NBER WORKING PAPER SERIES

THE TIME OF YOUR LIFE:
THE MORTALITY AND LONGEVITY OF CANADIANS

Kevin S. Milligan

Working Paper 33066
http://www.nber.org/papers/w33066

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2024

I have read the NBER disclosure policy and attest that the acknowledgments and my additional disclosure statement together disclose all sources of funding and all material and relevant financial relationships The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

The author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33066

The Time of Your Life: The Mortality and Longevity of Canadians
Kevin S. Milligan
NBER Working Paper No. 33066
October 2024
JEL No. I14, J14

ABSTRACT

I develop and implement a methodology for cohort life expectancy using a panel of administrative tax data on a large sample born between 1930 and 1964. Over these 35 years, cohort life expectancy after age 54 grew by 5 years for women and 7 years for men. The income-longevity gradient for the top vs. bottom five percent of incomes is 9 years of post-54 life for men and 7 years for women. The life expectancy improvements arise across the income distribution in Canada, unlike the United States. Large differences across neighbourhoods emerge which cannot be explained by income differences alone.

Kevin S. Milligan
Vancouver School of Economics
University of British Columbia
6000 Iona Drive
Vancouver, BC V6T 1L4
and NBER
kevin.milligan@ubc.ca

A Code repository is available at https://doi.org/10.5683/SP3/UZXMZB

Introduction

Economists study many types of inequality, ranging from the basic economic factors of income, consumption, and wealth to important social outcomes like access to education and exposure to crime. An underappreciated element of economic inequality belongs among these objects of inquiry: the study of inequality in longevity. Consider how much people value an extra year with loved ones; an extra month enjoying the vitality of grandchildren or imparting hard-earned wisdom to the next generation. At the extreme, in some circumstances we expend great resources on medical treatments to extend life by even a day or two[1]. Or, more conventionally, consumption inequality measured at a point in time might look very different when aggregated over lifespans of differing lengths.

Beyond thinking of longevity as a key element of economic inequality, longevity matters for understanding economic behaviour. For example, lower-earning Canadians retire earlier than higher-earning Canadians, as shown in Milligan and Schirle (2024). In the most basic lifecycle model we might expect earlier retirement from those with shorter expected lifespans—how much of a contribution does differential longevity play in these observed patterns of retirement?

Differences in longevity also matter for economic policy. Analysis of pension policy looks very different when low earners live much shorter lifespans than high earners, as dollars of contributions pay very different streams of future benefits for low and high earners for a given

[1] See Tanuseputro et al. (2015) estimate that about 10 percent of all government-funded health care cost is spent in the last year of life in their 2010-2013 Ontario sample.

retirement age. In addition, to the extent that access to quality health services or adaptation of positive health behaviours matter, understanding the contributions of these factors to longevity inequality can provide guidance on attenuating lifespan differences.

For all these reasons, longevity deserves attention from economists and policy makers. This lecture addresses the patterns of longevity in Canada. I start with the basic building block of period mortality, which counts how many people of a given age die in a given period—typically a year. I show how mortality across ages has shifted over time and use these shifts to motivate the focus of the paper on longevity. I then present a methodology for estimating cohort longevity and implement these methods using a longitudinal file of administrative income tax data. These data allow me to disaggregate the results into quantiles of lifetime income to explore the heterogeneity and inequality of longevity, and how it has shifted over time. The final piece of the analysis compares across geographies using the first three characters of postal codes to document spatial patterns of longevity.

This work builds on a growing Canadian and international literature on socio-economic status and longevity. The work most similar to what appears in this paper is Milligan and Schirle (2021) which develops a methodology for producing longevity estimates for birth cohorts using administrative data from the Canada Pension Plan. That methodology borrowed from Chetty et al. (2016) which used tax records to analyze differential longevity in the United States and is a pivotal paper in the literature. The new evidence presented in this lecture extends the timeframe of Canadian analysis to the 1960s birth cohorts using a dataset that allows for many more dimensions of analysis, including by geography.

An extensive literature review is offered in Milligan and Schirle (2021) but I note three additional Canadian papers here. First, Etches (2009) seems to be the first paper that used administrative tax data to study mortality in Canada; the same dataset I use in this lecture. Second, the same administrative tax data is used by AlFakhri and Compton (2023) to study heterogeneity in joint survival of couples in Canada. Most recently, Wolfson et al. (2024) show differences by geography, using 2011 Census data disaggregated to census tracts.

There are three main findings in this lecture. First, there is a strong income gradient of cohort life expectancy in Canada of 9 years for men and 7 years for women, comparing the bottom and top five percent of the income distribution. Second, over the period from 1930 to 1964 cohort life expectancy improved approximately uniformly across the income distribution, in contrast to the steepening income gradient in the United States. Third, there is substantial geographical variation in life expectancy across neighbourhoods within cities, and these cannot be explained by income differences alone.

The lecture is organized as follows. I first document trends in mortality from 1921-2021, noting that most of the gains in mortality over the last fifty years come at older ages. This motivates the focus on longevity. I then describe the methodology used to estimate cohort mortality using tax administrative data. The results across sex, birth cohort, and income fractile are then reported. The final analysis presents differences in longevity across neighbourhood, which is followed by a discussion and conclusion.

Mortality

I begin with the study of period mortality, looking at deaths cross-sectionally within a given period. Mortality data used as economic data is particularly interesting for two reasons. First is the ease of measurement. The outcome is binary, observable, and not subject to any form of subjectivity bias. The data can be compared without much concern across time and place. Related, the second advantage of mortality data as economic data is the availability of long consistent series. At the extreme, a consistent series of mortality data for Sweden is available from 1751 forward.[2] For Canada, available data cover 1921 forward.

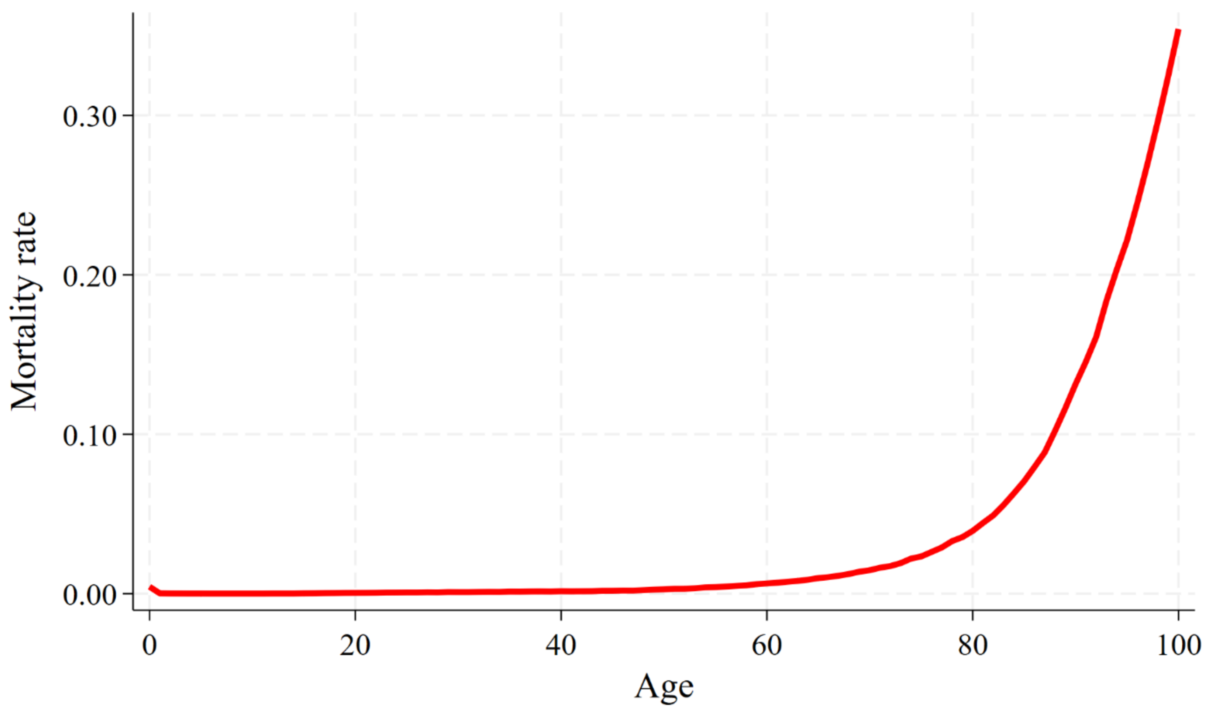
These properties of mortality data can be contrasted with survey health data. Survey assessments of health can reveal patterns and trends in subjective measures like self-assessed health or the incidence of chronic disease. While clearly valuable information, the time series for survey assessments is typically short.[3] Moreover, comparing across countries cross-sectionally is limited by comparability of responses to similar questions across cultures.[4] So, mortality data offer some potential for improvement on survey-based data.

Period mortality is measured by counting the number of deaths in a population divided by the number at risk. There can be some subtlety involved in ascertaining the number at risk (beginning population; ending population; some combination of these). However, these subtleties

[2] See the Swedish data available at the Human Mortality Database project, www.mortality.org.
[3] For example, the Canadian Community Health Survey begins in 2000, so only a bit more than 20 years of data are currently available.
[4] See Juerges (2007) for a cross-country comparison of self-reported health. Kapteyn et al. (2007) show differences in response to self-report questions versus vignettes in the United States and the Netherlands. Heger (2018) shows Canada-US differences in self-reported versus objective health measures matter.

only make material differences in mortality rates for populations with a large risk of death over the time period (typically newborns or the very old). For this paper, I make use of the mortality data put together for the Human Mortality Project and the Canadian component of this project, the Canadian Human Mortality Database (Université de Montréal 2024). These data come from a common methodology that allows for confident international and intertemporal comparisons. I present below a sequence of graphs showing all-sex mortality across ages and how it has evolved over time.

Figure 1: Mortality rate by age for 2021



Notes: Mortality rates across all sexes by age in 2021. Source is the Canadian Human Mortality Database, Université de Montréal (2024).

The first graph in Figure 1 shows the mortality rate in 2021 at each age up to 100. Differences across younger ages are difficult to distinguish in this data representation, but the mortality rate at age 20 is 0.0005; at age 50 is 0.0028, and at age 80 is 0.039. Mortality rates at ages over 80 accelerate sharply. As Figure 1 makes it hard to distinguish smaller changes across younger ages, for the next step in the analysis I take the natural logarithm of mortality to make these changes more readily observable.

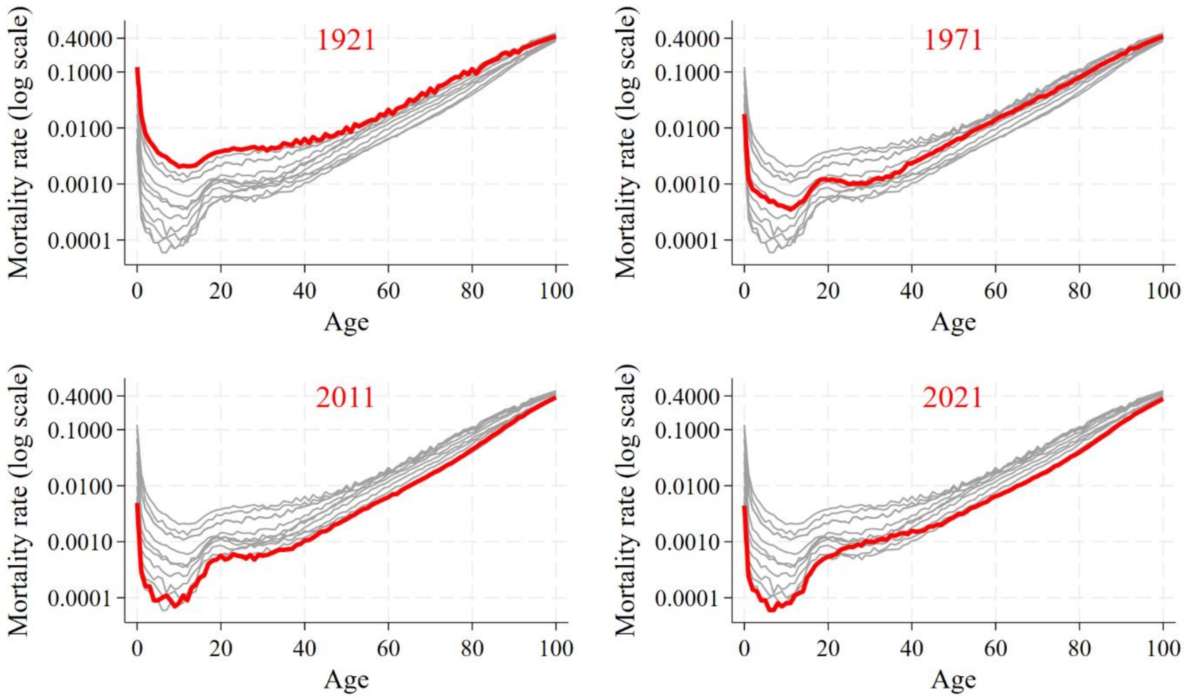
In Figure 2, I graph the mortality rate on a log scale by age across decades. Each of the light grey lines shows the mortality rates for one of 11 years in the sequence {1921,1931…,2021}. The four panels highlight the years 1921, 1971, 2011, and 2021. Several patterns emerge. First, there are substantial gains in mortality at younger to mid ages in the first fifty years from 1921 to 1971. Perinatal and child mortality improved substantially over this period. Second, in the next fifty years from 1971 to 2021 substantial gains in mortality can be seen after age 60. Third, two recent health developments can be examined in the bottom two panels, comparing 2011 and 2021. The impact of Covid-19 does not have a substantial impact when comparing 2011 to 2021. However, between ages 20 and 40 there is a large increase in mortality, rolling back gains by 20 to 30 years. This age range coincides with death patterns from opioids.[5]

An important caveat on these results is the outcome for Indigenous people. Feir and Akee (2019) provide an important contribution with the study of status First Nations mortality using a different source of administrative data, finding no evidence of improvements in mortality in that population over the 30-40 year time periods they study.

---

[5] See Public Health Agency of Canada (2024). Opioid deaths are concentrated among males with highest impact between ages 30-39. Medical assistance in dying also began over this time period, but amounted to only 10,000 of the 312,000 deaths in 2021; with 95 percent of these deaths occurring after age 55 (Health Canada 2022).

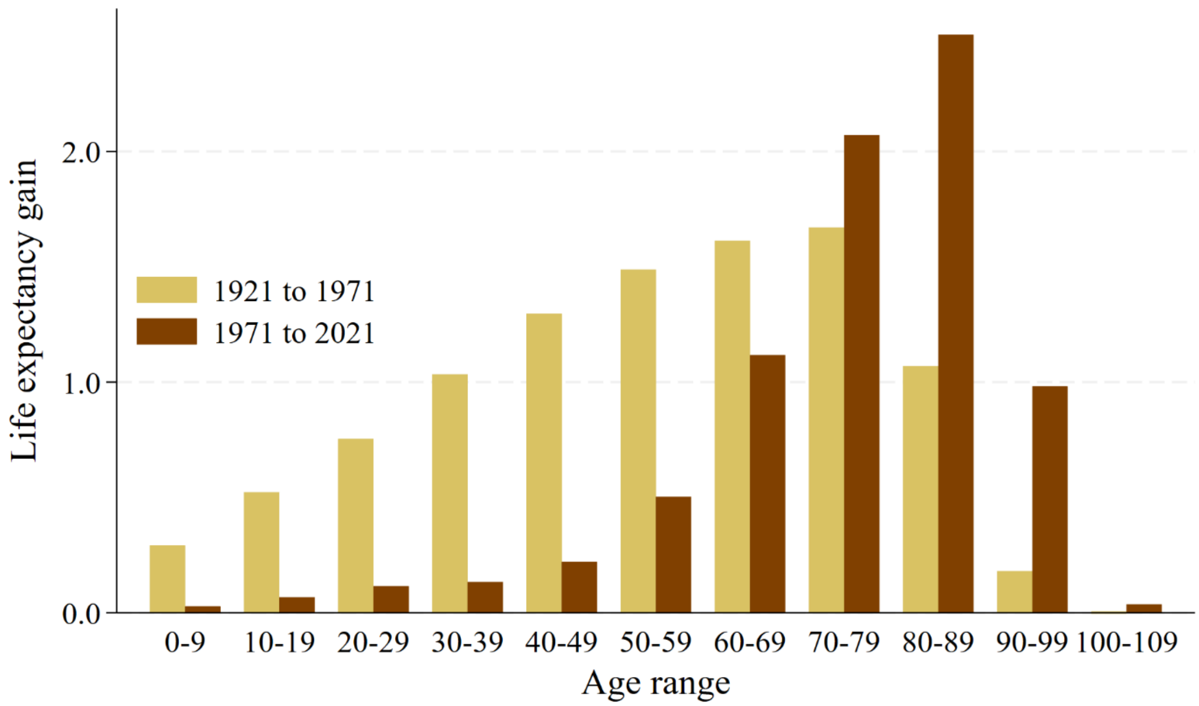
Figure 2: Mortality by age across years



Notes: Mortality rates on a log scale across all genders by age. Each line shows the mortality rate for a different year, with the four panels highlighting the noted years. Source is Canadian Human Mortality Database, Université de Montréal (2024).

The improvements in mortality after age 60 over the last fifty years provide the primary motivation for my concentration on longevity rather than all-ages mortality. To make this point more concretely, I aim to calculate how many years one might expect to live within each age decade (the teenage years, the 20s, the 30s etc.). I start by calculating a one-year survival rate as one minus the mortality rate at each age. Then these one-year survival rates are multiplied across ages from birth up to the age of interest to obtain the survival rate from birth to a given age. The sum of these survival rates across ages within an age-decade then yields the expected number of years lived within each age-decade. So, the sum of the survival rates at each age from 20 to 29 gives the expected number of years lived by those in their 20s. I repeat this calculation for the years 1921, 1971, and 2021 to see how the expected number of years lived in each age-decade has changed over the century from 1921 to 2021.

Figure 3: Life expectancy gains by age-decade



Notes: Survival rates at each age are added up for each age-decade (teens, 20s, 30s etc.) to yield the number of expected years lived by age-decade. The difference in these expected years lived by age-decades between 1921-1971-2021 are graphed. Source is Canadian Human Mortality Database, Université de Montréal (2024).

The result by age-decade is presented in Figure 3, with the lighter bars showing the change from 1921 to 1971 and the darker bars showing the change from 1971 to 2021. In the first 50 years from 1921-1971, 68 percent of the gains were under age 60. In sharp contrast, in the fifty years from 1971 to 2021, 84 percent of the gains are after age 60. Survival rates were already so high up to age 60 that continued improvement in mortality rates didn't yield much more in life expectancy at younger age ranges. For example, in 1921 the expected number of years lived from ages 20-29 was 9.1. By 1971, 20-29 year olds expected to live 9.8 years. This grew only a bit more to 9.9 years by 2021.

After 1971, the life expectancy gains from age 60 arise for two reasons. First, mortality improvements after age 60 mean that more people are surviving among those who have reached age 60. Second, because there were life expectancy gains before age 60 there are many more people who are arriving at age 60. These two factors combine to the large rise in the number of expected years lived in the 60s, 70s, 80s, and 90s. This evolution of where the life expectancy gains are happening provides a strong motivation for moving from the study of mortality to the study of longevity—understanding mortality patterns over the last fifty years largely involves understanding longevity.

Longevity

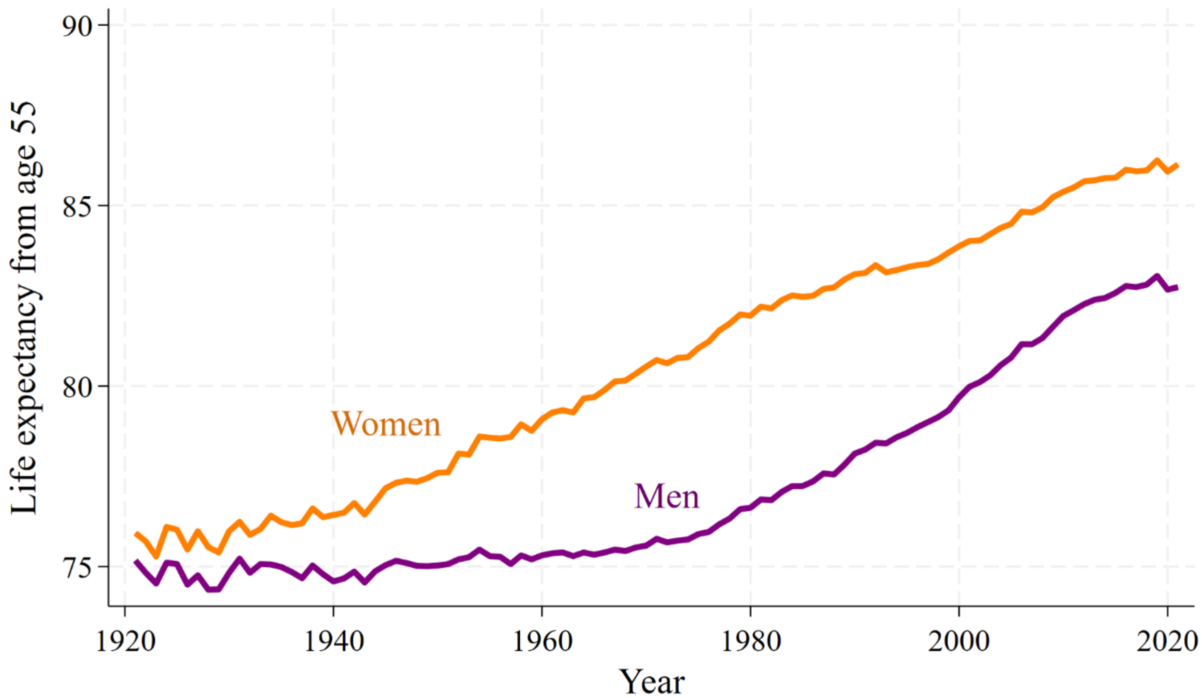
To investigate longevity, I move from looking at mortality rates in a given year to estimates of life expectancy. I begin with the standard life expectancy measure provided by most statistical agencies, including Statistics Canada. I then motivate the use of cohort life expectancies as a contrast to the standard measure. I present the main longevity results, followed by sensitivity tests of some important modeling assumptions.

Period and cohort life expectancy

The standard life expectancy calculation takes the set of mortality rates at each age for a given year, transforms the mortality rates into survival rates for a newborn making it to each age, and then sums these survival rates across ages to come to an estimate of life expectancy. This measure represents the lifespan expected by a newborn who experiences the age-by-age mortality rates observed in that year across the baby's lifespan. This is a cross-sectional or period life expectancy estimate.

To highlight longevity gains, I adjust the standard life expectancy from birth calculation to start at age 55, which yields the life expectancy given one has survived to age 55. I graph this life expectancy from age 55 for women and for men separately in Figure 4, for the years 1921 to 2021. Life expectancy for women after age 55 starts to grow after 1940, while for men life expectancy stagnated at around 75 until 1970. From then until 2020 there was steady growth in life expectancy until a slight reversal owing to the Covid-19 pandemic. For both men and women, this growth was substantial. In the fifty years from 1969 to 2019 women's life expectancy from 55 grew by 5.9 years, while for men the growth was 7.5 years.

Figure 4: Period life expectancy from age 55 by sex



Notes: Life expectancy from age 55 calculated from period mortality rates. Separate lines for women and for men. Source is Canadian Human Mortality Database, Université de Montréal (2024).

A shortcoming of the period life expectancy estimate is that it estimates the life expectancy of no actual person—the estimate assumes the age-specific mortality rates in a given year do not change as a newborn moves through life from age 0 to 100. Of course, mortality rates do change over time (and substantially so as shown in Figure 2) so this assumption of mortality rate stability over time is violated. Since no actual people experience the calculated life expectancy, period life expectancy is limited in the kind of questions it can answer. One obvious advantage for period life expectancy measures is the convenience in calculation, despite the difficulty in interpreting its meaning.

An alternative measure would summarize the life expectancy of actual people born in a given year. This kind of cohort life expectancy measure has clear advantages over a period life expectancy measure as the cohort measure can be used to evaluate the impact on lifespan of interventions that affect a given birth cohort over some part of its life. It also allows better tracking of gains in life expectancy over generations of people.

The major shortcoming of cohort life expectancy measures is the amount of time it takes to observe a cohort completing its lifespan. For people born in 1921, it would take 100 years to 2021 in order to observe and measure experienced cohort life expectancy up to age 100. For those born in 1971, we will need to wait until 2071 to observe cohort life expectancy up to age 100. This is too long to wait for answering important questions about policy impacts or demographic patterns.

However, it is possible to overcome this shortcoming of cohort life expectancy by using available information to estimate how much of each cohort survives to each age. I propose below a method for estimating cohort life expectancy. The method is primarily built on the method developed in Milligan and Schirle (2021), which borrows from Chetty et al. (2016).

The starting place for the method is an empirical regularity known as “Gompertz’s Law”. Observing Figure 2, the linear relationship between age and log mortality over most age ranges is evident. In his work 200 years ago (Gompertz 1825), English actuary Benjamin Gompertz documented the regularity that log mortality is linear in age. This finding has become known as

“Gompertz’s Law.” With observed age-specific mortality rates for a given birth cohort, future age-specific mortality rates can be estimated using Gompertz’s Law. For example, the 1950 birth cohort can be observed at age 55 in 2005, age 56 in 2006, and so on up to age 71 in 2021. Using a basic linear regression (specified below) of log mortality on age for the data up to age 71, one can project forward age-specific mortality rates for ages 72 onward. These mortality rates can then be transformed into survival rates to calculate life expectancies. Importantly, this can be done for any population for which age-specific mortality rates are available. This could be people from a specific year of birth cohort, of a certain sex, geography, income group, or family type.

I implement a cohort longevity estimation using the Longitudinal Administrative Databank (LAD). The LAD is a 20 percent sample of Canadian taxfilers who are followed longitudinally back to 1982 and forward to 2021. The sample universe includes those who have filed taxes at least once. The data report information found on income tax forms, including detailed and disaggregated measures of income, family characteristics, postal code of residence, year of birth and year of death. The LAD is an attractive data source because large samples are needed to study rare events like death in a given year, the depth of income information, and the fairly broad population coverage. Because of its longitudinal structure, LAD is often used for age-period-cohort modeling.[6] On the downside, LAD does not report education or any subjective or attitudinal information one might find in a survey.

[6] See for example Lehrer, Pan, and Finnie (2023).

My analysis is organized around longevity from age 55 onward. This requires that someone has survived until age 54. Given this age focus and the years available in the LAD, I can meaningfully analyze birth cohorts from 1930 to 1964.[7]

To form life expectancies from age 54 for a given population, I require a complete vector of mortality rates from ages 55 to 100. I form this vector of mortality rates for a given population in three parts. The first part is directly observed data. For example, for the 1950 birth cohort I observe actual mortality rates at each age from 55 (in 2005) to 71 (in 2021). Second, I use estimated mortality rates projected from the observed mortality rates using a Gompertz regression, as specified in more detail below. Third, from age 90 onward, I use sex- and age-specific mortality rates drawn from national lifetables for 2010. This choice to use national rates for age 90 onward results from poorer empirical performance in the Gompertz relationship after that age.[8] For most populations this assumption doesn't make a large difference to the estimated total life expectancy. Survival rates are then formed using the mortality rates at each age from 55 up to 100, and from these survival rates the life expectancy estimate can be formed. This methodology of using three ranges to form the mortality vector mirrors Milligan and Schirle (2021) following the model of Chetty et al. (2016).

The basic Gompertz relationship that could be used to estimate mortality for any group with data available for ages a is:

---

[7] An individual in the 1930 birth cohort is age 52 in 1982, so incomes and other characteristics from ages 52-54 can be observed in 1982. An individual in the 1964 birth cohort is age 57 in 2021, so mortality outcomes past age 54 can be observed for three ages (55, 56, and 57).
[8] See Gavrilov and Gavrilova (2011) and Gavrilova and Gavrilov (2014). This research documents the underestimation of mortality after age 90 using a Gompertz projection.

$$\log(mortality\ rate_a) = \beta_0 + \beta_1 age_a + e_a.$$

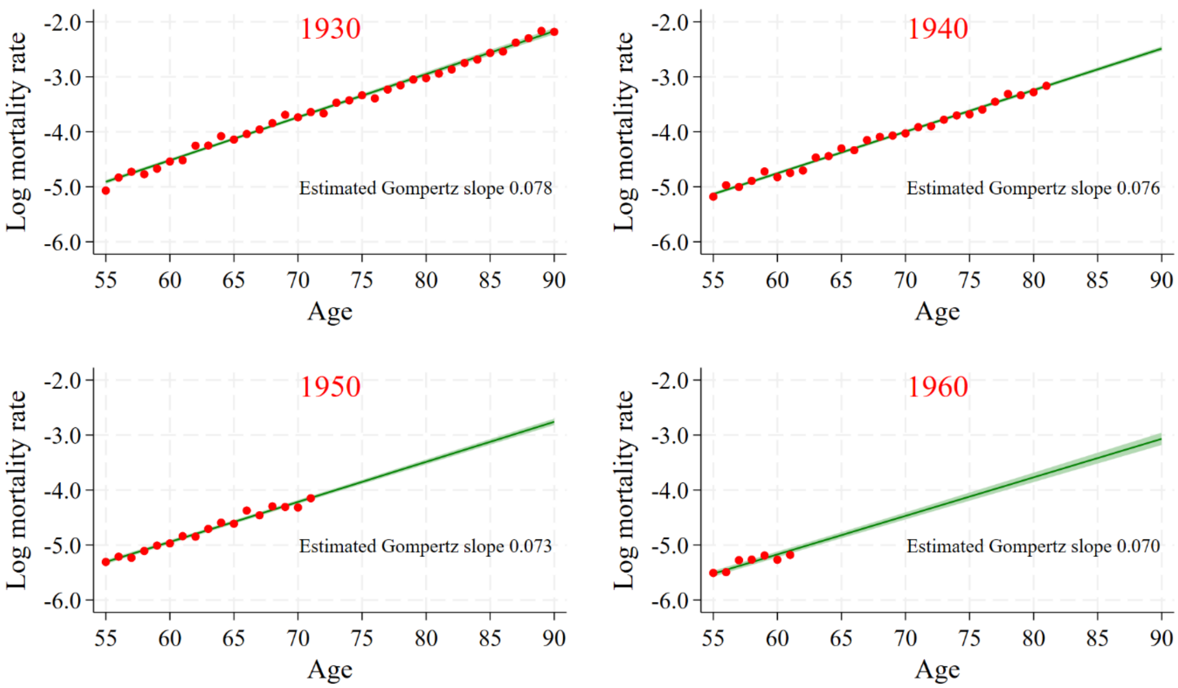
Here, β_1 is the estimate of the Gompertz slope term. The group might be defined by certain years of birth, sexes, places, or income groups. For younger cohorts, though, there are fewer data points available so the projections would be more highly leveraged off those few data points. For example, the 1960 birth cohort is observable at ages 55 through 61 when we have data up to 2021. Estimating a Gompertz equation on only a handful of data points would therefore lead to imprecise projections.

To improve the precision of the Gompertz projections, I pool the data across cohorts but allow each cohort to have its own intercept and further allow the Gompertz slope term to change linearly across cohorts. The equation that I implement therefore takes the following form for cells defined by ages *a* and birth cohorts *y*:

$$\log\left(mortality\ rate_{ay}\right) = \beta_0 + \sum \gamma_y + \beta_1 age_a + \beta_2 age_a \times coh_y + e_{ay}.$$

The γ_y terms are a set of cohort fixed effects which allows each cohort to have its own intercept. The linear coh_y term when multiplied by age_a allows for a linear drift of β_2 in the Gompertz slope. This pooled equation allows the projections sufficient flexibility to vary across cohorts both in terms of level and slope but delivers improved precision. For males, the estimated slope term is 0.078 (0.001) for the 1930 cohort with a small negative linear drift of 0.00028 per year of birth. This means that by 1960 the estimated slope falls to 0.070. For females the estimated slope term is 0.093 (0.001) for the 1930 cohort with a negative linear drift of 0.00095 per year of birth so that the 1960 estimated slope is 0.064.

Figure 5: Gompertz projections by birth cohort for males



Notes: Each dot is an observed log mortality rate for a given age and year of birth cohort. The line shows a forecast from a linear projection using the available data. The 95 percent confidence interval of the estimate is shaded around the line. The data are for males, taken from the Longitudinal Administrative Databank.

Figure 5 displays the data and the estimated log mortality-age relationship for males for birth cohorts from 1930, 1940, 1950, and 1960. The dots are the log mortality rates by age and the line shows the estimated relationship. Also shown in the figure is the estimated 95 percent confidence interval for the estimates, which is shaded around the line but also fairly small so it is hard to distinguish. The estimated Gompertz slope is noted in each of the four panels.

For the 1930 birth cohort, we have data up to age 91 in 2021. The tight log-linear relationship between mortality rates and age is very clear. For the other birth cohorts, data is only available up to a lower age cutoff so there are fewer data points available. This is where the pooled approach outlined above becomes important. The estimated slope falls from 0.078 per year of age for the 1930 birth cohort to 0.070 per year of age in 1960 which results from the small negative linear drift in the slope term. The standard error of the forecast is higher here, which is reflected in a wider confidence interval at higher age ranges. With the overall tight relationship between log mortality and age across cohorts, the main impression from these results is that the Gompertz relationship provides a very solid basis for projecting future mortality rates even for birth cohorts (like 1960) that currently have very few observed data points after age 55.

Main cohort life expectancy results

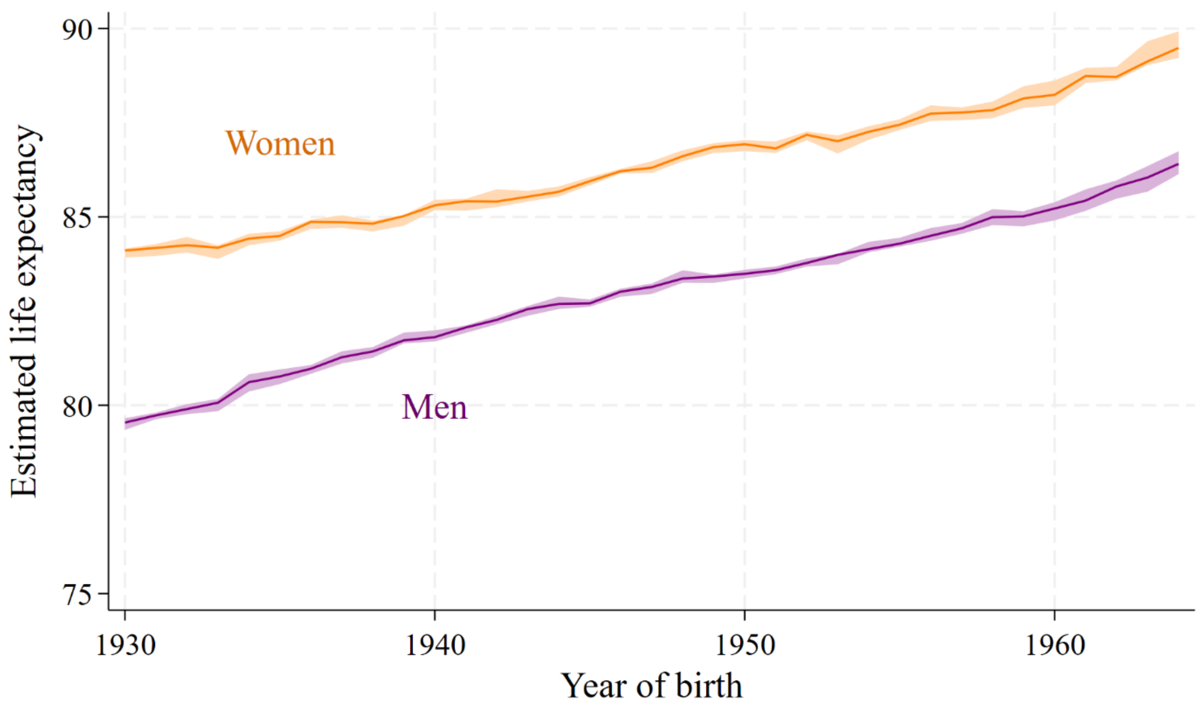
The estimates used to produce my main results are taken from a slightly augmented Gompertz regression which pools data across cohorts but allows cohort-specific slope and intercept terms in the following way, for age a and year of birth y. Age enters the regression linearly, while the year of birth enters as a vector of dummy variables YOB_y.

$$\log(mortality\ rate_{ya}) = \beta_0 + \beta_1 age_a \times YOB_y + \beta_2 YOB_y + e_{ya}.$$

In this regression, the estimated parameters β_1 and β_2 are vectors, with separate estimates for each year of birth. Using the projected mortality rates calculated from the estimates of these regressions along with the mortality rates in the data and the population mortality rates for ages

90 forward, I form a vector of survival rates for ages 55 to 100. These survival rates are then combined to form an estimate of the life expectancy. Simulated confidence intervals are formed by running 1000 simulations based on a perturbed vector of mortality rates where the observed mortality rate is replaced with a draw from a normal distribution using the estimated standard error.

Figure 6: Cohort life expectancy over time



Notes: Each line plots the estimated cohort life expectancy by year of birth, conditional on reaching age 54. Separate lines for women and for men. The simulated 95 percent confidence interval is shaded. The data are taken from the Longitudinal Administrative Databank.

Figure 6 shows the estimated life expectancy conditional on reaching age 54 for men and women across single-year birth cohorts from 1930 to 1964. For women, life expectancy grows by 5.4 years from 84.1 for the 1930 cohort to 89.5 for the 1964 cohort. The simulated confidence interval is wider for more recent cohorts owing to the greater variability of the estimates based on fewer observed mortality rates as was seen in Figure 5.[9] For men, the growth in life expectancy is 6.9 years from 79.5 for the 1930 cohort to 86.4 for the 1964 birth cohort.

Comparing the cohort life expectancies in Figure 6 to the period life expectancies in Figure 4 the estimated cohort life expectancies are clearly higher. This difference arises for two reasons. First, the sample of taxfilers used in the LAD for the cohort estimation is likely on average in better health than the sample of all Canadians used in the period life expectancy calculations. Second, the period life expectancy assumes no improvement in mortality rates over time. So, the period life expectancy assumes that the age 75 mortality rate of 2021's 65 year olds will be the same as the age 75 mortality rate of 2021's 75 year olds. Observing the levels of log mortality across the panels of Figure 5, the evident strong cross-cohort mortality improvements will not be captured in period life expectancy estimates. So, projecting forward the observed trajectory for a given birth year cohort leads to higher life expectancy estimates using cohort methods.

Life expectancy by income group

An advantage of the LAD is the ability to disaggregate the overall life expectancies by observable characteristics. Previous work by Milligan and Schirle (2021) using Canada Pension Plan administrative data revealed a strong gradient across lifetime earnings. I now proceed to

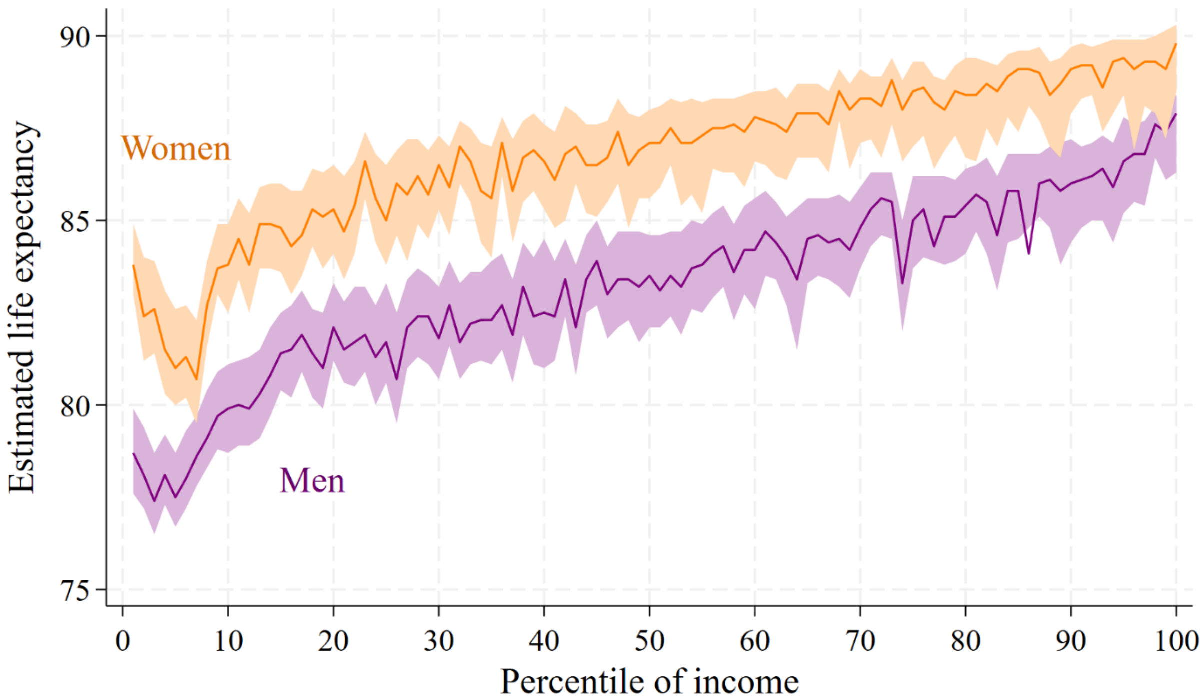
[9] These confidence intervals are not necessarily symmetric and in practice here the midpoint is below the estimate. Even with favourable draws on mortality, longevity cannot be pushed much higher for two reasons. First, high base mortality rates at older ages tend to cut off the long-term survival of a cohort even if favourable mortality draws occur at younger ages. Second, because I assume common mortality rates after age 90 favourable mortality draws have less impact. This is similar to the asymmetry of "hurricane cone" projections for the projected path of a hurricane that arise because certain paths of the hurricane run up against hard geographic barriers (like mountains or prevailing winds) that block the hurricane's path.

disaggregate each birth year cohort into income groups to see how the income gradient results using the LAD line up with the previous results in Milligan and Schirle (2021).

I measure income using after-tax family income, using an equivalence scale adjustment of the square-root of family size. I use the average of this measure of income over the ages 52-54, and put people in separate bins by percentile of this income distribution for the population of interest. The income measurement matters because individual income or earnings may not represent the living standard of a household which pools income across household members to fund its living standard. Milligan and Schirle (2021) used individual earned income; here with the LAD I have the ability to form a full range of income measures and compare the sensitivity of the results. Similarly, the age period for income averaging may matter if incomes are volatile; shorter periods may lead to more noise in the average as transitory elements are likely more present in shorter time spans.

For each of these assumptions (income definition and age-averaging period) I present sensitivity analysis below, but first I present the main results across income percentiles using family-size adjusted after-tax family income averaged over the age 52-54 period.

Figure 7: Life expectancy by income percentile



Notes: Each line plots the estimated cohort life expectancy by income percentile, conditional on reaching age 54. The income measure is family-size adjusted after-tax family income, averaged over the ages 52-54. Separate lines for women and for men. The simulated 95 percent confidence interval is shaded. The data are taken from the Longitudinal Administrative Databank.

For the income analysis, I pool the data across all years of birth between 1930 and 1964 but analyze men and women separately. The results are graphed in Figure 7. The lines show the projected life expectancy by income percentile, along with the simulated confidence intervals. The data reveal a strong, fairly linear gradient of life expectancy with income, but with several interesting elements. In the bottom ten percent of incomes, the life expectancy differences are non-monotonic. This result likely reflects a subpopulation that had limited labour force attachment, so their incomes reflect more the prevailing transfer income policies at the time of measurement than underlying aspects of the conditions they've lived over their lifespans. At the top of the income distribution, the data do not show strong evidence of convexity among the highest lifetime income percentiles. Further investigation within the top one percent would be interesting, but sample size limitations render such analysis difficult.

As a summary of this gradient, I take the averages over the bottom and top five percent. For women, the average life expectancy for the bottom five percent of income earners is 82.3, growing to 89.3 for the top five percent. For men, the change from the bottom to top five percent is 78.0 to 87.3. So, the female top-bottom difference is 7.1 years and the male difference is 9.3 years. These differences can be compared to previous work for Canada. Using Canada Pension Plan records with life expectancy from age 50, Milligan and Schirle (2021) find the top-bottom five percent difference is 2.4 years for women and 7.0 years for men. The Canada Pension Plan data, however, does not include those who don't have earned income so the LAD results presented in this paper have a broader sample of non-earners. This explains the steeper gradient presented here—especially for women—than in Milligan and Schirle (2021). Another comparison can be made to results from the United States in Chetty et al. (2016), which finds a top-bottom 5 percent difference of 7.9 years for women and 11.9 years for men using tax data and calculating life expectancy from age 40. While methodologies differ, the income gradients of life expectancy in the United States appear steeper.[10]

[10] The Chetty et al. (2016) paper uses household income but shows sensitivity to other measures of income. It focuses on age 40+ survival, using data only from 1999-2014 on a cross-sectional basis.

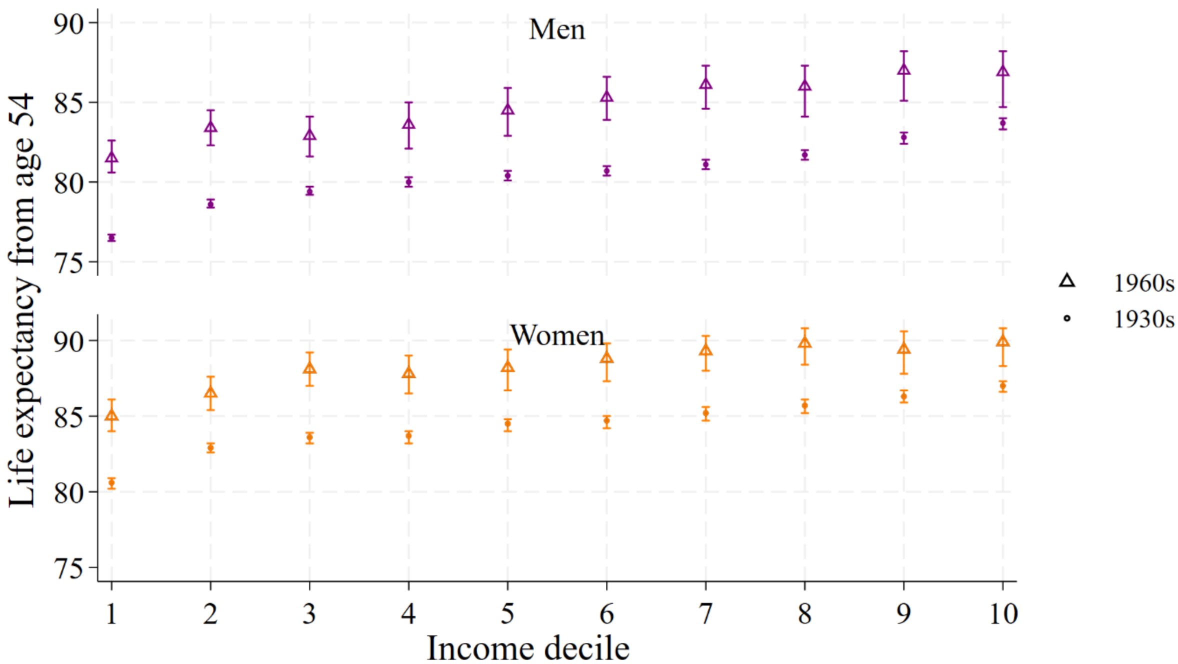
Life expectancy gains at low and high incomes

An important question is how this gradient in life expectancy has shifted over time. Milligan and Schirle (2021) show that the improvements in life expectancy across birth cohorts were uniform across lifetime earnings groups between year of birth cohorts from the 1920s compared to the 1940s. In the United States, National Academies of Sciences, Engineering, and Medicine (2015) find results that contrast sharply to the Canadian results—comparing birth cohorts from 1930 to 1960 there is no growth in life expectancy in the bottom two quintiles in the United States, but strong growth in the top quintiles.

In this paper with the LAD, I can check to see if the results in Milligan and Schirle (2021) continue to hold with newer birth cohorts now available and in a different data source. I now cut the data into income deciles and pool the data by decade of birth, comparing 1930s births to 1960s births.[11] Figure 8 presents the results. For women, life expectancy in the bottom decile grew by 4.4 years between 1930s and 1960s births while in the top decile the growth was 2.9 years. For men, the growth in the bottom decile was 5.0 years and 3.2 years in the top decile. There is no evidence of growth favouring those at the top; and the point estimates suggest inequality-reducing changes in life expectancy. However, with the 95 percent confidence intervals for the 1960s births being somewhat wide, the most conservative conclusion is to find no evidence against uniform growth across income deciles for both men and women.

[11] For the 1930s, all years between 1930 and 1939 are included. For the 1960s, only 1960-1964 are included.

Figure 8: Gains in life expectancy by income group



Notes: The data are divided into groups by sex, decade of birth, and decile of age 52-54 family-size adjusted after-tax family income. Each dot and triangle plots the average life expectancy from age 54 of the group, with the simulated 95 percent confidence interval indicated by the bars. The data are taken from the Longitudinal Administrative Databank.

This section of main results shows that useful and informative cohort measures of life expectancy can be calculated using administrative tax data. These calculations yield three important findings. First, there has been substantial growth in life expectancy from age 54 of about 5 years for women and 7 years for men between the 1930 and 1964 birth cohorts. Second, there is a strong gradient of longevity across family income quantiles, with a difference between the top and bottom five percent of 9 years for men and 7 years for women. Third, this income gradient has shifted fairly uniformly over time, with bottom earners increasing at least as much as high earners between 1930s and 1960s birth cohorts.

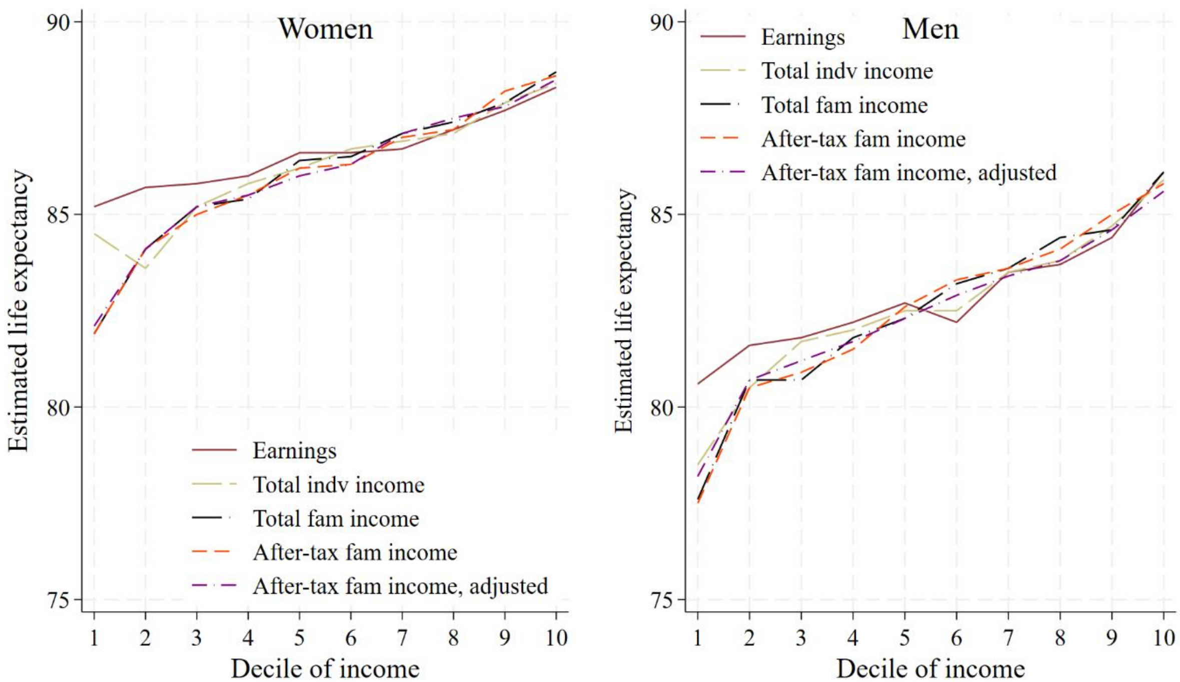
Sensitivity to income measurement and age averaging

Putting my results presented here in the context of previous evidence requires assessment of differences in methodology. A key advantage of the LAD is the availability of a broad set of income measures at both the individual and family levels. In Milligan and Schirle (2021) the Canada Pension Plan administrative data only reported employment income and wasn't aggregated at the family level. In particular, the results for women looking only at their own labour market earnings provided a limited view on female longevity because a substantial share of women in the birth cohorts under study did not participate fully in the labour market over their lives.

With the LAD, I construct five different income measures separately for women and men. The first measure is individual earnings. Then I use total individual income across all observed income types in the LAD. Next I use total family income, aggregating across all individuals in the LAD family. I then use the after-tax family income measure in the LAD, and then finally adjust for family size using a square-root equivalence scale.

Figure 9 shows the results for women and for men across income definitions. Here, I pool across all years of birth. For women, the gradient of longevity with respect to deciles of income is flatter for individual earnings. This is consistent with the flat gradient in Milligan and Schirle (2021) which also used individual earnings. Individual earnings for women may not be reflective of their living standard if they are partnered with a higher-earning partner. For all the other income measures however there is little difference to be seen for women (except for individual income in the first decile). For men, the pattern is similar; earnings shows a flatter gradient while broader income measures show sharper gradients across the first few deciles.

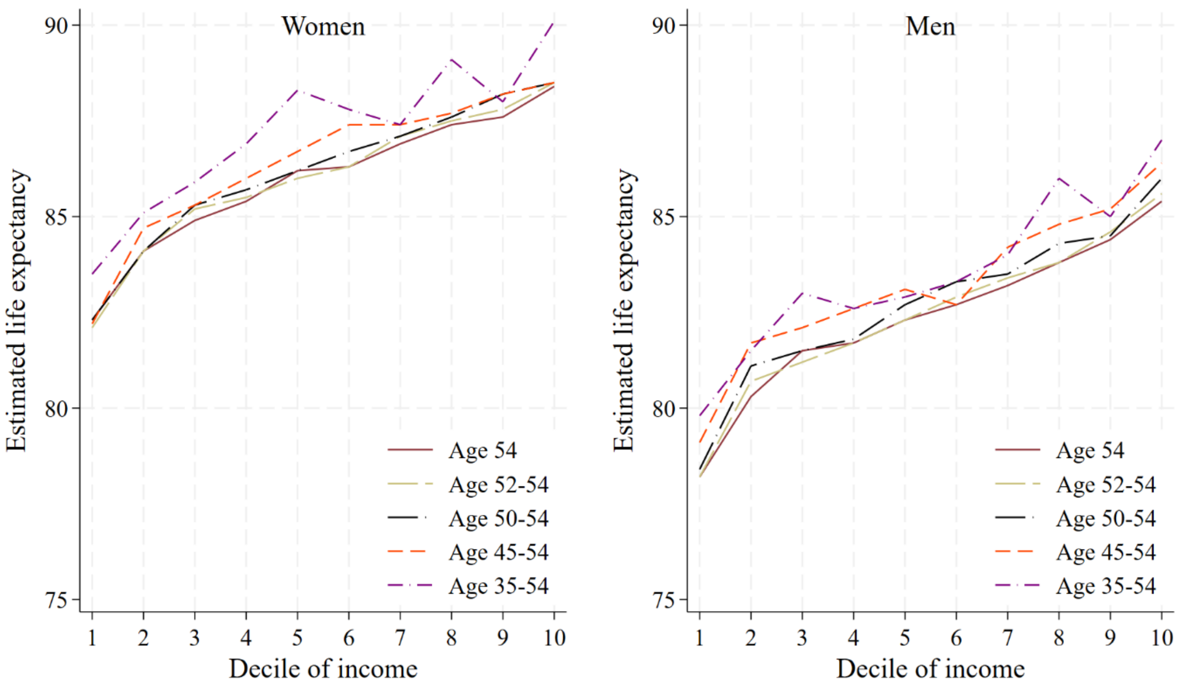
Figure 9: Sensitivity to income definitions



Notes: Each line shows the gradient of life expectancy from age 54 across income deciles for a different definition of income. Women are in the left-hand panel and men in the right-hand panel. The data are taken from the Longitudinal Administrative Databank.

These income definition sensitivity results are important for comparing longevity results across countries and data sources. While the results presented here suggest some caution in comparing individual earnings to broader income measures, little difference across total income measures is evident.

Figure 10: Sensitivity to age ranges



Notes: Each line shows the gradient of life expectancy from age 54 across income deciles for a different age range for income averaging. Women are in the left-hand panel and men in the right-hand panel. The data are taken from the Longitudinal Administrative Databank.

I also show sensitivity analysis with respect to the age range over which the income averaging is calculated. I estimate the gradient with respect to family-size adjusted after-tax family income averaged over 20 years (ages 35-54), 10 years (ages 45-54), 5 years (ages 50-54), 3 years (ages 52-54), and just age 54. Again, I pool across all available cohorts from 1930-1964 for this sensitivity analysis. One might expect that using shorter averaging periods introduces more transitory elements to income which might flatten the longevity-income gradient. However, the results displayed in Figure 10 show very little difference in the longevity gradient when using shorter or longer averaging periods. The longer windows have higher average mortality because they include only more recent cohorts for whom the long window of income is available.

This age-averaging sensitivity result matters because it gives confidence in using shorter averaging windows. A major downside of a longer averaging window is the data it requires—to do ages 35-54 for the averaging window limits me to beginning with year of birth cohort 1947 as the LAD only begins in 1982. The shorter averaging window allows earlier birth cohorts to be included.

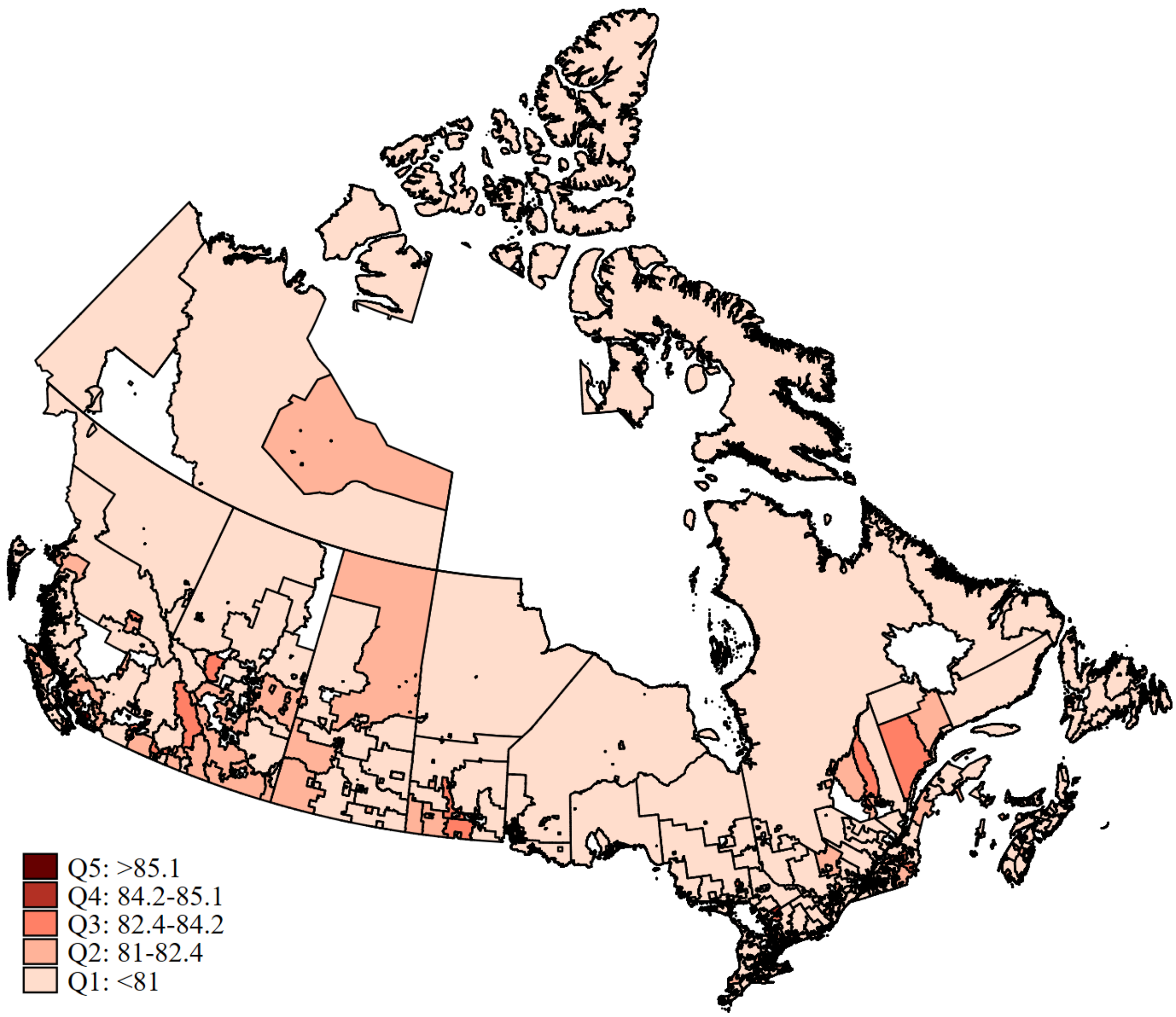
Longevity and geography

I now turn to the final analysis, comparing longevity across geography. For this analysis I pool the data across years of birth and sex but form separate samples by postal code. Using all six characters of the postal code does not provide enough sample to perform the cohort longevity projection, so I use the first three characters of the postal code (called the Forward Sortation Area, or FSA). I tag the FSA where someone lives at age 54 for this analysis.[12] There are about 1,600 FSAs across Canada and I impose a sample size restriction of at least 100 observations in the LAD, which allows sufficient sample to conduct the longevity estimation. About 10 percent of the FSAs are removed because of this restriction; mostly rural lower-populated areas.

There are two parts to the analysis below. First, I map the average longevity by FSA, showing quintiles of longevity across FSAs on a national and local scale. Second, I compare people of similar incomes who live in FSAs with on average high and low income in order to gain insight into the influence of location and neighbourhoods on longevity.

[12] I choose age 54 since that is the focal point for measuring income. Examining the role of residence at different points in one's life is potentially very interesting but left for future work.

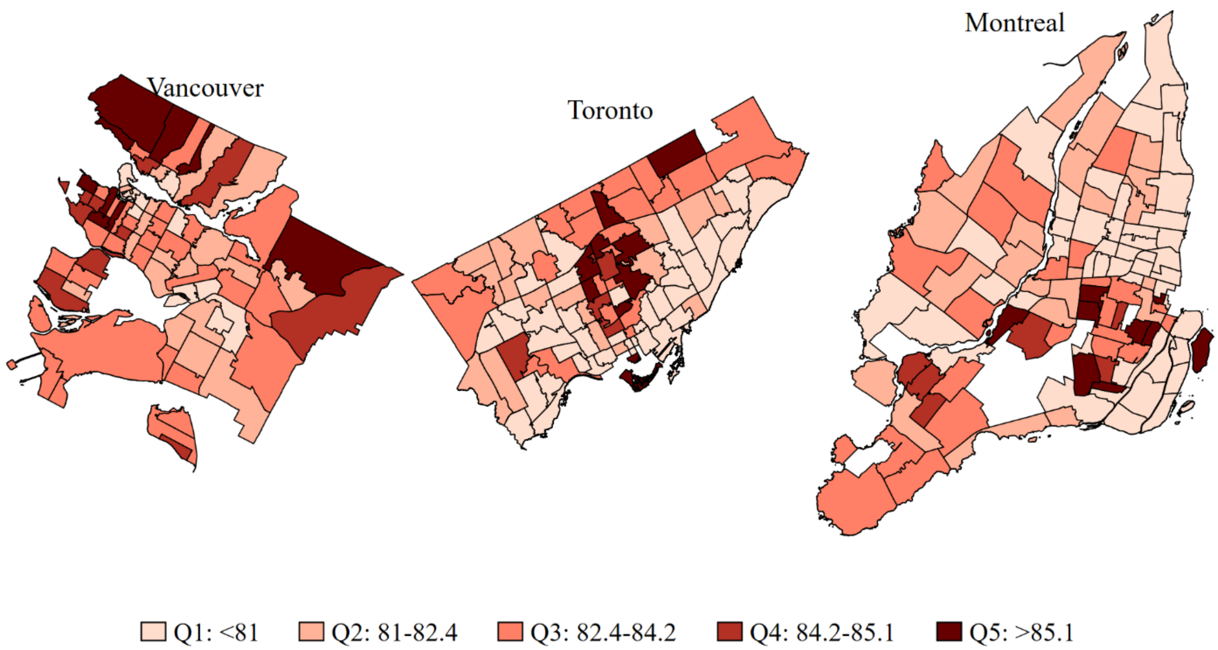
Figure 11: Life expectancy by postal code



Notes: Life expectancy from age 54 is mapped by Forward Sortation Area, with quintiles shaded. White areas are data suppressed because of limited sample size. The data are for men, pooled across years of birth 1930-1964. The data are taken from the Longitudinal Administrative Databank.

The maps shade FSAs by quintile of longevity, with the quintile cutpoints fixed nationally by sex. The first map in Figure 11 shows the pattern across FSAs for all of Canada. The dominant difference in longevity is urban-rural. The geographically-large FSAs in rural areas typically have lower longevity, which displays as lighter-shaded and tends to visually dominate the map. The higher longevity quintiles are concentrated in very compact FSAs in cities and so can’t be seen in the country-scale graphic.

Figure 12: Life expectancy within cities



Notes: Life expectancy from age 54 is mapped by Forward Sortation Area, with quintiles shaded. White areas are data suppressed because of limited sample size. The data are for men, pooled across years of birth 1930-1964. The data are taken from the Longitudinal Administrative Databank.

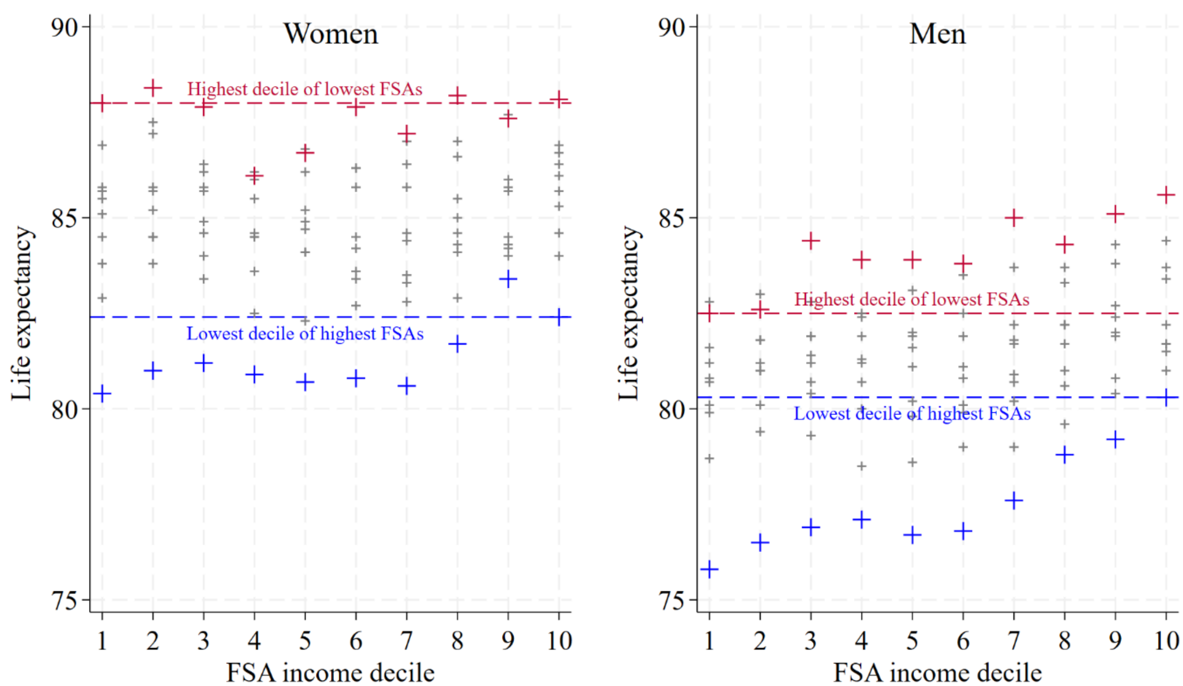
In Figure 12 I show the longevity patterns for the three largest cities in the country: Vancouver, Toronto, and Montreal. In all three cities there are stark differences across geography, which largely line up with patterns of income across these cities. For example, V6A (Vancouver-Downtown Eastside; a low-income part of Vancouver) has a life expectancy of 74.7 for men while V6M (Vancouver-Shaughnessy; a high-income part of Vancouver) is 87.6. In Toronto, life expectancy for women is 75.3 in M5G (Toronto-Downtown Core) but 89.6 in M4N (Toronto-Lawrence Park). Similar patterns across high and low socio-economic status neighbourhoods of Montreal are evident. These findings show large longevity differences of more than a decade in life just a few kilometres apart within the same cities.

The evidence points toward income as a very important factor in explaining longevity differences. I can dig into how important income might be by comparing high-income people in low-income neighbourhoods with high-income people in high-income neighbourhoods. If longevity is about the same in these different neighbourhoods that would suggest that income alone is a strong explanatory factor.

I explore this analysis in Figure 13 by looking across family incomes and neighbourhoods. Along the X-axis are ten groups of FSAs, arranged by the average income in each FSA. The bottom-income FSAs on the left in FSA income decile 1 has the 10 percent of Canadians living in the lowest-income FSAs. The highest-income FSAs on the right in FSA income decile 10 has the 10 percent of Canadians living in the highest-income FSAs. Along the Y-axis in Figure 13 I put a marker for the average longevity within each family income decile for that group of FSAs, where these family income deciles are formed using the national family-size adjusted after-tax family income pool.

The mean after-tax family income in the top income decile is $258,000 in 2021 dollars. In the left-hand panel for women, there is little difference in the longevity of women living in families who have top-decile incomes no matter what the average income of the FSA they live in. For the lowest decile of income, the average after-tax family income is $18,700. For low-income women in the highest-income FSAs the estimated life expectancy is 82.4, which is higher than the 80.4 for those in the lowest-income FSAs. So, there is more evidence of a difference across neighbourhoods for women who themselves have low income than for those who have high income.

Figure 13: Life expectancy by income and postal code



Notes: Forward Sortation Areas are grouped into ten deciles by average income across the X-axis, so that each decile contains 10 percent of Canadians by sex. Average life expectancy from age 54 is then taken by national income decile with each group of FSAs and graphed by tick marks horizontally. Data are pooled across years of birth 1930-1964. The data are taken from the Longitudinal Administrative Databank.

For men, there is a clearer gradient across all own-family-income groups. For men within this top income decile but living in a low-income FSA, the average longevity is 82.5. But in that same top income decile the average longevity in the 10th FSA decile is 85.6, which is three years longer. The difference for those with low family income is a bit stronger. A man in a low-income family in a low-income FSA has life expectancy of 75.8. But in the high-income FSA those men with the same low family income have life expectancy of 80.3. Put differently, someone of low income ($18,700) in the high-income neighbourhood has the same life expectancy of someone in the 4th decile of income ($66,700) in the low-income neighbourhood. Going the other way, you have to be in the highest income decile ($258,000) if you live in the low-income neighbourhood to have the same life expectancy as someone in the 5th household income decile ($79,300) in the high-income neighbourhood.

These income and neighbourhood results strongly suggest that where you live matters a lot for explaining longevity differences (at least for men), over and above the amount of family income. I pick up the discussion of what might be contributing to these geographic differences in the final section below.

Discussion

The analysis in this paper has shown that longevity varies substantially across cohorts, lifetime-income groups, and geography. In this section, I discuss potential underlying mechanisms.

The first to consider is the direct effect of income. Those with higher lifetime income can afford to consume more of everything, including potentially longevity-enhancing goods such as nutritious food, lifestyle advice, healthcare, and amenable housing. While income may explain some part of the differences in longevity the analysis presented above shows income does not explain all the differences—men with similar incomes live longer in higher-income neighbourhoods. In Milligan and Schirle (2021), the analysis shows that differences in longevity across birth cohorts are larger than can be explained by income growth across birth cohorts. So, income may be an important explanation, but not a complete explanation.

A second important factor may be access to quality health services. Across birth cohorts, health insurance coverage and the range and depth of health treatment expanded in Canada. Similarly, across geography the availability and quality of health services surely differs. At a basic level, those in rural areas may have to travel for hundreds of kilometers for specialized (or even basic) medical treatment while those in urban areas have more ready access to services. More subtly, the quality of health services could also vary by geography.

Research from other countries supports the importance of health care in understanding longevity trends across time, income, and space. Recent Dutch evidence in Danesh et al. (2024) finds that the onset of chronic diseases in the lower half of the income distribution arrives on average 15 years ahead of those in the top half of the income distribution. In Sweden, Hagen et al. (2024) look at trends in causes of death across the income distribution, finding that preventable and treatable deaths are more prevalent lower in the income distribution and posit that health behaviours and quality of health care both may play some role. In the United States, Badinski et al. (2023) find large differences in physician practices across regions which supports the notion that supply-side health factors matter for understanding geographical differences.

Beyond access to health services, a third explanation is the availability of other amenities that might vary by neighbourhood or across birth cohort. These amenities might be the environment, public services like parks and bike paths that facilitate recreation and exercise, or they might be private services like retail stores and professional services that vary in quality across neighbourhoods. For example, a voluminous literature in public health and urban planning (see for example Sallis et al. 2009) focuses on correlations between a neighbourhood's "walkability" and health outcomes. If these kind of neighbourhood amenities themselves—or the peer information provided from watching others engage in healthy activities—matter, then these kinds of neighbourhood effects could influence longevity. In addition, the cleanliness of the air can vary sharply across neighbourhoods and a large literature finds evidence of impacts on health (e.g. Currie and Walker 2011 on automobile exhaust and infant health) and life expectancy (e.g. Ebenstein et al. on air pollution in China).

Finally, the observed correlations discussed above could be driven mostly by unobservable characteristics. The correlations might arise because of unobserved "third factors" that affect both health and the decision about where you live, as emphasized by Fuchs (1982). In this explanation, there is not a causal relationship between income, health services, or neighbourhood that can be manipulated easily by policy because outcomes are largely driven by an immutable underlying 'type' that self-selects across neighbourhoods. To separate selection of this type from causal stories of course requires careful empirical work with credible identification—but much work remains to be done to understand the deep differences in longevity.

Conclusion

This paper studies the longevity of Canadians using tax administrative data. There is a century of documented improvements in average mortality in Canada, with most recent advances coming at older ages. To study these advances in longevity, a method for extrapolating life expectancy from age 54 by cohort is presented, and the method is shown to be robust to income measure and income averaging period.

Three main findings emerge from the analysis. First, there is a strong gradient of longevity with income for both men and women, with low-income Canadian men living about 9 years less than those at high-income levels, and low-income Canadian women living about 7 years less than those at high-income levels. Second, this longevity gradient shifts up across birth cohorts approximately uniformly across income groups. Canadians across all income groups gained in longevity. This is in strong contrast to results from the United States showing growth in longevity only in the top half of the income distribution. Third, sharp differences in longevity across geography are documented, which are only partially explained by income differences across areas. Within cities, average longevity can vary by a decade or more across neighbourhoods; substantial longevity differences remain across neighbourhoods even when comparing those with similar incomes.

This research leaves open several important and interesting questions. First, decomposing the growth in longevity across contributing factors like income, health care, and neighbourhood amenities can help guide policy that might aim to boost longevity where it lags. Second, understanding more about when in life exposure to incomes, health care, and neighbourhood amenities would be helpful: is it where you are born, where you are schooled, or where you live during your worklife that matters most? Further research can help answer these questions.

References

AlFakhri, M., and J. Compton (2023) “Life expectancy of couples in Canada,” *Canadian Public Policy* 49(S1), 76-93

Badinksi, I., A. Finkelstein, M. Gentzkow, and P. Hull (2023) “Geographic Variation in Healthcare Utilization: The Role of Physicians,” NBER Working Paper No. 31749

Public Health Agency of Canada (2024) “Opioid- and Stimulant-related Harms in Canada,” Federal, provincial, and territorial Special Advisory Committee on the Epidemic of Opioid Overdoses. https://health-infobase.canada.ca/substance-related-harms/opioids-stimulants/. Accessed July 31, 2024

Canadian Human Mortality Database. Department of Demography, Université de Montréal (Canada). [dataset] Available at www.demo.umontreal.ca/chmd/ (data downloaded on April 27, 2024)

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

Crow, K. 2006. "SHP2DTA: Stata module to converts shape boundary files to Stata datasets," Statistical Software Components S456718, Boston College Department of Economics, revised 17 Jul 2015.

Currie, J. and R. Walker (2011) “Traffic Congestion and Infant Health: Evidence from E-ZPass,” *American Economic Journal: Applied Economics* 3(1), 65-90.

Danesh, K., J.T. Kolstad, J. Spinnewijn, and W.D. Parker (2024) “The Chronic Disease Index: Analyzing Health Inequalities Over the Lifecycle,” NBER Working Paper No. 32577

Ebenstein, A., M. Fan, M. Greenstone, G. He, and M. Zhou (2017) “New evidence on the impact of sustained exposure to air pollution on life expectancy from China’s Huai River Policy,” *Proceedings of the National Academy of Science* 114(39), 10384-10389

Etches, J. (2009) “Economic inequality in adult mortality in Canada: Analyses of the Longitudinal Administrative Databank. Doctoral Dissertation, University of Toronto

Feir, D., and R. Akee (2019) “First Peoples lost: Determining the state of status First Nations mortality in Canada using administrative data,” *Canadian Journal of Economics* 52(2), 490-525

Fuchs, V. R. (1982) “Time Preference and Health: An Exploratory Study” in V.R. Fuchs (ed.) *Economic Aspects of Health*. Chicago: University of Chicago Press. 93-120

Gavrilov, L. A., and N. S. Gavrilova (2011) “Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File,” North American Actuarial Journal 15(3), 432–447

Gavrilova, N. S., and L. A. Gavrilov (2014) “Mortality trajectories at extreme old ages: A comparative study of different data sources on US old-age mortality,” in Living to 100 monograph, Society of Actuaries

Gompertz, B. (1825) “On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies,” Philosophical Transactions of the Royal Society of London 115, 513–583

Hagen, J., L. Laun, C. Lucke, and M. Palme (2024) “The Rising Income Gradient in Life Expectancy in Sweden over Six Decades. Unpublished paper, Stockholm University

Health Canada (2022) “Third annual report on medical assistance in dying in Canada 2021,” Catalogue No. H22-1/6E-PDF, Ottawa ON.

Heger, D. (2018) “Decomposing differences in health and inequality using quasi-objective health indices,” *Applied Economics* 50(26), 2844-2859.

Jürges, H. (2007) “True health vs response styles: exploring cross-country differences in self-reported health,” *Health Economics* 16(2), 163-178

Kapteyn, A., J.P. Smith, and A. van Soest (2007) “Vignettes and self-reports of work disability in the United States and the Netherlands,” *American Economic Review* 97(1), 461-473

Lehrer, S.F., Y. Pan, and R. Finnie (2023) “Evolution of Gender Patterns in Retirement Saving in Canada,” *Canadian Public Policy*, 49(S1), 6-31

Milligan, K., and T. Schirle (2021) “The Evolution of Longevity: Evidence from Canada,” *Canadian Journal of Economics* 54(1), 164-192

Milligan, K. and T. Schirle (2024) “Retirement incentives and decisions across the income distribution: Evidence in Canada,” unpublished manuscript

National Academies of Sciences, Engineering, and Medicine (2015) “The growing gap in life expectancy by income: Implications for federal programs and policy responses,” Technical report, National Academies of Sciences, Engineering, and Medicine, Washington DC

Pisati, M. 2007. "SPMAP: Stata module to visualize spatial data," Statistical Software Components S456812, Boston College Department of Economics, revised 18 Jan 2018.

Sallis, J.F., B.E. Saelens, L.D. Frank, T.L. Conway, D.J. Slymen, K.L. Cain, J.E. Chapman, J. Kerr (2009) "Neighborhood built environment and income: Examining multiple health outcomes," *Social Science and Medicine* 68(7), 1285-1293

Statistics Canada. 2021. Boundary Files, 2021 Census. [dataset]. Available at https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/boundary-limites/index2021-eng.cfm?year=21 (data downloaded April 4, 2024)

Tanuseputro, P., W. P. Wodchis, R. Fowler, P. Walker, Y.Q. Bai, S. E. Bronskill, and D. Manuel (2015) "The Health Care Cost of Dying: A Population-Based Retrospective Cohort Study of the Last Year of Life in Ontario, Canada," *PLOS ONE* 10(3): e0121759.

Université de Montréal (2024) "Canadian Human Mortality Database," Department of Demography. Available at http://www.bdlc.umontreal.ca/chmd/ (data downloaded on April 17, 2024)

Wolfson, M., D. Chapman, J.H. Lee, V. Bijelic, and S. Woolf (2024) "Extent and socioeconomic correlates of small area variations in life expectancy in Canada and the United States," *Health Reports*, Statistics Canada Catalogue no. 82-003-X.