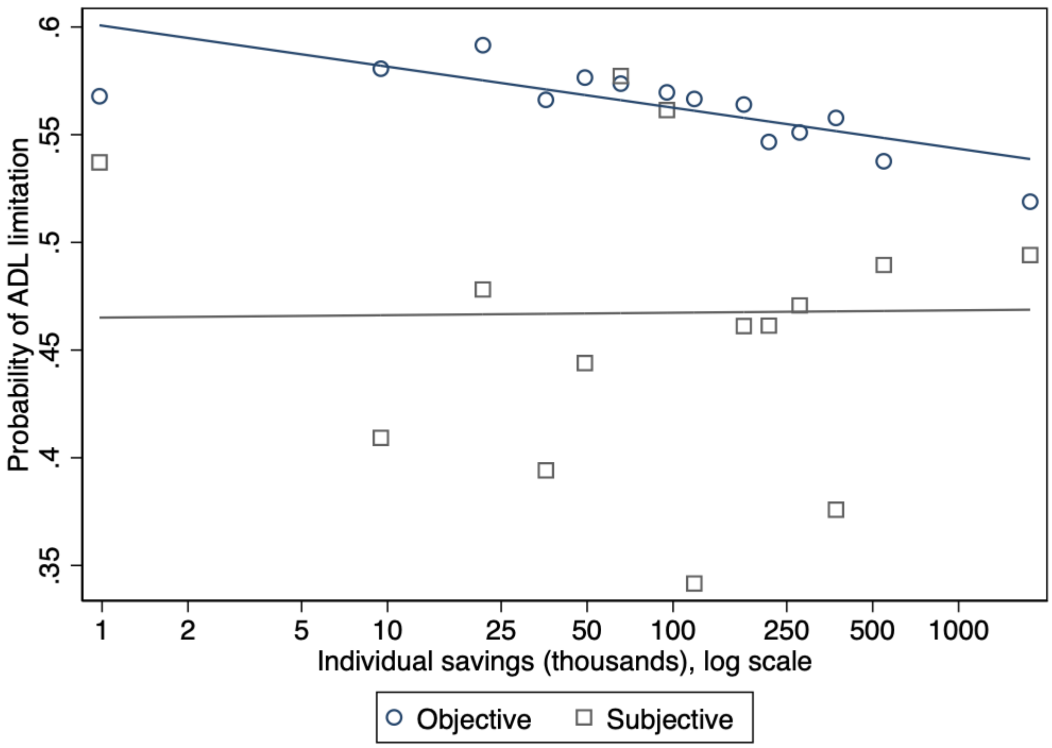
Source: Authors' calculations based on the RSI-01 data.
Note: Each point corresponds to a tercile of household total income (incomes under 1,000$ and above 2,000,000$ excluded), with average probabilities for the tercile on the Y-axis. Shaded areas represent 95% confidence intervals.

Figure A2: Probabilities of survival and individual savings



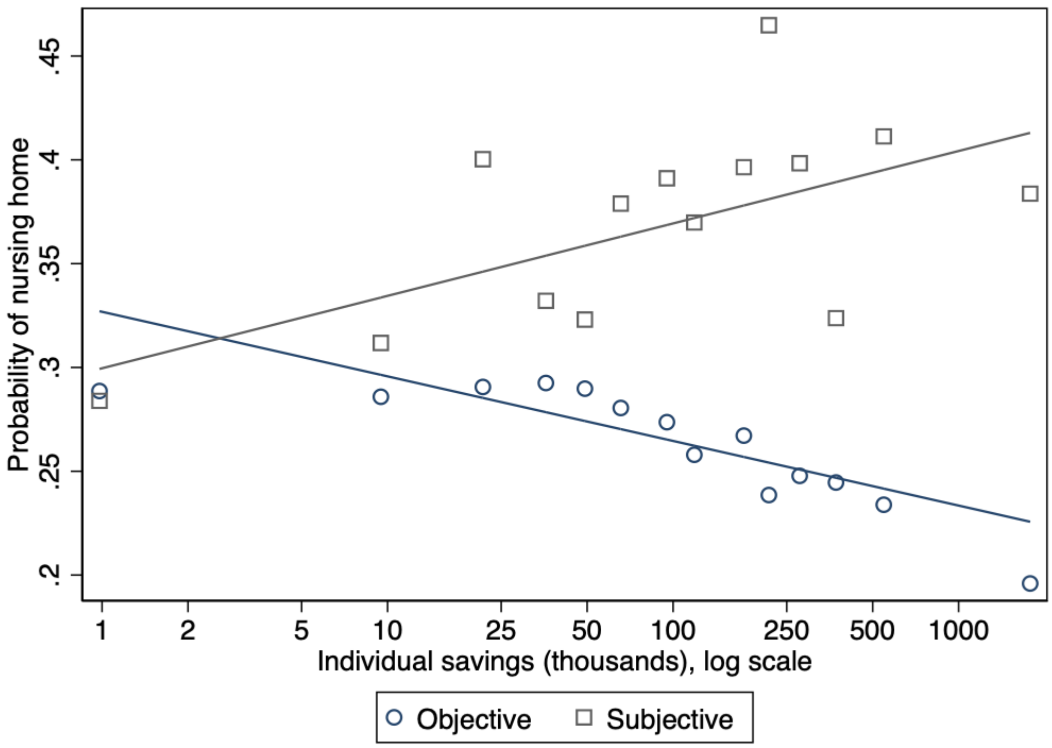
Source: Authors' calculations based on the RSI-01 data.
Note: Each point corresponds to a hexadecile of individual savings (savings of 0$ excluded), with average probability of survival for the bin on the Y-axis. X-axis is in log scale. Slopes of fitted lines are 0.013 (SE 0.002) and 0.005 (SE 0.004) for objective and subjective probabilities of survival, respectively.

Figure A3: Probabilities of ADL limitation and individual savings



Source: Authors' calculations based on the RSI-01 data.
Note: Each point corresponds to a hexadecile of individual savings (savings of 0$ excluded), with average probability of ADL limitation for the bin on the Y-axis. X-axis is in log scale. Slopes of fitted lines are -0.008 (SE 0.003) and 0.0005 (SE 0.009) for objective and subjective probabilities of ADL limitation, respectively.

Figure A4: Probabilities of being in nursing home and individual savings



Source: Authors' calculations based on the RSI-01 data.
Note: Each point corresponds to a hexadecile of individual savings (savings of $0 excluded), with average probability of being in nursing home for the bin on the Y-axis. X-axis is in log scale. Slopes of fitted lines are -0.014 (SE 0.004) and 0.015 (SE 0.005) for objective and subjective probabilities of being in nursing home, respectively.

B Appendix Tables

Table B1: Correlations between probability variables, household income and individual savings

	Survival		ADL limitations		Nursing home	
	Objective	Subjective	Objective	Subjective	Objective	Subjective
Survival, objective	1.000					
Survival, subjective	0.257	1.000				
ADL limitations, objective	0.221	0.044	1.000			
ADL limitations, subjective	0.035	-0.115	0.066	1.000		
Nursing home, objective	0.069	0.034	0.778	0.084	1.000	
Nursing home, subjective	0.096	-0.088	-0.002	0.465	-0.030	1.000
Total household income (log)	0.256	0.067	-0.087	-0.062	-0.188	0.141
Individual savings (log)	0.147	0.032	-0.173	-0.004	-0.176	0.069

Source: Authors' calculations based on the RSI-01 data.

Note: Each number corresponds to a pairwise correlation between the probability variable indicated in the column and the variable in the row. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$. Sample size varies across correlations due to missing values. Total household income and individual savings include imputed values and are measured in logs.

Table B2: Relationship between probability of survival and log income

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Log(income)	0.0461***	0.0437***	0.0239	0.0244
	(0.0067)	(0.0073)	(0.0123)	(0.0143)
Province of residence (ref.: Ontario)				
Québec	–	0.0052	–	0.0827***
		(0.0107)		(0.0175)
Marital status (ref.: married)				
Common-law	–	-0.0008	–	-0.0145
		(0.0113)		(0.0287)
Widowed	–	0.0115	–	0.0394
		(0.0230)		(0.0583)
Separated	–	-0.0206	–	0.0455
		(0.0504)		(0.0555)
Divorced	–	-0.0058	–	0.0127
		(0.0244)		(0.0284)
Never married	–	0.0075	–	0.0082
		(0.0247)		(0.0371)
Number of children (ref.: zero)				
One	–	-0.0101	–	-0.0096
		(0.0204)		(0.0330)
Two	–	-0.0016	–	0.0073
		(0.0167)		(0.0287)
Three or more	–	-0.0329	–	0.0015
		(0.0199)		(0.0351)
Contributes or receives benefits from an employer pension plan (ref.: no)				
Yes	–	0.0129	–	0.0580**
		(0.0107)		(0.0204)
Risk willing to take when saving or investing (ref.: substantial risk)				
Above average risk	–	-0.0147	–	0.0543
		(0.0243)		(0.0601)
Average risk	–	0.0077	–	0.0560
		(0.0211)		(0.0579)
Under average risk	–	-0.0292	–	0.0268
		(0.0254)		(0.0600)
Constant	0.1188	0.1493	0.4115**	0.2891
	(0.0769)	(0.0901)	(0.1396)	(0.1713)
R-squared	0.074	0.091	0.004	0.029
N	1,624	1,624	1,624	1,624

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

Table B3: Relationship between probability of survival and income tercile

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Income tercile (ref.: lowest)				
Second	0.0792***	0.0742***	0.0663**	0.0686*
	(0.0144)	(0.0148)	(0.0253)	(0.0279)
Highest	0.0967***	0.0945***	0.0688**	0.0756*
	(0.0155)	(0.0175)	(0.0256)	(0.0303)
Province of residence (ref.: Ontario)				
Québec	–	0.0050	–	0.0840***
		(0.0107)		(0.0174)
Marital status (ref.: married)				
Common-law	–	-0.0002	–	-0.0145
		(0.0111)		(0.0282)
Widowed	–	0.0115	–	0.0473
		(0.0245)		(0.0561)
Separated	–	-0.0199	–	0.0560
		(0.0460)		(0.0561)
Divorced	–	-0.0033	–	0.0222
		(0.0245)		(0.0285)
Never married	–	0.0086	–	0.0153
		(0.0239)		(0.0361)
Number of children (ref.: zero)				
One	–	-0.0111	–	-0.0097
		(0.0194)		(0.0318)
Two	–	-0.0021	–	0.0061
		(0.0158)		(0.0278)
Three or more	–	-0.0334	–	-0.0002
		(0.0188)		(0.0341)
Contributes or receives benefits from an employer pension plan (ref.: no)				
Yes	–	0.0108	–	0.0520*
		(0.0106)		(0.0202)
Risk willing to take when saving or investing (ref.: substantial risk)				
Above average risk	–	-0.0144	–	0.0533
		(0.0250)		(0.0596)
Average risk	–	0.0082	–	0.0572
		(0.0220)		(0.0574)
Under average risk	–	-0.0271	–	0.0321
		(0.0254)		(0.0598)
Constant	0.5745***	0.5811***	0.6323***	0.5126***
	(0.0133)	(0.0307)	(0.0209)	(0.0660)
R-squared	0.090	0.107	0.011	0.036
N	1,624	1,624	1,624	1,624

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sample restricted to observations with total annual household income between 1,000$ and above 2,000,000$, and with non-missing values of the subjective probabilities.

Table B4: Relationship between probability of ADL limitations and log income

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Log(income)	-0.0101**	-0.0082*	-0.0252	-0.0425*
	(0.0031)	(0.0035)	(0.0199)	(0.0204)
Province of residence (ref.: Ontario)				
Québec	–	0.0092	–	-0.0962***
		(0.0050)		(0.0249)
Marital status (ref.: married)				
Common-law	–	0.0135	–	0.0290
		(0.0073)		(0.0366)
Widowed	–	-0.0060	–	0.0509
		(0.0110)		(0.0524)
Separated	–	0.0092	–	-0.1026
		(0.0101)		(0.0890)
Divorced	–	-0.0013	–	-0.1008*
		(0.0092)		(0.0476)
Never married	–	0.0056	–	0.0294
		(0.0112)		(0.0526)
Number of children (ref.: zero)				
One	–	0.0111	–	0.0612
		(0.0100)		(0.0472)
Two	–	0.0087	–	0.0522
		(0.0067)		(0.0397)
Three or more	–	0.0077	–	-0.0137
		(0.0077)		(0.0402)
Contributes or receives benefits from an employer pension plan (ref.: no)				
Yes	–	-0.0045	–	0.0041
		(0.0050)		(0.0249)
Risk willing to take when saving or investing (ref.: substantial risk)				
Above average risk	–	0.0035	–	-0.0322
		(0.0114)		(0.0725)
Average risk	–	0.0209*	–	-0.0003
		(0.0106)		(0.0698)
Under average risk	–	0.0149	–	-0.0377
		(0.0118)		(0.0750)
Constant	0.6705***	0.6241***	0.7600***	0.9776***
	(0.0356)	(0.0402)	(0.2272)	(0.2463)
R-squared	0.014	0.026	0.003	0.033
N	1,271	1,271	1,271	1,271

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

Table B5: Relationship between probability of ADL limitations and income tercile

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Income tercile (ref.: lowest)				
Second	-0.0078	-0.0066	-0.0305	-0.0624
	(0.0062)	(0.0065)	(0.0367)	(0.0371)
Highest	-0.0239***	-0.0197**	-0.0468	-0.0795*
	(0.0064)	(0.0070)	(0.0383)	(0.0391)
Province of residence (ref.: Ontario)				
Québec	–	0.0091	–	-0.0954***
		(0.0050)		(0.0251)
Marital status (ref.: married)				
Common-law	–	0.0130	–	0.0275
		(0.0073)		(0.0367)
Widowed	–	-0.0059	–	0.0558
		(0.0109)		(0.0557)
Separated	–	0.0099	–	-0.0999
		(0.0102)		(0.0911)
Divorced	–	-0.0015	–	-0.1025*
		(0.0091)		(0.0475)
Never married	–	0.0052	–	0.0307
		(0.0110)		(0.0522)
Number of children (ref.: zero)				
One	–	0.0116	–	0.0647
		(0.0100)		(0.0470)
Two	–	0.0084	–	0.0525
		(0.0067)		(0.0400)
Three or more	–	0.0078	–	-0.0149
		(0.0077)		(0.0403)
Contributes or receives benefits from an employer pension plan (ref.: no)				
Yes	–	-0.0045	–	0.0022
		(0.0050)		(0.0253)
Risk willing to take when saving or investing (ref.: substantial risk)				
Above average risk	–	0.0036	–	-0.0365
		(0.0113)		(0.0726)
Average risk	–	0.0200	–	-0.0030
		(0.0106)		(0.0698)
Under average risk	–	0.0141	–	-0.0388
		(0.0118)		(0.0751)
Constant	0.5680***	0.5417***	0.5044***	0.5537***
	(0.0049)	(0.0125)	(0.0324)	(0.0825)
R-squared	0.017	0.029	0.002	0.032
N	1,271	1,271	1,271	1,271

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

Table B6: Relationship between probability of nursing home and log income

	Objective probability		Subjective probability	
Explanatory variables	(1)	(2)	(3)	(4)
Log(income)	-0.0243***	-0.0240***	0.0517***	0.0388***
	(0.0042)	(0.0047)	(0.0132)	(0.0134)
Province of residence (ref.: Ontario)				
Québec	–	-0.0076	–	-0.0570**
		(0.0072)		(0.0200)
Marital status (ref.: married)				
Common-law	–	0.0215	–	0.0181
		(0.0114)		(0.0316)
Widowed	–	0.0115	–	0.1099*
		(0.0214)		(0.0492)
Separated	–	0.0088	–	-0.0759
		(0.0272)		(0.0544)
Divorced	–	0.0139	–	-0.0131
		(0.0115)		(0.0381)
Never married	–	-0.0138	–	0.0052
		(0.0113)		(0.0440)
Number of children (ref.: zero)				
One	–	0.0347**	–	-0.0252
		(0.0115)		(0.0390)
Two	–	0.0239*	–	0.0113
		(0.0093)		(0.0315)
Three or more	–	0.0302**	–	-0.0431
		(0.0114)		(0.0324)
Contributes or receives benefits from an employer pension plan (ref.: no)				
Yes	–	-0.0042	–	0.0141
		(0.0078)		(0.0221)
Risk willing to take when saving or investing (ref.: substantial risk)				
Above average risk	–	-0.0010	–	0.0011
		(0.0186)		(0.0531)
Average risk	–	0.0267	–	-0.0015
		(0.0174)		(0.0498)
Under average risk	–	0.0265	–	-0.0481
		(0.0191)		(0.0533)
Constant	0.5344***	0.4922***	-0.2248	-0.0519
	(0.0471)	(0.0563)	(0.1498)	(0.1578)
R-squared	0.031	0.056	0.019	0.035
N	1,322	1,322	1,322	1,322

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

Table B7: Relationship between probability of nursing home and income tercile

Explanatory variables	Objective probability (1)	(2)	Subjective probability (3)	(4)
Income tercile (ref.: lowest)				
Second	-0.0151	-0.0151	0.0706*	0.0460
	(0.0091)	(0.0091)	(0.0287)	(0.0277)
Highest	-0.0505***	-0.0482***	0.0851**	0.0542
	(0.0092)	(0.0099)	(0.0286)	(0.0287)
Province of residence (ref.: Ontario)				
Québec	–	-0.0071	–	-0.0590**
		(0.0072)		(0.0201)
Marital status (ref.: married)				
Common-law	–	0.0197	–	0.0196
		(0.0113)		(0.0318)
Widowed	–	0.0150	–	0.0985
		(0.0207)		(0.0515)
Separated	–	0.0132	–	-0.0855
		(0.0281)		(0.0545)
Divorced	–	0.0161	–	-0.0178
		(0.0112)		(0.0376)
Never married	–	-0.0128	–	0.0003
		(0.0111)		(0.0437)
Number of children (ref.: zero)				
One	–	0.0366**	–	-0.0277
		(0.0114)		(0.0388)
Two	–	0.0231*	–	0.0122
		(0.0093)		(0.0313)
Three or more	–	0.0299**	–	-0.0397
		(0.0116)		(0.0320)
Contributes or receives benefits from an employer pension plan (ref.: no)				
Yes	–	-0.0057		0.0183
		(0.0078)		(0.0221)
Risk willing to take when saving or investing (ref.: substantial risk)				
Above average risk	–	-0.0008	–	0.0055
		(0.0189)		(0.0548)
Average risk	–	0.0251	–	-0.0011
		(0.0178)		(0.0514)
Under average risks	–	0.0269	–	-0.0523
		(0.0195)		(0.0546)
Constant	0.2845***	0.2458***	0.2998***	0.3470***
	(0.0067)	(0.0203)	(0.0232)	(0.0601)
R-squared	0.032	0.057	0.013	0.030
N	1,322	1,322	1,322	1,322

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

Table B8: First stage (log income explained by home ownership)

	Sample					
	Probability of survival		Probability of ADL limitations		Probability of nursing home	
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)
Own primary residence	0.6101***	0.2987***	0.6900***	0.3866***	0.7260***	0.4007***
	(0.0607)	(0.0584)	(0.0750)	(0.0698)	(0.0731)	(0.0679)
R-squared	0.089	0.266	0.112	0.292	0.123	0.313
N	1,624	1,624	1,271	1,271	1,322	1,322
F-stat	101.01	26.12	84.65	30.71	98.65	34.80
Sociodemographic controls	NO	YES	NO	YES	NO	YES

Source: Authors' calculations based on the RSI-01 data.

Note: Robust standard errors in parentheses, ***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$. This table presents the coefficients on the dummy variable indicating if the respondent owns their primary residence in the first stage regressions, in which log income is explained by the instrument. Since the sample varies slightly across the different probabilities due to non response, three sets of results are presented. First-stage F-statistics are presented at the bottom of the table. Columns 1, 3, and 5 are without sociodemographic controls, and columns 2, 4, and 6 include them. Sociodemographic control variables are dummy variables for the province of residence, marital status, number of children, whether the respondent contributes or receives benefits from an employer pension plan, and for the risk preferences. Sample restricted to observations with total annual household income between 1,000$ and 2,000,000$, and with non-missing values of the subjective probabilities.

C Survey questionnaire

Long-Term Care Insurance Survey (Paper Version of Questionnaire for Internet Survey)

Introduction

For purposes of this survey, when we use the term 'long-term care,' we are referring to assistance with personal care needs such as dressing, bathing, getting in and out of bed, using the bathroom or eating. A long-term care home or assisted living facility refers to a facility that offers board, meals and other basic care services for persons who need long-term care. The facility also offers medical services. It is therefore distinct from a retirement home, where no or limited care is offered.

Section 1: Long-Term Care Insurance

Q1 This survey is going to ask you questions about long-term care insurance. Which of the following best describes your current knowledge about this type of insurance?
1 A lot
2 A little
3 None at all

Q2 For purposes of this survey, we define long-term care insurance as a type of insurance that helps to pay for extended stays in a long-term care home or assisted living facility, or for personal or medical care in your home. It is typically separate from your health insurance and requires paying separate premiums. Do you have a long-term care insurance policy?
1 Yes
2 No
3 Don't Know

IF Q2==3 (Don't know) GOTO Q6
ELSE IF Q2==2 (No)

> Q3a Why don't you have a long-term care insurance policy? Choose the main reason.
> 1 I have never thought about buying one, and I have never been offered one (for instance by a financial advisor).
> 2 I have thought about buying one, but I have not (yet) made a decision.
> 3 I used to have such a policy, but I let it lapse.
> 4 Such insurance policies are too expensive for me.
> 5 Such insurance policies do not cover my needs.
> 6 I do not think I will need such a policy.
> 7 I don't know what that is.
> 8 Other, open...
> GOTO Q6

ELSE IF Q2==1 (Yes)

> Q3b How did you come to purchase that insurance policy?

1 I was offered a long-term care policy
2 I searched myself for a long-term care policy
3 Other, open …

Q4 What is the monthly premium on that policy, including taxes?
Numeric
9999 Don't know
IF Q4==9999
Q4a Is it more than $200 1 Yes 2 No 8888888 Refuse to answer
IF Q4a==1
Q4b Is it less than $400 1 Yes 2 No 8888888 Refuse to answer
ELSE IF Q4a==2
Q4c Is it more than $100 1 Yes 2 No 8888888 Refuse to answer
END IF
END IF
Q5 What is the amount of the benefit the insurance would pay out (monthly)?
Numeric
9999 Don't know
IF Q5==9999
Q5a Is it more than $2,500 1 Yes 2 No 8888888 Refuse to answer
IF Q5a==1
Q5b Is it less than $3,500 1 Yes 2 No 8888888 Refuse to answer
ELSE IF Q5a==2
Q5c Is it more than $1,500 1 Yes 2 No 8888888 Refuse to answer
END IF
END IF
END IF

Q6 Do you have life insurance for which you currently pay a premium (or that is in force)?
1 Yes
2 No
3 Don't Know

Section 2: Background

Q7 At the present time, do you smoke cigarettes daily, occasionally or not at all?
1 Daily
2 Occasionally
3 Not at all
IF Q7==1 GOTO Q8
ELSE IF Q7==2,3
Q7a Have you ever smoked cigarettes daily?
1 Yes
2 No

IF Q7a==1 GOTO Q8
ELSE IF Q7a==2
Q7b Have you smoked 100 cigarettes or more in your life?
1 Yes
2 No
IF Q7b==1 GOTO Q8
ELSE IF Q7b==2
Q7c Have you ever smoked a whole cigarette?
1 Yes
2 No
END IF
END IF
END IF

Q8 What is the highest degree, certificate or diploma you have obtained?
1 Less than high school diploma or its equivalent
2 High school diploma or a high school equivalency certificate
3 Trade certificate or diploma
4 College, CEGEP or other non-university certificate or diploma (other than trades certificates or diplomas)
5 University certificate or diploma below the bachelor's level
6 Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
7 University certificate, diploma, degree above the bachelor's level

Q9 What is your marital status?
1 married
2 living common-law
3 widowed
4 separated
5 divorced
6 single, never married

Q10 Do you have children?
1 Yes
2 No
IF Q10==1
Q10a How many children do you have?
Numeric (>0)
END IF

Q11 For 2016, what is your best estimate of the total income received by all members of your household, from all sources, before taxes and deductions?
Numeric
9999999 Don't know or prefer not to say
IF Q11==9999999
Q11a Is it more than $60,000 1 Yes 2 No 8888888 Refuse to answer

IF Q11a==1
Q11b Is it less than $120,000 1 Yes 2 No 8888888 Refuse to answer
ELSE IF Q11a==2
Q11c Is it more than $30,000 1 Yes 2 No 8888888 Refuse to answer
END IF
END IF

Q12 Do you consider yourself retired?
1 Yes
2 No
IF Q12==2
Q12a What is your best estimate of what total income received by all members of your household will be once you are fully retired, as a fraction of your current income?
Numeric (0%-200%)
9999999 Don't know
IF Q12a==9999999
Q12b Is it more than 50%? 1 Yes 2 No 8888888 Refuse to answer
IF Q12b==1
Q12c Is it less than 75%? Yes 2 No 8888888 Refuse to answer
ELSE IF Q12b==2
Q12d Is it more than 25%? 1 Yes 2 No 8888888 Refuse to answer
END IF
END IF
END IF

Q13 Do you own your primary residence?
1 Yes
2 No

IF Q13==1
Q13a What is the current market value of your residence?
Numeric
9999999 Don't know
IF Q13a==9999999
Q13b Is it more than $300,000? 1 Yes 2 No 8888888 Refuse to answer
IF Q13b==1
Q13c Is it less than $600,000? 1 Yes 2 No 8888888 Refuse to answer
ELSE IF Q13a==2
Q13d Is it more than $150,000? 1 Yes 2 No 8888888 Refuse to answer
END IF
END IF

Q14 How much do you still carry as a mortgage, as a proportion of the current market value of your residence?

1 Less than 20%
2 Between 20 and 40%
3 Between 40 and 60%
4 More than 60%
5 Don’t know
END IF

Q15 – We are interested in your pension plan and its nature, if you have one. Do you currently contribute to, or receive benefits from, an employer provided pension plan?
1 Yes
2 No
3 Don't Know

IF Q15==1
Q15a Is your pension plan a defined-benefit or a defined-contribution plan? A defined-benefit plan is one where you receive fixed income in retirement for as long as you live and you don’t get to decide how much is contributed and how it is invested. A defined contribution plan is one where you decide how the contributions are invested and you receive at retirement the amount accumulated from your contributions.
1 Defined-benefit
2 Defined-contribution
3 Other
4 Don't Know
END IF

Q16 What is your best estimate of how much you have accumulated in Registered Retirement Savings Plans (RRSPs), Tax-Free Savings Accounts (TFSAs) and other savings accounts?
Numeric
9999999 Don’t know or prefer not to say
IF Q16==9999999
Q16a Is it more than $50,000? 1 Yes 2 No 8888888 Refuse to answer
IF Q16a==1
Q16b Is it less than $200,000? 1 Yes 2 No 8888888 Refuse to answer
ELSE IF Q16a==2
Q16c Is it more than $10,000? 1 Yes 2 No 8888888 Refuse to answer
END IF
END IF

Q17 Looking at the following list of health conditions, has a doctor ever told you you had:
[Check any of:]
1 Heart disease
2 Stroke
3 Lung disease

4 Diabetes
5 Hypertension
6 Depression or other mental health problems
7 Cancer

Section 3: Risk Perception

Q18 On a scale of 0 to 100, where 0 is absolutely no chance and 100 is absolutely certain, what do you believe is the percent chance you will live to age 85 or more?
Numeric (0-100)
9999999 Don't know

Q19 On a scale of 0 to 100, where 0 is absolutely no chance and 100 is absolutely certain, what do you believe is the percent chance you will live more than 1 year during your lifetime with two or more limitations in activities of daily living? Activities of daily living include eating, bathing, getting dressed, walking about one's home and getting in and out of bed.
Numeric (0-100)
9999999 Don't know
IF Q19>0
    Q19a 2 or more years?
    Numeric (Range 0 – Answer to Q19)
    9999999 Don't know
    IF Q19a>0
        Q19b 4 or more years?
        Numeric (Range 0 – Answer to Q19a)
        9999999 Don't know
    END IF
END IF

Q20 Of course nobody wishes to go to a long-term care home, but sometimes this becomes necessary. On a scale of 0 to 100, what do you believe is the percent chance that you will have to move to a long-term care home because of important limitations in your activities of daily living?
Numeric (0-100)
9999999 Don't know

Q21 On a scale of 0 to 100, what do you believe is the percent chance that your family would take up the responsibility of taking care of you if you had important limitations in activities of daily living?
Numeric (0-100)
9999999 Don't know

Formal care refers to that provided by qualified caregivers who are usually paid and unrelated to the person receiving care; **informal** care refers to that usually provided for free by relatives. Please keep these definitions in mind for the following questions.

Q22 **Formal** care refers to that provided by qualified caregivers who are usually paid and unrelated to the person receiving care; **informal** care refers to that usually provided for free by relatives.

Do you agree with the following statements? (Answers: 1 Strongly Agree; 2 Agree; 3 Disagree; 4 Strongly Disagree; 5 Don’t know)
Q22a It is the responsibility of the family, when feasible, to take care of elderly parents
Q22b Parents should set aside money to leave to their children or heirs once they die, even when it means somewhat sacrificing their own comfort in retirement
Q22c It is children's duty to provide their parents with informal long-term care or to pay for their formal long-term care, should the need arise.

Q23 **Formal** care refers to that provided by qualified caregivers who are usually paid and unrelated to the person receiving care; **informal** care refers to that usually provided for free by relatives.

If you found yourself in a situation where you needed long-term care, which type of care would you prefer to receive: formal or informal?
1 Formal
2 Informal
3 Don’t know

Section 4: Literacy and Knowledge

Now we would like to ask some questions about your familiarity and comfort with financial concepts. Please answer these questions the best you can.

Q24 Suppose you have $100 in a savings account, the interest rate is 2% per year and you never withdraw money. After 5 years, how much will you have in this account in total?
1 More than $110
2 Exactly $110
3 Less than $110
4 Don’t know

Q25 True or false? You should invest most of your money in a single stock that you select rather than in lots of stocks or in mutual funds.
1 True
2 False
3 Don’t know

Q26 Suppose the chances of someone aged 50 living to age 85 are 60%. What do you think the chances are that this same person will live to age 60?

1 Fewer than 60%
2 More than 60%
3 Don’t know

Q27 Which of the following statements comes closest to describing the amount of financial risk that you are willing to take when you save or make investments?
1 I am willing to take substantial financial risks expecting to earn substantial returns
2 I am willing to take above average financial risks expecting to earn above average returns
3 I am willing to take average financial risks expecting to earn average returns
4 I am willing to take under average financial risks expecting to earn under average returns

IF PROV = QC
    Q28 In 2016, what is the average monthly cost of staying in a private, unsubsidized long-term care home (CHSLD) if you are uninsured (for a private room)? This would include the cost of room and board as well as that of all personal and nursing care.
    Numeric
    9999999 Don’t know
    IF Q27==9999999
        Q27a Is it more than $3,000? 1 Yes 2 No 8888888 Refuse to answer
        IF Q27a==1
            Q27b Is it less than $5,000? 1 Yes 2 No 8888888 Refuse to answer
        ELSE IF Q27a==2
            Q27c Is it more than $1,000? 1 Yes 2 No 8888888 Refuse to answer
        END IF
    END IF
END IF

TEXT

IF PROV = QC: $HOME = subsidized long-term care homes (CHSLD)
IF PROV = ON: $HOME = long-term care homes
FOR PROGRAMMER and TRANSLATOR: [A CHSLD is a Centre d’hébergement et de soins de longue durée. The acronym seems to be used by native english speakers in Quebec as well.]

Q29 Are [$HOME] free to the user?
1 Yes
2 No
IF Q29==2
    Q29a In 2016, what is the monthly fee that you think you would have to pay in [$HOME] for a private room?
    Numeric

9999999 Don't know

Q29b Is there a reduced user contribution if you have low personal resources (income and assets)?

1 Yes

2 No

Q29c If you receive benefits from a long-term care insurance, how does that affect the user contribution you have to pay in [$HOME] if you have low personal resources?

1 It increases my fee

2 It decreases my fee

3 It does not affect my fee

4 Don't know

END IF

Q30 Is there a waiting period to obtain a room in a [$HOME]?

1 Yes

2 No

IF Q30==1

Q30a On average, how many months do you think the wait is in your province?

Numeric (>0)

9999 Don't know

END IF

Q31 If you purchase a long-term care insurance policy and you stop paying premiums after having paid them for several years, do you generally get reimbursed for what you already paid?

1 Yes

2 No

3 Don't know

Section 5: Preferences for Insurance Products

(Not reported here)