Three Essays on the Economics of Health and Well-Being

by

Paul J. Spin

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
April 2018

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation offers three vignettes on the determinants of health and well-being over the life course. The first essay uses multiple Canadian census files to document the long-term effects of *potential* in utero exposure to the 1918 flu pandemic on educational attainment. This study finds that those who were in utero during the peak of the pandemic, particularly in their first or second trimesters, experienced long-term deficits in their educational attainment. The second study explores the potential impact of spousal institutionalization in nursing homes/residential care facilities on elderly financial security. It shows that the absence of fully funded universal long-term care insurance (like Canadian medicare) places married seniors at risk of significant losses in their material standards of living and low income status. The third paper examines the impact of online communication and social media use on subjective well-being (SWB). In one empirical approach, I find that those who communicate online or use social media report lower levels of SWB. This is especially true for older adults and social media. In a separate quasi-experimental analysis that exploits variation in access to and use of social media by time, age group, and access to personal computers, I find that social media may be responsible for increased political engagement and social trust.

List of Abbreviations Used

AB	Alberta
BC	British Columbia
CCRI	Canadian Century Research Infrastructure
CHT	Canadian Health Transfer
CI	Confidence Interval
CPP	Canadian Pension Plan
DDD	Differences-in-Differences-in-Differences
GIS	Guaranteed Income Supplement
GSS	General Social Survey
HY	Half-Year of Birth
IPUMS	Integrated Public Use Microdata Series
LIM	Low Income Measure
LTC	Long-Term Care
MN	Manitoba
NB	New Brunswick
NL	Newfoundland and Labrador
NS	Nova Scotia
OAS	Old Age Security
OLS	Ordinary Least Squares
ON	Ontario
PE	Prince Edward Island
RLTC	Residential Long-Term Care
SD	Standard Deviation
SES	Socio-economic status
SLID	Survey of Labour and Income Dynamics
SWB	Subjective well-being
U.S	United States
WWI	World War I
WWII	World War II
YOB	Year of Birth

Acknowledgements

Words cannot describe how grateful I am to my supervisor, Shelley Phipps, whose wisdom and kindness inspire me as an economist and a person. I could not have completed this dissertation without her motivation and support. Her selfless dedication to bringing out the most in others sets an example that I can only hope to follow.

I owe a debt of gratitude to Courtney Ward for her collaboration, encouragement, and continuous feedback. I thank Peter Burton for his immeasurable contribution to my supervisory committee over the last several years. I am thankful to Martha MacDonald for opening her home and her well-stocked supply of coffee for many hours of policy discussion and brainstorming. Casey Warman deserves a special thank you for reviewing my dissertation and providing valuable feedback throughout my PhD studies.

I thank Heather Hobson for painstakingly vetting all of my research output, often on short notice. Special thank you to Lars Osberg, Mutlu Yuksel, Mevlude Akbulut-Yuksel, and Yigit Aydeded for their comments and suggestions over the years.

I am thankful to my mother for her unconditional support, and to my late father, who would have been proud. I owe much to my second father, Bill Montelpare, for his generosity, love of British ale, and enthusiasm for statistics. Thanks to Nancy and Erin for welcoming me into their family.

A special thank you to my fellow comrades and friends, Angela, Arefin, Mook, and Nancy, for making the last several years so enjoyable.

I am deeply thankful for Krista, who, for reasons that I will never understand, took a chance on a destitute graduate student. Her love and patience helped me complete this thesis, while her humour, openness, and commitment to serving others remind me how to be a better human

Chapter 1: Introduction

This dissertation offers three vignettes that document the effects of environmental circumstances, government policy, and technology in shaping health and well-being over the life course. Chapter 2 explores the role of the fetal environment in shaping long-term human capital generation and earnings. Chapter 3 considers the potential financial repercussions associated with spousal institutionalization for long-term care. Chapter 4 considers the transformative role of the internet by examining the impact of online communication and social media on subjective well-being.

Numerous studies have argued that educational attainment is causally related to conditions that arise early in life, particularly in utero (Almond & Currie, 2011). Epidemiological models of human development have stressed that fetal stressors (e.g. maternal stress, nutritional deficiency) can trigger permanent metabolic "programming" that trigger irreversible damage to long-term health. The potentially dire consequences associated with permanent programming of poor health have inspired economists to apply these theories to advance our understanding of the role of early life conditions in long-term economic outcomes and human capital accumulation.

Among the most influential of these economic contributions was a seminal study of Douglas Almond (2006), who demonstrated that children who were in utero at the time of 1918 flu pandemic experienced permanent reductions in their long-term educational and earnings outcomes compared to unexposed controls. While the 1918 flu pandemic is generally seen as an exogenous effect, recent evidence suggests that Almond's findings could be explained by the U.S. WWI mobilization, which diverted young fertile men (via conscription) away from their families toward the war effort.

Chapter 2 of this dissertation re-examines the legacy of the 1918 flu pandemic in the context of Canada, which is demographically and economically similar to the U.S. My analysis uses recently distributed Canadian historical census files, which allow for the use of control variables that capture cohort-level averages of parental characteristics in cohorts born around the time of the pandemic. Unlike the U.S. literature (Brown and Thomas 2011), I show that the effect of prenatal overlap with peak pandemic conditions have adverse effects on long-term education, even after conditioning on cohort-level parental characteristics. I also document adverse effects in French and English Canadians (and in Quebec-born and Ontario-born populations), which had markedly different rates of military enlistment during WWI. This finding re-affirms Almond's (2006) seminal findings as well as those that have been documented internationally for countries in Europe, Asia and South America.

My third chapter examines the potential impact of spousal institutionalization on elderly financial security. Although Canada provides universal public health insurance for acute health care, publicly subsidized long-term care programs typically feature means-tested user fees. This study shows that "potential" copayments, i.e. the cost that would arise if a person's spouse were to enter care, range from 40-55% of equivalent disposable income, depending on the province. Such payments are implicated as a potentially significant source of spousal impoverishment and loss of material standard of living. I show that protections against such losses can be implemented by providing larger spousal allowances or more favourable divisions of family income. While many studies of elderly financial security have focused on the role of spousal mortality, this study shows that long-term illness remains a serious threat to the financial well-being of older adults, despite the availability of universal medical care and a robust retirement income security system.

The final chapter estimates the effect of social media and online contact on subjective well-being (SWB) nationally representative data from multiple waves of the General Social Survey. While the impact of social media on well-being and public health is controversial, it has received limited study in population-level datasets.

I use two approaches which differ in their ability to account for confounding variables. The first approach involves cross-sectional comparisons of SWB of users and non-users of social media and online communication. The results indicate that people who communicate online report lower SWB, though the effect decreases as the frequency of contact increases. The negative effects on SWB are especially significant in older adults. These effects are larger for social media use than for online contact. For non-seniors I find that SWB is largely uncorrelated with online contact or social media use.

My second approach exploits variation in social media use by age and pre-social media computer ownership rates using a continuous triple-differences regression. I show that the impact of regional rates of computer ownership on post-2004 social media utilization is larger for non-seniors than for seniors. Using this source of variation, I find no corresponding effect on life satisfaction. However, I report evidence of a beneficial effect of social media access on voter turnout and social trust. These findings suggest that social media likely has limited causal effect on self-reported happiness. Global measures of life satisfaction may lack the specificity necessary to capture the impact of social media use. That voter turnout and social trust exhibit positive associations with social media use suggests that the latter may be helpful in activating pro-social interactions and facilitating political engagement (e.g. via news circulation, group coordination, online volunteering, etc.).

Chapter 2: The 1918 Flu Pandemic in Canada: Revisiting the Long Run Effects on Educational Attainment

2.1 Introduction

The fetal origins hypothesis states that intrauterine adversity can produce long-term negative effects on postnatal health (Barker, 1997).[1] While initially limited to epidemiology, this hypothesis has since been used by social scientists to model human capital accumulation, health production, cognitive development and skill formation (Almond & Currie, 2011; Cunha, Heckman, Lochner, & Masterov, 2006). In the empirical literature, a central question for economists is whether fetal adversity is simply a sign of postnatal adversity, such as low family income. If so, then these factors may account for any observed long term statistical correlations between fetal conditions and adult outcomes. To correct for this bias, most studies have used natural experiments to isolate the fetal environment from other omitted confounders (see Almond & Currie, 2011 for a review).

As an example of this approach, this study exploits the Canadian episode of the 1918 flu pandemic to identify exogenous variation in fetal conditions. The pandemic was an unanticipated, short-term and large-scale health shock. As shown in Figure 2.1, the pandemic arrived in October 1918 and generated a pronounced three-month spike in flu

[1] Many studies (Barker, 2012; Almond & Mazumder, 2011; Painter, et al., 2006; de Rooij, et al., 2010; Victora, et al., 2008; Chen & Zhou, 2007; Stein, et al., 2006; Roseboom, et al., 2011) have implicated fetal malnutrition as a risk factor of heart disease, metabolic disorders (e.g. high blood pressure, impaired glucose response), disability, Type-II Diabetes, increased adiposity, decreased stature. Other risks (Galler, et al., 2013; Susser & St Clair, 2013; Christian, et al., 2010; Victora, et al., 2008; St Clair, et al., 2005; Susser & Lin, 1992) may include adverse personality traits, mental illness and schizophrenia. These effects may be triggered by adverse and irreversible prenatal metabolic and neurological programming (see Hales and Barker, 1992; Bale et al., 2010). Recent scholarship has extended this hypothesis to include epigenetic (e.g. gene-environment) reconfigurations of gene expression (Cao-Lei et al., 2014).

and pneumonia mortality. The cause of this spike was an abrupt mutation in the genes of the virus, known as antigenic drift, which seriously weakened human immune responses to infection (Taubenberger and Morens, 2006). Because genetic mutations in the flu virus are inherently random, the pandemic was completely unanticipated. Furthermore, even after the pandemic arrived, media censorship in the warring countries (e.g. Canada, Britain, U.S., Germany) constrained public knowledge of its severity (in terms of infection and/or mortality risk) and geographic distribution.

Such traits are useful for identification purposes. For example, the sudden and short-term nature of the pandemic rule out parental avoidance of the pandemic via fertility postponement or out-migration (i.e. to less affected regions). If such avoidance mechanisms were available and correlated with parental determinants of child success (e.g. wealth, income, education), then the long-term effects of *in utero* exposure to the pandemic would remain confounded by omitted variables. A further advantage is that the long-term effects of fetal exposure to the pandemic are clearly predicted by the fetal origins hypothesis. In particular, it predicts that those who were *in utero* during the peak of the pandemic should have worse health and lower education on average. This prediction can be easily tested using data with time of birth and place of birth identifiers, such as the Canadian census.

A number of previous studies have exploited the 1918 pandemic as a natural experiment in settings other than Canada, with most generating supportive results (Lin and Liu, 2014; Brown and Thomas, 2011; Richter and Robling, 2013; Neelsen and Stratmann, 2011; Nelson, 2010; Garthwaite, 2008; Almond, 2006; Almond and Mazumder, 2005). For example, in the U.S., fetal exposure to the pandemic is associated with less education, lower self-reported health, increased disability, and higher risks of diabetes and cardiovascular disease (Garthwaite, 2008; Almond, 2006; Almond and Mazumder, 2005). Similar results have been reported for data in other countries and other pandemics.[2] Yet, the experimental validity of the 1918 flu pandemic in the United States may be compromised by its proximity to WWI mobilization. For example, Brown and Thomas (2011) argue that this seriously confounds the long-term estimates associated with fetal exposure to the pandemic. Central to their case is the fact that U.S. mobilization began in late 1917, only one year before the pandemic, and was supported by a comprehensive national draft that enlisted relatively young and healthy males, and, apparently, men of relatively high socio-economic status (SES).[3] Brown and Thomas (2011) show that these aspects of U.S. mobilization led to a contemporaneous and discontinuous decrease in parental SES among children who were also fetally exposed to the pandemic. Furthermore, using the same data as Almond (2006), they report no adverse educational effects from *in utero* exposure to the pandemic in models that control for observable parental confounders.

While Canada experienced similarly adverse epidemiological consequences during the pandemic, its experience is less confounded by WWI. Unlike the U.S., Canada entered the war in 1914. Consequently, several birth cohorts were exposed to WWI conditions but not the pandemic. Furthermore, prior to June 1917, when a draft

---

[2] Examples of other countries include Taiwan (Lin and Liu, 2014), Sweden (Richter and Robling, 2013), Switzerland (Neelsen and Stratmann, 2011) and Brazil (Nelson, 2010). Examples of other pandemics include the 1957/58 Asian Flu Pandemic, in the case of the England (Kelly, 2011).

[3] In the U.S. and Canada, grounds for exemption (or deferment) from the draft included old age, disability, presence of dependent family members and risk of family impoverishment (if enlisted). Additional grounds for exemption in Canada included employment in essential occupations (farming, munitions).

(conscription) was imposed, Canada relied exclusively on volunteers, most of whom were from English-majority regions (e.g. Ontario) rather than French regions (e.g. Quebec).[4] Card and Lemieux (2001), who observe similar linguistic discrepancies in enlistment during WWII, point out that French-speaking enlistment was likely to be constrained by two factors: 1) the presence of a predominantly English-speaking armed forces; and 2) relative weakness of family ties between French-speaking Canadians to France compared to English-speaking Canadians and England, owing to a slowing of French migration to Canada by the mid-nineteenth century. Thus, as a robustness check, we examine the long-term effects of *in utero* exposure to the pandemic in various provinces of birth, under the assumption that parental characteristics in low enlistment provinces were less influenced by WWI conditions, particularly in Quebec. Finally, because the draft was largely limited to "young men" aged 20-24 (Sharpe, 2015, p. 43), a group that accounted for fewer than 10 percent of live births in 1921, it likely had a minimal effect on aggregate fertility behaviour.

2.2 Background

2.2.1 1918 Flu Pandemic as a Natural Experiment

The 1918 flu pandemic provides a useful opportunity to test the fetal origins hypothesis. The pandemic was a severe health shock that generated 50-100 million deaths and more than 500 million infections - for a global infection rate of about one-third (Taubenberger and Morens, 2006; Johnson and Mueller, 2002). Figure 2.1 shows that the pandemic led

[4] During WWI, Canada was still a colony under the British empire, and its foreign policy was dictated largely by England. When England declared war on Germany, all of its colonies, including Canada, also entered the war. Many historians have attributed the low rate of Enlistment in Quebec to apathy or opposition toward the war, at least relative to English Canadians (Auger, 2008; Sharpe, 2015; Stacey, 1981). Furthermore, the eventual imposition of conscription, despite widespread opposition among French Canadians, has long been acknowledged as a source of ethnic conflict along linguistic lines.

to a drastic rise in the flu and pneumonia mortality rate (i.e. deaths per 1,000 residents) in Ontario and Quebec for three months beginning in October 1918. The mechanism behind this was a random shift in the genes of the virus, known as antigenic drift (see Taubenberger and Morens, 2006). This limited the effectiveness of viral antibodies acquired for previous infections, limiting human immune responses that protect against future infections. Because these shifts were random, the severity of the pandemic was unanticipated. Additionally, there were likely uncertainties about the geographic scale and distribution of the pandemic in real time because (a) the flu was not a reportable disease and (b) media censorship restricted newspapers from reporting on known infections and deaths in warring areas.

These features are crucial for identifying the long-term effects of fetal exposure to the pandemic. Because the pandemic was unexpected and abrupt, parents had no opportunity to avoid exposing their offspring to illness by postponing fertility or migrating to less impacted regions. This ensures that those exposed to the pandemic are likely to be similar to those who were unexposed. In theory, this allows us to treat fetal exposure to the pandemic exogenously. Furthermore, according to the fetal origins hypothesis, those exposed to the pandemic are predicted to experience long-term reductions in their health and socio-economic status. This paper tests this using census data on year and place of birth information.

2.2.2 Previous Literature

Several past studies have used the 1918 flu pandemic to test the fetal origins hypothesis (Lin and Liu, 2014; Brown and Thomas, 2011; Richter and Robling, 2013; Neelsen and Stratmann, 2011; Nelson, 2010; Garthwaite, 2008; Almond, 2006; Almond and Mazumder, 2005). All except Brown and Thomas (2011) show that fetal exposure to pandemic conditions produce adverse long-term health and economic effects. The seminal study of Almond (2006) shows that *in utero* exposure to the 1918 flu pandemic led to lower education, earnings and health. These effects are statistically and economically significant. For instance, he finds that men born in 1919 (i.e. exposed during their second or third trimester to the peak of the pandemic) were 14-16 percent less likely to complete high school and had 5-9 percent lower employment earnings, after controlling for a quadratic function of year of birth. He finds similar results for many other outcomes. He also shows similar results using a difference-in-differences specification that combines cohort-level exposure to pandemic conditions with cross-cohort state variation in pandemic virulence, as measured by the rate of maternal mortality.

One threat to Almond's identification strategy, elucidated by Brown and Thomas (2011), stems from the fact that the U.S. entered WWI only one year before the arrival of the pandemic. They show that this close historical overlap, coupled with the fact that the U.S. generally drafted younger and healthier men, while providing exemptions to men from poor families, led to a discontinuous change in parental characteristics. For instance, babies who were exposed *in utero* to the pandemic, because they were more likely to be fathered by men who avoided the draft, typically had older, less educated and lower income fathers. After controlling for cohort-specific parental characteristics, they find that fetal exposure to the 1918 flu pandemic has no significant association with long-term schooling or earnings outcomes. On the other hand, adverse effects of fetal exposure to the 1918 flu pandemic have been reported internationally across a diverse set of countries that vary in their WWI participation and exposure. The findings of Almond (2006) have been reproduced in samples from Taiwan (Lin and Liu, 2014), Brazil (Nelson, 2010), Switzerland (Neelsen and Stratmann, 2011) and Sweden (Richter and Robling, 2013), despite differences in WWI conditions, as well as pandemic severity, disease environment and health care institutions (such as access to health care). Although Taiwan was a colony of Japan's, it did not participate militarily in WWI. Likewise, as neutral countries, both Sweden and Switzerland avoided armed conflict. The effects observed in Taiwanese and Brazilian samples are larger than those of Almond's (2005), possibly reflecting the poorer disease environment that existed in developing countries. On the other hand, Almond's (2006) estimate of the effect of fetal exposure to the 1918 flu pandemic on high school completion is seven times larger than Neelsen and Stratmann's (2011) estimate using Swiss data (e.g. -14 versus -2 percentage points). Neelsen and Stratmann (2011) speculate that this difference may be explained by greater access to medical care in Switzerland than in the U.S. at the time of the pandemic.

2.2.3 The Canadian Case

This study examines the Canadian influenza pandemic. Canada offers a better comparison to the U.S. because of their similarities in economic structure, demographic, culture and epidemiological characteristics.[5] Moreover, like the U.S., Canada did go to war, but did so almost four years before the U.S., in August 1914. Since the pandemic began in October 1918 and the war ended in November 1918, all persons born in Canada from August 1914 to September 1919 were exposed *in utero* to WWI conditions, whereas only those born from October 1918 to August 1919 were exposed *in utero* to the

[5] For example, in 1921, the infant mortality rate in Canada was 102 (deaths prior to age 1 divided by number of live births), compared to 91 in the United States (Leacy et al., 1983; Department of Commerce, 1924).

pandemic. Therefore, the main advantage of using Canadian data is that they allow for a clean separation of the effects of WWI conditions from the effects of the pandemic.

Unlike the U.S. (Brown and Thomas, 2011), there is less reason to suspect that WWI led unambiguously to worse parental characteristic. This is partly because Canada relied extensively on a voluntary recruitment.[6] One possibility, given the build-up of excess productive capacity in 1914 (Lew and McInnis, 2006, p. 3), is that enlistment was more likely among men with comparatively few employment prospects. In this case, children of non-enlistees may have had better parental SES than children born either before or after WWI. The demands of the warring British empire for staple foods and raw materials (lumber, steel) also helped expand certain sectors, such as agriculture and munitions (Lew and McInnis, 2006), potentially improving the economic status of workers in these industries.

Canadian military enlistment rates also varied regionally due to political and linguistic differences. Support for the war, including military participation, was much more fervent among English-speaking Canadians, especially those who had family ties to Britain. Table 2.1 shows that enlistment rates in Quebec, Canada's only French-majority province, were half those of Ontario and well below the national average. Such differences reflect the fact that British-born military-aged men were concentrated in Ontario[7], and Quebec's more ambivalent stance toward Canada's military effort.[8] If it is assumed that the low enlistment rate in Quebec was due primarily to political motives

[6] 75 percent of its recruits were volunteers, as opposed to draftees (conscripts).
[7] About 40 percent of such men were living in Ontario at the time of the 1911 census.
[8] Quebec had much stronger opposition to conscription. They voted overwhelmingly against conscription in a national plebiscite, whereas every other province voted in favor of it.

rather than health or economic reasons, then fetal exposure to the pandemic in Quebec should be less prone to confounding from WWI-related effects.

Finally, while Canada eventually imposed a draft,[9] it was much less comprehensive than that of the U.S. Table 2.1 reports the draft rate in the U.S. and Canada. The draft rate measures the fraction of the military-aged population (18-45) that was enlisted under the draft. The draft rate was more than 40 percent higher in the U.S. than in Canada (11.8 percent versus 8.1 percent). An estimated 37 percent of Canadian conscripts faced combat (or 3 percent of all males aged 18-45), most of whom were aged 20-24, which was below the usual age of male fertility. For instance, men aged 20-24 only accounted for 10 percent of births in 1921.[10] Thus, their diversion to the military likely had a small impact on fertility outcomes.

If the draft reduced fertility, then this should be observable in aggregate fertility outcomes. Figure 2.2 examines whether this is the case, using data on live births in Ontario. Each data point on the graph captures the average number of live births that occurred in a region in any given month from the pre-conscription era to the post-war era.[11] The number of live births increased, albeit very slightly, during the conscription period, then began falling in linear fashion for ten months beginning in in October 1918. The series then abruptly returns to pre-pandemic levels in September 1919, ten months after the peak month of the pandemic. This evidence suggests a rather muted effect of

[9] The draft was announced in June 1917. The first call-up was in October 1917.
[10] This figure is from the first volume of the Dominion Bureau of Statistics (DBD) *Vital Statistics* report, which covered the year 1921. Similar nationally comprehensive data are not available for earlier years. Further, the 1921 data does not include Quebec. The earliest data for which similar information is available for Quebec is in 1926. The fraction of births attributable to males 20-24 in Quebec in 1926 was 9.8 percent.
[11] Similar data for other provinces was not available.

conscription on fertility or family planning. However, the pattern of live births observed during and after the pandemic period is suggestive of an increase in prenatal mortality resulting from pandemic conditions.

2.3 Methodology

2.3.1 Model Specification

We begin by setting up a model that compares differences in outcomes by birth cohort. The primary objective of this approach is to ascertain whether cohorts born around the time of the pandemic have significantly worse long run outcomes. A common approach (Almond, 2006) is to estimate the impact of being born in 1919 ($D1919_{ic}$) on a long run outcome (y_{ic}), conditional on a quadratic of year of birth (YOB_{ic}):

(2-1) $$\boldsymbol{y_{ic} = \beta_0 + \beta_1 D1919_{ic} + \beta_2 YOB_{ic} + \beta_3 YOB_{ic}^2 + u_{ic}}$$

where i denotes a census individual and c their birth cohort. The 1919 birth cohort is comprised of individuals born in January-September 1919, who were *in utero* during the peak of the pandemic (i.e. between October and December 1918), and those born between October and December 1919, who were conceived after the pandemic departed. The effect of being born in 1919 on long run outcomes is captured by β_1, which measures the difference in the mean outcome for the 1919 birth cohort and the combined mean outcome for cohorts 1912-1918 and 1920-1922. Because we control for a quadratic in year of birth, β_1 is the departure from trend in outcome y_{ic}. The fetal origins hypothesis predicts $\beta_1 < 0$.

Because the cohort dummy ($D1919_{ic}$) combines exposed and unexposed sub-cohorts, OLS estimates of β_1 will be biased upward relative to the effect predicted by the fetal origins hypothesis (i.e. biased toward no effect). One way to address this is to use more narrowly defined cohort bins (e.g. month of birth, quarter of birth, half-year of birth). As already mentioned, we can construct half-year cohort variables, but only in the 1971 to 1991 census samples. Almond's (2006) analysis of U.S. census data shows that the negative effects of the pandemic were concentrated among individuals born in the first two quarters of 1919, or the first half of 1919. If similar effects took place in Canada, they should be evident in outcomes observed for half-year cohorts. This conjecture motivates estimating the effect on $\mathrm{y_{ic}}$ of being born in the first half of 1919 (D1919H1)

(2-2) $\mathbf{y_{ic}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}\mathbf{D1918H2_{ic}} + \boldsymbol{\beta_2}\mathbf{D1919H1_{ic}} + \boldsymbol{\beta_3}\mathbf{D1919H2_{ic}} + \boldsymbol{\beta_4}\mathbf{YOB_{ic}} + \boldsymbol{\beta_5}\mathbf{YOB_{ic}^2} + \mathbf{u_{ic}}$

The coefficient β_2 measures the individual effect of being born in the first half of 1919 on outcome y_{ic}, conditional on a quadratic function of year of birth and a half-year dummy (HY_{ic}), which controls for seasonal influences on fetal health. The specification also contains dummy variables for being born in the second halves of 1918 and 1919 (i.e. 1918H2 and 1919H2). The effect of the 1919H2 dummy, β_3, may be upward biased (i.e. toward no effect), as it combines exposed and unexposed monthly cohorts. For example, the 1919H2 cohort (born June-December 1919) includes individuals born in the second half of 1919 (June-December 1919), and thus exposed prenatally to the pandemic, as well as individuals who were conceived after the pandemic departed (January to March 1919), and thus were unexposed to the pandemic. Likewise, a similar argument applies to the effect of the 1918H2 dummy (β_1), which combines the effect of being exposed postnatally (June September 1918) with the effect of being exposed in late gestation (i.e. third trimester, October-December 1918).

To account for potential confounding caused either by WWI or by conscription, we add a vector of proxies for family background characteristics, x'_{cp}, and a vector of individual-level controls, z'_{ip},. This leads to our preferred specification:

(2-3) $$\mathbf{y_{ic}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} \boldsymbol{D}\mathbf{1918H2_{ic}} + \boldsymbol{\beta_2}\boldsymbol{D1919H1}_{\mathbf{ic}} + \boldsymbol{\beta_3}\boldsymbol{D1919H2}_{\mathbf{ic}} + \boldsymbol{\beta_4}\boldsymbol{YO}\mathbf{B_{ic}} + \boldsymbol{\beta_5}\boldsymbol{YO}\mathbf{B^2_{ic}} + \boldsymbol{\pi'_1 x'_{cp}} + \boldsymbol{\pi'_2 z'_{ic}} + \mathbf{u_{ic}}$$

2.3.2 Data Sources

The main data files include the 1971, 1981 and 1991 census samples. The 1971 file is a 33 percent sample, while the 1981 and 1991 files are 20 percent samples. Each dataset contains several education and economic outcomes. The primary dependent variable is the individual's highest grade of primary/secondary schooling – or grade attainment. We also use three additional binary schooling outcomes: completion of at least eight years of school; completion of at least ten years of school; and college attendance. The former two variables are cut-off indicators based on grade attainment (whether the respondent completed at least grade 8 or at least grade 10), whereas the latter is based on years of post-secondary training (e.g. university, non-university, vocational/trades).

The final set of dependent variables include earnings, income and employment status. Earnings and income are self-reported estimates of annual wage income and total income accrued over the previous calendar year (i.e. 1970), and employment status variable is a dummy indicating whether an individual was employed during the same period. This analysis is limited to the 1971 male sample because employment rates (and thus the fraction of the sample with zero earnings) are low for females in all samples and for males in 1981 and 1991. For example, the sample-wide employment rate falls from 59 percent in 1971 to 35 percent in 1981 and 8 percent in 1991. Much of this decline is consistent with retirement behaviour.[12]

The cohort dummies of interest (e.g. 1918H2, 1919H1, 1919H2) are derived from census-provided date of birth variables. All three samples come with year of birth and approximate half-year of birth variables. Half-year of birth is approximated using the census-provided census split variable. The census split variable indicates if a person was born before or after the census split date, usually in the first week of June. Consequently, the first half of the year covers January 1 to the census split date (approximately 5 months) and the second half covers the census split date to December 31 (approximately 7 months). We use this information to assign each person into a half-year cohort (i.e. a year of birth and a corresponding half within that year). Based on this, we define the "potentially" exposed (*in utero*) group as all individuals born in the second half of 1918 (1918H2), first half of 1919 (1919H1) or second half of 1919 (1919H2). All of these cohorts include some individuals who were exposed *in utero* to the pandemic. However, gestational age at exposure varies by birth cohort.

Table 2.2 lists trimesters of pregnancy at exposure to the peak month of the pandemic (October 1918) for each of the exposed cohorts. 1919H1 includes individuals born between January 1 1919 and May 31 1919. These individuals were *in utero* during their first or second trimester when the pandemic arrived in October 1918. 1918H2 includes individuals born between June 1 1918 and December 31; of these individuals, only those born between October 1 and December 31 were exposed (i.e. about half of the

---

[12] Retirement normally occurs around age 60-65. The sampled cohorts, born in 1912-1922, were aged 59-69 in 1981 and 69-79 in 1991.

composite monthly cohorts). Finally, 1919H2 includes individuals born between June 1 1919 and December 31. In this group, only individuals born in June 1 1919 were *in utero* in October 1918, and this during the first trimester of pregnancy. So, of the three exposed half-year cohorts, 1919H1 was most affected by the pandemic. A key advantage to constructing the birth cohorts in this way is that we are able to group the first two quarterly birth cohorts from 1919, which incidentally are the same birth cohorts identified by Almond (2006) as most adversely affected by the pandemic.

2.3.3 Control Variables

The vector z'_{ic} is based on observables in the 1971-1991 censuses. These include parental nativity/migration status, mother tongue, ethnicity, province of birth and WWII veteran status. To the extent that migrants were poorer than non-migrants, the parental nativity variable proxies for SES. Furthermore, since WWI service was concentrated among English men, children of English backgrounds that were born during WWI may have been more likely to have fathers who were previously screened out of the military due to age, health or disability. Thus, mother tongue picks up cohort-specific selection biases induced by military recruitment during WWI.

Following Neelsen and Stratmann (2011), we include province of birth dummies to account for potential biases arising from pandemic-related shifts in the regional distribution of births across birth cohorts. They argue that the pandemic may have resulted in excess mortality in poor regions, where educational attainment may be lower for all cohorts. Furthermore, if such mortality is greater for those exposed *in utero* to the pandemic (rather than postnatally), then the birthplace distribution among fetally exposed birth cohorts may have shifted to regions with better economic prospects. As a result, failure to include birthplace dummies could create positive bias (toward a non-effect) in OLS estimates of the effects of fetal exposure to pandemic conditions.

The last individual covariate is a WWII veteran status dummy. This variable is used for the male sub-samples. For men, WWII enlistment is a significant source of potential mortality. Furthermore, if less educated men were more likely to be assigned into combat roles, then WWII survivors should have somewhat higher education levels. Such bias would have been reinforced in Canada under various WWII veteran programs, including educational remediation programs. Previous research (e.g. Card and Lemieux 2002) shows the latter programs to be especially effective in increasing human capital attainment among veterans. In our sample, those exposed *in utero* to the pandemic were significantly more likely to report being WWII veterans. To the extent veteran status correctly captures actual enlistment rates, this suggests that models that do not control for WWII veteran status will understate the effects of fetal exposure to the pandemic among males. Alternatively, if the decision to enlist is affected by available job opportunities, such that those with less human capital are more likely to volunteer for service, then failure to control for WWII status would overstate the effects of fetal exposure to the pandemic.

The cohort-level controls, x'_{cp}, are populated from the 1921-1951 censuses. Cohort controls are measured as cohort-by-province of birth cell means of childhood characteristics. These data files were recently digitized by the Canadian Century Infrastructure Project (CCRI) and were made available through Statistics Canada's Research Data Centres (RDCs). Observables taken from the 1931 census include father's age, literacy and occupational status, homeownership and urban/rural status. Paternal characteristics are computed based on unambiguous child-father links. Each child-father relationship is unambiguous if a child is related to the household head as their biological son or daughter, if the household head is a male and if the child lives in a single-headed household (i.e. multi headed households are dropped). Parental attrition via mortality or marital break-up led to non-matches between children and parents in some instances. While we accounted for this by conditioning on the average father-child match rate for a cohort, the match rate did not significantly differ between cohorts exposed *in utero* to the pandemic and unexposed cohorts (p-value=0.33).

Note, however, that the birth cohorts are constructed using age at last birthday, as year of birth was not collected in any of these surveys (see Table 2.3). Because age is determined relative to the census day (usually early June), the birth cohorts are defined over a June-to-May interval. For instance, a person of age 12 on the census day in 1931 (June 1) is assumed to have been born between June 1 1918 and May 31 1919, and their birth cohort is labelled June/May 1919 (this includes the half year cohorts, 1918H2 and 1919H1). Thus, to match the historical covariates to the individual data, the individual-level cohorts must first be collapsed into June/May annual cohorts. Consequently, adjacent off-year cohorts in the 1971-1991 censuses (e.g. 1918H2 and 1919H1) will share common covariate values.

2.4 RESULTS

2.4.1 Descriptive Statistics

To assess the experimental validity of the pandemic, we begin by exploring whether childhood background characteristics (e.g. household characteristics, paternal characteristics) differ for those who were exposed *in utero* to the pandemic relative to non-exposed adjacent cohorts. To identify the cohorts as children, we use the 1931 CCRI census sample. As noted above, a key feature of this sample is that it can be used to construct child-father matches, which can then be used to assess changes in paternal characteristics across birth cohorts.

Table 2.4 reports summary statistics of each covariate for exposed (born in 1918/19) and unexposed cohorts (born in 1916/18 or 1919/1921). The table reports the sample average of each covariate in each group, along with the averages and t-statistics for group differences. The results display similarities between exposed and unexposed cohorts across a wide range of covariates.[13] For instance, fathers of the pandemic-exposed children had similar literacy rates,[14] labour force outcomes and religious affiliation as fathers of unexposed cohorts. A further issue, given the proximity of WWI and the pandemic, is whether pandemic-exposed cohorts differ in paternal WWI veteran status. This is important because veteran status may be an indicator of paternal health or of access to post-war veterans' benefits. However, precluding such a comparison is the fact that the 1931 census did not ask about WWI veteran status.

To circumvent this problem, we adopt an indirect proxy of paternal WWI enlistment, which is based on an imputed paternal WWI veteran rate for each child-father match in the 1931 census. To impute this rate, the mean WWI veteran rate was calculated for each age-province cell in the 1951 census. These rates were then matched to each father-child match in the 1931 census, based on the father's age and child's province of

[13] Most of the t-statistics for group differences in covariates fall below 1.96, meaning that the groups are not significantly different from one another at a significance level of 0.05.
[14] Interestingly, fathers were reported to have, on average, a high ability to write (>0.9) and a low ability to read (<0.1). Enumerators simply asked whether a given individual could read or write without attempting to gauge the degree of literacy. The writing response may reflect, for example, whether a person was able to write their name.

birth.[15] As with the other characteristics, the imputed WWI veteran rates are similar for exposed and unexposed cohorts.

However, exposed cohorts do differ from unexposed cohorts in several respects. For example, the 1918/19 cohort was more rural, had fathers who were older and more likely to be working as farmers, and lived in larger households. Brown and Duncan (2013) suggest that greater paternal age and larger household sizes are indicative of low parental SES, and thus constitute potential sources of bias in long-term estimates of fetal exposure to the pandemic in favor of the fetal origins hypothesis. However, unlike this study, they do not report any change in rurality or parental farming among those with fetal exposure to the pandemic. This is a crucial distinction since, as noted earlier, farmers' living standards may have increased during WWI. Therefore, individuals who were *in utero* during the pandemic, because they were more likely to be born into farming households, may have gained extra fetal protection from positive war-related income shocks. Additionally, rural areas may have been somewhat more protected from the spread of influenza due to low population density.

In addition to this, the 1918/19 cohort was more likely to be in school during the reference period (September 1930-June 1931). Cohort differences in educational attainment could reflect compulsory schooling laws at the time. According to Oreopoulos (2006), the minimum school-leaving age in Canada in the 1930s was 12 or 13, depending on the province. A preliminary analysis of the 1931 census revealed a discontinuous drop in school attendance for the 1917/1918 and older birth cohorts. The 1918/19 cohort,

[15] Because this rate is based on the child's location of birth, it accounts for regional differences in exposure to the draft.

which was aged 12 in the 1931 census, likely had above average attendance rates because they were not yet old enough to leave school. However, mandated school entry and exit ages did not change during our sampling period. As a result, our identification strategy is unlikely to be impeded by simultaneous, cohort-specific changes in compulsory schooling.

Overall, these results indicate that exposed and unexposed are similar in terms of most, but not all, observable characteristics. Furthermore, it is not obvious whether, to what extent, and in what direction, observable differences between these two groups are likely to bias OLS estimates of the effects of *in utero* exposure to the pandemic. Consequently, our regression results are presented with and without control variables.

2.4.2 OLS Results: Education

The first set of results present the effects of *in utero* exposure to the 1918 flu pandemic, conditional on a birth year trend and a seasonal half-year dummy. Figure 2.3 plots mean grade attainment by sex and cohort using the 1971 sample. The figure shows three main patterns. First, grade attainment is increasing secularly over time, meaning that younger cohorts have higher grade attainment. Second, mean grade attainment follows a quadratic trend, suggesting that the linear and quadratic year of birth variables used in equation (2) are appropriate for these data. Finally, there is a clear departure from trend in mean grade attainment among cohorts that had some fetal exposure to the peak of the 1918 flu pandemic – that is, the 1918H2, 1919H1 and 1919H2 cohorts. For these deviations to reflect pandemic-related fetal shocks, cohorts that were exposed *in utero* to the pandemic should have similar characteristics as those who were unexposed. Results from the previous section are somewhat supportive of this assumption.

Panel A of Table 2.5 reports OLS estimates of the parameters β_1, β_2, and β_3 for males, again using the 1971 sample. For the grade attainment equation, the coefficients represent the departure from trend among the exposed cohorts. Consistent with the fetal origins hypothesis, *in utero* exposure to the peak of the pandemic is associated with less education. For instance, males born in 1919H1 attained 0.1 fewer grades of primary/secondary schooling than unexposed cohorts, conditional on seasonality and yearly cohort trends. This coefficient is approximately equal to 3.6 percent of the standard deviation of grade attainment. Similar effects are reported for men born in 1918H2 and 1919H2, suggesting that the findings are not overly sensitive to trimester of exposure. However, note that since the 1918H2 cohort consists partly of individuals born prior to the peak of the pandemic (June-September 1918), we are not able to rule out long-term adverse effects from early postnatal exposure to the pandemic (e.g. exposure in the first six months after birth). Essentially the same cohort patterns and effects are reported for completion of 8 and 10 years of schooling. Likewise similar coefficient signs are reported for college attendance, though the effects are generally not significantly different from zero at conventional levels of significance.

Panel B of Table 2.5 adds controls for parental nativity, ethnicity, mother tongue, province of birth and WWII veteran status in the male sample. The addition of these controls has essentially no effect on the coefficient estimates. For the grade attainment equation, the coefficient for the 1919H1 cohort is almost identical to the one derived from models without individual controls, i.e. Panel A (-0.1051 versus -0.1089). Overall, men born in 1919 experienced about a 0.080 to 0.105-point reduction in grade attainment (3-4 percent of a standard deviation in grade attainment). The next three columns, which report the same cohort coefficients for the other three education outcomes, show similar increases in coefficient magnitudes relative to the results in Panel A.

Table 2.5, Panel C adds in parental proxy controls, as measured as cohort-by-province of birth cell means of childhood characteristics. For grade attainment, the coefficients are somewhat smaller than in the prior specifications, but they remain consistent with the fetal origins hypothesis (i.e. p values<0.05). Similar observations are noted for completion of 10 years of schooling. However, *in utero* exposure to the pandemic is not associated with completion of 8 years of schooling or college attendance for males (i.e. p-values>0.05). A final point is that male and female results are very similar for all specifications. Like males, females who were *in utero* during the pandemic experienced long-term educational deficits and the addition of controls has little bearing on the long-term effects experienced by females (Table 2.6).

We next turn to our provincial sub-sample results. Table 2.7 reports the results for grade attainment in males. Estimates based on cohort-level controls were dropped due to the high degree of collinearity between province-cohort cell means and the time of birth controls in the sub-sample models.[16] Consequently, Table 2.7 only reports results for models without control variables and with individual control variables. Column [1] of Table 2.7 repeats the results for the full-sample (all provinces of birth). Results are reported separately for males born in Ontario (Column [2]), Quebec (Column [3]), or other provinces (Column [4]). In most estimations we find that being *in utero* during the

[16] Attempts were made to estimate such regressions, but typically resulted in numerous omitted coefficient (due to collinearity).

peak of the 1918 flu pandemic has negative consequences for long-term educational attainment.

Column [3] reports the results for the Quebec-born sub-sample. Since WWI enlistment rates were lower in Quebec for political reasons, these results should be less confounded by WWI-related changes in parental characteristics. In Panel A, that is the results without individual controls, there is no difference in grade attainment for two of three cohorts. Only males born in the latter half of 1918 had lower grade attainment than unexposed cohorts. However, after adding individual controls, the results indicate that males born in the second half of 1918 and the first half of 1919 had lower grade attainment than non-exposed cohorts. Note, as well, that the pattern of results by province of birth and cohort are strikingly similar for both males (Table 2.7) and females (Table 2.8). For Quebec-born females, however, we note that only those born in the first half of 1919 have educational deficits compared to surrounding cohorts.

2.4.3 OLS Results: Labour Market Outcomes

Table 2.9 reports estimates for a set of labour market outcomes, including earnings, income and employment status. These results are limited to the male sample because most women were non-earners and employment rates among women were quite low. Annual wages were $62-$202 lower for males who were *in utero* during the peak of pandemic. These effects vary by time of exposure and model specification (i.e. presence of control variables). Earnings losses are statistically significant in models with no controls (Panel A) and in models with individual controls (Panel B). In both models, the effects are larger for the 1919H1 cohort than for the other two exposed cohorts. However, earnings losses fall to $62-$125 after controlling for all covariates (Panel C). Except for the 1919H1 cohort, pandemic-related earnings losses in this model are not statistically significant. For the 1919H1 cohort, the estimated earnings losses ($125) are significant only at the 10% level. Income effects, shown in Column [2], are somewhat larger than the wage effects. For instance, annual incomes were $60-$280 lower for men who were exposed prenatally or immediately after birth to the peak of the pandemic, conditional on trend and seasonality (Panel A). Adding individual controls (Panel B) increases both the size and significance of each coefficient. However, after controlling for cohort and individual covariates, only the 1919H1 cohort has lower earnings (β_2 = -$180; $p < 0.05$). The table also reports OLS results for employment status, which indicates whether a person was ever employed during the year prior to census day. This outcome is unaffected by fetal exposure to the pandemic, which suggests that the earnings losses are due to lower wages or reduced hours/weeks of work[17] rather than long-term non-employment.

2.4.4 Economic Significance

The estimated exposure coefficients represent the average long-term individual effect of being *in utero* during the peak of the pandemic. Almond (2006) combines this with an estimate of the fraction of pregnant women who were infected in the U.S. (28 per cent) to generate an estimate of the average effect of maternal infection. Specifically, the average effect of maternal infection equals the effect of *in utero* exposure to the pandemic (e.g. β_2) divided by the maternal influenza infection rate.[18] If it is assumed that the maternal infection rate for Canada was the same as that of the U.S., then the estimated effect of

[17] Reductions in hours/weeks of work could be due to part-time employment or temporary unemployment during the reference period.
[18] This is akin to an Average Treatment Effect on the Treated (ATET).

maternal infection on the grade attainment of males born in 1919H1 is -0.39 — i.e. 0.39 fewer years of primary/secondary schooling.[19] This is equivalent to 4 percent of mean grade attainment, 12 percent of a standard deviation in grade attainment and 75 percent of the total increase in mean grade attainment between the 1912 and 1922 birth cohorts. These effects are also comparable (in absolute value) to the effects of compulsory schooling laws on education. For instance, Oreopoulos (2006) estimates that a one-year increase in the school-leaving age increased grade attainment by 0.22-0.93 units. Our results are at the lower to middle end of this range.

The caveat of these results is that they are based on a U.S. maternal infection rate. Almond (2006) takes this value from a previously published rate based on a group of pregnant women from Maryland who were hospitalized for flu-related reasons. Obviously, this may not be generalizable to Canada. Applying a larger (smaller) infection rate would lead to smaller (larger) effects of maternal infection. Because there is no published estimate of the infection rate in Canada we cannot determine whether the rate is higher or lower than in the U.S. However, because influenza-related mortality rates during the pandemic period were similar in Canada and the U.S., it seems sensible to use the U.S. infection rate for back of the envelope calculations.

2.4.5 Selective Mortality

Our analysis up until this point has been supportive of the fetal origins hypothesis. Yet, this analysis is conditional on observation up to 50 years after the shock. Life expectancy for WWI cohorts was about 60 years. Thus, by 1971, much of the cohort may have

[19] This is calculated as $\frac{\beta_2}{0.27}$. Substituting β_2 = -0.11 from Panel C yields the indicated value of -0.39.

perished. Furthermore, the pandemic may have led to increased premature mortality, either due to direct infection or fetal exposure to the pandemic.

Recall that Figure 2.2 documents aggregate live births around the time of the pandemic, showing that live births decreased during the pandemic period, particularly 5-9 months after the peak of the pandemic (October 1918). One possibility is that the pandemic increased the rate of miscarriage or stillbirth. We assess this possibility using data on stillbirths (see Figure 2.4).[20] These data show that the stillbirth rate increased by about one-third in the first (Quebec) or second month (Ontario) of the pandemic ($p < 0.001$).[21] Finally, Figure 2.5 shows that, in 1971, fetally exposed cohorts, particularly the 1919H1 cohort, was smaller in size than surrounding, unexposed cohorts.[22] This suggests potentially lower life expectancy among those who were fetally exposed relative to those who were not.

If hardiness predicts survival from the pandemic, then those surviving to subsequent census days (in 1971 or later) despite being exposed in utero to pandemic conditions are more likely to be found in the upper tails of their cohort-specific health and education distribution. Under these conditions, OLS estimates of the educational effects of fetal exposure to the pandemic likely understate the true effects. Furthermore, if subsequent mortality after 1971 follows the same pattern, then coefficient estimates

[20] Miscarriages and stillbirths are distinguished by the timing of gestational death. Miscarriages include deaths prior to 20 weeks of gestation. It is not clear whether this distinction was applied to fetal deaths in the early 1900s. Thus, Figure 4 could potentially understate the full extent of fetal mortality.

[21] The pandemic also increased prenatal mortality in the U.S. (Almond, 2006) and Japan (Nishiura 2009).

[22] The graph covers the full sample, that is males and females born in Canada between 1912-1922.

should increase with each subsequent census year. To assess this, Table 2.10 (males) and 2.11 (females) display estimates of β_1, β_2 and β_3, using the 1971, 1981 and 1991 census samples.

Table 2.10 reports results for the male samples, while Table 2.11 reports results for the female samples. In each case, the dependent measure is highest grade attainment. In almost all models, effects that are significant in 1971 are rendered non-significant or marginally significant in the 1981 or 1991 censuses. Despite this, the coefficients are of the same sign and of similar magnitudes. More importantly, tests of equality for coefficients estimated from the 1971 and 1991 census fail to reject the null hypothesis for both males and females. The same is true when comparing coefficients for 1971 and 1981. If selective mortality is a reasonable hypothesis, the results presented here would imply that the bulk of pandemic-related mortality is likely to have occurred before 1971.

2.5 Conclusion

This paper has examined the long run effects of *in utero* exposure to the 1918 Flu Pandemic in Canada. On balance, we report modest evidence of long-term negative effects on years of schooling and high school completion rates. We find that Quebec-born women and Ontario-born men who were born in the first half of 1919 accumulate less schooling from birth to age 62. We find no evidence of pre-1919 WWI-related changes in schooling outcomes and, additionally, we find that parents of the 1919 birth cohort had similar socio-economic characteristics as parents of adjacent birth cohorts. This rules out the possibility that our findings are driven by the effect of WWI on selection into fertility, a key criticism of U.S.-based research on the 1918 flu pandemic.

Tables and Figures

Table 2.1: WWI Enlistment in the United States and Canadian Provinces

	Men Aged 18-45	% Enlisted in WW1		% Drafted	% Volunteer
Region of Enlistment	Total	All of WWI	After Adopting Draft	All of WWI	All of WWI
US: WWI Entry April 1917	23,626,884	20.3	20.3	11.9	8.4
Canada: WWI Entry August 1914	1,537,172	31.6	14.3	8.1	23.5
ON	536,169	45.3	n/a	7.1	38.2
QC	376,232	23.4	n/a	7.4	16.0
NS & PE	112,584	35.0	n/a	7.9	27.1
NB	68,097	39.7	n/a	10.2	29.5
MN	108,536	61.1	n/a	10.7	50.4
SK	130,250	32.0	n/a	8.2	23.8
AB	93,375	52.4	n/a	9.8	42.6
BC	109,448	50.8	n/a	7.1	43.7

Notes: The U.S. adopted a draft at the onset of war, in April 1917, whereas Canada adopted one in October 1917. This table is populated from various sources. U.S. population estimates were derived from the 1910 census (IPUMS 1 percent sample) based on men aged 11-38 (18-45 in 1917). U.S. conscription numbers are from the U.S. selective service system (https://www.sss.gov/About/History-And-Records/Induction-Statistics). Another two million U.S. soldiers had volunteered for service. Canadian and provincial population and enlistment (conscripts and volunteers) are from Sharpe (2015, Table VI). The percentage of eligible males who enlisted after the draft began was estimated using monthly enlistment figures published by the Canadian Great War Project (http://www.canadiangreatwarproject.com/dl/downloads.asp).

Table 2.2: Identification of potential fetal exposure to flu pandemic using half-year cohorts in the 1971-1991 censuses

Year (y)	Half (h)	Cohort Label	Birthday	Trimester, October 1918	Potentially Exposed
1918	1	1918H1	Jan 1-May 31, 1918	Born before	No
1918	2	1918H2	Jun 1-Dec 31, 1918	Third	Yes
1919	1	1919H1	Jan 1-May 31, 1919	First or second	Yes
1919	2	1919H2	Jun 1-Dec 31, 1919	First	Yes
1920	1	1920H1	Jan 1-May 31, 1920	Conceived after	No

Notes: Year halves are constructed from a census split variable, which groups birthdays into unbalanced bins; e.g. January 1-May 31 (first half) and June 1-December 31 (second half).

Table 2.3: Identification of potential fetal exposure using age at last birthday in the 1921-1951 censuses

Age	Cohort	Birthday	Trimester, October 1918	Potentially Exposed
23	1917/18	June 1, 1917-May 31, 1918	Born before	No
22	1918/19	June 1, 1918-May 31, 1919	First, second or third	Yes
21	1919/20	June 1, 1919-May 31, 1920	Conceived after	No

Notes: Cohorts are defined as census year minus age at last birthday prior to census day, which varies from mid-May to early June. This table refers to the 1941 census. The census day in this case was June 1, 1941.

Table 2.4: Summary statistics of treated and control cohorts

	Unexposed (Born 1917/18 and 1920/21)	Potentially Exposed *in Utero* (Born 1918-1919)	Difference
	Mean [St. Dev]	Mean [St. Dev]	Mean [t-statistic]
Child characteristics			
Son/daughter (match rate)	0.9798	0.9791	-0.0007
	[0.1437]	[0.1462]	[-0.3398]
Other relative of household head	0.0133	0.0140	0.0007
	[0.1172]	[0.1201]	[0.3778]
Non-relative	0.0069	0.0069	0.0001
	[0.0843]	[0.0846]	[0.0516]
Can read	0.0632	0.0589	-0.0044
	[0.2481]	[0.2406]	[-1.2081]
Can write	0.9782	0.9814	0.0032
	[0.1493]	[0.1387]	[1.5133]
Attends school	0.9170	0.9480	0.0311
	[0.2816]	[0.2273]	[8.697]
Months in school	8.0362	8.3198	0.2837
	[2.6358]	[2.1832]	[8.3287]
Child works	0.0009	0.0000	-0.0009
	[0.0312]	[0.0021]	[-4.3879]
Non-White	0.0182	0.0185	0.0003
	[0.1366]	[0.138]	[0.1576]
Female	0.4896	0.4943	0.0046
	[0.5106]	[0.5118]	[0.6045]
Born in Quebec	0.3071	0.2950	-0.0122
	[0.4698]	[0.4652]	[-1.7515]
Father's characteristics			
Age	45.3835	45.6551	0.2716
	[7.602]	[7.4399]	[2.4261]
Age heap (age is multiple of 5)	0.2068	0.2083	0.0015
	[0.4137]	[0.4157]	[0.2473]
Imputed WWI enlistment	0.1596	0.1570	-0.0026
	[0.1185]	[0.1159]	[-1.4879]
Married	0.9703	0.9670	-0.0032
	[0.1738]	[0.1834]	[-1.1924]
French	0.3596	0.3524	-0.0072
	[0.4895]	[0.4882]	[-0.9822]

Continued on next page

Table 2.4 Continued from last page

	Unexposed (Born 1917/18 and 1920/21) Mean [St. Dev]	Potentially Exposed *in Utero* (Born 1918-1919) Mean [St. Dev]	Difference Mean [t-statistic]
English	0.8545	0.8550	0.0005
	[0.36]	[0.3601]	[0.0881]
Other language	0.1362	0.1368	0.0006
	[0.3502]	[0.3514]	[0.1119]
Non-White	0.0177	0.0183	0.0007
	[0.1346]	[0.1373]	[0.3186]
Catholic	0.4323	0.4376	0.0053
	[0.5058]	[0.5076]	[0.7044]
Born in Quebec	0.2936	0.2889	-0.0047
	[0.4642]	[0.4627]	[-0.6818]
Reads	0.0692	0.0634	-0.0058
	[0.2587]	[0.249]	[-1.5558]
Writes	0.9260	0.9317	0.0057
	[0.2678]	[0.2588]	[1.4685]
Employed June 1 1931	0.8353	0.8262	-0.0091
	[0.3835]	[0.3941]	[-1.1435]
Sick in 1930/31	0.0333	0.0304	-0.0029
	[0.1827]	[0.1753]	[-1.0823]
Absent from work 1930/31	0.2224	0.2166	-0.0058
	[0.4238]	[0.4206]	[-0.9271]
Ever unemployed 1930/31	0.1487	0.1464	-0.0023
	[0.3622]	[0.3606]	[-0.4239]
Ever laid off 1930/31	0.1478	0.1453	-0.0025
	[0.3613]	[0.3594]	[-0.464]
Farmer	0.2984	0.3117	0.0133
	[0.4673]	[0.4738]	[1.7608]
Labourer	0.1298	0.1307	0.0009
	[0.3418]	[0.3432]	[0.1585]
Professional occupation	0.0353	0.0386	0.0033
	[0.1878]	[0.1962]	[1.0678]
Professional/managerial	0.1297	0.1290	-0.0007
	[0.3419]	[0.3415]	[-0.1291]
Semi-skilled trades	0.2617	0.2416	-0.0201
	[0.4474]	[0.4364]	[-2.8588]
Semi-skilled services	0.0200	0.0147	-0.0053
	[0.1415]	[0.121]	[-2.6441]

Continued on next page

Table 2.4 Continued from last page			
	Unexposed (Born 1917/18 and 1920/21)	Potentially Exposed in *Utero* (Born 1918-1919)	Difference
	Mean [St. Dev]	Mean [St. Dev]	Mean [t-statistic]
Public service	0.0070	0.0062	-0.0008
	[0.0849]	[0.0796]	[-0.6465]
Other occupation	0.1180	0.1274	0.0094
	[0.3298]	[0.3414]	[1.7318]
Household Characteristics			
Household Size	7.4058	7.4896	0.0839
	[2.6874]	[2.6723]	[2.0996]
Rural area	0.5126	0.5321	0.0195
	[0.5104]	[0.5105]	[2.5569]
Home is owned	0.6668	0.6779	0.0111
	[0.4797]	[0.4762]	[1.5598]
Radio in home	0.3112	0.3012	-0.0100
	[0.4728]	[0.4694]	[-1.4226]
Sample Size	22,730	5,590	

Sample: 1931 census.

Table 2.5: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Education (1971 Census; Males)

Cohort/Trimester of Exposure	Highest Grade	8 Years	10 Years	Any College /University
[A] Baseline, no controls				
1918H2=Exposed				
3rd Trimester (β_1)	-0.0960	-0.0078	-0.0204	-0.0048
	[0.0243]***	[0.0039]*	[0.0044]***	[0.0033]
1919H1=Exposed				
1st/2nd Trimester (β_2)	-0.1089	-0.0116	-0.0206	-0.0068
	[0.0289]***	[0.0046]*	[0.0051]***	[0.0039]+
1919H2=Exposed				
1st Trimester (β_3)	-0.0559	-0.0018	-0.0129	-0.0031
	[0.0237]*	[0.0038]	[0.0044]**	[0.0033]
Adjusted R^2	0.0045	0.0028	0.0036	0.0017
[B] Individual controls				
1918H2=Exposed				
3rd Trimester (β_1)	-0.1079	-0.0090	-0.0229	-0.0067
	[0.0218]***	[0.0035]**	[0.0041]***	[0.0033]*
1919H1=Exposed				
1st/2nd Trimester (β_2)	-0.1051	-0.0107	-0.0214	-0.0083
	[0.0261]***	[0.0041]**	[0.0049]***	[0.0038]*
1919H2=Exposed				
1st Trimester (β_3)	-0.0862	-0.0063	-0.0190	-0.0066
	[0.0214]***	[0.0034]+	[0.0041]***	[0.0033]*
Adjusted R^2	0.1941	0.2012	0.1119	0.0264
[C] Individual/Cohort Controls				
1918H2=Exposed 3rd Trimester (β_1)	-0.0645	-0.0052	-0.0140	-0.0038
	[0.0256]*	[0.0041]	[0.0049]**	[0.0039]
1919H1=Exposed 1st/2nd Trimester (β_2)	-0.0750	-0.0077	-0.0138	-0.0069
	[0.0289]**	[0.0046]+	[0.0055]*	[0.0043]
1919H2=Exposed				
1st Trimester (β_3)	-0.0532	-0.0016	-0.0151	-0.0064
	[0.0224]*	[0.0035]	[0.0043]***	[0.0035]+
Adjusted R^2	0.1959	0.2218	0.1219	0.0303

Notes: The sample includes 322,250 males. This table reports the estimated deviation from trend in educational outcomes among individuals that were in utero during the peak of the 1918 flu pandemic (October 1918). Each column-panel represents a regression. Robust standard errors are reported in parentheses. Panel [A] regressions control for a quadratic trend in year of birth and a half-year (i.e. census split) dummy. Panel [B] add individual-level controls for province of birth, ethnicity (French, English, other), mother tongue (French, English, unofficial language), parental nativity (mother foreign born, father foreign born, mother and father foreign born, both native born) and WWII veteran status. Panel [C] regressions add provincial-cohort parental controls. These include, from the 1931 census, various paternal characteristics (age, literacy, employment status, number of children under age 18, percentage non-white, percentage catholic), the fraction of children within the cohort who could not be linked to a father in 1931, and several household characteristics (radio ownership, homeownership). We also control for the fraction of the cohort that was born in an urban or rural area using the 1941 census and for the father's potential WWI enlistment rate. The latter was derived by calculating the WWI veteran status rate for each age-province cell in the 1951 census and then matching these rates to 1931 census children by the child's province of birth and father's age. *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Table 2.6: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Education (1971 Census; Females)

Cohort/Trimester of Exposure	Highest Grade	8 Years	10 Years	Any College/ University
[A] Baseline, no controls				
1918H2=Exposed				
3rd Trimester (β_1)	-0.0887	-0.0085	-0.0165	-0.0015
	[0.0224]***	[0.0037]*	[0.0043]***	[0.0032]
1919H1=Exposed				
1st/2nd Trimester (β_2)	-0.1074	-0.0113	-0.0191	-0.0026
	[0.0270]***	[0.0044]*	[0.0051]***	[0.0037]
1919H2=Exposed				
1st Trimester (β_3)	-0.0579	-0.0059	-0.0076	0.0053
	[0.0226]*	[0.0037]	[0.0043]+	[0.0032]+
Adjusted R^2	0.0024	0.0011	0.0015	0.0002
[B] Individual controls				
1918H2=Exposed				
3rd Trimester (β_1)	-0.0733	-0.0059	-0.0147	-0.0015
	[0.0198]***	[0.0032]+	[0.0039]***	[0.0031]
1919H1=Exposed				
1st/2nd Trimester (β_2)	-0.1069	-0.0107	-0.0188	-0.0029
	[0.0238]***	[0.0039]**	[0.0047]***	[0.0037]
1919H2=Exposed				
1st Trimester (β_3)	-0.0580	-0.0064	-0.0084	0.0046
	[0.0200]**	[0.0032]*	[0.0040]*	[0.0032]
Adjusted R^2	0.2153	0.2215	0.1446	0.0340
[C] Individual and cohort controls				
1918H2=Exposed				
3rd Trimester (β_1)	-0.0478	-0.0048	-0.0098	-0.0012
	[0.0234]*	[0.0038]	[0.0047]*	[0.0038]
1919H1=Exposed				
1st/2nd Trimester (β_2)	-0.0828	-0.0101	-0.0141	-0.0029
	[0.0264]**	[0.0042]*	[0.0053]**	[0.0042]
1919H2=Exposed				
1st Trimester (β_3)	-0.0646	-0.0065	-0.0104	0.0039
	[0.0209]**	[0.0033]*	[0.0042]*	[0.0034]
Adjusted R^2	0.2155	0.2422	0.1511	0.0370

Notes: The sample includes 330,25 females. This table reports the estimated deviation from trend in educational outcomes among individuals that were in utero during the peak of the 1918 flu pandemic (October 1918). Model specifications are the same as in Table 2.5, except here we do not control for WWI veteran status. *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Table 2.7: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Grade Attainment by Province of Birth (1971 Census; Males)

	[1]	[2]	[3]	[4]	[5]	[6]
Cohort/Trimester of Exposure		Province of Birth			Mother Tongue	
[A] Baseline, no controls	Full Sample	ON	QC	Other Prov.	English	French
1918H2=Exposed						
3rd Trimester (β_1)	-0.0960	-0.1193	-0.1175	-0.0657	-0.0924	-0.0915
	[0.0243]***	[0.0421]**	[0.0452]**	[0.0340]+	[0.0269]***	[0.0451]*
1919H1=Exposed						
1st/2nd Trimester (β_2)	-0.1089	-0.1283	-0.0740	-0.0798	-0.1296	-0.0795
	[0.0289]***	[0.0498]*	[0.0540]	[0.0407]*	[0.0320]***	[0.0538]
1919H2=Exposed						
1st Trimester (β_3)	-0.0559	-0.1069	0.0289	-0.0996	-0.0896	0.0355
	[0.0237]*	[0.0400]**	[0.0450]	[0.0336]**	[0.0260]***	[0.0444]
Adjusted R^2	0.0045	0.0040	0.0045	0.0077	0.0051	0.0056
[B] Individual controls						
1918H2=Exposed						
3rd Trimester (β_1)	-0.1079	-0.1229	-0.1439	-0.0762	-0.0965	-0.142
	[0.0218]***	[0.0406]**	[0.0434]***	[0.0318]*	[0.0264]***	[0.0444]**
1919H1=Exposed						
1st/2nd Trimester (β_2)	-0.1051	-0.1229	-0.1252	-0.0759	-0.1296	-0.1381
	[0.0261]***	[0.0480]*	[0.0520]*	[0.0385]*	[0.0313]***	[0.0531]**
1919H2=Exposed						
1st Trimester (β_3)	-0.0862	-0.1113	-0.0183	-0.1298	-0.1107	-0.0244
	[0.0214]***	[0.0389]**	[0.0432]	[0.0315]***	[0.0256]***	[0.0439]
Adjusted R^2	0.1941	0.0833	0.0782	0.1446	0.0452	0.0313
Observations	321,895	91,765	99,200	130,930	192,950	100,420
Mean of Dep. Var	8.7	9.6	7.4	8.9	9.5	7.2
SD of Dep. Var	2.8	2.6	2.9	2.6	2.4	2.9

Notes: This table reports the estimated deviation from trend in grade attainment among individuals that were in utero during the peak of the 1918 flu pandemic (October 1918), by province of birth. Each column-panel combination represents a regression. Robust standard errors are reported in parentheses. Panel [A] regressions control for a quadratic trend in year of birth and a half-year (i.e. census split) dummy. Panel [B], Columns [1]-[4] include individual-level controls for province of birth, ethnicity (French, English, other), mother tongue (French, English, unofficial language), parental nativity (mother foreign born, father foreign born, mother and father foreign born, both native born) and WWII veteran status. Panel [B], Columns [5] and [6] include the same covariates except mother tongue. *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Table 2.8: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Grade Attainment by Province of Birth (1971 Census; Females)

	[1]	[2]	[3]	[4]	[5]	[6]
Cohort/Trimester of Exposure			Province of Birth		Mother Tongue	
[A] Baseline, no controls	Full Sample	ON	QC	Other Prov.	English	French
1918H2=Exposed						
3rd Trimester (β_1)	-0.0887	-0.1093	-0.0164	-0.1156	-0.1016	-0.0365
	[0.0224]***	[0.0378]**	[0.0396]	[0.0317]***	[0.0241]***	[0.0392]
1919H1=Exposed						
1st/2nd Trimester (β_2)	-0.1074	-0.1394	-0.1021	-0.1145	-0.0882	-0.1185
	[0.0270]***	[0.0460]**	[0.0480]*	[0.0379]**	[0.0290]**	[0.0479]*
1919H2=Exposed						
1st Trimester (β_3)	-0.0579	-0.1214	-0.0068	-0.051	-0.0689	-0.0721
	[0.0226]*	[0.0388]**	[0.0397]	[0.0317]	[0.0243]**	[0.0393]+
Adjusted R^2	0.0024	0.0033	0.0018	0.004	0.0031	0.0022
[B] Individual controls						
1918H2=Exposed						
3rd Trimester (β_1)	-0.0733	-0.0981	-0.0154	-0.1049	-0.0992	-0.0375
	[0.0198]***	[0.0367]**	[0.0383]	[0.0295]***	[0.0241]***	[0.0391]
1919H1=Exposed						
1st/2nd Trimester (β_2)	-0.1069	-0.1430	-0.0968	-0.0871	-0.0861	-0.1215
	[0.0238]***	[0.0449]**	[0.0464]*	[0.0352]*	[0.0289]**	[0.0478]*
1919H2=Exposed						
1st Trimester (β_3)	-0.0580	-0.1258	-0.0192	-0.0402	-0.0679	-0.0731
	[0.0200]**	[0.0376]***	[0.0383]	[0.0298]	[0.0243]**	[0.0392]+
Adjusted R^2	0.2153	0.0571	0.0622	0.1461	0.0051	0.0046
Observations	330,255	94,430	104,895	130,930	195,270	106,965
Mean of Dep. Var	8.9	9.9	7.5	9.4	9.9	7.4
SD of Dep. Var	2.7	2.5	2.7	2.4	2.2	2.6

Table 2.9: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Labour Market Outcomes (1971 Census; Males)

Cohort/Trimester of Exposure	Wages	Income	Employed	WWII Veteran
[A] Baseline, no controls				
1918H2=Exposed				
3rd Trimester (β_1)	-124.80	-154.50	-0.0009	0.0285
	[53.95]*	[66.06]*	[0.0032]	[0.0044]***
1919H1=Exposed				
1st/2nd Trimester (β_2)	-186.94	-263.96	-0.0043	0.0236
	[60.09]**	[71.48]***	[0.0038]	[0.0052]***
1919H2=Exposed				
1st Trimester (β_3)	-67.34	-110.52	0.0011	0.0420
	[53.12]	[65.58]+	[0.0031]	[0.0043]***
Adjusted R^2	0.0042	0.0025	0.0043	0.0410
[B] Individual controls				
1918H2=Exposed				
3rd Trimester (β_1)	-151.87	-185.22	-0.0011	0.0321
	[52.81]**	[65.08]**	[0.0031]	[0.0042]***
1919H1=Exposed				
1st/2nd Trimester (β_2)	-202.48	-280.50	-0.0038	0.0284
	[58.77]***	[70.10]***	[0.0037]	[0.0049]***
1919H2=Exposed				
1st Trimester (β_3)	-130.40	-176.63	-0.0003	0.0428
	[51.99]*	[64.57]**	[0.0030]	[0.0041]***
Adjusted R^2	0.0489	0.0345	0.0351	0.1384
[C] Individual and cohort controls				
1918H2=Exposed				
3rd Trimester (β_1)	-62.88	-60.85	-0.0007	0.0276
	[61.32]	[77.77]	[0.0037]	[0.0050]***
1919H1=Exposed				
1st/2nd Trimester (β_2)	-125.74	-185.51	-0.0051	0.0267
	[65.62]+	[81.03]*	[0.0041]	[0.0055]***
1919H2=Exposed				
1st Trimester (β_3)	-82.20	-81.90	0.0033	0.0360
	[54.25]	[67.04]	[0.0032]	[0.0043]***
Adjusted R^2	0.0640	0.0450	0.0415	0.1401
Observations				
Mean of Dep. Var	6,203	7,902	0.82	0.44
SD of Dep. Var	6,088	7,878	0.38	0.50

Notes: This table reports the estimated deviation from trend in the given outcomes among individuals that were in utero during the peak of the 1918 flu pandemic (October 1918), by province of birth. Each column-panel combination represents a regression. Model specifications are the same as Table 2.5. *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Table 2.10: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Grade Attainment (1971-1991 Censuses; Males)

Cohort/Trimester of Exposure	1971	1981	1991
[A] Baseline, no controls			
1918H2=Exposed			
3rd Trimester (β_1)	-0.0960	-0.0502	-0.1228
	[0.0243]***	[0.0690]	[0.0469]**
1919H1=Exposed			
1st/2nd Trimester (β_2)	-0.1089	-0.1393	-0.0962
	[0.0289]***	[0.0790]+	[0.0527]+
1919H2=Exposed			
1st Trimester (β_3)	-0.0559	0.0662	-0.1293
	[0.0237]*	[0.0639]	[0.0441]**
Adjusted R^2	0.0045	0.0068	0.0037
[B] Individual controls			
1918H2=Exposed			
3rd Trimester (β_1)	-0.1079	-0.023	-0.0948
	[0.0218]***	[0.0622]	[0.0431]*
1919H1=Exposed			
1st/2nd Trimester (β_2)	-0.1051	-0.0693	-0.0632
	[0.0261]***	[0.0701]	[0.0490]
1919H2=Exposed			
1st Trimester (β_3)	-0.0862	0.0686	-0.118
	[0.0214]***	[0.0604]	[0.0408]**
Adjusted R^2	0.1941	0.1531	0.1522
[C] Individual and cohort controls			
1918H2=Exposed			
3rd Trimester (β_1)	-0.0645	-0.0164	-0.0714
	[0.0256]*	[0.0721]	[0.0513]
1919H1=Exposed			
1st/2nd Trimester (β_2)	-0.0750	-0.0672	-0.0542
	[0.0289]**	[0.0784]	[0.0558]
1919H2=Exposed			
1st Trimester (β_3)	-0.0532	0.0958	-0.0774
	[0.0224]*	[0.0625]	[0.0424]+
Adjusted R^2	0.1959	0.1536	0.1525
Observations	322,250	153,570	106,400

Notes: This table reports the estimated deviation from trend in the given outcomes among individuals that were in utero during the peak of the 1918 flu pandemic (October 1918), by province of birth. Each column-panel combination represents a regression. Model specifications are the same as Table 2.5 (note that the 1981 and 1991 regressions do not control for WWII status, which is unobserved). *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Table 2.11: Effects of *In Utero* Exposure to 1918 Flu Pandemic on Grade Attainment (1971-1991 Censuses; Females)

Cohort/Trimester of Exposure	1971	1981	1991
[A] Baseline, no controls			
1918H2=Exposed			
3rd Trimester (β_1)	-0.0887	-0.1033	-0.0988
	[0.0224]***	[0.0460]*	[0.0390]*
1919H1=Exposed			
1st/2nd Trimester (β_2)	-0.1074	-0.1305	-0.0293
	[0.0270]***	[0.0637]*	[0.0440]
1919H2=Exposed			
1st Trimester (β_3)	-0.0579	-0.0434	-0.0712
	[0.0226]*	[0.0469]	[0.0384]+
Adjusted R^2	0.0024	0.0031	0.0018
[B] Individual controls			
1918H2=Exposed			
3rd Trimester (β_1)	-0.0733	-0.0836	-0.0721
	[0.0198]***	[0.0423]*	[0.0349]*
1919H1=Exposed			
1st/2nd Trimester (β_2)	-0.1069	-0.1123	0.0009
	[0.0238]***	[0.0587]+	[0.0394]
1919H2=Exposed			
1st Trimester (β_3)	-0.0580	0.0144	-0.0443
	[0.0200]**	[0.0454]	[0.0346]
Adjusted R^2	0.2153	0.1796	0.1862
[C] Individual and cohort controls			
1918H2=Exposed			
3rd Trimester (β_1)	-0.0478	-0.0394	-0.0387
	[0.0234]*	[0.0501]	[0.0416]
1919H1=Exposed			
1st/2nd Trimester (β_2)	-0.0828	-0.0895	0.0299
	[0.0264]**	[0.0626]	[0.0448]
1919H2=Exposed			
1st Trimester (β_3)	-0.0646	0.049	-0.0197
	[0.0209]**	[0.0471]	[0.0358]
Adjusted R^2	0.2155	0.1799	0.1863
Observations	330,655	175,745	144,095

Notes: This table reports the estimated deviation from trend in the given outcomes among individuals that were in utero during the peak of the 1918 flu pandemic (October 1918), by province of birth. Each column-panel combination represents a regression. Model specifications are the same as Table 2.5, except here none of the regressions control for WWII veteran status. *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Figure 2.1: Influenza Mortality During the 1918 Flu Pandemic and Adjacent Months

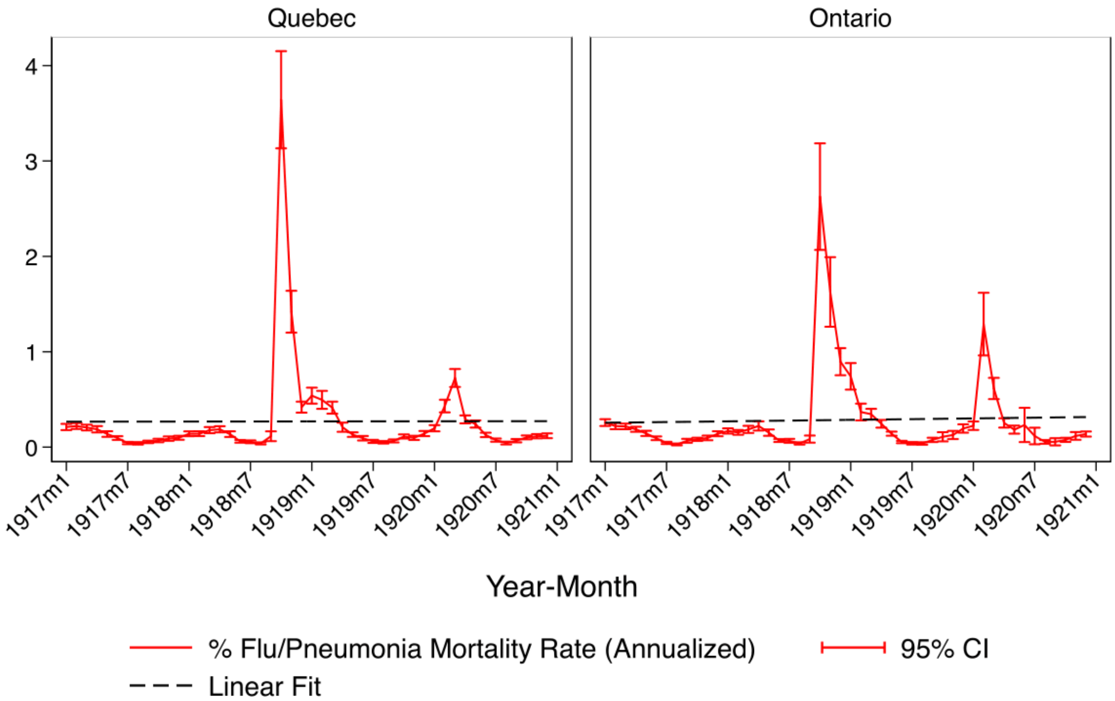
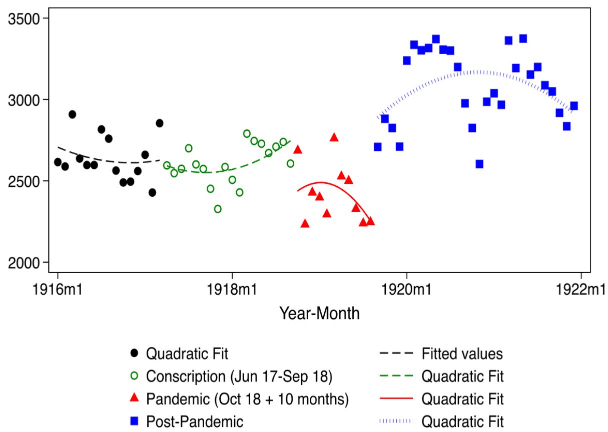
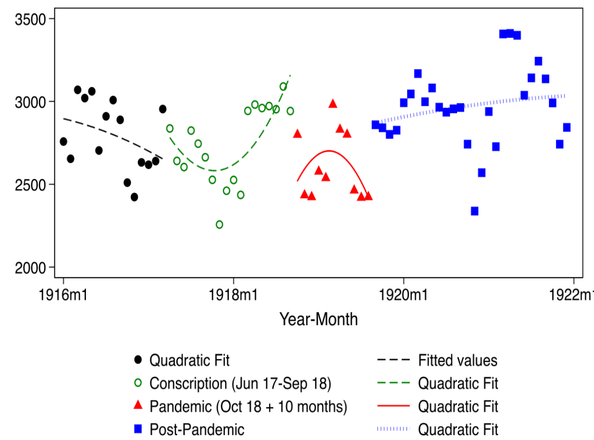


Figure 2.2: Live Births During Conscription, Pandemic and Adjacent Months in the Province of Ontario

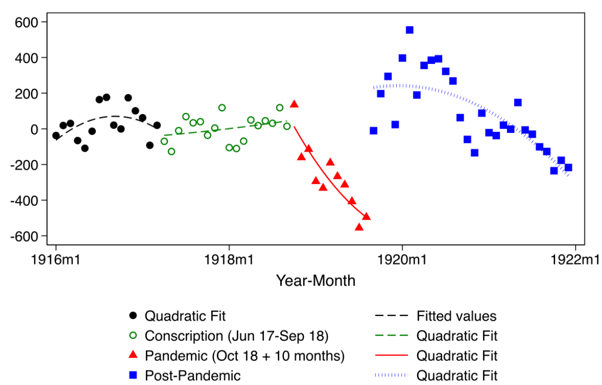
[a] Live Births – Urban Areas (Observed)



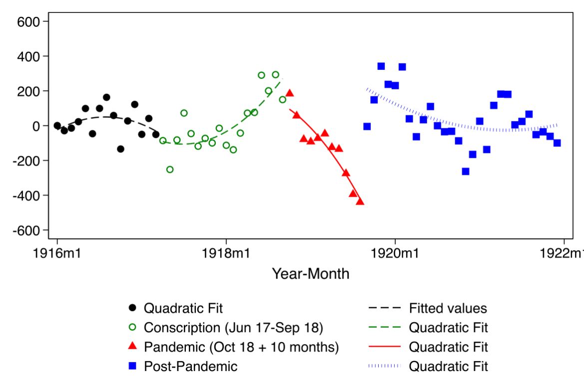
[b] Live Births – Rural Areas (Observed)



[c] Live Births – Urban Areas (Adjusted)



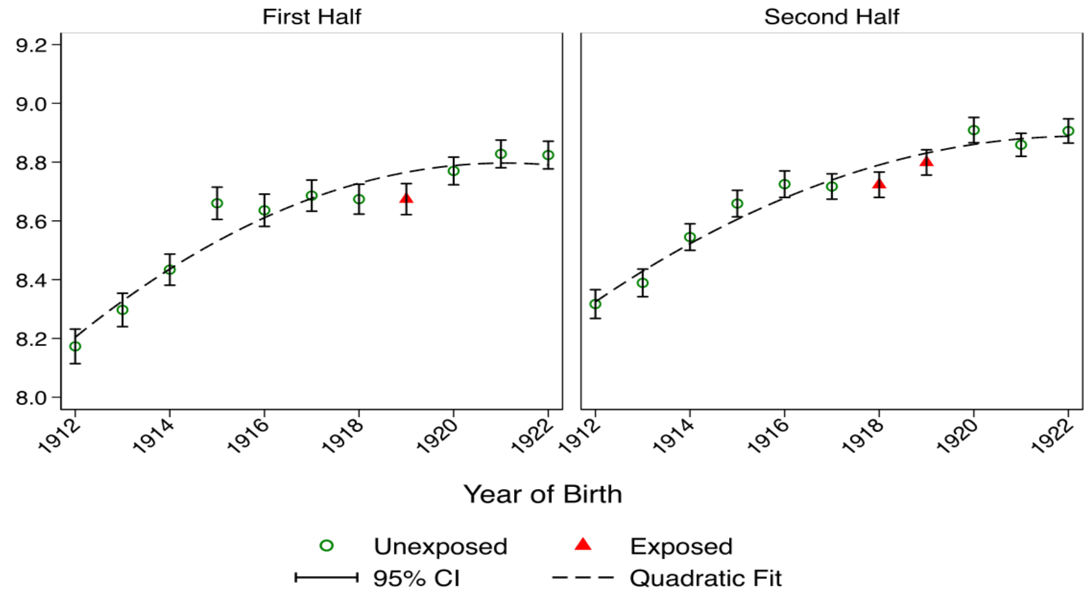
[d] Live Births – Rural Areas (Adjusted)



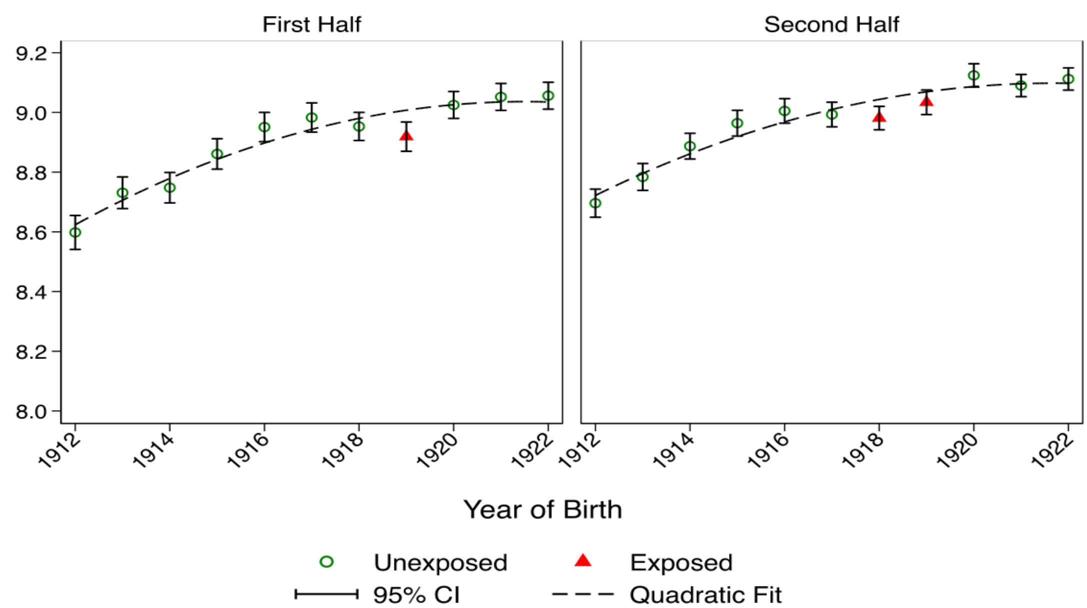
Notes: This figure displays aggregate fertility in Ontario during the pandemic and adjacent time periods (1916-1921). Live births are aggregated by month and region (i.e. urban and rural regions). Urban regions include cities and towns, while county data are treated as rural. The observed series includes the total number of live births that were reported in each month and region. The adjusted series include the predicted residuals of live births from an OLS regression of live births on a quadratic time trend and a set of month dummies. To illustrate how conscription and the pandemic affected fertility levels and trends, the series is divided into four sub-periods: the pre-conscription WWI period, the pre-pandemic conscription period, the pandemic period (defined as October 1918-August 1919) and the post-WWI period. Piece-wise quadratic time trends are displayed for each sub-period.

Figure 2.3: Grade Attainment by Pandemic Exposure Status: 1971 Census

[a] Males



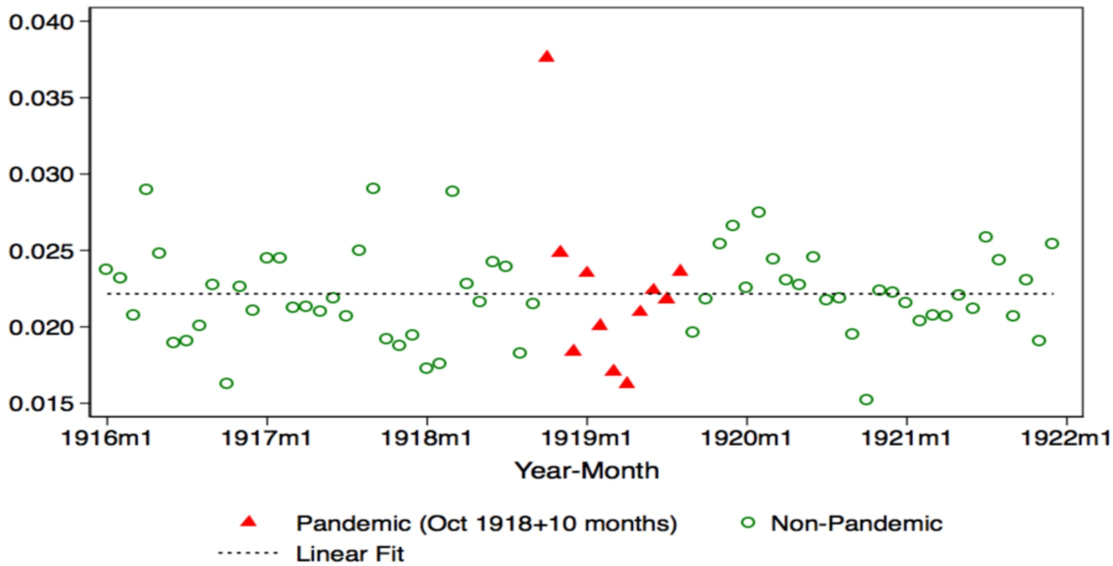
[b] Females



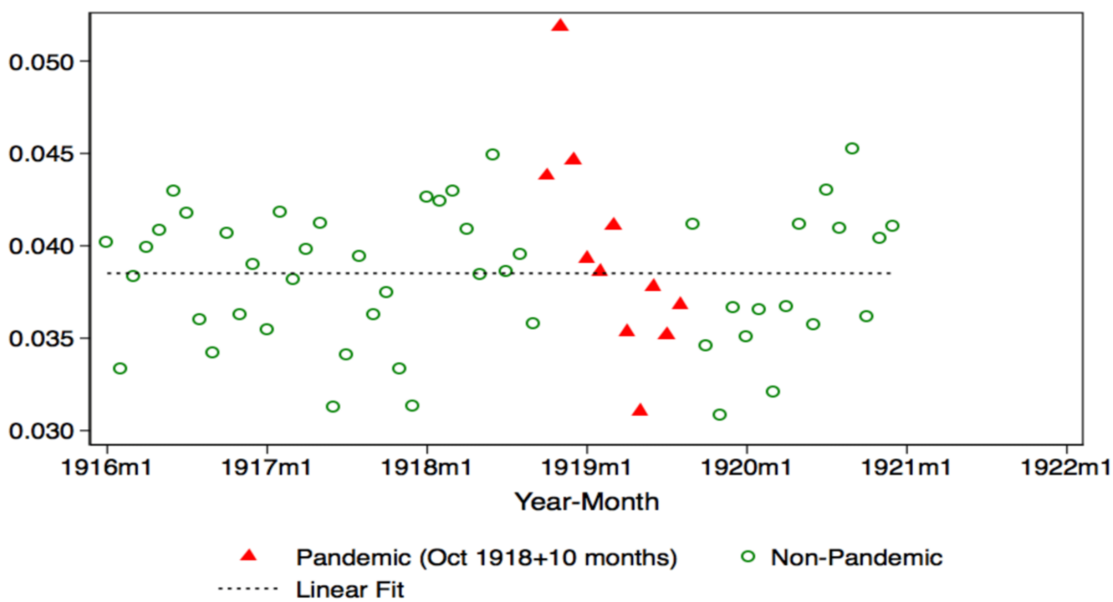
Notes: This figure displays mean grade attainment by half-year birth cohort and sex using the 1971 census. Grade attainment is defined as the highest grade of primary/secondary schooling ever attained. Each series is fitted to a quadratic birth year trend.

Figure 2.4: Prenatal Mortality by Pandemic Exposure Status and Province of Birth

[a] Quebec



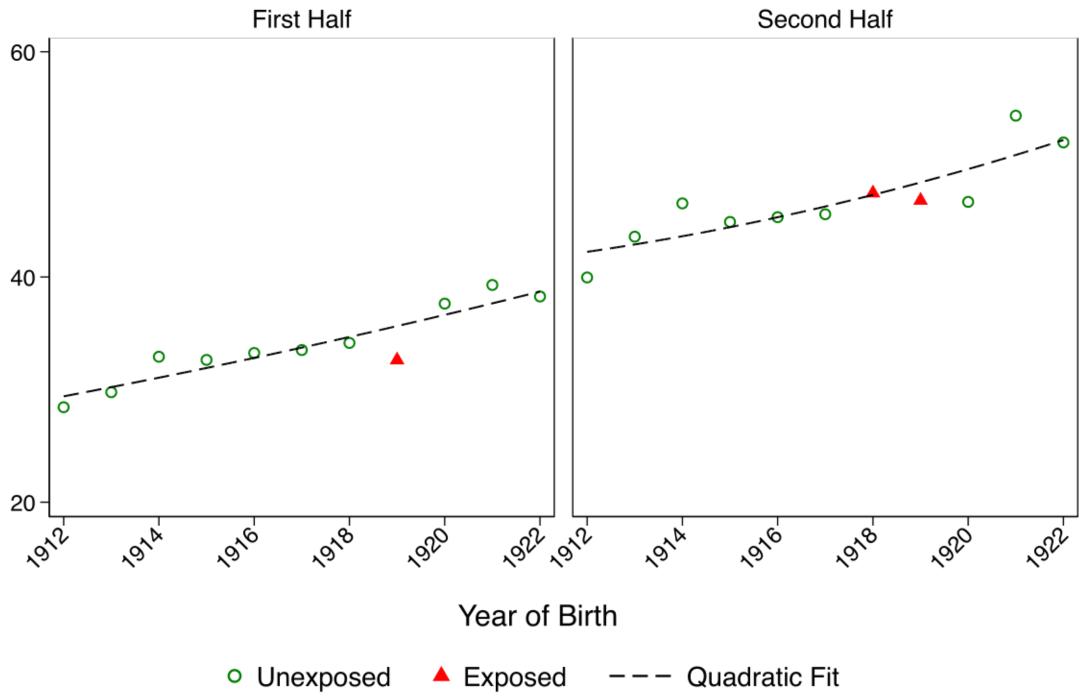
[b] Ontario



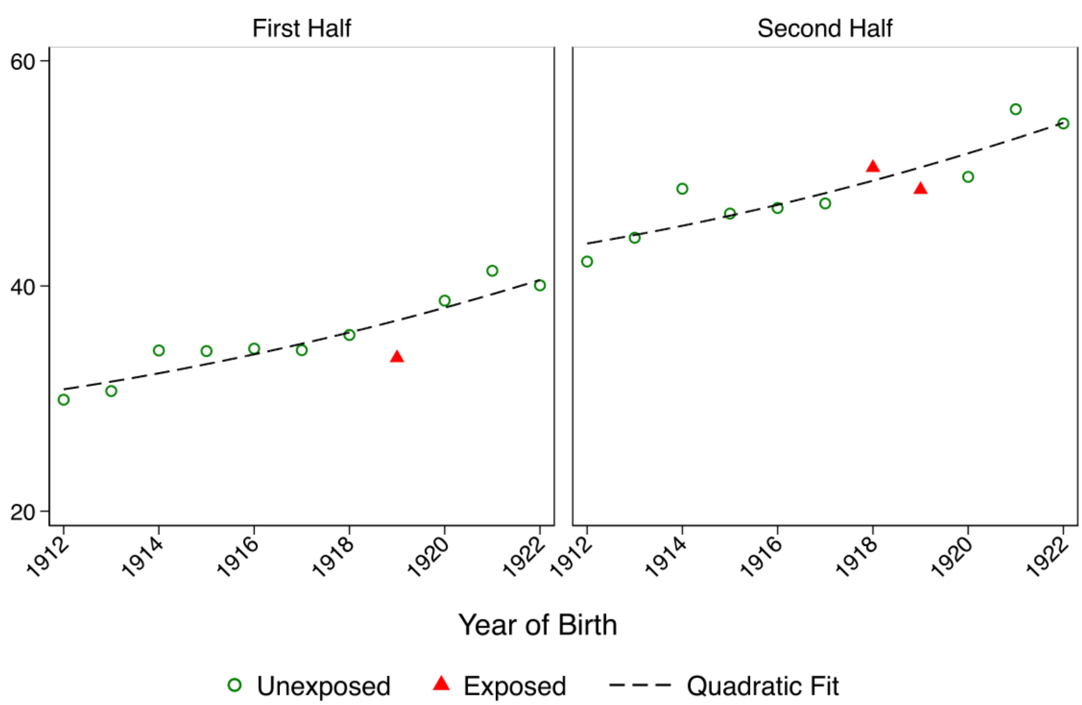
Notes: This graph presents the residual time series variation that remains after regressing the prenatal mortality rate on a quadratic time trend and a set of month dummies. That is, we plot $r_t = \widehat{e_t} + \bar{y}$, where $\widehat{e_t}$ is the predicted residual for month-year t and $\bar{y}$ is the mean mortality rate across all t. These regressions were estimated separately for each province. Note that da4ta4are not reported for Ontario after 1920.

Figure 2.5 Cohort Sizes by Pandemic Exposure Status and Gender: 1971 Census

[a] Males



[b] Females



Notes: This figure displays the estimated population size in each half-year birth cohort by sex in the 1971 census. Each series is fitted to a quadratic trend.

Chapter 3: Elderly Financial Well Being: The Role of Nursing Home Costs in Canada

3.1 Introduction

This paper examines the impact of residential long-term care (RLTC) on the financial well-being of older adults in Canada. Although the Canada Health Act prohibits user fees for medically physician and hospital care, it allows them for extended health care services, which include RLTC. Moreover, because RLTC programs are designed and delivered provincially/territorially, fees for RLTC vary regionally. In Canada, copayments for RLTC represent 20-95% of the median disposable income of seniors, depending on province of residence and marital status (MacDonald, 2015; Fernandes and Spencer, 2010), as well as 23% of total payments to RLTC facilities.[23]

Our analysis focuses on married seniors, for whom RLTC may necessitate an involuntary separation that diverts marital income/assets from the spouse who remains at home.[24] Specifically, we simulate the *potential* change in discretionary income that would occur if the spouse of an older adult were to receive RLTC. We term the difference in discretionary income as the potential financial hardship from RLTC, which has the advantage of focusing on the potential impacts of the RLTC choice under different programs while abstracting from the care decision itself and obviating the need for data on RLTC recipients. To isolate program-related variations in financial hardship from those generated by regional differences in financial resources, we calculate RLTC copayments, given income, for each province and period (from 2000-2010) using a fixed

[23] Statistics Canada, CANSIM Table 107-5508.
[24] This diversion is particularly pertinent to older women, who are more likely than their male counterparts to experience spousal institutionalization (Braithwaite, 2002) and whose financial status is more closely linked to that of a spouse (McDonald and Robb, 2004).

national sample of older married seniors from the Survey of Labour and Income Dynamics. We then use the resulting estimates to simulate the financial hardship caused by out-of-pocket RLTC expenses.

This paper makes several contributions to the literature on the financial well-being of older adults. First, it documents the provincial means tests that were used to determine copayments for RLTC over the period 2000-2010. This period is longer than has been covered in past research and permits a retrospective evaluation of the financial consequences of program changes in RLTC. Second, whereas prior studies have simulated copayments for RLTC (MacDonald, 2015; Fernandes and Spencer, 2010), our emphasis on income net of copayments provides a more direct income-based measure of the potential material impoverishment that can occur when a spouse enters care. Third, we provide an alternative policy analysis of the financial well-being of seniors within a literature that has traditionally emphasized retirement income security programs. Finally, to the extent that RLTC expenses vary by province, our assessment of their associated financial hardships are informative about regional economic inequality among seniors.

Consistent with previous work on the size of RLTC copayments, we find that simulated potential copayments are large compared to income. On average, they represent 30-40% of combined disposable incomes (e.g. total incomes minus taxes and transfers) of elderly couples. After adjusting combined incomes for family size, the ratios increase to 42- 55%. Conditional on provincial differences in income, the potential impact of nursing home costs on average discretionary incomes (of spouses who remain at home) varies provincially from a net gain of $2,500-$5,000 (Saskatchewan) to a net loss of almost $12,500 (New Brunswick). These impacts are highly variable. On average, spouses lose more income in the Maritimes (especially before 2005) and Manitoba. These results implicate nursing home care as a potential source of spousal impoverishment, based on traditional income-based "relative poverty" thresholds (e.g. the Low-Income Measure, commonly known as the LIM).[25] However, we also show that policy variables that mitigate the financial hardship of RLTC include spousal allowances, the division of spousal incomes, implicit tax rates, and maximum/standard user fees.

3.2 Background

3.2.1 Residential Long-Term Care

Long-term care (LTC) includes social or medical care given to individuals (usually seniors) with chronic health problems that make independent living and self care difficult.[26] Unlike acute medical care, LTC manages, rather than cures, chronic illness, so care often occurs indefinitely and commonly as a part of end-of-life (palliative) care. This paper focuses on publicly-funded residential LTC (e.g. nursing homes, personal care homes), which typically provide 24-hour nursing/health care, room and board, and meals (note that there is some variation across provinces in the scope of "care" that receives public funding; see MacDonald (2015)).[27]

[25] Canada does not officially use the term "poverty" when characterizing relative income. In this context and throughout the paper, "poverty" should be interpreted as "low income."

[26] Typically, care services are categorized into help with Activities of Daily Living (e.g. bathing, dressing, moving about), Instrumental Activities of Daily Living (e.g. cooking, household chores, transportation) and supervised medical assistance.

[27] This excludes home care and home support, as well as care provided in hospitals (e.g. geriatric wards, alternate-level of care (ALC) beds), lower-level care facilities (e.g. assisted living centres) and retirement homes. Here, we refer interchangeably between nursing homes, residential LTC and residential care.

In general, nursing home utilization rises in advanced age. According to data from Statistics Canada, nearly 10% of seniors aged 80 or older lived in nursing homes in 2009, representing more than 60% of all residents in long-term care facilities.[28] Given the intensity of care involved and the age of clients, most RLTC recipients have severe activity limitations and health conditions that necessitate institutional care. According to the Canadian Institute for Health Information (CIHI 2018), musculoskeletal (e.g. arthritis), neurological (e.g. dementia), and heart/circulatory (e.g. hypertension) conditions were present in a majority (>50%) of assessed residents of RLTC facilities in 2015. Likewise, 57% of residents were classified as extensively to fully dependent on others in their ability to perform basic activities of daily living (e.g. personal hygiene, toileting, locomotion, and eating).

The bulk of RLTC expenditures is financed from out-of-pocket expenditures and public. financing.[29] LTC, like that of general medical care, is mentioned in the Canada Health Act (CHA), but regulatory requirements relating to public funding of LTC are limited. For instance, the CHA requires comprehensive universal insurance for medically necessary health care services (e.g. hospital and physician care), but not LTC.[30,31]

---

[28] This ratio is derived as the ratio of residents living in residential care facilities (CANSIM Table 107-5504) to the estimated population of individuals aged 80 and older.
[29] Private insurance payments are limited. For example, in Canada, the fraction of private insurance payments in total LTC expenditures is only 0.5% (Colombo et al., 2011).
[30] Technically, extended health care services include nursing home intermediate care, adult residential care (including nursing homes), home care and ambulatory care.
[31] Health care is delivered by the provinces, with financing coming from provincial and federal revenue sources. Federal contributions include equalization payments, as well as cash transfers and transferred tax points, which are allocated according to the Canada Health Transfer (CHT). In 2014, the federal government was responsible for about 50% of total health care expenditures (inclusive of transferred tax points). The federal government can withhold contributions to provinces that are found to be in non-compliance with the CHA; an example would be if a province imposed user fees for medically necessary health care. Such penalties do not apply in the case of “extended health care services”, which include LTC.

Consequently, each province has a unique and autonomous public funding program for LTC under which copayments are usually means tested.[32] Most provinces use income tests, while a subset of provinces (Newfoundland, New Brunswick and Quebec) also use liquid asset tests,[33] which exclude the value of a principal residence, unless sold. On average in Canada, public subsidies account for 74% of revenues to residential care facilities.[34] However, since means testing varies regionally, public support varies by province and income/wealth level.[35]

3.2.2 Residential Care and Elderly Financial Well Being

When considering elderly financial well-being, many studies focus on income or various measures of income deprivation or low-income status (Myles, 2000; Osberg, 2001; Milligan, 2008; Veall, 2008; Milligan and Wise, 2013; Milligan and Schirle, 2013; Bernard and Li, 2006). Based on this work, there is a broad consensus that Canadian seniors have adequate incomes, largely due to maintenance payments available from the retirement security system, which ensure that most seniors are above the low income threshold.[36] Despite this, both immigrants and unmarried women remain particularly

---

[32] One exception is Nunavut, which provides nursing home care free of charge.
[33] Prior to 2005, the Maritime provinces also used asset tests. These were abolished sequentially from 2005-2007. In late 2015, New Brunswick announced that it would be introducing a different financial asset test.
[34] We estimated this figure as the proportion of total revenues accrued to RCF facilities from governmental sources, as reported in the Residential Care Facilities (RCF) Survey for the 2009-2010 fiscal year (Statistics Canada, CANSIM Table 107-5508). In total, Overall, $10.2 of revenues were sourced from various levels of government. The remaining revenues were from resident insurance, self-pay and differential fees for preferred accommodation (e.g. private rooms), and sundry earnings.
[35] Using the same RCF data, public subsidy rates range from 66% (Ontario) to 80% (Newfoundland). Note, however, that these estimates do not control for provincial differences in elderly income or wealth.
[36] Although Canadian seniors have lower incomes than non-seniors, their incomes are higher than seniors in other countries. Moreover, the percentage of seniors whose income falls below the low

vulnerable to impoverishment in old age. Moreover, among the latter, widowhood is frequently implicated as a critical income-related financial risk, particularly for women (Veall, 2008; Bernard and Li, 2006).[37]

However, elderly well-being is shaped not just by income, but also by expenditures, which may change in response to changes in health. For instance, expenses associated with the care of a spouse could divert financial resources from personal consumption (e.g. food, clothing, transportation, recreation) and fixed household expenses (e.g. utilities, rent/- mortgage, property taxes), thereby reducing well-being. Such expenditures are potentially significant; for example, previous simulations, based on actual income tests, have assessed nursing home copayments at 20-70% of joint incomes for married seniors, depending on province of residence (MacDonald, 2015; Fernandes and Spencer, 2010). Given that these costs tend to arise near the end of life, previous estimates of the income effects of widowhood likely understate the total financial impact of spousal mortality. Thus, the basic argument of this study is that assessments of elderly financial well-being are likely to be enhanced by the inclusion of the cost of end-of-life care. A further issue, given that costs can potentially vary by province, is that nursing home expenditures could generate regional health inequalities (in terms of expenditure risks). From a policy standpoint, this could undermine the inequality-reducing goals of the retirement security system, particularly the OAS/GIS component.

---

income threshold (defined as 50% of median income for all families) is 9%, which compares favourably to non-seniors, children and seniors from other countries.

[37] Both studies use tax filer data from the Longitudinal Administrative Database to derive the incomes after taxes and transfers of seniors before and after becoming widows. They find that female widows experience an 8% drop in equivalent disposable income within six years of the death of their spouse.

Therefore, the focus of this paper is on the financial impact of spousal institutionalization, with emphasis on how such impacts vary regionally and over time. Specifically, we simulate the potential change in discretionary financial resources that could occur if a spouse were to enter a public LTC facility. This differs from previous related work in several ways. First, while previous studies have simulated nursing home copayments for married seniors, none have predicted associated impacts on spousal discretionary incomes. Second, in addition to evaluating potential changes in spending money, our approach also allows us to evaluate the extent to which married seniors become impoverished after a spouse enters care. By considering the various provincial income tests, we are also able to assess the extent to which certain policy features protect against spousal impoverishment. We also exploit a broader set of policy cycles (2000-2010) to evaluate the impact of province-specific policy changes on elderly well-being.

3.3 METHODS

This study uses two data sources: (1) income data from the Survey of Labour and Income Dynamics (SLID); and (2) provincial income-testing policies. Given a couple's joint marital income, copayments for nursing home care of a spouse depends on five factors: (1) the basic fee structure (e.g. minimum and maximum fees); (2) personal allowances, which go to the nursing home resident[38]; (3) the implicit tax rate on excess income (over the allowance); (4) sharing rules for combined marital incomes; and (5) spousal allowances (i.e. a minimum level of income for the spouse of a nursing home resident). Using both income and these province-specific income tests, we simulate annual potential

[38] These are like income disregards in the case of social assistance benefits.

copayments for nursing home care as well as the amount of discretionary income that would remain after an individual's spouse enters care.

3.3.1 Policy Data

Information on the income tests is obtained from provincial statutes on long-term care, health care and social assistance, and correspondence with provincial health and social welfare departments. Where possible, this information is validated against that published in the literature or on related websites. Appendix A provides further details on our data collection. Depending on their income and the income of their spouse, nursing home residents are charged a stated minimum fee up to a maximum fee. The minimum fees, shown in Table A2, range from $0 (if not stated) to $12,640, which is less than the maximum OAS/GIS benefit for a single elderly person. As shown in Table A3, maximum fees per year range from $10,300 (Alberta, in 2000) to $48,472 (Nova Scotia, in 2004), representing 50-120% of median elderly income based on incomes reported in the 1999 SLID (authors' estimates). In general, these fees cover the cost of accommodation (e.g. rent, food and related custodial services), while the provinces often fully subsidize the cost of care. Exceptions to the latter include the Atlantic Canadian provinces before 2005 (when resident fees included the cost of care and accommodation), so that, from 2000 to 2004, maximum fees were about 50% higher in Atlantic Canada than in the next-costliest province.[39] Note, however, that several policies moderate the impact of income on

[39] For example, in 2004, the lowest maximum annual fee among the Atlantic provinces was $34,065 (Newfoundland and Labrador), compared to $24,163 in British Columbia. From 2005 to 2007, each of the Maritime provinces introduced coverage for the medical care received in nursing homes, leaving residents responsible for room and board. Newfoundland and Labrador maintained its nominal maximum fee at $34,064 for the duration of our study period so that its real maximum fee declined annually with inflation.

copayments. For example, nursing home residents retain a minimum personal allowance regardless of their income.[40] Also, in addition to this amount, residents contribute a fraction of excess income (over the personal allowance), up to the maximum fee. The implicit tax on income varies by province and income (from 50-100%), with the resident (or couple) retaining the residual percentage of excess income.

Further adjustments to the cost of care are made if the resident has a spouse living at home. In this case, fees also depend on the treatment of spousal income. In general, fees are calculated from combined marital income (Newfoundland, New Brunswick, Manitoba, British Columbia before 2010), individual income (Ontario, British Columbia starting in 2010) or a share of combined income (Prince Edward Island, Nova Scotia, Saskatchewan and Alberta). The percentage of combined income is usually 50%. These policies affect the share of marital income retained by the non-institutionalized spouse. For instance, if copayments are based on combined spousal income, then the spouse who lives at home retains, at most, the difference between the couple's combined income and the assessed copayment (including, in this case, the basic personal allowance). If they are based on half of combined income, the spouse retains at least the other half. If individual income is used, the spouse retains their own income, which may be more or less than 50% of combined income depending on each spouse's contribution to the couple's combined income.

Additionally, most provinces also provide spousal allowances or hardship waivers.[41] These policies allow spouses to retain a greater share of combined income

[40] The personal allowance is a flat amount or pegged to the difference between the OAS/GIS for a single elderly and the minimum annual fee.
[41] Saskatchewan is the only exception.

when their proportion of assessed income falls below a particular income threshold. Several provinces provide a flat spousal allowance, usually set above the OAS/GIS rate for a single elderly person. Others instead apply a financial hardship test. In the latter case, a spouse is granted temporary relief if they can demonstrate financial hardship owing to their spouse's copayment. Although this is somewhat ambiguous, some provinces define financial hardship more concretely. For example, Prince Edward Island defines it as having a disposable income below the maximum combined OAS/GIS benefit for a single person. Since this seems to be a reasonable definition, we set the spousal allowance to the OAS/GIS rate (single) in all provinces/years for which a financial hardship test was in effect.[42]

Uniquely, Newfoundland uses a modified spousal allowance consisting of a flat spousal amount ($10,800 per year), an amount for travel to visit the spouse under care, and a variable expense component. The expense component allows spouses to retain more of the institutionalized spouse's income to pay for various "allowable" household expenses (e.g. rent, mortgage, utilities, cable, telephone). Using data from the Survey of Household Spending on approximately similar spending categories, we find that such expenses account for a variable share of joint marital income for older married seniors. This percentage ranges from 51% in the bottom income decile to 22% in the top income decile (Table A8). Thus, given a couple's position in the after-tax income distribution, we set their spousal allowance under Newfoundland's income test equal to the flat amounts

[42] In terms of our simulations, this assumption may lead to an understatement of the financial impacts of spousal institutionalization when hardship waivers are applied more stringently (e.g. if the spousal income threshold is set below the OAS/GIS benefit).

plus the variable share of income, up to a maximum of their combined after-tax incomes minus the resident's minimum personal allowance.

3.3.2 Income Data

Our analysis employs the cross-sectional public-use microdata files of the Survey of Labour and Income Dynamics (SLID). The SLID is a nationally representative household survey that collects a rich set of income and socio-demographic variables for each individual of a sampled household. The main variables used in our analysis include age, sex and marital status, and one of two income variables: total income and income after taxes and transfers (i.e. total income less taxes paid), which we call disposable income.[43] Because the SLID is a household survey, it contains unique family identifiers that allow us to identify a common set of variables for cohabitant spouses. We use these identifiers to conduct our analysis on a sample of couples rather than individuals.

We limit the sample to married or common-law spouses in which both spouses are aged 65 or older. To better target older adults who are vulnerable to spousal institutionalization (e.g. due to their age), we further restrict the sample to couples in which at least one spouse is aged 80 or older. Next, we drop couples that have a combined disposable income less than the combined maximum of OAS and GIS that would be payable to the couple if both spouses had no other income.[44,45] Finally, because

[43] Total income includes wages and salaries, self-employment or business income, investment income, public and private pension income, and other government transfers (e.g. GIS, social assistance, child benefits).
[44] Specifically, we use the yearly average of maximum benefits payable over the period 2000 to 2010.
[45] This exclusion has the advantage of ignoring social assistance benefit calculations that would arise in couples who do not qualify for the maximum OAS and GIS entitlements. In effect, this excludes immigrants that lack sufficient residency in Canada and seniors who do not apply for benefits.

each cycle of SLID data contains a rather small number of elderly individuals, we pool all 15 cycles (1996-2010) together. Overall, the analytic sample contains 5,210 couples or 10,420 older adults.

For each couple, we designate one spouse for "potential institutionalization," and assume the other spouse remains at home. Unfortunately, the SLID lacks consistent predictors of LTC use (e.g. health, disability) over time, so we assign the institutionalized spouse as either the older spouse (again reflecting the age dependency of residential care use) or by randomization when spouses are of the same age.[46] Because husbands tend to be older than wives, males account for 69% of the seniors assigned for institutionalization. Using this designation, we calculate potential nursing home costs for each institutionalized spouse and post-institutionalization potential discretionary income for the spouse who remains at home.

To ensure that our results capture regional and temporal policy differences rather than income differences, a potential copayment is assigned to each couple under each provincial income test over the period 2000-2010, regardless of the survey year. We adjust for secular growth in observed incomes across survey years by deflating incomes to 1998 constant dollars, then re-inflating them back to each "policy year" using growth in the Consumer Price Index. Thus, our results are derived by exposing a fixed income distribution (observed from 1996-2010) to each provincial income test over 2000-2010.

[46] We assign each spouse a randomly generated real number from 0 to 1 and then designate the spouse with the higher number for potential institutionalization. This procedure applies to 43% of couples in our sample.

3.3.3 Potential Copayments

Given the income of an elderly individual and their spouse, potential annual copayments equal the assessed cost of RLTC, based on a provincial income test plus the comfort allowance of a given province. The comfort allowance guarantees the institutionalized person a nominal portion of marital income for discretionary spending even when their care is fully subsidized. We define the copayment formula as:

(3-1) $$\boldsymbol{C = min(max(0, t(Y_0 - I_a) - S_a), P_m) + I_a}$$

where the copayment (C), is imputed from assessed income (Y_0), the personal allowance (I_a), the maximum nursing home fee (P_m), the implicit tax (t) on excess income and the spousal allowance (S_a). Spouses at home receive a spousal allowance when their remaining discretionary income, net of the copayment, falls below the spousal income threshold for their province.[47] The personal allowance is a minimum level of family income available to the nursing home resident.

The following describes the basic intuition of the copayment formula. If assessed income is below the personal allowance, then the institutionalized spouse is assigned a copayment of \$0. If it is above the personal allowance and no spousal allowance is applicable, then they pay a proportion, *t,* of their assessed excess income up to the maximum fee. Thus, copayments are proportional to income for incomes between the personal allowance up to the point where the maximum fee is charged, and regressive after this point. Depending on the size of the maximum fee, fee structures will vary in their degree of regressivity. If the fee is high relative to the average of assessed income for the target population, then potential copayments will be proportional to income over a

[47] We assume that the spouse at home retains combined marital income net of the copayment and personal allowance.

relatively wide range of incomes. Finally, if a couple is also eligible for the spousal allowance, then the assessed copay is reduced by the assigned allowance, down to a minimum of $0.

As stated earlier, “assessed” income varies by province and may include individual income, combined marital income or a share of combined income. Moreover, each province uses a particular (tax-based) definition of income (e.g. gross income, net income, or net after-tax income). Except in Alberta and Saskatchewan, assessed income excludes taxes payable.[48] All regions calculate income using official tax returns (Notice of Assessment). While SLID incomes are imputed from tax records 80% of the time, we do not observe actual tax data. Instead, we use “total income” to proxy for gross or net income, and disposable income (total income after tax and transfers) for net after-tax income.[49]

3.3.4 Potential Financial Well-being of Spouses

To assess the potential impact of nursing home costs on a spouse’s financial well-being, we consider a simple “thought experiment” in which one spouse enters care and the other remains at home. In this case, the spouse who stays at home retains the couple’s combined disposable income minus their spouse’s copayment (inclusive of the personal allowance). Therefore, discretionary income, given the disposable income of the wife (Yw) and husband (Yh), is calculated as

(3-2) $$\boldsymbol{Y_d = Y_w + Y_h - C}$$

[48] Many provinces assess income as net (line 236 from a person’s Notice of Assessment) or gross (line 150) income minus taxes payable.
[49] The difference between the tax-based definitions of net and gross income consists mostly of work-related deductions. Because less than 1% of our sample worked in the reference year, the difference between net and gross income for this sample is likely to be quite small.

This variable measures the maximum amount of discretionary income available to an elder for household expenses and personal consumption after their spouse enters care.[50] To quantify the financial impact of spousal institutionalization, we compare a senior's potential discretionary income after their spouse enters care (i.e. Y_d) to their observed level of financial well-being before institutionalization.

To account for economies of scale in household consumption,[51] pre-institutionalization financial well-being is defined using family equivalent disposable income, i.e. combined disposable income divided by the square root of family size (in this case, 2), As a result, the predicted individual-level potential financial impact of a nursing home admission (of a spouse) is the difference between the individual's simulated post-institutionalization discretionary income and their observed equivalent disposable income. Changes in financial well-being that stem from a spouse's entry into care reflect both a potential loss of discretionary income arising from increased RLTC expenses, and a decrease in the divisor of financial well-being (from 2 to 1). In this case these two components will have opposing effects on well-being.

We also use our results to measure the potential change in incidence impoverishment that occurs after a spouse enters care. We use the Low-Income Measure (LIM) to define the low income population. The rate of low income under the definition of the LIM is the proportion of couples whose equivalent after-tax income falls below one-half of median equivalent disposable income for the Canadian population. Since each

[50] Note that here we are assuming no behavioural responses by the non-institutionalized spouse. This rules out, for example, the possibility of downsizing to a smaller, less expensive home.

[51] For example, a couple may use their collective income to purchase in bulk at discounted prices (e.g. food) or other jointly consumed household goods (e.g. home heating, cable and internet services, cleaning services).

sampled individual has an income no less than the maximum OAS/GIS benefits for married seniors, the baseline LIM for our sample is zero. Therefore, the post-institutionalization LIM, derived from an individual's remaining discretionary income, directly captures the potential additional impoverishment generated from out-of-pocket RLTC spending.

3.4 Results

3.4.1 Summary Statistics

Table 3.1 reports sample summary statistics using the full sample of couples. Recall that each couple has one spouse designated for institutionalization (the "institutionalized"), while the other remains at home (the "spouse"). Across all couples, 69% of the "institutionalized" are male, so the same percentage of "spouses" are female.[52] The mean spousal age is 77, while 43% of spouses are 80 or older.[53] Regionally, more than half of the sample resides in Ontario, 9% in Atlantic Canada, 10.5% in the Prairies, 11% in Alberta, and 16% in British Columbia.

The sample mean of disposable income (e.g. total income net of taxes and transfers) among those designated for institutionalization is $25,877. [54] Their spouses have slightly lower incomes ($23,141), partly because they are disproportionately female. The overall sample mean of equivalent disposable income (combined income divided by $\sqrt{2}$) is $34,661. Since we exclude couples with combined disposable incomes below the

[52] One interpretation of this figure is that 69% of involuntary marital separations arise from a husband's admission into care.

[53] Age is top-coded at 80, so there is no variation in age in those 80 or older.

[54] Since this is a post-tax and transfer definition of income, it is inclusive of deductible medical expenditures and any other provincial/federal medical expense tax credits.

maximum OAS and GIS benefit available to a couple, the sample-wide pre-institutionalization LIM is zero.[55]

Neither of the income nor low income indicators given so far illustrate the risk of impoverishment that could potentially arise if a couple were to experience out-of-pocket medical or LTC expenses. To provide a baseline measure of this form of potential financial insecurity, Table 1 also reports the fraction of non-poor couples whose income falls within $5,000 of the low income threshold, which is 20%. In other words, one-in-five "spouses" are at risk of impoverishment from out-of-pocket medical or LTC expenditures over $5,000.

3.4.2 Potential Copayments by Sex and Year

Table 3.2 documents the sample means of simulated potential copayments under each provincial RLTC program, holding the income distribution constant. To keep the table legible and highlight the main policy changes that have occurred over the study period, we limit this portion of our analysis to RLTC programs in effect in 2000 and 2010. Here, potential copayments refer to the potential expense that a couple would incur in their province of residence if the "institutionalized" spouse were to enter care in a given year.

The sample mean of potential annual copayments is $18,954 using RLTC programs for 2000. This value drops slightly to $18,583 in 2010. Holding income constant, lower copayments over time reflect changes in RLTC programs only. Pertinent policy variables that affect potential RLTC expenses include maximum fees, which declined in the Atlantic provinces, and spousal allowances, which increased in Nova

[55] The official LIM-based low income rate for elderly couples averaged 5.5% over our study period (CANSIM Table 202-0804).

Scotia. Notice as well that the sample mean of copayments of an institutionalized wife is $2,000 (in 2000) to $3,000 (in 2010) below that of an institutionalized husband (see Columns 3-6 of Table 3.2). Since women have lower incomes than men, their copayments tend to be lower, particularly in provinces where copayments are assessed using individual rather than combined marital income. A transition from combined to individual income assessments in British Columbia, which came into effect in 2010, is partly responsible for the widening of the gender gap in average potential annual copayments.

Finally, if we divide individual copayments by the couple's equivalent disposable income, then the resulting ratios average 41-42% overall, 44-45% in couples with institutionalized husbands, and 35-37% in couples with institutionalized wives. As with copayment levels, the ratio of potential annual copayments to equivalent disposable income decreased between 2000 and 2010, especially among institutionalized wives. We thus conclude that, holding income fixed, nursing home care has become less costly over time.

3.4.3 Potential Copayments by Income Level

Since copayments are income tested, the above results mask important differences in the cost of care at different levels of disposable equivalent income. To address this, Figure 3.1 reports sample means of copayments by equivalent disposable income percentiles, across all couples.[56] Average annual copayments range from $14,000 in the bottom of the income distribution to $23,500 in the top. Relative to the couple's equivalent disposable

[56] The mean cell size per percentile is 52.

income, they fluctuate from 35-70% (Figure 3.2). Moreover, relative copayments are highest in the lower-middle range of the income distribution (e.g. 60-70%) than in the very low or middle-upper range of the income distribution (e.g. 30-40%). Also, as predicted earlier, copayments are proportional to income over moderate income ranges, and regressive over higher income ranges. Notice, as well, that the aforementioned decline in mean and relative copayments is particularly pronounced among middle and lower income seniors (e.g. from ventiles 1-13).

3.4.4 Provincial Differences in Potential Copayments

One of our objectives is to evaluate whether provincial policies correspond with regional differences in potential nursing home costs. In doing so, however, we are limited by the fact that provincial differences in the sample mean of copayments are likely to be affected by similar differences in the level and distribution of income.[57] For example, copayments may be small in a province simply because its residents are relatively poor. To control for such income differences, we simulate each "institutionalized" individual's copayment under each provincial income test over the study period, then take the sample mean of copayments by year and "province." By estimating province-by-year means using the same sample, provincial variation in potential RLTC costs can be attributed to regional differences in policies rather than income.[58] As illustrated in Figure 3.3, our simulations predict markedly lower potential annual copayments for Saskatchewan and Alberta than in other provinces. In Saskatchewan, potential copayments average slightly

[57] This is especially problematic because regions where elderly incomes are low (e.g. Atlantic Canada) have historically used more stringent means tests.
[58] Another issue is that some provincial sub-samples are quite small so that provincial differences in means are likely to be imprecisely estimated.

more than $10,000 per year throughout the study period. In Alberta, they are $15,000 from 2000-2002 and a little more than this afterwards. Elsewhere, copayments are $3,000-$10,000 higher depending on the province and year. Before 2005, the highest cost provinces (approximately $25,000 per year) include New Brunswick, Nova Scotia, Manitoba, and British Columbia. For RLTC programs for 2000, potential copayments are nearly $15,000 more per year under the most stringent income test than under the most generous one. This difference represents 44% of average disposable equivalent income.[59]

Figure 3.4 summarizes regional differences in income-related variation in potential annual copayments over time. Although copayments rise with income in all regions and years, it is evident that the Atlantic provinces are the costliest region for affluent couples before 2005. However, these costs fall dramatically from 2005 to 2007, reflecting declines in the maximum fee in each Maritime province. After 2007, the simulated cost of care for high-income couples under the policies of the Atlantic region are comparable to those generated for Manitoba and British Columbia. Based on policies for 2010, high-income couples incur the lowest annual nursing home costs in Ontario, Saskatchewan, and Alberta ($19,000-$21,000) and the highest in Newfoundland, Nova Scotia, New Brunswick, and British Columbia ($30,000-$34,000). Several policy factors keep fees relatively low in Ontario, Saskatchewan, and Alberta. For example, Alberta has the lowest maximum fee, and Saskatchewan applies an implicit tax rate of 50% on assessed income rather than the 80-100% rate used elsewhere.

[59] $14,000 represents the difference in mean copayments for New Brunswick ($25,000) and Saskatchewan ($11,000).

Although costs are high in the Atlantic regions for affluent couples, they are comparatively moderate for low-income seniors. For couples in the bottom quartile of disposable incomes, potential out-of-pocket expenses in the Atlantic region are large compared to Saskatchewan ($6,500), but small compared to British Columbia and Manitoba ($17,000). Manitoba and British Columbia, both of which assess fees from joint income using a 100% implicit tax rate, are ranked by our simulations as the least affordable locations for RLTC from the perspective of low-income seniors. In Newfoundland, costs are kept low by the provision of generous spousal allowance, which varies with allowable household expenses (e.g. food, rent, mortgage). Indeed, our simulations place Newfoundland as the second most affordable location for residential care, after Saskatchewan. Nova Scotia is also comparatively affordable after 2006, having adopted a more generous spousal allowance. For seniors in the bottom quartile of disposable incomes, in 2010 Nova Scotia ranks as the fourth most affordable province for RLTC. Our simulations for Saskatchewan predict an 19 average potential copay of $6,500 per year for low-income couples, making it the most affordable jurisdiction.

3.4.5 Potential Discretionary Income After Spouse Enters Care

We now explore the potential financial hardship of institutional care from the perspective of older adults who stay at home after their spouses enter care. To reiterate, we define discretionary income as the couple's combined disposable income, minus the potential copay of the institutionalized spouse and the minimum personal allowance in a given province.

Table 3.3 reports the main results for the beginning and end of the study period, thus highlighting the impact of policy changes between 2000 and 2010. The first row reports province-specific estimates of mean equivalent disposable incomes. In general, average disposable incomes are higher outside of Atlantic Canada (except in Saskatchewan). The subsequent rows report the sample-wide means of potential discretionary incomes available after the institutionalized spouse enters care in a given province. The fifth column reports the results for Ontario. The sample mean of equivalent disposable income among older adults who live in Ontario is $35,677. If a husband enters care for a year and is exposed to Ontario's income test, then his spouse's discretionary income would be around $28,650. If the wife enters care, then her husband retains $35,400. This sex gap is due to the fact that Ontario assesses copayments based on individual income; since husbands earn more than wives on average, and thus face higher copayments, wives retain less discretionary income when their spouse enters care.

Consistent with our estimates of regional differences in potential copayments, our simulations project larger post-institutionalization dicretionary incomes in Alberta and Saskatchewan than in the Maritimes, Manitoba, and British Columbia. This is true whether it is the wife or husband who enters care, and, in the Maritimes, when using the year-2000 policies. For example, if a husband were to enter care, his spouse would retain between $23,000-$25,000 if they lived in one of the Maritime provinces (using the 2000 policies) or Manitoba, compared more than $32,000 in Alberta or Saskatchewan. The regional gap in discretionary income is smaller when using the 2010 policies, as the Maritime provinces lowered their fees by 50% from 2005-2007. An increase over time in post-institutionalization discretionary incomes is particularly evident in Nova Scotia, mainly because the spousal and personal allowances were increased in 2005. The opposite applies to Alberta, which raised its maximum fee in 2002 and 2007.

Retained income also depends on whether spousal income is included in the copay formula. If copayments depend on individual income (NL, ON, BC in 2010), then husbands' copayments are generally higher (because their incomes are higher), so wives retain less discretionary income. This is also the case when copayments are based on a choice of individual or combined income (AB, SK). If they are instead strictly assessed from combined income (MN, NB) or half of combined income (NS, PE), then wives and husbands retain similar amounts of discretionary income after their spouse enters care.

Figure 3.5 converts the tabular results into a graph, which shows that post-institutionalization income typically falls below the sample average of equivalent disposable income. The only exception to this is in Saskatchewan, where equivalent income increases *after* a spouse's admission into care. In Saskatchewan, the spouse at home retains at least 50% of joint income plus the excess of the institutionalized spouse's assessed income over their simulated copayment. Since copayments are low, spouses who stay at home retain a higher average level of equivalent income after institutionalization than before. Figure 3.6 plots the percentage loss of discretionary income.[60] On average, the potential cost of care for an institutionalized spouse is associated with a 10-40% loss in discretionary income, except in Saskatchewan where it results in a 15% increase. Manitoba, British Columbia (after 2002) and New Brunswick (before 2007) each post income losses of 30-40%.

[60] That is, the post-institutionalization level of discretionary income of the spouse who stays at home minus baseline equivalent disposable income, expressed as a percentage of baseline income.

3.4.6 Potential Impacts on Relative Income

We have shown that the potential cost of care for a spouse can reduce discretionary income, potentially making it harder to finance household expenses and personal consumption. A limitation of our analysis, however, is that discretionary income is an absolute measure of financial well-being that does not capture relative income changes stemming from spousal institutionalization. For instance, in addition to experiencing a loss of income, a senior may also fall in the (discretionary) income distribution after their spouse enters care. This observation is informative about the potential stress associated with out-of-pocket RLTC expenses. For example, we know from prior work that there is a positive correlation between relative income and subjective well-being (i.e. happiness), even when controlling for observed income (see Clark et al., 2008, for a review). This fact points to relative income changes as a possible contribution to the distress that afflicts seniors after a spouse enters care.

Figure 3.7 illustrates this point. It reports, for each percentile of disposable equivalent income, the resulting percentile ranking of discretionary income after one's spouse enters care, holding constant the living arrangements of seniors in other income percentiles. Discretionary incomes are ranked against the pre-institutionalization income distribution, which has the advantage of isolating the relative loss of income for seniors in a particular percentile of the disposable income distribution. As is evident from the diagram, spousal institutionalization almost always results in a decline in a senior's percentile income ranking. 22 For example, consider the case of Manitoba. Here, an individual in the 50th percentile of pre-institutionalization disposable equivalent incomes falls into the bottom of the income distribution after their spouse enters care. The solid black lines document the percentile ranking of post-institutionalization discretionary incomes under the policies that existed at the beginning and end of our study period. In almost all provinces (except for Saskatchewan) and years, our simulations predict that a nursing home admission would generate a significant decline in relative income for the spouse at home. Such impacts are notably significant for spouses in the middle and lower parts of the pre-institutionalization disposable income distribution. In Prince Edward Island, Manitoba, and British Columbia (especially in 2000), this observation applies to virtually all seniors in the bottom half of the income distribution. We find similar impacts for Nova Scotia in 2000. In 2010, these effects are muted, again reflecting the adoption of a more generous spousal allowance.

In general, the size of the loss of relative income decreases as we move up the disposable income distribution. This pattern reflects the regressivity implied by the fee structures. Since fees are subject to a maximum, the proportion of income devoted to potential copayments falls as income increases beyond a critical point.[61] This, in turn, softens the potential decline in relative income that would occur if a spouse were to enter a residential care facility.

3.4.7 Potential Impacts on Low Income Status

We turn now to our final indicator: the Low Income Measure (LIM), defined as the proportion of seniors whose income is less than one-half of the median equivalent disposable income of all households. The LIM-based low income threshold is $17,814. The official LIM for seniors is 10.5%. Because the sample is restricted to seniors with a

[61] This "critical point" is the exact level of income at which a resident would be expected to pay the maximum fee.

combined disposable income above the combined maximum of OAS and GIS available to couples with no other income, our sample LIM is 0%. Despite this, 20% of the sample is within \$5,000 of the low income threshold, while 55% of the poorest 40% of seniors fall within the same radius. This suggests that out-of-pocket RLTC expenses, which usually exceed \$5,000, constitute a significant financial risk for a large fraction of elderly couples, particularly for those already "relatively poor." Since the baseline LIM is nil, our post-institutionalization LIM estimates capture the direct impoverishment brought about from a spouse's admission into care.

We project a potential post-institutionalization LIM of 0-62% depending on province and year (Table 3.2 and Figure 3.8). These findings are consistent with the simulated impacts on discretionary income. The low income impacts are largest in Manitoba (simulated LIM rates of 53-62%), where fees are based on a couple's combined income and a high implicit tax rate (approximately 100%). Impacts on the LIM are negligible or non-existent in Newfoundland (0-28%), Alberta (7-11%) and Saskatchewan (0%). In the latter two provinces, low fees (Alberta) and low implicit tax rates (Saskatchewan) reduce the average of simulated copayments, thus protecting the spouse against impoverishment following an admission into care. In Newfoundland, fees are high, but spousal allowances are generous. In Nova Scotia, the simulated LIM falls by half between 2000 and 2010, reflecting decreased fees and higher personal and spousal allowances.

The potential scope for spousal impoverishment is constrained when copayments are based on a share of joint income, low fees (Alberta), low implicit tax rates (Saskatchewan) or high allowances (Nova Scotia, Newfoundland). The former two policies ensure that nursing home costs remain low, while the latter furnishes a baseline consumption floor for spouses living at home. Programs that calculate copayments using joint income, but without generous spousal allowances (as in Newfoundland), generate LIM rates from 20% (New 24 Brunswick) to 55% (Manitoba). When fees are based on individual income, husbands face less impoverishment than wives. This is because women, who have less income on average, face lower copayments. This is evident in Ontario and British Columbia (2010). Since the well-being of women is more dependent on changes in spousal incomes, supplementation of individual income tests with large spousal allowances or with cost-of-living adjustments appears to be an effective way of reducing the risk of spousal impoverishment.

Notice, as well, that the depth of low income[62] is also quite large in some of the provinces. In Prince Edward Island, the mean percentage shortfall of income relative to the low income threshold is 20-24%. In Manitoba, again, it is much higher than in other provinces, at 15-40% depending on the year of potential institutionalization. Expressed in dollars, the average shortfall of discretionary income after a spouse enters care is \$3,000-\$4,000 depending on the year of potential institutionalization.

3.5 CONCLUSION

While Canada's retirement security system has been effective in raising the financial well-being of seniors, the prospect of potentially unforeseen medical expenses in old age also poses a serious financial risk. This is particularly true in the case of long-term care, which is income-tested. This study departs from the literature by exploring the potential

[62] Defined as the average percentage shortfall of discretionary income after a spouse enters care, relative to the low income threshold.

implications of the cost of long-term care for the disposable incomes of seniors if their spouse enters residential LTC.

Our analysis finds that, among married seniors, annual “potential” copayments represent 42-55% of equivalent disposable income. Moreover, copayments generated by a spouse’s “hypothetical” entry into care are associated with lower discretionary incomes and higher rates of impoverishment among spouses who remain at home after care is initiated.

However, such financial losses are larger under provincial income tests characterized by large maximum fees or high implicit taxes on income. The same is true when copayments are assessed using combined marital income, rather than a share of combined income or individual income. Indeed, spousal impoverishment is moderated by higher spousal allowances, higher shares of combined income assigned to non-institutionalized spouses and lower implicit income tax rates. For instance, Nova Scotia presently provides comparatively generous spousal allowances, allows couples to retain 60% of joint income, and assesses the remaining 40% of income (over the personal allowance) at an 85% implicit tax rate. According to our simulations, the introduction of this policy, in 2006, is associated with a dramatic reduction in simulated income and effects on LIM.

Our analysis is subject to a few caveats, which we leave to future research. First, our sample includes non-institutionalized seniors rather than institutionalized seniors. It also lacks information on health-related predictors of long-term institutionalization in seniors. These issues are critical because socio-economic differences in health or institutionalization could generate incongruence between the potential and actual income-related differences in outof- pocket RLTC expenses. A lack of variables on perceived well-being or financial stress also preclude us using our estimates as predictors for the psycho-social well-being of older adults. The Canadian Longitudinal Study on Ageing, a new survey that contains measures of financial insecurity, income, and long-term care use, offers a critical source of data that could be used to address each of these issues. Finally, we note that our simulations are likely to be error-prone because of discrepancies between the SLID definition of income and the tax-based definitions that are typically used to administer the income tests. Approximately 20% of SLID respondents provide self-reported incomes in lieu of giving Statistics Canada access to their tax files for the purposes of survey completion. This could be addressed by using taxfiler data, such as the Longitudinal Administrative Databank.

Tables and Figures

Table 3.1: Sample Summary Statistics, Senior Couples

Variables	Mean	SD	Min	Max
Disposable Income (Institutionalized)	25,877	15,034	3,583	178,124
Disposable Income (Spouse)	23,141	15,433	5,549	193,584
Equivalent Disposable Income (Couple)	34,661	15,912	19,548	166,569
Excess Equivalent Disposable Income over Low Income Threshold	16,847	15,912	1,734	148,755
Percentage Ratio of Excess Equivalent Disposable Income to Low Income Threshold	94.6	89.3	9.7	835.0
Proportion of Couples within $5,000 of Low Income Threshold	0.199	0.399	0	1
Proportion Female (Institutionalized)	0.311	0.463	0	1
Age (Spouse)	77.5	3.3	65	80
Proportion Aged 80 or Older (Spouse)	0.432	0.495	0	1
Newfoundland and Labrador	0.014	0.117	0	1
Prince Edward Island	0.006	0.080	0	1
Nova Scotia	0.040	0.195	0	1
New Brunswick	0.032	0.177	0	1
Ontario	0.538	0.499	0	1
Manitoba	0.056	0.230	0	1
Saskatchewan	0.049	0.215	0	1
Alberta	0.107	0.310	0	1
British Columbia	0.157	0.364	0	1
Sample Size = 5,210 Couples				

Notes: Monetary values are in 2010 constant dollars. Disposable income is total income minus taxes plus transfers. The data are from the public-use versions of the SLID, 1996-2010 cycles. The sample includes married and common-law couples in which both spouses are 65 or older, at least one spouse is 80 or older, and in which the couple has a combined disposable income of at least twice the combined maximum benefit from Old Age Security and Guaranteed Income Supplement that would be available to a married senior with no other income source. For each couple, one spouse is designated as “potentially institutionalized,” while the other is assumed to live at home. The institutionalized spouse is the older of the two spouses, or if both spouses have the same reported age, chosen at random. The institutionalized spouse’s age is unreported as it is top-coded at 80 and does not vary. The low income threshold is defined using the Low-Income Measure concept, which sets the threshold at 50% of median equivalent disposable income of all households. We set this value to $17,814, which is the average after-tax LIM threshold for 2000-2010, as reported in CANSIM Table 206-0091. All estimates are estimated with normalized survey weights that sum to 1.

Table 3.2: Simulated Potential Annual Nursing Home Copayments by Sex

			Sex of Institutionalized Spouse			
Sub-Samples	Full-Sample		Male		Female	
Panel A: Sample Summary Statistics	Mean	SD	Mean	SD	Mean	SD
Disposable Income (Institutionalized)	25,877	15,034	29,291	15,485	18,330	10,626
Disposable Income (Spouse)	23,141	15,433	18,833	12,137	32,662	17,537
Combined Disposable Income	49,018	22,503	48,124	22,487	50,993	22,418
Equivalent Disposable Income (Couple)	34,661	15,912	34,029	15,901	36,057	15,852
Excess Equivalent Disposable Income (% Low Income Threshold)	94.6	89.3	91.0	89.3	102.4	89.0
Proportion of Non-Poor within $5,000 of Low Income Threshold	0.20	0.40	0.21	0.41	0.17	0.38
Panel B: Simulated Variables (2000 Policies)						
Potential Annual Copayment	18,954	5,721	19,574	5,114	17,583	6,676
Copay Divided by Combined Disposable Income	0.42	0.14	0.45	0.13	0.37	0.14
Copay Divided by Equivalent Disposable Income	0.60	0.20	0.63	0.19	0.52	0.20
Post-Institutionalization Income and Low Income Status						
Discretionary Income	30,140	20,868	28,657	20,871	33,419	20,490
Low Income Rate	0.30	0.46	0.34	0.48	0.20	0.40
Depth of Low Income (% of Low Income Threshold)	19.9	9.3	20.1	9.3	19.3	9.2
Excess Discretionary Income (% of Low Income Threshold)	107.0	121.3	103.3	125.1	113.7	113.9
Proportion of Non-Poor within $5,000 of Low Income Threshold	0.22	0.41	0.24	0.43	0.17	0.38
Panel C: Simulated Variables (2010 Policies)						
Potential Annual Copayment	18,583	5,666	19,564	5,074	16,452	6,270
Copay Divided by Combined Disposable Income	0.41	0.13	0.44	0.13	0.35	0.12
Copay Divided by Equivalent Disposable Income	0.58	0.19	0.62	0.18	0.49	0.17
Post-Institutionalization Income and Low Income Status						
Discretionary Income	30,444	20,685	28,712	20,359	34,205	20,893
Low Income Rate	0.27	0.44	0.32	0.47	0.16	0.37
Depth of Low Income (% of Low Income Threshold)	18.8	12.2	20.1	12.4	13.1	9.6
Excess Discretionary Income (% of Low Income Threshold)	104.3	119.7	99.7	120.7	112.3	117.7
Proportion of Non-Poor within $5,000 of Low Income Threshold	0.23	0.42	0.27	0.44	0.17	0.37
Number of Couples	5,210		3,587		1,622	

Notes: All monetary values are in constant 2010 dollars. This table reports the sample means of simulated copayments by sex of the institutionalized spouse and the assumed policy period (2000, 2010). Copayments are calculated using the income testing procedures in the household's province of residence. Copayments are also expressed as a percentage of combined or equivalent disposable income. Equivalent income is combined income divided by $\sqrt{2}$. SD=Standard Deviation.

Table 3.3: Simulated Potential Impact of Nursing Home Copayments on the Discretionary Income of Spouses

	NL	PE	NS	NB	ON	MN	SK	AB	BC
Equivalent Disposable Income (Couples)	27,539	30,111	32,010	30,791	35,677	32,375	31,687	34,614	35,230
	[10,370]	[10,881]	[13,580]	[14,434]	[17,794]	[11,951]	[12,469]	[11,725]	[14,569]
Number of Couples in Each Province	72	34	207	169	2804	293	253	559	819
Panel A	Husband/Male Institutionalized, Wife/Female at Home (Pooled Sample=3,587 Couples)								
2000 Policies	27,966	24,768	24,600	23,890	28,603	23,214	36,126	32,937	26,042
	[13,475]	[14,985]	[14,389]	[13,238]	[21,487]	[20,089]	[18,101]	[21,317]	[20,945]
2010 Policies	26,628	27,207	28,410	26,001	28,710	24,445	36,161	32,017	24,746
	[15,498]	[19,008]	[15,667]	[17,268]	[21,309]	[18,761]	[17,538]	[20,230]	[19,087]
Panel B	Wife/Female Institutionalized, Husband/Male at Home (Pooled Sample=1,622 Couples)								
2000 Policies	25,734	26,122	25,921	24,229	35,515	25,494	41,801	37,107	28,546
	[13,407]	[14,260]	[13,656]	[12,330]	[20,272]	[20,013]	[19,117]	[20,901]	[20,899]
2010 Policies	25,320	29,014	29,834	27,370	35,377	26,235	41,514	36,034	33,572
	[17,324]	[20,046]	[16,586]	[18,296]	[21,101]	[19,821]	[19,843]	[21,093]	[19,916]

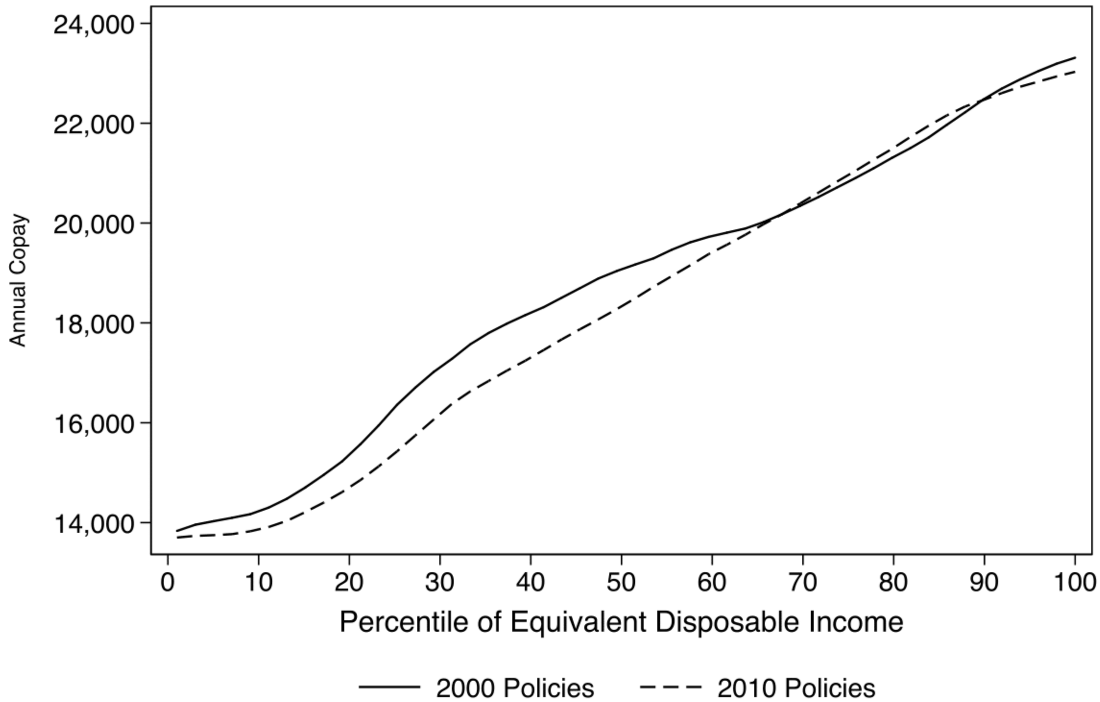
Notes: All monetary values are in 2010 constant dollars. This table reports the sample average of simulated discretionary income after a spouse is admitted into a nursing home for a year. Discretionary income is the couple's combined disposable income (total income minus taxes plus transfers) minus the husband or wife's simulated copayment and personal allowance in a province, and plus any spousal transfer from the nursing home resident to the spouse at home. The first row reports the mean of equivalent disposable income in each province. The sample mean of disposable income for all couples is $34,661. Standard deviations are reported in parentheses.

Table 3.4: Simulated Potential Impact of Nursing Home Copayments on the Low Income Status of Spouses

	NL	PE	NS	NB	ON	MN	SK	AB	BC
Number of Couples in Each Province	165	195	417	371	1922	549	502	483	606
Number of Couples in Each Province after Weighting	72	34	207	169	2804	293	253	559	819
Panel A	Husband/Male Institutionalized, Wife/Female at Home (Pooled Sample=3,587 Couples)								
2000 Policies	0.00	0.33	0.33	0.21	0.37	0.61	0.00	0.11	0.51
	[0.00]	[0.47]	[0.47]	[0.41]	[0.48]	[0.49]	[0.00]	[0.32]	[0.50]
2010 Policies	0.02	0.32	0.15	0.21	0.36	0.62	0.00	0.10	0.46
	[0.15]	[0.47]	[0.36]	[0.41]	[0.48]	[0.48]	[0.00]	[0.30]	[0.50]
Panel B	Wife/Female Institutionalized, Husband/Male at Home (Pooled Sample=1,622 Couples)								
2000 Policies	0.00	0.28	0.28	0.17	0.11	0.53	0.00	0.07	0.44
	[0.03]	[0.45]	[0.45]	[0.37]	[0.31]	[0.50]	[0.00]	[0.25]	[0.50]
2010 Policies	0.28	0.30	0.14	0.19	0.12	0.56	0.00	0.07	0.14
	[0.45]	[0.46]	[0.34]	[0.39]	[0.32]	[0.50]	[0.00]	[0.26]	[0.35]

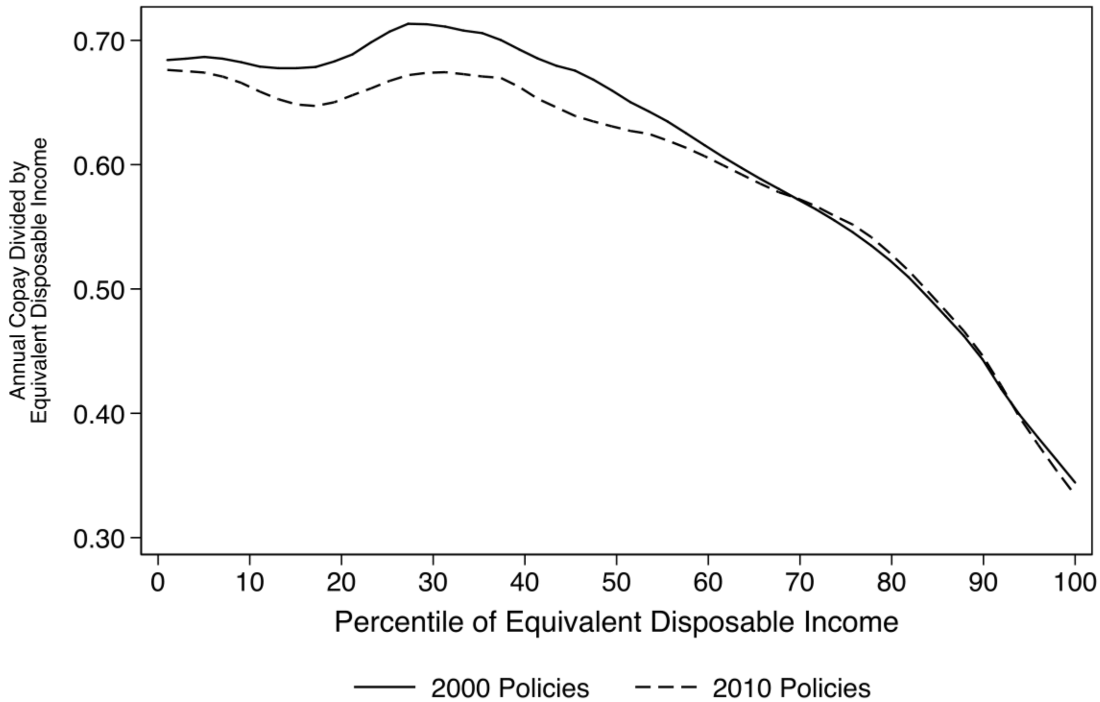
Notes: This table reports the sample-wide simulated Low Income Measure (LIM) rate after a spouse is admitted into a nursing home for one year. The rate represents the proportion of married seniors whose simulated discretionary income (as defined in the notes to Table 3.3) falls below 50% of the median equivalent after-tax income for families in the SLID ($17,814 in 2010 constant dollars). Since the sample is restricted to senior couples with a combined income of no less than the combined maximum benefit from OAS and GIS, the sample LIM is 0%.

Figure 3.1: Simulated Potential Annual Copayments by Income Percentile



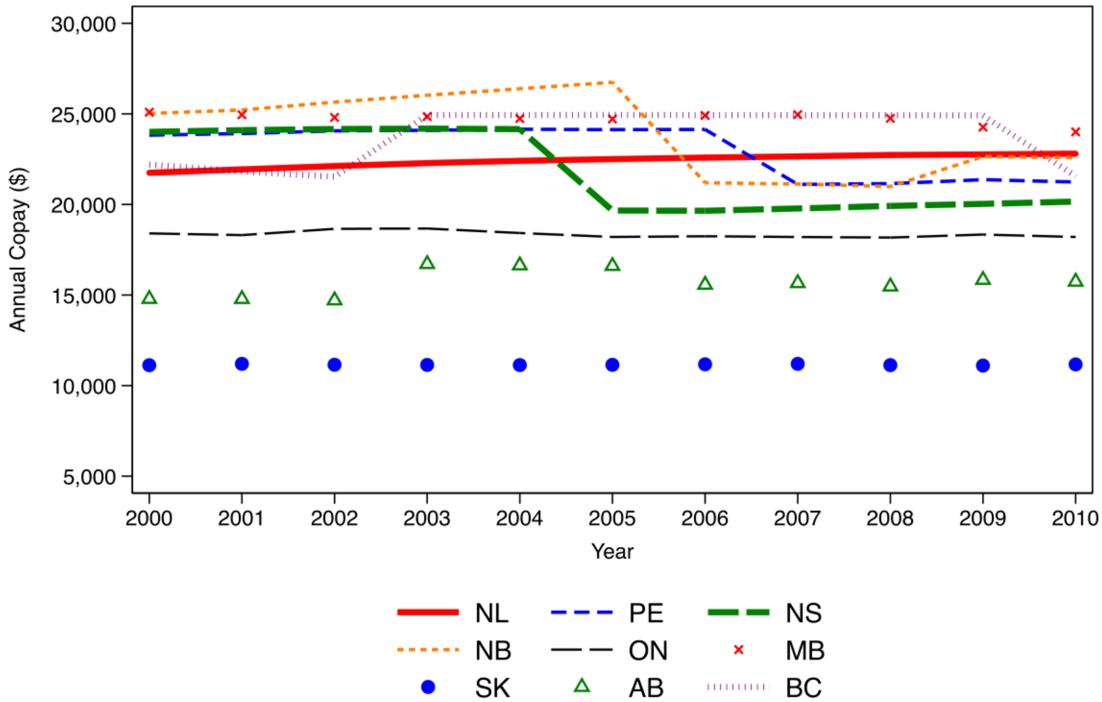
Notes: Copayments are reported in 2010 constant dollars. The figure reports the weighted means of simulated potential annual copayments of married seniors in each equivalent disposable income percentile. Copayments are calculated using the income test in each couple's province of residence for the 2000 and 2010 policy years.

Figure 3.2: Simulated Ratio of Copayments to Equivalent Disposable Income by Income Percentile



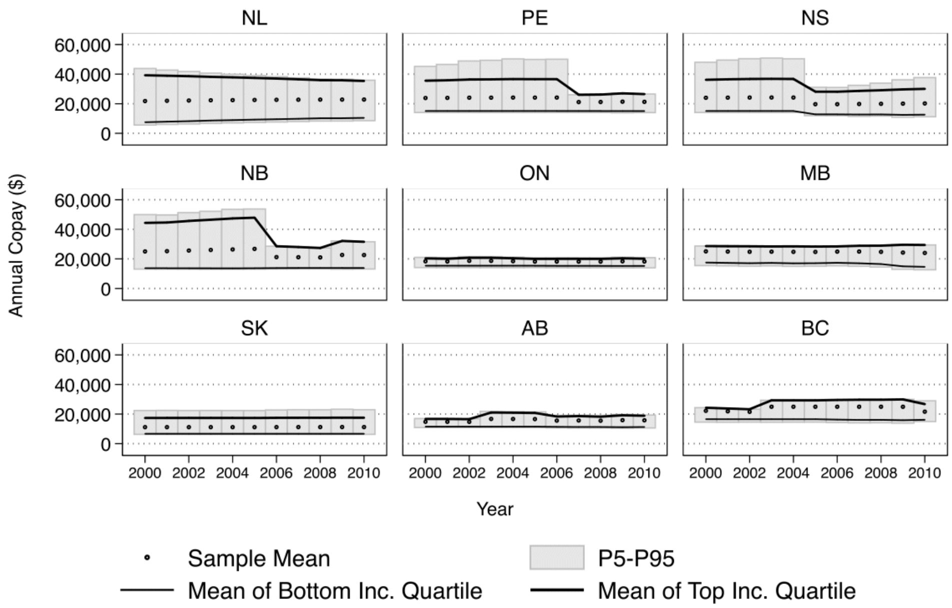
Notes: See Figure 3.1 notes. Here, mean copayments are divided by the mean of equivalent disposable income for each income percentile.

Figure 3.3: Mean of Simulated Potential Copayments under Provincial RLTC Programs



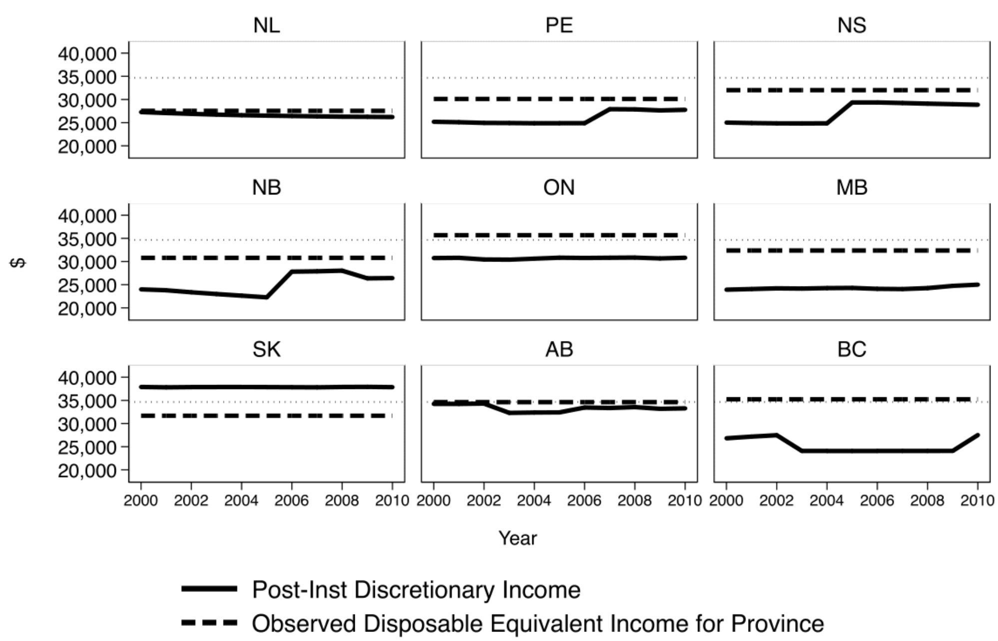
Notes: Copayments are reported 2010 constant dollars. The figure reports the weighted means of simulated potential annual copayments under each provincial RLTC program for the period 2000-2010.

Figure 3.4: Mean of Simulated Potential Copayments over the Income Distribution



Notes: See Figure 3.3 notes. The figure reports the weighted mean of simulated potential annual copayments for different sections of the overall "pre-institutionalization" disposable income distribution.

Figure 3.5: Simulated Discretionary Income After Spouse Enters Care



Notes: Income is reported in 2010 constant dollars. The horizontal dotted lines denote the sample mean of "pre-institutionalization" disposable equivalent incomes for all couples.

Figure 3.6: Simulated Percentage Change in Disposable Equivalent Income After Spouse Enters Care

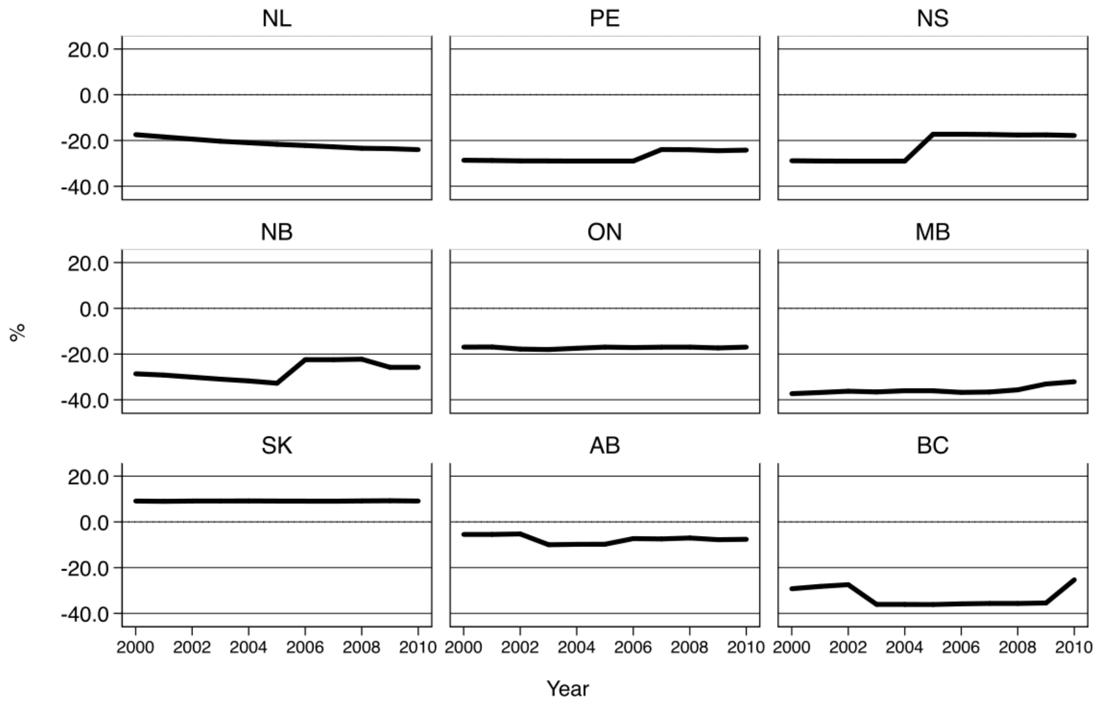


Figure 3.7: Simulated Change in Percentile Income Ranking After Spouse Enters Care

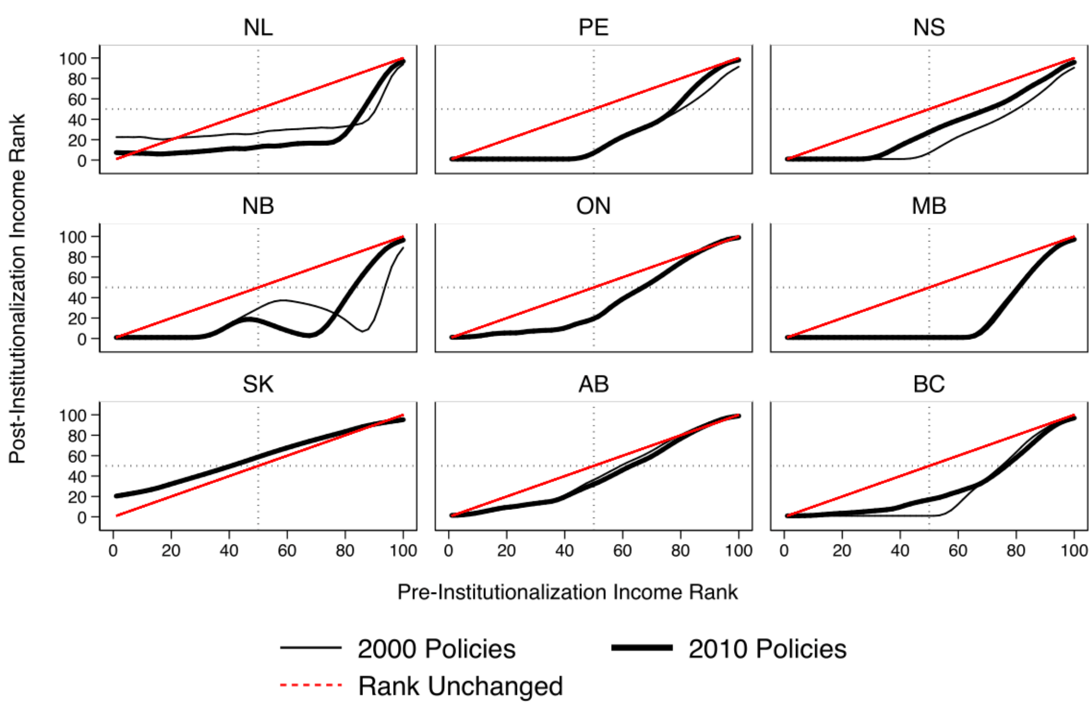
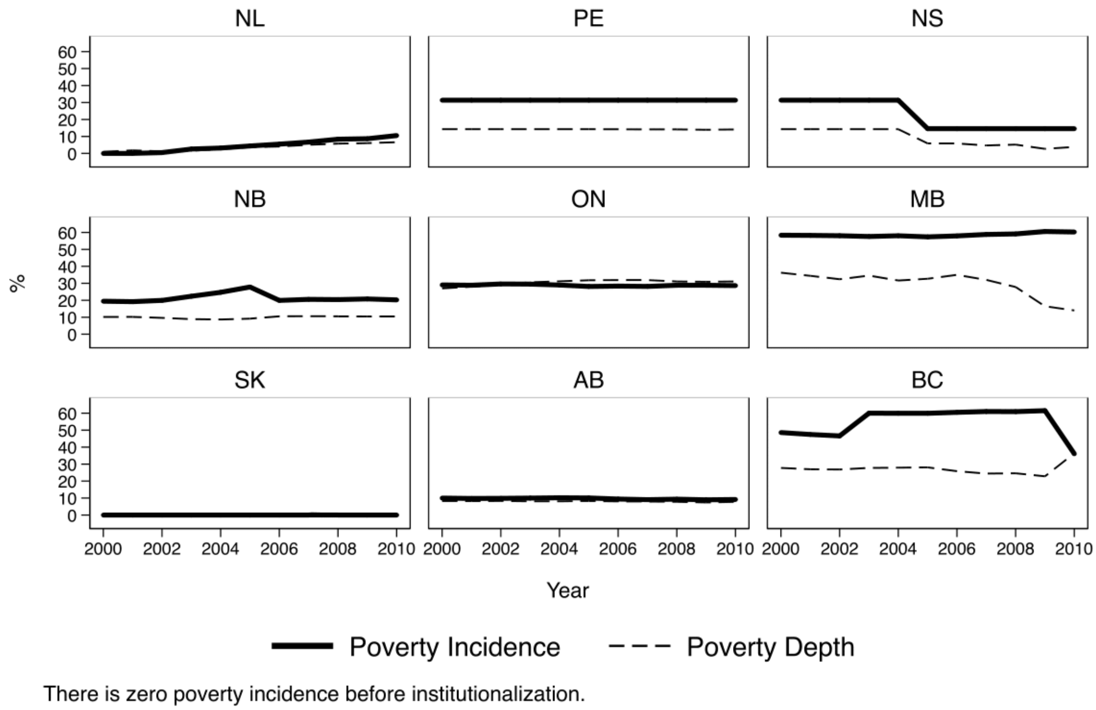


Figure 3.8: Simulated Change in Low Income Status and Depth of Low Income After Spouse Enters Care



Notes: Recall that the LIM rate (and depth of low income) is nil before institutionalization. See Table 3.1 notes and text for further details.

Chapter 4: Online Communication and Subjective Well-Being: The Impact of Social Media on Life Satisfaction

4.1 Introduction

Despite the ubiquity of online communication, its association with subjective well-being (SWB) is poorly understood. While numerous economic studies have examined outcomes related to human happiness, such as wages or income, employment and consumption (Akerman et al., 2015; Ivus and Boland, 2015; Hong, 2013; Kolko, 2012; Koutroumpis, 2009; Forman et al., 2012), and marriage and sexual/reproductive behaviour (Bellou, 2015; Chan and Ghose, 2014; Bhuller et al, 2013), most evidence on the direct impacts of internet and online contact on SWB is based on small observational studies of student populations of limited statistical power and generalizability. In this paper, I examine the relationship between SWB and online communication with friends and family using large, nationally representative samples from the General Social Survey (GSS).

Since commercial internet first arrived in the early 1990s – initially by dial-up connections and then by broadband (or high-speed internet) – online communication has sky-rocketed. As of 2012 in Canada, 50% of workers used e-mail daily (Statistics Canada, 2012).[63] Online communication accelerated around 2004 following the creation of Facebook and subsequent social media websites (e.g. Twitter, Reddit) as well as other user-centred digital platforms (e.g. comment forums, blogs, Wikis). According to the 2013 General Social Survey, 70% of adults communicate weekly with friends and family

[63] This figure represents the fraction of employees who reported using e-mail in their jobs on a daily basis.

by e-mail, internet, social media or text message, while nearly 60% have a social networking account (Statistics Canada, 2013).[64]

By making communication easier and more convenient, these platforms have arguably made it easier for people to maintain and strengthen ties with family and friends. They also store considerable information about people that can aid in establishing new social circles.[65] Social networking websites aid in this regard both by providing information about its users and allowing people to retain or discard ("unfriend"/"unfollow") contacts. For example, focusing on romantic relationships, Bellou (2015) finds that broadband internet increases the rate of marriage, partly through greater use of dating websites, which offer similar information and screening functions. Additionally, through media sharing, liking and comment functions, online communication can potentially supplement friendships by expanding opportunities for the consumption of relational goods and services (e.g. news, television shows, spectator sports, political discussion). Beyond expanding opportunities for communication, the Internet has also created new markets for goods and services, which may also enhance SWB.

However, while the internet can be used to find friends, online friends may offer less companionship. Consistent with this, Helliwell and Huang (2013) find that offline friendships are more effective than online friends in producing individual happiness. Thus, substituting online friends for offline friends could, all else equal, generate a loss in

---

[64] According to Facebook, nearly three-quarters of its users log on at least once per day (Canadian Press, 2014).

[65] In economics terminology, the informational advantages offered by the internet may reduce the frictional search cost of generating social relationships.

SWB. Online communication may also be negative in nature, as in the case of cyber-bullying, which could produce social isolation and psychological distress (Schneider et al., 2012). Theories of emotional contagion, which suggest that human emotions are spread through interpersonal interactions, have been extrapolated to online settings whereby people can become unhappy or happy through the sharing or viewing of positive or negative messages (Kramer et al., 2014; Coviello et al., 2014). Finally, others have argued that the social transparency brought about by social media websites regarding the material well-being of others could manifest in low self-esteem, envy and heightened perceptions of material deprivation (Lohmann, 2015; Vogel et al., 2014).

Using cross sectional data from the General Social Survey, I examine the impact of online social contact on SWB. SWB is defined as an individual's self-reported life satisfaction, as derived from the question: "how do you feel about life as a whole right now?" Life satisfaction is a global measure of happiness or SWB. Previous work shows that it is sensitive to a wide range of economic, social capital and psycho-social characteristics (Helliwell et al., 2004; Kahneman and Krueger, 2006; Helliwell et al., 2009). This is important since, as discussed above, online contact is likely to have an array of social, economic and psychological effects. Online social contact is probed using questions about the frequency with which one has internet or e-mail contact with their friends and relatives[66], as well as use of a social networking website (e.g. Facebook, Twitter, Instagram).

[66] In the 2013 GSS, respondents were asked to include incorporate online social networking into their answers.

My analysis estimates the impact of online communication and social media use on life satisfaction, after conditioning on socio-demographic and social capital variables. To account for biases from observed confounders, I combine OLS regression with matching on observables (covariates). Specifically, I implement an entropy weighting strategy (Hainmueller 2012) that guarantees covariate balance between treatment (has monthly online contact) and control groups (less than monthly or no contact) prior to OLS estimation. A key practical advantage of entropy weights is that it eliminates the need to estimate the conditional probability of treatment, which is a requirement of alternative methods such as propensity score estimation.

In the absence of controls, SWB and online communication exhibit a weak but positive association. However, after adding controls, I find that monthly online contact lowers SWB, whereas daily online contact enhances it. This suggests that frequent internet contact influences well-being, whereas internet access (for which monthly online contact is a plausible proxy) does not. For social media use, I find a consistent, negative association with SWB.

To account for the endogeneity of social media use, I exploit the arrival of social media (as marked by the launch of Facebook in 2004) as an exogenous shock. Specifically, I estimate a triple-differences model that identifies variation in social media use by age group and pre-2004 rates of computer ownership. I demonstrate that the effect of aggregate regional computer ownership rates on the post-2004 change in social media use is larger for non-seniors than for seniors. While I find no evidence of corresponding differential changes in SWB, I find some evidence that implicates social media as a source of political engagement and increased social trust in non-seniors.

4.2 WELL-BEING AND ONLINE COMMUNICATION

While several studies have examined the relationship between SWB and online contact, these studies generally differ in their definitions of well-being (e.g. composite scores from factor analyses models; related psychological outcomes, such as loneliness) and online contact (internet access (Nie and Hillygus 2002), online communication/chatting (Valkenburg and Jochen 2007), social media use (Elison et al 2007; Steinfeld et al 2008). Most of these are based on small samples of younger populations, particularly college students, which limits their generalizeability. For instance, in a meta-analysis, Huang (2010) reports negative associations between internet use and a variety of different measures of psychological well-being and internet use. Of the 40 studies included in their review, only 8 had samples of >500 participants. Furthermore, almost all of them looked at adolescents and young adults. A limitation of this demographic focus is that the functionality of internet use may shift over the life course in ways that have different implications for well-being. For example, older adults may find the internet useful for keeping in touch with children and grandchildren, which could increase their SWB.

Other studies focus more specifically on social media use (Elison et al 2007; Steinfeld et al 2008; Burke et al 2011; Krasnova et al 2013; Vogel et al 2014; Verduyn et al 2015; Sabitini and Sarracino 2016). Unlike e-mail and instant messaging, social media is a broader tool that offers a variety of information and communication services, from e-mail and instant messaging to status updates, media sharing, information consumption and comment forums. Most social media platforms, including Facebook, Twitter, Instagram, Reddit and others, offer at least two of these functionalities. This suggests that the effects of social media may be heterogeneous depending on how it is used. Thus, studies that focus on social media use in general may be less informative than those that separately explore the communication and informational flows of social media use.

Using Facebook user data, Burke et al (2011) find that online self-disclosure (e.g. communication by broadcasting one's status updates) is positively associated with both SWB and social capital. Lab-based experiments with undergraduate students have shown that *active* social media use (e.g. direct communication with Facebook "friends") has no association with psychological well-being (Krasnova et al 2013; Verduyn et al 2015). In contrast, *passive* social media utilization[67] has been linked to lower SWB (Verduyn et al 2016). Using a method known as experience sampling – where information is drawn about individual social media experiences and well-being repeatedly over a two-week period -- Verduyn et al (2015) implicate envy[68] as a mediator in an inverse relationship between passive social media use and well-being.[69] A related set of studies have examined whether *passive* Facebook usage predicts increased envy or decreased self-worth via upward social comparisons (Vogel et al 2014; Lohmann, 2015; Verduyn et al., 2015). Vogel et al. (2014) and Verduyn et al (2015) report a negative association between self-esteem and Facebook usage. They find that this effect is mediated in part by upward social comparisons. i.e. negative self-evaluations made relative to people with positive characteristics (e.g. many friends, disclosures of success). These studies, while confined to small samples and narrowly defined populations (e.g. students), have been reinforced by Lohmann (2015), who takes a population-based approach. He exploits the internet as a

[67] Passive social media "involves consuming information without direct exchanges (e.g., scrolling through news feeds, viewing posts)" (Verduyn et al 2015, p.480).
[68] Elicited by questionnaire: "How envious have you been of others since the last time we asked?"
[69] Using a series of structural equations, they find that passive Facebook utilization predicts envy, which predicts declines in well-being.

social transparency shock that increases the amount of information about the economic status of people in one's social reference group. Using a household fixed effects regression, he finds that high speed internet predicts increased income aspirations (e.g. the amount of income a household would like to have) by 7% in households with computers, after controlling for actual income. His study focuses on internet access in general, rather than specific communication services such as social media.

Sabitini and Sarracino (2016) argue that the internet has brought forth a new wave of status consumption, which they liken to "keeping up with the 'e-Jones'." Like Lohmann, they hypothesize that upward social comparisons could produce frustration about one's own consumption levels, thus decreasing satisfaction with available income. Using nationally representative data for Italy, they find a negative relationship between social media use and self-reported income satisfaction. To address the endogeneity of social media use, they use spatial measures of broadband density to instrument social media utilization.[70] Notably, in another study based on the same data and a similar methodology, they report negative effects on life satisfaction and social trust (Sabitini and Sarracino 2017).

This paper offers several contributions to the literature. First, I study the relationship between online contact and well-being in a much larger sample than most other studies – excluding the few population-based studies mentioned above. Specifically, I use multiple waves of General Social Survey data, a nationally representative survey for Canada, which contains over 50,000 observations. Second, I

[70] They attempt to rule out reverse causality by using spatial indicators of broadband infrastructure density as an instrumental variable for social media utilization.

separately investigate the impacts of online communication and social media use. This is important because the impacts of active communication may differ from those of pure informational flows (for example, from passive Facebook use) accrued to social media participants. To my knowledge these issues have not yet been examined jointly in large or nationally representative samples. Finally, I treat the arrival of Facebook in 2004 as a source of exogenous time variation in access to social media. I combine this with respondent age and regional computer ownership rates to estimate a triple-difference model of the effect of social media on SWB.

4.3 Data

The data includes four waves of the General Social Survey (GSS; 2000, 2003, 2008, 2013 cycles).[71] Each sample includes people aged 15 and older, excluding those living in the territories and residents of full-time institutions (e.g. nursing homes, prisons).[72] The GSS is carried out each year using themed questionnaires that repeat approximately every five years. The 2000 wave, Information-Communication Technology, was not repeated. The 2003, 2008 and 2013 surveys shared a common theme of social engagement and social networks. All four surveys contain questions about online contact with friends and family. The last three surveys asked about life satisfaction and a broad range of questions about demographics, socio-economic status, health and social capital questions. To focus on those with influence over household decisions about internet access and computer ownership, the sample is limited to adults aged 25 and older.

[71] I use the public-use microdata files.
[72] The unit of analysis is the respondent.

4.3.1 Online Communication and Social Media

Each survey provides respondent-level variables about internet/e-mail communication with friends and family.[73] The 2000 survey asks only about e-mail communication, while the 2003 and 2008 surveys add in general internet communication without specifying the method of contact (e.g. social media, chat, instant messaging). The 2013 questionnaire again includes e-mail/internet contact, but specifically instructs respondents to include communication via social networking websites (e.g. Facebook, Twitter). While the earlier questionnaires appear limited to e-mail, this was a dominant mode of internet-mediated communication at the time. Nevertheless, all regressions include year dummies. In addition to controlling for aggregate trends in online communication, year dummies also account for potential questionnaire changes that influence responses to these and other questionnaire items.

The 2013 GSS contains a separate module about online social networking. Respondents are asked whether they use Facebook, Twitter, MySpace, Google+ and LinkedIn.[74] Except for LinkedIn, which arrived in late 2003, these websites were launched in 2004 or later. As documented in Figure 4.1 using Google Search trends, these and other social networking websites gained popularity only after 2005.[75] Therefore it is assumed throughout this paper that social media utilization was negligible during the 2000 and 2003 surveys. Since the 2008 survey did not ask about social media use despite

[73] Appendix B reports the exact wording of each survey question.

[74] They can also state "other" and specify which website they use. The PUMF identifies general participation on online social networks rather than website-specific usage.

[75] Note that the Google Search volume index for a given keyword (e.g. Facebook) is first indexed against all searches, then normalized against itself. Thus, it measures relative rather than absolute interest. An index value of zero indicates that the keyword had not been searched during the specified period (one month).

it being available at the time, this sample is excluded from all models containing social networking variables. In summary, social media access is recorded as nil in 2000 and 2003, as missing in 2008 and as observed in 2013.

Social networking and online contact variables are specified as dummies and categorical variables. The dummies are coded as 1 (contact: yes) and 0 (contact: no), whereas the categorical variables are coded by frequency of contact/participation: less than monthly/none, monthly, weekly or daily. Variables coded as "valid skip" because the respondent does not use the internet are re-coded to "less than monthly/none."

4.3.2 Life Satisfaction

Subjective well-being is measured by life satisfaction. Life satisfaction is a global measure of psychological well-being. Several studies have shown that it is sensitive to changes in psycho-social states (e.g. loneliness, depression), income, unemployment and health status. Helliwell and colleagues (e.g. Helliwell et al., 2004, 2009) show that it is also sensitive to numerous social capital and civic engagement variables (e.g. number of friends, social interactions, social trust). For these reasons, it is well suited for studying the effects of rising online communication, which, like the internet, is likely to have complex impacts across a range of domains (e.g. health, social, economic).

In the 2003 and 2008 cycles, respondents are asked "On a scale of 1 to 10, where 1 means 'very dissatisfied' and 10 means 'very satisfied', how do you feel about your life as a whole right now?" Respondents who reported "no opinion" are excluded from the sample. The 2013 survey asks the same question but modifies the scale to range from 0 to

10. To ensure that the scales are the same in each period, I recode the zeroes to 1.[76] My use of year dummies also account for possible biases from scale changes. The 2000 survey does not ask about life satisfaction, so it is not used in my formal analysis, although I do use it in some cases to plot trends in general online communication. Figure 4.2 shows that the distribution of life satisfaction responses is stable over time.

4.3.3 Control Variables

The GSS provides a wide range of covariates that help adjust for observed sources of bias in the relationship between well-being and online communication. Each model includes a core set of demographic controls: age, sex, marital status, household size, and presence of own children in the household. Survey year dummies capture group-invariant trends in online communication. To account for time-invariant and unobserved spatial determinants of well-being and online communication (e.g. fixed spatial determinants of high speed access, such as topography), I include a set of region dummies, defined as the interaction of province and rural/urban status. In analyses containing all three cycles, region dummies are interacted with survey year to account for unobserved, region-specific linear confounders (e.g. growth in broadband infrastructure, computer prices, Internet service prices).

Both well-being and online contact could be influenced by a person's stock of social capital or sociability. For example, friendships can make people happier, while a person with many friends may find it easier to maintain their relationships online rather than offline. Therefore, my most comprehensive set of controls includes a set sociability and social capital variables that might influence both well-being and online

[76] My results are robust to a variety of transformations to life satisfaction.

communication. I use offline communication (e.g. monthly, weekly, and daily in-person communication with family and friends) as a proxy for general sociability. Such controls are important since offline friendships could predict both well-being (Helliwell and Huang 2013) and online communication (Penard and Poussing 2010). Social capital controls include participation in civic groups, number of close friends and family, provision (or receipt) of favours from one's neighbour, and sense of belonging to one's local community. Finally, a categorical variable of religious attendance is used to capture potential religious-based social capital.

4.4 Descriptive Statistics

This section describes the raw data. As illustrated in Figure 4.3, online communication has increased over time. The proportion of survey respondents who communicated with friends or family tripled from 5% in 2000 to 16% in 2013, reflecting both an increase in internet access and growth in software and applications suited to online communication. Reducing the threshold of communication to weekly increases these figures to 17% and 55%, respectively. While rural areas and older populations report lower levels of online contact, the trends in online communication for these groups were similar to that of the overall sample (Figure 4.4).

Panel A of Table 4.1 reports average life satisfaction by frequency of online contact with friends and family. The estimation sample includes the 2003, 2008 and 2013 surveys (because life satisfaction is not available in 2000). For simplicity, the first row of the panel combines family and friend contacts into a single measure. SWB exhibits a positive relationship with online communication, so that respondents with daily online contact report the highest average level of life satisfaction. Similar results are shown in the next two rows for specifications that separate family and friend contact. Using OLS regressions that condition on survey year dummies, Panel B of Table 4.1 shows a positive association between frequency of online contact and life satisfaction. In general, the point estimates are larger for contact with relatives than with friends.

Table 4.2 provides summary statistics for a wide set of control variables, by frequency of online contact. For simplicity, I again combine the friend and family variables into a single measure of online social contact. Online communication is indeed associated with a wide range of observable characteristics. For instance, respondents who communicate less than monthly are more likely to be elderly, retired, and living in a rural area. Respondents who reported having daily online contact also reported greater numbers of close friends and family as well as higher levels of social capital and socio-economic status (e.g. education, household income).

4.5 Models and Estimation

4.5.1 Regression and Matching

To address biases from observed confounders (e.g. household income, age, employment status, number of relatives), I combine standard regression analysis with matching on observables (covariates). Regression is used to condition on a set of covariates, while matching ensures that “controls” (e.g. no online communication) and “treatment” observations (e.g. online communication) are observationally equivalent on average, in terms of observable covariates.

I use entropy balancing as a matching algorithm. Entropy balancing is a recently popularized technique that specifies covariate balance requirements (defined here as equality of covariate means between treatment and control groups), then estimates optimal respondent-level weights subject to those requirements. By construction, the weighted mean of each covariate is equalized across comparison groups. This method is computationally convenient compared to other matching techniques, particularly propensity score methods, which require iterating over alternative parametric models (typically logistic regressions) for “treatment” until sufficient balance is achieved. The default algorithm implemented in Stata produces weights equal to 1 for the treated units and generally <1 for the control units. A control unit with a weight of 1 would be considered observationally equivalent, on average, to a randomly selected member of the control group. Optimal weights are chosen to be as close to 1 as possible, to maximize the information (entropy) retained for estimation. Identification hinges on the “unconfoundedness” or “conditional independence” assumption; that is, conditional on observed covariates, assignment to “treatment” and “controls” are assumed to be random.

Hainmueller (2012) introduced entropy balancing as a solution to endogenous selection into binary treatments, implementing it in Stata (Heinmueller and Xu, 2013). Others have since used the method to estimate the effect of democratization on carbon emissions (Mayer 2017), the effect of spousal unemployment on mental health (Marcus 2013), and the effect of political endorsements on voting outcomes (Hainmueller 2012). One constraint of entropy balancing, like most other matching techniques, is that it is designed for binary rather than multi-valued treatments. Therefore, this study considers the following binary transformations of the dichotomous variable of interest, frequency of online communication:

1) $O^{monthly}$: At least monthly online communication (treatment) versus less than monthly (control);

2) O^{weekly}: At least weekly online communication (treatment) versus less than weekly; and

3) O^{daily}: Daily online communication (treatment) versus less than daily;

I use these variables sequentially to estimate the effects of increasingly frequent online contact.

Figure 4.5 displays the covariate balance property before and after re-weighting treatment and control groups within survey years. After weighting, the covariates should have equal sample means. I test this using a series of t-statistics for each variable. The first row of diagrams report t-statistics for differences in sample means across comparison groups, under the null hypothesis of equal means. This shows that there are significant differences in observables before weighting. The diagram in column 2 of the first row show the results after weighting by the entropy weights. Here the differences in means are not significantly different. This is replicated again in the second row of figures using p-values for each t-statistic. After weighting, each p-value exceeds 0.10 by a wide margin (most are over 0.90).

My primary specification is the following equation, estimated by OLS:

(4-1) $$\mathbf{y_{irt} = \beta_0 + \beta_1 O^j_{irt} + \beta_2' X_{irt} + \delta_r + \delta_t + \delta_{rt} + u_{irt}}$$

where y_{irt} is the stated life satisfaction of respondent i who lives in region r at the time t of the survey. O^j_{irt} is a dummy variable indicating online communication of a given frequency (j=monthly, weekly, daily) with friends or family, and β_1 captures its impact on life satisfaction. X_{irt} is a vector of socio-demographic controls; for example, age group, sex, household size, education, household income, and self-reported health. Since SWB and online participation could be jointly influenced by sociability variables, I also estimate specifications that augment X_{irt} with several social capital and political engagement variables (e.g. number of close friends, relatives, social trust, belonging to the community, voting history, civic participation).

The period following the arrival of Facebook was one of rapid growth in technological innovation (e.g. the arrival of smart phones) and ownership of consumer durables (e.g. smart phones, tablets). Several social media platforms also changed rather dramatically; for example, Facebook shifted from a "social networking" platform to one that includes networking, e-mail, instant messaging, video chat, advertising, and news media. It stands to reason that the initial impact of these platforms could theoretically differ from their longer-term impacts. To account for this, I estimate an augmented version of (4-2) that treats the 2008 and 2013 survey years as distinct post-treatment observation periods.

4.5.2 Triple-Differences Estimation

One challenge with this estimation strategy is that online communication and social media use may be affected by unobserved predictors of well-being (e.g. wealth, personality). Reverse causality (from well-being to online contact) could result in spurious correlations even after conditioning on observables. To address these endogeneity issues, I estimate a triple-differences (DDD) model that exploits the arrival of Facebook and other social networking websites as a plausibly exogenous shock to online communication.

Facebook arrived in 2004, followed by numerous competitor platforms (e.g. Twitter, Instagram, Snapchat) and by complementary computer hardware (e.g. laptops, tablets, smart phones) and software innovations that permit "always-on", 24-7 access to online social networks and social media content. In addition to providing e-mail and chat functions, social media companies generally provide a range of other services, including the ability to "broadcast" life experiences (e.g. Facebook posts or tweets; sharing of personal photographs and videos), share (or create) media and relational goods and services (e.g. online video games), and engage in online market transactions.

I use Facebook's launch in 2004 to identify a post-treatment period during which social media was widely available to people with computers and an internet connection. In the GSS sample, the "treatment group" includes respondents of the 2008 and 2013 surveys, with respondents to 2003 survey serving as controls.[77] I interact this post-dummy by an indicator variable for non-seniors, yielding a differences-in-differences (DD) component. The coefficient on such a term would capture the relative difference in well-being arising from age-related differences in social media use during the post-2004 period. The identifying assumption for a DD model in this context is that well-being would not have changed over time by age group in the absence of Facebook's arrival. I relax this assumption by interacting the DD expression by a continuous measure of computer ownership that prevailed in each region from 1997 to 2000, giving a continuous differences-in-differences-in-differences (DDD) term. This allows me to identify

[77] Recall that respondents to the 2000 survey are dropped because it did not ask about life satisfaction.

variation in social media use across age groups by pre-Facebook regional computer penetration. [78]

My primary specification is the following

(4-2) $$\begin{aligned} \mathbf{y_{irt}} &= \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\bar{\boldsymbol{z}}_r - \bar{\boldsymbol{z}}) * \mathbf{Post_t} * \mathbf{Age2564_{irt}} \\ &+ \boldsymbol{\beta_2}(\bar{\boldsymbol{z}}_r - \bar{\boldsymbol{z}}) + \boldsymbol{\beta_3'}\mathbf{X_{irt}} \\ &+ \boldsymbol{\delta_{at}} + \boldsymbol{\delta_a} + \boldsymbol{\delta_r} + \boldsymbol{\delta_t} + \mathbf{u_{irt}} \end{aligned}$$

where i is a survey respondent, a is their age category, r is their region of residence[79] and t is the survey year. $\bar{z}_r$ denotes the aggregate rate of computer ownership (averaged over 1997 2000) in a respondent's region of residence. The vector $\mathrm{X_{irt}}$ contains the same individual-level covariates as in (4-1). $\mathrm{Age2564_{irt}}$ is a binary indicator that is set to one if the individual is aged 25-64.[80] $\bar{z}_r$ is demeaned by the national rate of computer ownership, with the result multiplied by 10. Therefore, the DDD parameter (β_1) is an estimate of the differential post-2004 change in well-being of seniors, relative to non-seniors, for each 10% increase in a region's rate of computer ownership.

I control for a complete set of region (δ_r), survey year (δ_t), and age category (δ_a) fixed effects. The region fixed effects account for time-invariant unobserved regional confounders (e.g. fixed telecommunications infrastructure, demographic composition, geographic and topographical characteristics, proximity to internet-providing public buildings (e.g. schools, libraries)). The year effects serve two functions: they absorb the primary *post-treatment* effect while also accounting for aggregate trends in well-being.

[78] Computer ownership is the households who reported having a home computer in the 1997-1999 Surveys of Consumer Finances and 2000 Survey of Household Spending. This rate was first estimated by province and rural/urban status, which collectively determine the regional unit of analysis, then matched to GSS respondents by their "region" of residence.
[79] Regions are province-by-rural/urban interactions.

The age categories account for age-related differences in well-being, such as the well-documented u-shaped relation between well-being and age (Steptoe et al. 2015). I allow for this pattern to vary regionally by including age-by-region fixed effects. Since the triple-difference parameter of interest is derived from an age interaction, any source of age-specific changes in well-being (e.g. unemployment) could threaten identification. To account for this possibility, I include a complete set of age-by-time fixed effects (δ_{at}). Finally, I control for region-by-time fixed effects, which account for changes in region-specific unobservables (e.g. smartphone penetration, wireless internet).

4.6. Regression and matching Results

4.6.1 Well-Being and Online Contact

Table 4.3 reports unweighted OLS estimates[81] of the impact of online social contact on life satisfaction using data from the 2003, 2008 and 2013 surveys. Each row represents a regression coefficient for online contact, conditional on socio-demographic characteristics, health, and province and year fixed effects. The first row displays the results for life satisfaction. Column 1 reports the estimated conditional difference in life satisfaction between respondents with monthly online contact with their friends or family (*Monthly Contact*) and respondents in the base group (*Less than Monthly Contact*). Columns 2 and 3 show the same estimates for *Weekly Contact* and *Daily Contact*, respectively (again using *Monthly Contact* as the base category). For comparison, in column 4, I report the impact of being in fair or poor health instead of good, very good or excellent health.

[81] That is, these results are derived from standard, unweighted regressions, without entropy balancing (matching). Inclusion of survey weights do not impact my findings.

The main result is that *Monthly Contact* is associated with lower life satisfaction (compared to the base group of less than monthly contact), while *Daily Contact* is associated with higher life satisfaction. Both estimates are statistically significant. In contrast, respondents who communicate weekly report similar levels of life satisfaction as the base group, conditional on the control variables. However, the effects of online contact are small relative to the effects of fair/poor health. For instance, the coefficient on *Daily Contact* (0.1207) is, in absolute values, roughly one-tenth that of fair/poor health (-1.29).

Considering that these effects are likely confounded by individual differences in sociability, the remaining rows repeat the same regression using a variety of social capital and sociability outcomes. The frequency of online communication is positively associated with belonging to community, civic participation, social trust, voter turnout, offline social contact with friends and family, and number of close friends and close family. In the case of "close friends", the coefficient of daily (weekly) online communication is 20 (18) times that of monthly contact.[82] Table 4.4 show the estimated associations between online communication and well-being after controlling for social capital and sociability variables. Panel A, Column [1] reports unweighted results for binary indicators of online communication: *At Least Monthly; base group: Less than Monthly*, *At Least Weekly; base group: Less than Weekly* and *Daily Contact; base group: Less than Daily*. Both *monthly* and *weekly* online contact communication exhibit statistically significant inverse relationships with life satisfaction. In contrast, the coefficient on *Daily Contact* is small and positive but statistically insignificant. In all

[82] The same is true for close relatives.

cases, magnitudes of associations are comparable to those reported in Table 4.3 and remain smaller than the estimated impact of fair/poor health.

The estimated coefficients for weekly and daily contact are either positive or less negative than that of *At Least Monthly Contact.* I next formally explore the possibility of a dose-response function in which the relationship between well-being and online contact depends both on the presence of contact and the frequency of contact. Panel B of Table 4.4 replicate the specification from Table 4.3, reporting coefficients *At Least Monthly Contact* and interaction terms between the latter and weekly or daily contact. The interaction effects capture the additional impact of more frequent online communication (among those who communicate at least monthly). Consistent with the Panel A findings, *At Least Monthly* contact bears a significant inverse relationship with well-being. The interaction term for weekly contact is not significantly different from zero in any specification. However, the interaction for daily is positive and significant. These findings are consistent with a dose-response function in which the impact of online contact on well-being is positive only in those who communicate daily.

4.6.2 Entropy-Weighted Results

Table 4.5 summarize the entropy-weighted regression results, where the weights constructed from the entropy balancing algorithm ensure covariate balance on average. For *At Least Monthly* contact (Panel A), the unweighted and weighted results are quite similar. Column [2] reports the weighted coefficients for contact with friends or family, which are marginally larger than the unweighted coefficients. When considering family or friend contact separately, the entropy-weighted coefficients are 50-70% as large as the unweighted ones. The entropy-weighted impact of *At Least Weekly* contact bear no relationship with well-being, whereas *Daily* contact with friends or family (or just family) promotes wellbeing.

4.6.3 Well-Being and Social Media

One difficulty in interpreting these results is that the nature of online contact has changed considerably over time. The arrival and spread of Web 2.0[83] technologies (such as social media) expanded online contact from strictly e-mail and chat-like modes of contact to include social networking, video chatting and other more passive modes of contact (e.g. commenting on news articles, sharing media over Facebook). Although online social networking is specifically mentioned in the 2013 GSS questionnaire as an example of online communication with friends and relatives, its arrival may have changed the fundamental relationship between online contact and well-being.

I examine this issue by estimating equation (4-1) with a dummy variable for social media use instead of online communication. The dummy is set to 1 for those who report using social media in 2013 and zero otherwise. Social media-related questions were not included in either of the 2003 or 2008 surveys. However, I assume it was largely non-existent in 2003 (Facebook had only arrived in 2004) and set the dummy to zero for all respondents to this survey. As noted previously, the 2008 survey is dropped since social media was available but unobserved. Table 4.6 (column 2, Panel A) reports entropy-weighted OLS coefficient for *At Least Monthly* use of social networking websites. The coefficient is -0.1679, which is almost three times as large as the corresponding coefficient for *At Least Monthly* online contact. This coefficient is

[83] Web 2.0 (O'Reilly 2005) consists of web services that emphasize user content, ease of use and compatibility with computer hardware, software, and other web products (e.g. the ability to embed news articles into Facebook posts).

somewhat less negative than the unweighted estimates. Notice that the coefficients for *At Least Weekly* (Panel B) or *Daily* (Panel A) social networking use are of similar magnitudes as *At Least Monthly* utilization. Among those who used social media at least monthly, there is no evidence of correlation between frequency of social media use and well-being (F-Test = 0.95; p-value=0.38).

I next use Facebook's arrival in 2004 to test whether the subsequent growth in social networking platforms altered the association between SWB and online communication. Using a standard Oaxaca decomposition, I estimate the proportion of the change in the well-being-online contact relationship that can be explained by social media access. Table 4.7 presents the basic set of results. The main effect of online communication, reported in Column [1], is statistically insignificant. The interaction between online contact and the 2008 and 2013 survey dummies are each negative and significant (Column [2]). Thus, a negative relationship between SWB and online contact is evident only after the arrival and growth of social media in 2004. As a baseline, the interaction term for the 2013 survey and online contact is approximately -0.2177 ($p<0.001$). The coefficient falls to -0.141 ($p<0.01$) after controlling for social media use. This result suggests that social media utilization accounts for about one-third of the total decline in the impact of online contact on well-being that occurred from 2003 to 2013.

4.6.4 Estimated Impacts by Gender and Age Group

Thus far the analysis has pooled respondents by age and sex, yet associations between well-being and online communication may vary by demographic groups for a variety of reasons. Men and women may use online communication differently. The potential benefits (or risks) associated with online communication and social media use may vary by gender or by age. For instance, online connections may help to develop or reinforce supportive kinships that are protective in older populations or in times of stress (e.g. family illness). This could matter more for older female seniors, who are more likely to survive their spouse.

Figure 4.6 reports the predicted life satisfaction by online communication status and gender-by-age groups. Each plot is derived from an entropy-weighted regression of life satisfaction on a complete set of communication-by-age-by-gender interactions. The results for online communication, reported in Panel A, show that that the negative impact of online contact on well-being is limited to older populations, specifically seniors. The effects of social media use exhibit a similar pattern, but with steeper age curves. Overall, the SWB of younger and middle-aged adults (from 25-44) is largely uncorrelated with online contact and social media use of any given frequency.

4.7 Triple-Difference Results

4.7.1 Changes in Online Communication after the Launch of Social Media

I first document patterns of online communication and social media use by age group and pre-2004 computer ownership rates. Table 4.8 presents DDD estimates from equation (4-2), with weekly online contact (with friends or family) and weekly social media use as dependent variables.[84] Column [1] reports the results for online contact. The coefficient for the non-senior dummy variables are positive and significant. On average non-seniors were approximately 43 percentage points more likely to communicate online weekly. The post-2004 year-by-non-senior interactions are also positive; combining the interaction term for the 2013 survey with the main non-senior effect implies that, by 2013, non-

[84] Findings are robust to alternative frequencies of online contact and social media use (e.g. monthly or daily).

seniors were approximately 50 percentage points more likely than non-seniors to communicate online. The age-by-computer ownership is small and non-significant, suggesting that, before 2004, the effect of computer ownership on online communication was similar for seniors and non-seniors.

Multiplying the last expression by post-2004 survey year dummies yields the triple-difference term of interest. The point estimate is -0.04. Note that the differences-in-differences (DD) coefficients for age and survey years suggest that the senior-non-senior gap in online contact widened over time. The DDD estimates imply a weakening of this effect in regions with relatively high rates of computer ownership. For example, the point estimates indicate that each 10% increase in regional computer ownership produced a 4-percentage point increase in the rate of weekly online contact by seniors beyond that of non-seniors. This effect equates roughly 40% of the estimated average relative decrease in online contact by seniors compared to non-seniors.

Columns [2] and [3] report similar findings for social media use. Column [2] reports the DDD coefficient without controlling for online contact. The coefficient is positive but non-significant. The DDD estimate, which is picking up its negative association with online contact, is likely biased downward. In Column [3] I control for online contact. In this case, the triple-difference term is positive and statistically significant. The point estimate is 0.029; conditional on having online contact, each 10% increase in regional computer ownership translates into a 3-percentage point increase in the probability of social media utilization for non-seniors relative to seniors.

4.7.2 Life Satisfaction

The SWB results are summarized in Table 4.9. Panel A, Column [1] controls for the fixed effects variables. Column [2] adds economic variables (e.g. education, income categories), column [3] demographic controls (e.g. marital status, household size) and self-reported health, and column [4] social capital and sociability variables. In all cases, the DDD estimates are statistically equal to zero. Panel B reports very similar results with controls for online communication. By controlling for online communication, these results isolate age variation in social media utilization arising from regional differences in pre-social media computer penetration.

Table 4.10 repeats the estimation using different binary transformations of life satisfaction. For brevity, I report only the triple-difference parameters. Panel A reports results from a Linear Probability Model for "Low" SWB (e.g. life satisfaction scores of 0-4), while Panels B and C report similar estimates for "Moderate" [life satisfaction 5-7) and "High" (life satisfaction 8-10) SWB, respectively. As before, the DDD coefficients remain statistically equal to zero in most cases. In the case of "Low SWB", two of the coefficients are statistically significant ($p<0.05$), in models that control for fixed effects and demographic variables. The coefficients represent about 25% of the mean probability of having low SWB (4%) and 5% of the corresponding standard deviation. The estimates are statistically equivalent to zero after adding controls for economic and social capital variables.

4.7.3 Alternative Outcomes

Table 4.11 reports DDD estimates for several social capital and political engagement variables, which typically promote higher SWB. Each column represents a separate regression that conditions on fixed effects, demographic controls, economic controls, and online communication. The estimated coefficient for federal voter turnout is 0.04 ($p<0.05$). The implied impact of a 10-percentage point increase in a region's rate of computer ownership is a 4-percentage point increase in average voter turnout, or roughly 20% of mean voter turnout of adults aged 25-64. This finding conforms with several prior studies on the effects of broadband internet on voter turnout (e.g. Larcinese & Miner, 2017, Jaber, 2013; Czernich, 2012).[85]

I find similar results for social trust. The only other significant effect is for the "close family" variable. In this case, the effect is negative. Despite this, there is no evidence that the launch of social media affected the probability of offline contact with friends or with family (see Columns [8] and [9]). The remaining coefficients are statistically insignificant.

4.9 Conclusion

This study provides nationally representative estimates of the relationship between SWB and both online contact and social media use. Cross sectional comparisons reveal that, after matching on observable characteristics, infrequent online contact is negatively associated with SWB, especially among seniors. These effects are generally larger (more negative) for social media use than for general online contact. The arrival of social media also accounts for about 33% of the estimated decrease in the SWB-online contact relationship that is observed in the years following Facebook's launch in 2004. These findings are consistent with other studies that have implicated online social networking as a possible threat to social trust, self-esteem, and psychological well-being.

[85] Note that a few other studies have produced opposite findings (e.g. Campante, Durante, & Sobbrio, 2017; Falck, Gold, & Heblich, 2014).

Yet, in separate models that focus on quasi-experimental variation in social media access, I report evidence in which social media access has, at worst, no effect on SWB. In triple-difference estimations that identify variation in social media by age group and pre-Facebook regional computer penetration rates, I show that growth in social media use among younger adults in regions with high rates of computer ownership correspond with increased levels of social trust and voter turnout. I also observe potential declines in numbers of close family, but no decrease in offline contact with family.

These findings suggest that online contact and social media use are likely to be endogenous variables. Confounding by unobserved variables (e.g. loneliness) could explain the inverse relationship between psychological well-being and online contact documented in the cross-sectional relationships, despite the rich set of covariates included in the regressions. Furthermore, measures that are global in nature, such as life satisfaction, may lack the specificity necessary to capture the true effect of online contact on social or psychological well-being. Measures of social trust and voter turnout could capture more specific mechanisms that highlight the role of social media in facilitating group interaction and coordination as well as curating and distributing political information.

Tables and Figures

Table 4.1: Correlation Between Online Contact and Life Satisfaction

	< Monthly		Monthly		Weekly		Daily		
(A) Summary Statistics:									
Life Satisfaction	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Contact with Friends OR Family	7.80	1.91	7.85	1.64	7.93	1.60	7.99	1.64	
Friends	7.81	1.90	7.89	1.62	7.95	1.56	7.98	1.63	
Family	7.80	1.88	7.99	1.60	7.98	1.60	8.07	1.66	
(B) OLS Coefficients	β_0	SE	β_{monthl}	SE	β_{weekly}	SE	β_{daily}	SE	$Adj\ R$
Contact with Friends	7.7885	0.01		0.02	0.1209	0.01	0.1776	0.02	0.00
OR Family	***	52	0.0399	28	***	89	***	45	16
	7.7893	0.01	0.0743	0.02	0.1378	0.01	0.1629	0.02	0.00
Friends	***	46	***	18	***	93	***	71	16
	7.7867	0.01	0.0780	0.02	0.1763	0.01	0.2640	0.03	0.00
Family	***	44	***	03	***	97	***	28	25

Notes: N=51,022. Data Source: General Social Survey (2003, 2008, 2013 cycles). SD = Standard Deviation.

Panel (A) reports the sample mean of life satisfaction for each level of online communication with friends and family. Panel (B) reports OLS coefficients for monthly, weekly, and daily contact relative to the base group, which has no contact or less than monthly contact.

Table 4.2: Sample Summary Statistics

	Full Sample		By Frequency of Computer Contact							
			<Monthly		Monthly		Weekly		Daily	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Web Contact, <Monthly	0.41	0.49								
Web, Monthly	0.16	0.37								
Web, Weekly	0.30	0.46								
Web, Daily	0.13	0.34								
In-Person, Weekly	0.76	0.43	0.74	0.44	0.66	0.48	0.79	0.41	0.86	0.34
Relatives	3.30	1.13	3.14	1.16	3.24	1.08	3.44	1.07	3.57	1.15
Friends	3.11	1.10	2.91	1.15	3.02	1.01	3.26	1.01	3.50	1.08
Belonging to Community	2.03	0.83	2.02	0.85	2.08	0.82	2.02	0.81	2.00	0.84
Favours to Neighbour	0.68	0.47	0.64	0.48	0.68	0.47	0.71	0.45	0.73	0.45
Favours from Neighbour	0.64	0.48	0.59	0.49	0.64	0.48	0.68	0.47	0.68	0.47
Civic Group	0.66	0.47	0.54	0.50	0.70	0.46	0.75	0.43	0.76	0.42
Civic Group, Online	0.31	0.46	0.32	0.47	0.24	0.43	0.31	0.46	0.36	0.48
Social Trust	0.55	0.50	0.46	0.50	0.60	0.49	0.61	0.49	0.60	0.49
Voted, Federal	0.77	0.42	0.78	0.41	0.78	0.41	0.77	0.42	0.75	0.44
Religion	0.82	0.39	0.87	0.34	0.79	0.41	0.79	0.41	0.76	0.43
Fair/Poor Health	0.14	0.35	0.21	0.41	0.10	0.30	0.09	0.28	0.10	0.30
Female	0.55	0.50	0.54	0.50	0.50	0.50	0.57	0.50	0.63	0.48
Senior	0.21	0.41	0.35	0.48	0.12	0.33	0.12	0.32	0.10	0.29
Married	0.64	0.48	0.57	0.49	0.71	0.45	0.69	0.46	0.63	0.48
Immigrant	0.22	0.42	0.18	0.38	0.25	0.43	0.26	0.44	0.26	0.44
Child at home	0.36	0.48	0.28	0.45	0.44	0.50	0.42	0.49	0.41	0.49
Household Size	2.42	1.26	2.19	1.21	2.64	1.28	2.57	1.25	2.56	1.27
Hhld Inc <$40,000	0.18	0.38	0.28	0.45	0.12	0.32	0.10	0.30	0.12	0.33
Hhld Inc $40,000-$60,000	0.28	0.45	0.30	0.46	0.27	0.44	0.26	0.44	0.25	0.43
Hhld Inc >$60,000	0.43	0.50	0.25	0.43	0.54	0.50	0.58	0.49	0.55	0.50
Less than HS	0.16	0.37	0.31	0.46	0.06	0.24	0.05	0.21	0.05	0.22
High School	0.26	0.44	0.30	0.46	0.25	0.43	0.22	0.42	0.22	0.41
Diploma/Certificate	0.31	0.46	0.27	0.44	0.34	0.48	0.33	0.47	0.32	0.47
Bachelor Degree	0.19	0.39	0.09	0.29	0.23	0.42	0.27	0.44	0.27	0.44
Graduate Degree	0.09	0.28	0.03	0.16	0.11	0.31	0.13	0.34	0.14	0.35
Employed	0.61	0.49	0.47	0.50	0.71	0.45	0.70	0.46	0.71	0.46
Retired	0.23	0.42	0.36	0.48	0.15	0.36	0.15	0.36	0.12	0.33
Other Activity	0.16	0.36	0.17	0.37	0.14	0.34	0.14	0.35	0.17	0.38
Observations	51,882		21,407		8,240		15,324		6,911	

Notes: Data Source: General Social Survey (2003, 2008, 2013 cycles). SD = Standard Deviation.

Table 4.3: Unweighted OLS Impacts of Online Contact on Well-Being, Social Capital, and Sociability Meausres

	Online Communication						
Dependent Variables	Monthly	Weekly	Daily	Fair or Poor Health	$Adj\ R^2$	Mean	SD
Life Satisfaction	0.1207***	0.0091	-0.0661***	-1.2917***	0.1505	7.87	1.75
	[0.0209]	[0.0156]	[0.0157]	[0.0271]			
Sense of Belonging	-0.0663***	-0.0352***	0.0284***	0.2127***	0.0805	2.03	0.83
	[0.0109]	[0.0080]	[0.0081]	[0.0118]			
Voted in Federal Election	0.0221***	0.0286***	0.0315***	-0.0341***	0.1574	0.77	0.42
	[0.0053]	[0.0039]	[0.0039]	[0.0054]			
Trust	0.0422***	0.0648***	0.0526***	-0.0964***	0.1125	0.55	0.50
	[0.0063]	[0.0048]	[0.0048]	[0.0064]			
Did Favour for Neighbour	0.0539***	0.0422***	0.0084	-0.0441***	0.0444	0.68	0.47
	[0.0059]	[0.0045]	[0.0046]	[0.0064]			
Received Favour from Neighbour	0.0487***	0.0507***	0.0151**	-0.0307***	0.0398	0.64	0.48
	[0.0062]	[0.0047]	[0.0048]	[0.0066]			
Frequency of Religious Observance	0.0394	0.0030	0.0569***	0.1419***	0.1330	3.62	1.81
	[0.0232]	[0.0171]	[0.0173]	[0.0232]			
Participates in Civic/Volunteer Group	0.0774***	0.0841***	0.0518***	-0.0548***	0.1293	0.66	0.47
	[0.0056]	[0.0044]	[0.0044]	[0.0064]			
Number of Close Friends	0.3678***	0.2230***	0.0219*	-0.1786***	0.0811	3.11	1.10
	[0.0143]	[0.0104]	[0.0103]	[0.0153]			
Number of Close Relatives	0.2349***	0.1852***	0.0163	-0.1519***	0.0738	3.30	1.13
	[0.0151]	[0.0108]	[0.0109]	[0.0155]			

Notes: This table reports OLS-estimated impacts of online contact on life satisfaction and social capital/sociability outcomes (N= 51,022). Controls include respondent age in ten-year categories (25-34, 35-44 up to 75 and older), sex, marital status, education, household income categories, household size, presence of own children at home, immigration status, main activity during the reference year (retired, employed, other), region dummies and survey year dummies. Household income is interacted with survey year dummies to adjust for inflationary changes in the real values of the end points of the household income categories. Statistical significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$.

Table 4.4: Unweighted OLS Impact of Online Communication on Well-Being

	[1]	[2]	[3]
	Panel A: Separate Models by Frequency of Contact		
≥ Monthly	-0.0812***	-0.0510**	-0.0713***
	[0.0171]	[0.0158]	[0.0164]
Adjusted R^2	*0.2023*	*0.2021*	*0.2022*
≥ Weekly	-0.0428**	-0.0174	-0.0346*
	[0.0159]	[0.0160]	[0.0161]
Adjusted R^2	*0.2020*	*0.2020*	*0.2020*
Daily	0.0271	0.0618*	0.0246
	[0.0208]	[0.0285]	[0.0235]
Adjusted R^2	*0.2020*	*0.2020*	*0.2020*
	Panel B: Single Model		
≥ Monthly	-0.0858***	-0.0629***	-0.0835***
	[0.0206]	[0.0189]	[0.0205]
≥ Monthly * Weekly	-0.0104	0.0030	0.0037
	[0.0182]	[0.0199]	[0.0209]
≥ Monthly * Daily	0.0539*	0.0881**	0.0550*
	[0.0215]	[0.0312]	[0.0277]
Adjusted R^2	*0.2024*	*0.2022*	*0.2014*
Online Contact With			
Friends	Yes	No	-
Family	No	Yes	-
Family or Friends	No	No	Yes

Notes: This table reports OLS-estimated impacts of online contact on life satisfaction and social capital/sociability outcomes (N= 51,022). Controls include respondent age in ten-year categories (25-34, 35-44 up to 75 and older), sex, marital status, education, household income categories, household size, presence of own children at home, immigration status, main activity during the reference year (retired, employed, other), region dummies and survey year dummies. Household income is interacted with survey year dummies to adjust for inflationary changes in the real values of the end points of the household income categories. Social capital and sociability controls include social trust, number of friends, number of relatives, belonging to community, favours to and from neighbours, civic participation, online civic participation, categorical indicators of offline interactions with friends and family, and a dummy indicating that the respondent voted in the last federal election. Statistical significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$.

Table 4.5: Entropy-Weighted Impact of Online Communication on Well-Being

	Family or Friends		Family Only		Friends Only	
	[1]	[2]	[3]	[4]	[5]	[6]
	Panel A: At Least Monthly Contact					
β	-0.0812***	-0.0672**	-0.0510**	-0.0244	-0.0713***	-0.0502**
	{0.0171}	{0.0213}	{0.0158}	{0.0173}	{0.0164}	{0.0185}
Adjusted R^2	0.2023	0.1870	0.2021	0.1807	0.2022	0.1845
	Panel B: At Least Weekly Contact					
β	-0.0428**	-0.0064	-0.0174	0.0087	-0.0346*	-0.0053
	{0.0159}	{0.0173}	{0.0160}	{0.0161}	{0.0161}	{0.0172}
Adjusted R^2	0.2020	0.1832	0.2020	0.1806	0.2020	0.1812
	Panel C: Daily Contact					
β	0.0271	0.0429*	0.0618*	0.0688*	0.0246	0.0364
	{0.0208}	{0.0210}	{0.0285}	{0.0281}	{0.0235}	{0.0241}
Adjusted R^2	0.2020	0.1911	0.2020	0.1902	0.2020	0.1957
Weights	None	Entropy	None	Entropy	None	Entropy

Notes: N=51,022. Each regression controls for demographic and socio-economic characteristics, health status, social capital/sociability variables, and region and survey dummies (see Table 3 notes for details). Statistical significance: *** p < 0.001; ** p < 0.01; * p < 0.05; + p <0.10.

Table 4.6: Entropy-Weighted Impact of Social Media Use on Well-Being

	[1]	[2]
	Panel A: Use At Least Monthly	
β	-0.2649***	-0.1679***
	{0.0269}	{0.0360}
Adjusted R^2	0.2033	0.2006
	Panel B: Use At Least Weekly	
β	-0.2261***	-0.1348***
	{0.0265}	{0.0317}
Adjusted R^2	0.2027	0.2009
	Panel C: Use Daily	
β	-0.1990***	-0.0980**
	{0.0271}	{0.0301}
Adjusted R^2	0.2022	0.1999
Weights	None	Entropy

Notes: This table reports OLS-estimated impacts of online social networking on life satisfaction (N= 34,452), using the same set of controls as in Tables 4.5. The social media sample excludes the 2008 cycle. Social media use is observed is 2013 and assumed unavailable in 2003 (see the text and Figure 4.1 for justification of this assumption). Since each regression controls for a year 2013 dummy, β represents the impact of social networking after the arrival of social media (in 2004). Statistical significance: *** p < 0.001; ** p < 0.01; * p < 0.05; + p <0.10.

Table 4.7: Online Contact and Social Media Use

	[1]	[2]	[3]	*Oaxaca Decomposition* [3]-[2]
At Least Monthly Contact	0.0445	0.0409	0.0416	-
	[0.0279]	[0.0279]	[0.0279]	-
Monthly Contact *2008	-0.1063*			
	[0.0458]			
Monthly Contact *2013	-0.2241***	-0.2177***	-0.1410**	0.0767***
	[0.0486]	[0.0480]	[0.0529]	[0.0109]
Social Media Use				
Monthly			-0.1478+	
			[0.0795]	
Weekly			-0.138+	
			[0.0736]	
Daily			-0.1872***	
			[0.0457]	
Adjusted R^2	0.1838	0.1877	0.1889	
Observations	51,022	34,452	34,452	34,452
F Test of Equality of Interaction Terms	4.70*	-	-	
Exclude 2008 Survey	No	Yes	Yes	Yes
Entropy Weights	Yes	Yes	Yes	Yes

Notes: Statistical significance: *** p < 0.001; ** p < 0.01; * p < 0.05; + p <0.10.

Table 4.8: Triple-Difference Effect of Social Media Launch on Online Contact and Social Media Utilization

	Weekly Online Contact	Weekly Social Media Use	
	[1]	[2]	[3]
Comp9700*Age2564*Post	-0.0404***	0.0167	0.0296**
	{0.0101}	{0.0101}	{0.0095}
Comp9700*Age2564	-0.0676	-0.0269	-0.015
	{0.0457}	{0.0429}	{0.0404}
Comp9700*Post	0.0122	-0.1075**	-0.1082**
	{0.0517}	{0.0413}	{0.0396}
Age2564*Post	0.0717***	0.3673***	0.3556***
	{0.0114}	{0.0091}	{0.0085}
Post	0.1776***	0.1910***	0.1445***
	{0.0415}	{0.0343}	{0.0316}
Age 25-64	0.4309***	0.1578***	0.0589*
	{0.0285}	{0.0296}	{0.0270}
Web Monthly			0.0988***
			{0.0050}
Web Weekly			0.2096***
			{0.0050}
Web Daily			0.2776***
			{0.0065}
Adjusted R^2	0.2238	0.4238	0.4709
Observations	51,022	34,452	34,452
Mean	0.43	0.23	0.23
SD	0.49	0.42	0.42

Notes: Each regression controls for a complete set of economic and demographic controls as well as self-rated health (see Table 4.3 for a detailed list). Additional controls include age-by-year, age-by-region, region-by-year, region and time fixed effects as well as region-by-year rates of household internet access (any internet connection) and high speed access.
Statistical significance: *** p < 0.001; ** p < 0.01; * p < 0.05; + p <0.1.

Table 4.9: Triple-Difference Effect of Social Media Launch on Life Satisfaction

N=51,022	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Comp9700*Age2564*Post	0.012	-0.014	-0.006	0.0166	0.026	-0.0104	-0.006	0.0124
	{0.0529}	{0.0500}	{0.0500}	{0.0488}	{0.0528}	{0.0500}	{0.0500}	{0.0488}
Comp9700*Age2564	0.2422	0.2221	0.1981	0.1079	0.2769	0.2313	0.2005	0.1065
	{0.2081}	{0.1980}	{0.1957}	{0.1910}	{0.2076}	{0.1982}	{0.1959}	{0.1911}
Comp9700*Post	-0.1856	-0.0909	-0.0672	-0.0559	-0.1835	-0.0978	-0.0774	-0.0623
	{0.2076}	{0.1944}	{0.1912}	{0.1830}	{0.2067}	{0.1944}	{0.1913}	{0.1830}
Age2564*Post	-0.2889***	-0.1938***	-0.2215***	-0.2349***	-0.2874***	-0.2057***	-0.2295***	-0.2333***
	{0.0509}	{0.0488}	{0.0514}	{0.0503}	{0.0509}	{0.0488}	{0.0514}	{0.0502}
Post	0.2779*	0.1707	0.4995**	0.4528*	0.1999	0.1528	0.5040**	0.4697**
	{0.1181}	{0.1125}	{0.1873}	{0.1797}	{0.1183}	{0.1126}	{0.1875}	{0.1797}
Age 25-64	-0.2493	-0.4116**	-0.251	-0.0424	-0.3534**	-0.4604***	-0.2647*	-0.0138
	{0.1285}	{0.1297}	{0.1309}	{0.1275}	{0.1290}	{0.1304}	{0.1313}	{0.1279}
Web Monthly					0.1459***	-0.0326	-0.0840***	-0.1042***
					{0.0231}	{0.0218}	{0.0222}	{0.0217}
Web Weekly					0.2571***	0.0535**	-0.0125	-0.0913***
					{0.0200}	{0.0190}	{0.0197}	{0.0194}
Web Daily					0.3270***	0.1787***	0.1169***	-0.0288
					{0.0252}	{0.0242}	{0.0246}	{0.0243}
Adjusted R^2	0.0101	0.1345	0.1512	0.2031	0.0147	0.1357	0.1521	0.2036
Controls								
Fixed Effects	X	X	X	X	X	X	X	X
Health/Demographics		X	X	X		X	X	X
Economic Variables			X	X			X	X
Social Capital/Sociability				X				X

Notes: Fixed effects include age-by-year, age-by-region, region-by-year, region and time fixed effects. Each regression also controls for average region-by-year household internet access and high speed internet access. Statistical significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

Table 4.10: Linear Probability Model Estimates of Social Media Launch on High, Moderate, and Low SWB

Levels of SWB (N=51,022)	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Low SWB 1-4 (Mean=0.04, SD=0.20)								
Comp9700*Age2564*Post	0.0101	0.0115*	0.0092	0.0085	0.0088	0.0109*	0.0091	0.0087
	{0.0055}	{0.0054}	{0.0055}	{0.0054}	{0.0055}	{0.0054}	{0.0055}	{0.0054}
Adjusted R^2	0.0014	0.054	0.0703	0.0896	0.0045	0.0546	0.0705	0.0896
Panel B: Moderate SWB 5-7 (Mean=0.29, SD=0.45)								
Comp9700*Age2564*Post	-0.0218	-0.0164	-0.0164	-0.0201	-0.0241	-0.0174	-0.0166	-0.0197
	{0.0129}	{0.0126}	{0.0126}	{0.0126}	{0.0128}	{0.0126}	{0.0126}	{0.0126}
Adjusted R^2	0.0077	0.0552	0.0600	0.0844	0.0099	0.0562	0.0607	0.0846
Panel B: High SWB 8-10 (Mean=0.67, SD=0.47)								
Comp9700*Age2564*Post	0.0116	0.0049	0.0072	0.0116	0.0153	0.0065	0.0076	0.011
	{0.0134}	{0.0129}	{0.0129}	{0.0127}	{0.0134}	{0.0129}	{0.0129}	{0.0127}
Adjusted R^2	0.0089	0.0993	0.1097	0.1479	0.0135	0.101	0.1107	0.1482
Controls								
Online Contact					X	X	X	X
Fixed Effects	X	X	X	X	X	X	X	X
Health/Demographics		X	X	X		X	X	X
Economic Variables			X	X			X	X
Social Capital/Sociability				X				X

Notes: See Table 4.9 notes. Statistical significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

Table 4.11: Triple-Difference Effect of Social Media Launch on Social Capital and Political Engagement

N=51,022	Voted [1]	Trust [2]	Favour to Nbr [3]	Favour from Nbr. [4]	Civic Group [5]	Close Friends [6]	Close Family [7]	Offline Con. w/ Friends [8]	Offline Con. w/ Family [9]
Comp9700*Age2564*Post	0.029**	0.030*	0.000	0.011	0.013	0.025	-0.112***	0.000	-0.011
	{0.009}	{0.014}	{0.014}	{0.014}	{0.014}	{0.034}	{0.033}	{0.014}	{0.014}
Comp9700*Age2564	-0.012	-0.046	-0.012	-0.06	0.008	0.041	0.124	0.005	0.097
	{0.037}	{0.053}	{0.050}	{0.051}	{0.050}	{0.138}	{0.127}	{0.054}	{0.054}
Comp9700*Post	0.001	-0.019	-0.014	-0.03	-0.09	-0.105	-0.102	0.006	-0.077
	{0.044}	{0.054}	{0.050}	{0.052}	{0.049}	{0.129}	{0.129}	{0.055}	{0.057}
Age2564*Post	-0.020*	-0.059***	-0.011	-0.051***	-0.014	0.074*	0.070*	-0.042**	-0.005
	{0.010}	{0.014}	{0.014}	{0.014}	{0.014}	{0.034}	{0.033}	{0.015}	{0.015}
Post	0.017	-0.009	-0.023	0.027	-0.023	-0.104	0.052	-0.007	0.021
	{0.042}	{0.047}	{0.046}	{0.048}	{0.046}	{0.108}	{0.110}	{0.048}	{0.048}
Age 25-64	-0.384***	-0.124***	0.043	0.012	-0.071*	-0.138	-0.030	0.071	0.030
	{0.025}	{0.037}	{0.035}	{0.036}	{0.035}	{0.094}	{0.090}	{0.037}	{0.037}
Adjusted R^2	0.1579	0.1161	0.0461	0.0412	0.1318	0.0838	0.0754	0.0604	0.0546
Mean	0.78	0.55	0.68	0.64	0.66	3.12	3.31	0.59	0.45
SD	0.42	0.50	0.46	0.48	0.47	1.10	1.13	0.49	0.50

Notes: Model specification corresponds to Column [7] of Table 4.10.
Statistical significance: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.1$.

Figure 4.1: Google Search Trends for Social Media Websites

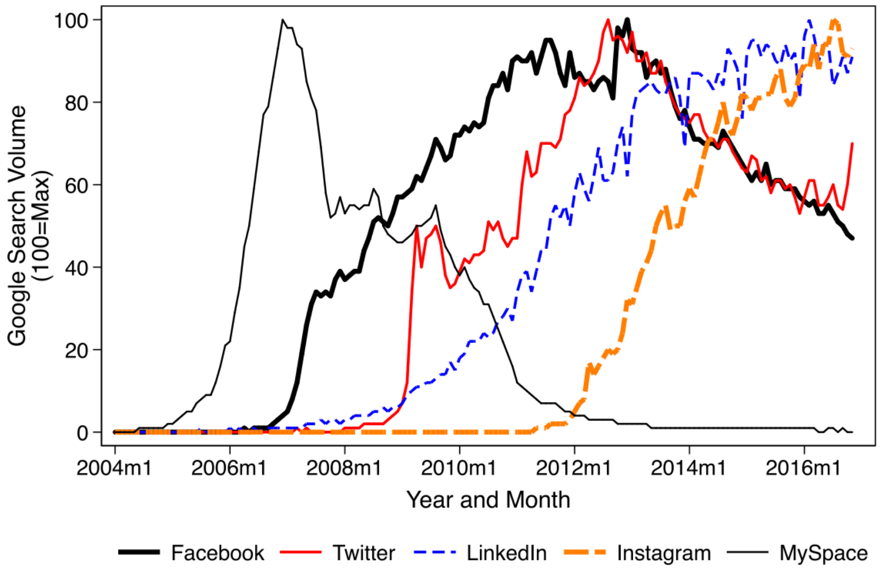


Figure 4.2: Trends in the Distribution of Life Satisfaction: 2003-2013

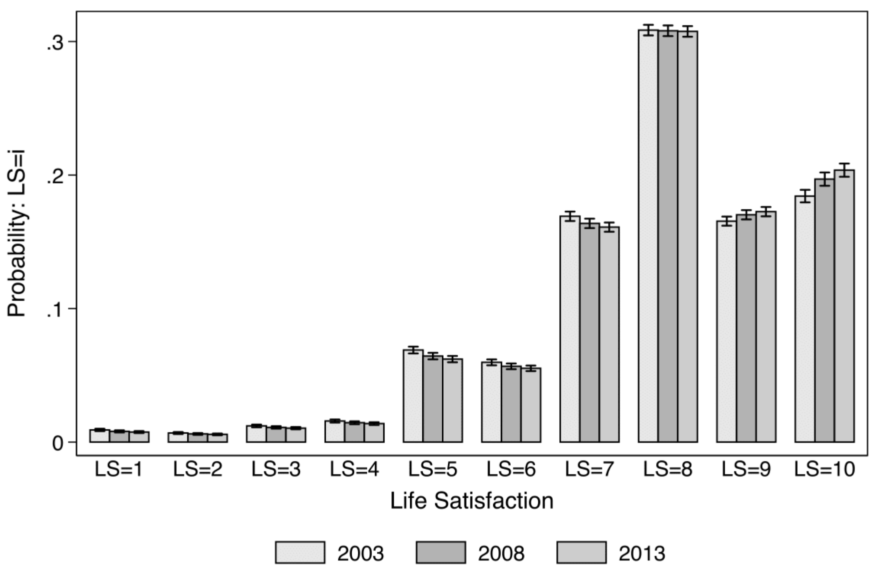


Figure 4.3: Trends in Online Communication: 2003-2013

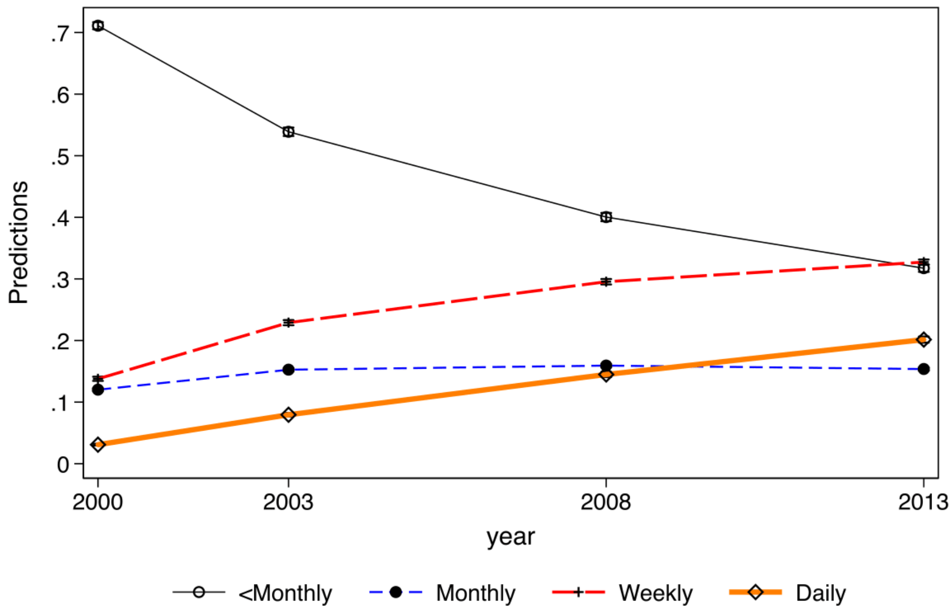


Figure 4.4: Trends in Online Communication by Age Group and Rurality: 2003-2013

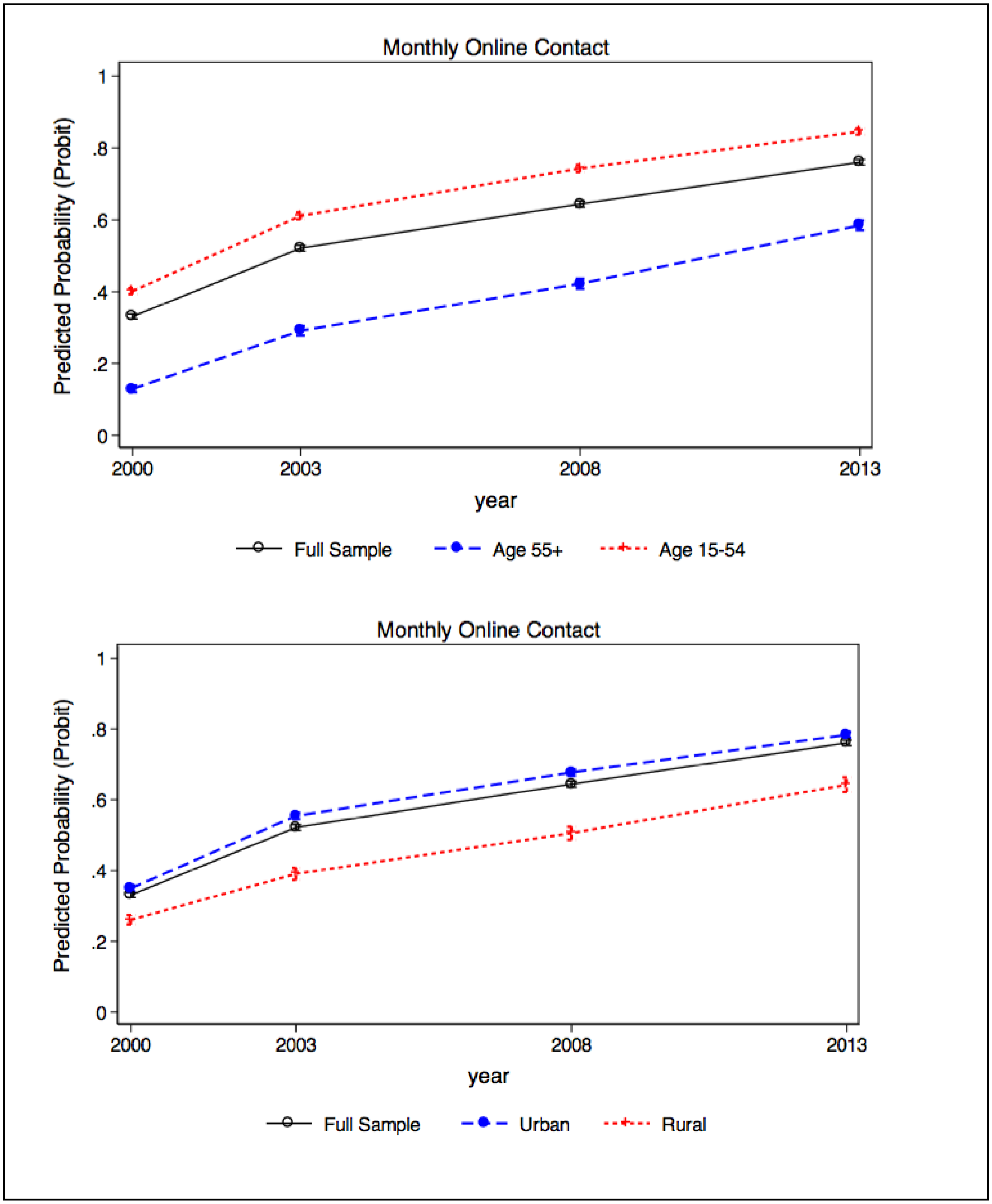


Figure 4.5: Covariate Balance Before and After Entropy Balancing

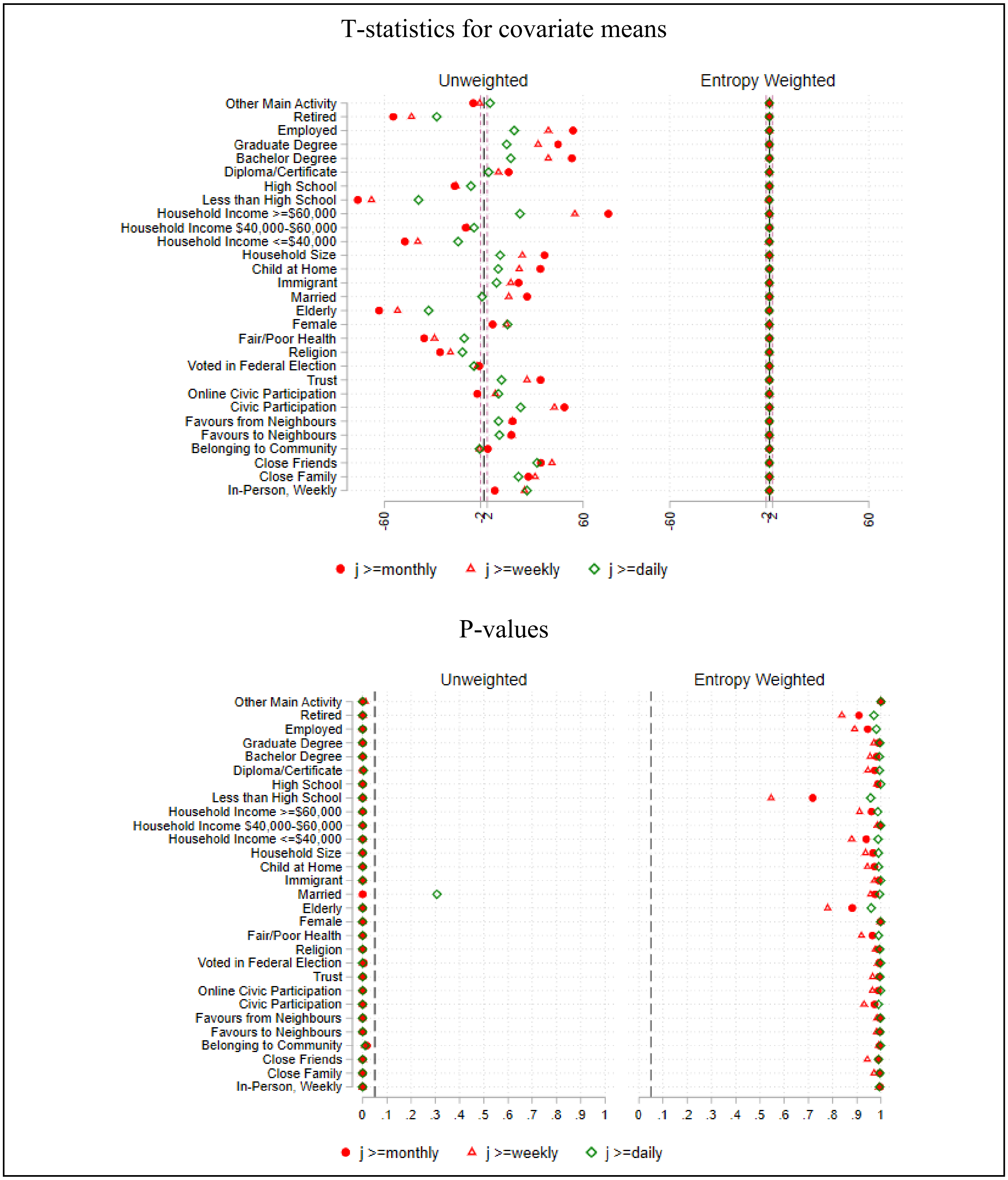
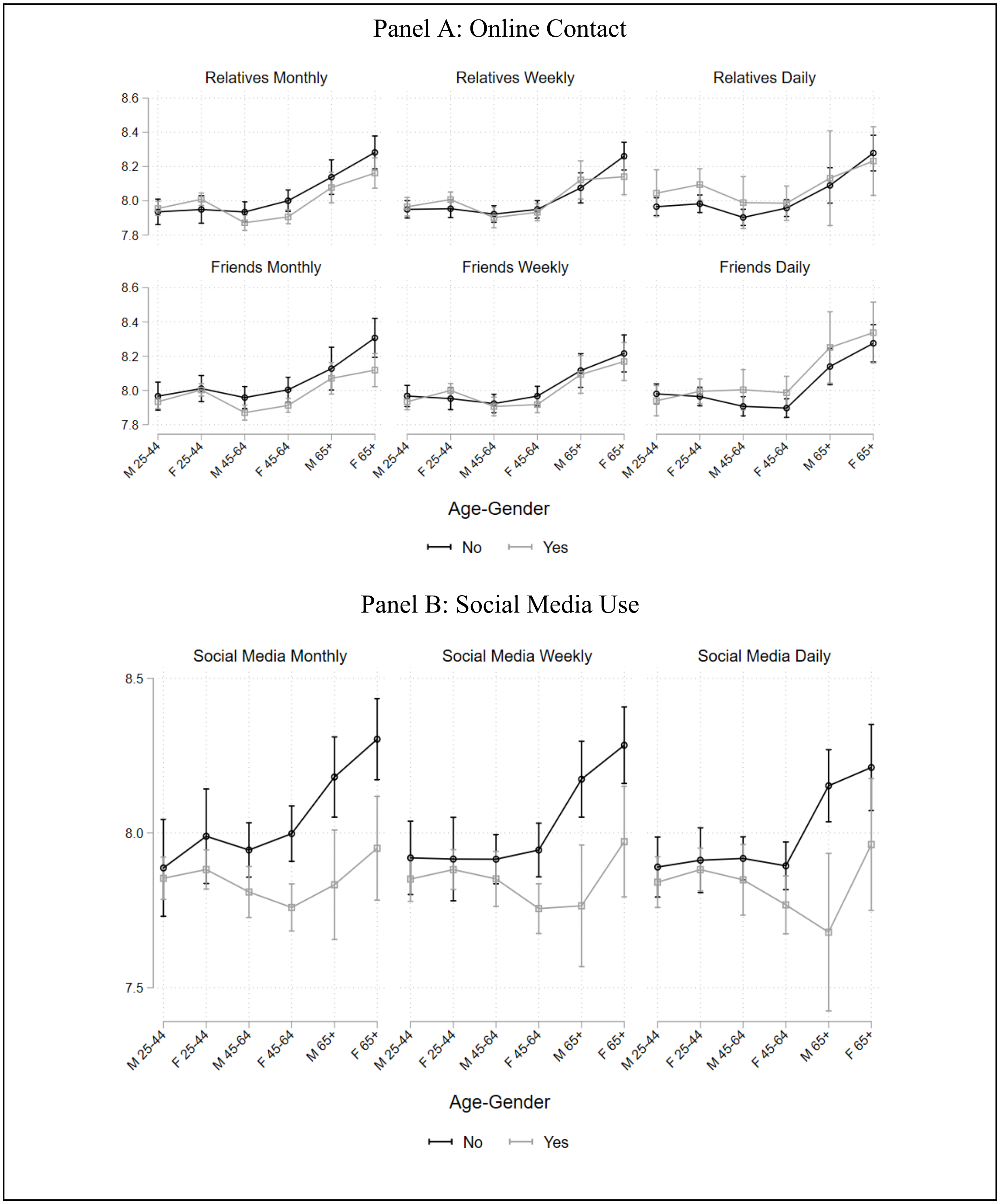


Figure 4.6: Associations between well-being, social media use and online contact by age and gender.



Notes: Each plot is derived from an OLS regression of life satisfaction on online contact/social media use interacted with age-by-gender dummies

Chapter 5: Conclusion

This dissertation examines the roles of environment, policy, and technology in shaping health and well-being in adulthood. The role of the environment is documented in Chapter 2, in which I show that education and earnings are negatively associated with adverse fetal shocks. The effects appear to be on par with the educational effects of compulsory schooling laws in Canada. That the effects arise from a high morbidity, low mortality illness (influenza) again establish the fetal period as a highly sensitive stage of human development.

In Chapter 3 I show that spousal institutionalization is a critical financial risk to older seniors. While Canada's retirement security system and public health insurance protect against loss of income and acute illness, seniors are not fully protected against out of pocket expenditures associated with long-term care. I show that residential long-term care can account for 40-55% of equivalent disposable income of a couple, suggesting that spousal institutionalization could result in a significant loss of income and impoverishment. Given the ageing of Canada's population and the tendency of ageing individuals to encounter chronic illness before death, occurrences of spousal impoverishment due to long-term care could rise in the future.

In Chapter 4, I provide quasi-experimental evidence that social media use has limited impact on SWB, but a positive impact on voter turnout and social trust. These findings suggest that social media could be effective in bridging diverse groups through pro-social or positive online interactions. The impact on voter turnout could reflect the importance of social media as a platform for coordinating political engagement and disseminating political news.

Bibliography

Akerman, A., Gaarder, I., & Mogstad, M. (2015). The Skill Complementarity of Broadband Internet. *The Quarterly Journal of Economics, 130*, 1781-1824.

Almond, D. (2006). Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population. *Journal of Political Economy, 114*, 672-712.

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives, 25*, 153-172.

Almond, D., & Mazumder, B. (2005). The 1918 Influenza Pandemic and subsequent health outcomes: An analysis of SIPP data. *American Economic Review, 95*(2), 258-262.

Almond, D., & Mazumder, B. A. (2011). Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy. *American Economic Journal: Applied Economics, 3*, 56-85.

Auger, M. F. (2008). On the brink of civil war: The Canadian government and the suppression of the 1918 Quebec Easter riots. *Canadian Historical Review, 89*, 503-540.

Bale, T. L., Baram, T. Z., Brown, A. S., Goldstein, J. M., Insel, T. R., McCarthy, M. M., . . . Nestler, E. J. (2010). Early life programming and neurodevelopmental disorders. *Biological Psychiatry, 68*, 314-319.

Barker, D. (2012). Developmental origins of chronic disease. *Public Health, 126*, 185-189.

Barker, D. J. (1997). Maternal nutrition, fetal nutrition and disease in later life. *Nutrition, 13*, 807-813.

Bellou, A. (2015). The impact of Internet diffusion on marriage rates: Evidence from the broadband market. *Journal of Population Economics, 28*, 265-297.

Bernard, A., & Li, C. (2006). *Death of a spouse: The impact of income for senior men and women.* Analysis in Brief, Statistics Canada.

Bhuller, M., Havnes, T., Leuven, E., & Mogstad, M. (2013). Broadband Internet: An Information Superhighway to Sex Crime? *The Review of Economic Studies, 80*, 1237-1266.

Braithwaite, D. O. (2002). ``Married widowhood": Maintaining couplehood when one spouse is living in a nursing home. *Southern Journal of Communication, 67*, 160-179.

Brown, R., & Thomas, D. (2011). *On the long term effects of the 1918 US influenza pandemic*. Unpublished Manuscript.

Burke, M., Kraut, R., & Marlow, C. (2011). Social Capital on Facebook: Differentiating Uses and Users. *Proceedings at the Annual Conference on Human Factors in Compuiting Systems*, (pp. 571-580). Retrieved from https://www.cameronmarlow.com/media/burke-2011-social.pdf

Campante, F., Durante, R., & Sobbrio, F. (2017). Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation. *Journal of the European Economic Association*.

Canadian Institute for Health Information. (2018). *CCRS Profile of Residents in Continuing Care Facilities 2016-2017.* Retrieved from https://www.cihi.ca/en/quick-stats

Cao-Lei, L., Massart, R., Suderman, M. J., Machnes, Z., Elgbeili, G., Laplante, D. P., . . . King, S. (2014). DNA methylation signatures triggered by prenatal maternal stress exposure to a natural disaster: Project Ice Storm. *PLoS One, 9*, e107653.

Card, D. & Lemieux, D. (2001). Education, earnings, and the 'Canadian G.I. Bill.' *Canadian Journal of Economics*, 42(2), 313-344.

Chan, J., & Ghose, A. (2014). Internet's dirty secret: Assessing the impact of online intermediaries on HIV transmission. *MIS Quarterly, 38*, 955-976.

Chen, Y., & Zhou, L.-A. (2007). The long-term health and economic consequences of the 1959--1961 famine in China. *Journal of Health Economics, 26*, 659-681.

Christian, P., Murray-Kolb, L. E., Khatry, S. K., Katz, J., Schaefer, B. A., Cole, P. M., . . . Tielsch, J. M. (2010). Prenatal micronutrient supplementation and intellectual and motor function in early school-aged children in Nepal. *Journal of American Medical Association, 304*, 2716-2723.

Clark, A., Frijters, P., & Shields, M. (2008). Relative income, happiness, and utility: An explanation for the Eastern Paradox and other puzzles. *Journal of Economic Literature, 46*(1), 95-144.

Colombo, F., Llena-Nozal, A., Mercier, J., & Tjadens, F. (2011). *Help wanted: Providing and paying for long-term care.* OECD: Paris.

Coviello, L., Sohn, Y., Kramer, A. D., Marlow, C., Franceschetti, M., Christakis, N. A., & Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PloS one, 9*, e90315.

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education, 1*, 697-812.

Czernich, N. (2012). Broadband Internet and Political Participation: Evidence for Germany. *Kyklos, 65*(1), 31-52.

de Rooij, S. R., Wouters, H., Yonker, J. E., Painter, R. C., & Roseboom, T. J. (2010). Prenatal undernutrition and cognitive function in late adulthood. *Proceedings of the National Academy of Sciences, 107*, 16881-16886.

Department of Commerce. (1924). *Mortality statistics 1921: Twenty-second annual report.* Department of Commerce: Washington.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook ``friends:" Social capital and college students´ use of online social network sites. *Journal of Computer-Mediated Communication, 12*, 1143-1168.

Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review, 104*(7), 2238-2265.

Fernandes, N., & Spencer, B. G. (2010). The private cost of long-term care in Canada: Where you live matters. *Canadian Journal on Aging, 29*, 307-316.

Forman, C., Goldfarb, A., & Greenstain, S. (2012). The internet and local wages: A puzzle. *American Economic Review, 102*, 556-575.

Galler, J. R., Bryce, C. P., Zichlin, M. L., Waber, D. P., Exner, N., Fitzmaurice, G. M., & Costa, P. T. (2013). Malnutrition in the first year of life and personality at age 40. *Journal of Child Psychology and Psychiatry, 54*, 911-919.

Garthwaite, C. (2008). *The effect of in-utero conditions on long term health: Evidence from the 1918 Spanish flu pandemic.* Tech. rep., Working Paper University of Maryland, Department of Economics.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis, 20*(1), 25-46.

Hainmueller, J., & Xu, Y. (2013). Ebalance: A Stata Package for Entropy Balancing. *Journal of Statistical Software, 54*(7), 1-18.

Hales, C. N., & Barker, D. J. (1992). Type 2 (non-insulin-dependent) diabetes mellitus: The Thrifty Phenotype Hypothesis. *Diabetologia, 35*, 595-601.

Helliwell, J. F., & Huang, H. (2013). Comparing the happiness effects of real and on-line friends. *PloS one, 8*, e72754.

Helliwell, J. F., Barrington-Leigh, C. P., Harris, A., & Huang, H. (2009). *International evidence on the social context of well-being.* Tech. rep., National Bureau of Economic Research.

Helliwell, J. F., Putnam, R. D., & others. (2004). The social context of well-being. *Philosophical Transactions-Royal Society of London Series B: Biological Sciences*, 1435-1446.

Hong, S.-H. (2013). Measuring the Effect of Napster on Recorded Music Sales: Difference-in-Differences Estimates Under Compositional Changes. *Journal of Applied Econometrics, 28*, 297-324.

Internet use and psychological well-being: A meta-analysis. (2010). *Cyberpsychology, Behavior, and Social Networking, 13*(3), 241-249.

Ivus, O., & Boland, M. (2015). The Employment and Wage Impact of Broadband Deployment in Canada. *Canadian Journal of Economics, 48*, 1803-1830.

Jaber, A. (2013). *Broadband Internet and Political Behavior: Evidence from the United States.* Available at SSRN. Retrieved from https://ssrn.com/abstract=2353549 or http://dx.doi.org/10.2139/ssrn.2353549

Johnson, N. P., & Mueller, J. (2002). Updating the accounts: global mortality of the 1918-1920 ``Spanish" influenza pandemic. *Bulletin of the History of Medicine, 76*, 105-115.

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The Journal of Economic Perspectives, 20*, 3-24.

Kelly, E. (2011). The scourge of Asian flu in utero exposure to pandemic influenza and the development of a cohort of British children. *Journal of Human Resources, 46*, 669-694.

Kolko, J. (2012). Broadband and local growth. *Journal of Urban Economics, 71*, 100-113.

Koutroumpis, P. (2009). The economic impact of broadband on growth: A simultaneous approach. *Telecommunications Policy, 33*.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*, 8788-8790.

Krasnova, H., Wenniger, H., Widjaja, T., & Buxmann, P. (2013). Envy on Facebook: A hidden threat to users' life satisfaction? *Papeer presented at the 11th International Conference on Wirtschaftsinformatik (WI).* Leipzig Germany. Retrieved from https://www.ara.cat/2013/01/28/855594433.pdf

Larcinese, V., & Miner, L. (2017). *Elections, The Political Impact of the Internet in US Presidential.* Available at EconPapers. Retrieved from https://econpapers.repec.org/paper/cepstieop/63.htm

Leacy, F. H., Canada, S., & of Canada, S. S. (1983). *Historical statistics of Canada [Online Version].* Statistics Canada: Ottawa.

Lew, B., & McInnis, M. (2006). *Guns and butter: World War I and the Canadian economy.* Working Paper.

Lin, M.-J., & Liu, E. M. (2014). Does in utero exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment. *Journal of Health Economics, 37*, 152-163.

Lohmann, S. (2015). Information technologies and subjective well-being: Does the Internet raise material aspirations? *Oxford Economic Papers, 67*, 740-759.

MacDonald, M. (2015). Regulating individual charges for long-term residential care in Canada. *Studies in Political Economy, 95*, 83-114.

Marcus, J. (2013). The effect of unemployment on the mental health of spouses - Evidence from plant closures in Germany. *Journal of Health Economics*, 32(3), 546-558

Mayer, A. (2017). Will democratization save the climate? An entropy-balanced, random slope study. *International Journal of Sociology*, 47(2), 81-98.

McDonald, L., & Robb, A. L. (2004). The economic legacy of divorce and separation for women in old age. *Canadian Journal on Aging, 23*, S83--S97.

Milligan, K. (2008). The evolution of elderly poverty in Canada. *Canadian Public Policy, 34*, S79-S94.

Milligan, K., & Schirle, T. (2013). *The retirement income system and the risks faced by Canadian seniors.* Working Paper, Canadian Labour Market and Skills Researcher Network.

Milligan, K., & Wise, D. A. (2013). *How do public pensions affect retirement incomes and expenditures? Evidence over five decades from Canada.* Retirement Research Center, National Bureau of Economic Research.

Myles, J. (2000). The maturation of Canadaś retirement income system: Income levels, income inequality and low income among older persons. *Canadian Journal on Aging, 19*, 187-316.

Neelsen, S., & Stratmann, T. (2011). Effects of prenatal and early life malnutrition: Evidence from the Greek famine. *Journal of Health Economics, 30*, 479-488.

Nelson, R. E. (2010). Testing the fetal origins hypothesis in a developing country: Evidence from the 1918 influenza pandemic. *Health Economics, 19*, 1181-1192.

Nie, N. H., & Hillygus, D. S. (2002). The impact of Internet use on sociability: Time-diary findings. *It & Society, 1*, 1-20.

Nishiura, H. (2009). Excess risk of stillbirth during the 1918--1920 influenza pandemic in Japan. *European Journal of Obstetrics \& Gynecology and Reproductive Biology, 147*, 115.

Oreopoulos, P. (2006). The compelling effects of compulsory schooling: Evidence from Canada. *Canadian Journal of Economics, 39*, 22-52.

Osberg, L. (2001). Poverty among senior citizens: A Canadian success story. In P. Grady, & A. Sharpe (Eds.), *The State of Economics in Canada Festschrift in Honour of David Slater.* Queens University: Kingston.

Painter, R. C., de Rooij and Susanne, R., Bossuyt, P. M., Simmers, T. A., Osmond, C., Barker, D. J., . . . Roseboom, T. J. (2006). Early onset of coronary artery disease after prenatal exposure to the Dutch famine. *The American Journal of Clinical Nutrition, 84*, 322-327.

Penard, T., & Poussing, N. (2010). Internet use and social capital: The strength of virtual ties. *Journal of Economic Issues, 44*(3), 569-595.

Richter, A., & Robling, P. O. (2013). *Multigenerational effects of the 1918-19 influenza pandemic in Sweden.* Working Paper.

Roseboom, T., Painter, R., van Abeelen, A., Veenendaal, M., & de Rooij, S. (2011). Hungry in the womb: What are the consequences? Lessons from the Dutch Famine. *Maturitas, 70*(2), 141-145.

Sabitini, F., & Sarracino, F. (2016). *Keeping up with the e-Joneses: Do Online Social Networks Raise Social Comparisons.* Munich Personal RePEc Archive 69201.

Sabitini, F., & Sarracino, F. (2017). Online Networks and Subjective Well-Being. *Kyklos, 70*(3), 456–48.

Schneider, S. K., Ońdonnell, L., Stueve, A., & Coulter, R. W. (2012). Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health, 102*, 171-177.

Sharpe, C. (2015). Enlistment in the Canadian Expeditionary Force 1914-1918. *Canadian Military History, 24*, 17-60.

St Clair, D., Xu, M., Wang, P., Yu, Y., Fang, Y., Zhang, F., . . . He, L. (2005). Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961. *Journal of the American Medical Association, 294*(5), 557-562.

Stacey, C. P. (1981). *Canada and the age of conflict: A history of Canadian external policies.* University of Toronto Press.

Stadnyk, R. (2002). *The status of Canadian nursing home care: Universality, accessibility, and comprehensiveness.* Tech. rep., Atlantic Centre of Excellence for Womenś Health.

Statistics Canada. (2012). *Programme for the International Assessment of Adult Competencies, 2012.* Public-Use MIcrodata File.

Statistics Canada. (2013). *General Social Survey. Cycle 17: Social Engagement.* Public-Use Microdata File.

Stein, A. D., Zybert, P. A., Van der Pal-de Bruin, K., & Lumey, L. H. (2006). Exposure to famine during gestation, size at birth, and blood pressure at age 59 y: Evidence from the Dutch Famine. *European Journal of Epidemiology, 21*, 759-765.

Steinfield, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology, 29*, 434-445.

Steptoe, A., Deaton, A., Stone, A.A. (2015). Subjective wellbeing, health, and ageing. *The Lancet*, 385(9968), 640-648.

Susser, E. S., & Lin, S. P. (1992). Schizophrenia after prenatal exposure to the Dutch Hunger Winter of 1944-1945. *Archives of General Psychiatry, 49*, 983-988.

Susser, E., & St Clair, D. (2013). Prenatal famine and adult mental illness: Interpreting concordant and discordant results from the Dutch and Chinese Famines. *Social Science Medicine, 97*(November), 325-330.

Taubenberger, J. K., & Morens, D. M. (2006). 1918 Influenza: The mother of all pandemics. *Emerging Infectious Disease, 12*, 15-22.

The Canadian Press (2014). *10 million Canadians use Facebook on mobile every day, some 20+ times a day.* Retrieved from https://www.theglobeandmail.com/technology/10-million-canadians-use-facebook-on-mobile-daily/article16976434/

Valkenburg, P., & Jochen, P. (2007). Online communication and adolescent well-being: Testing the stimulation versus the displacement hypothesis. *Journal of Computer-Mediated Communication, 12*(4), 1169-1182.

Veall, M. R. (2008). Canadian seniors and the Low Income Measure. *Canadian Public Policy, 34*, 47-58.

Verduyn, P., Lee, D., Park, J., Shablack, H., Orvell, A., Bayer, J., . . . Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental and Longitudinal Evidence. *Journal of Experimental Psychology: General, 144*(2), 480-488.

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., . . . others. (2008). Maternal and child undernutrition: Consequences for adult health and human capital. *The Lancet, 371*, 340-357.

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture, 3*, 206.

Appendix A: Data Appendix to Chapter 3

Table A1: Data Sources for Nursing Home Fees and Allowances

Province	Fees	Personal Allowance	Spousal Allowance
NL	Department of Health and Community Services (personal correspondence)		
PE	Health PEI (personal correspondence)		Subsection 8(3) General Regulation of the *Long-Term Care Subsidization Act*
NS	Department of Health and Community Services (personal correspondence), Stadnyk (2002)		
NB	Department of Social Development (personal correspondence)	Stadnyk (2002), New Brunswick Seniors' Coalition and subsection 21(4), General Regulation of the *Family Income Security Act*.	Standard Family Contribution Policy (public document)
ON	Table 3, General Regulation of the *Nursing Homes Act*	Subsection 42(1), General Regulation of the *Ontario Works Act*	Ministry of Health and Long-Term Care (personal correspondence)
MN	Schedule B, Personal Care Services Insurance and Administration Regulation of *The Health Services Insurance Act*	Maximum OAS+GIS (single) less minimum posted nursing home fee	Subsection 6.4, Personal Care Services Insurance and Administration Regulation of *The Health Services Insurance Act*
SK	Subsections 4(1) and 9(1) of the Special-Care Homes Rate Regulation of *The Housing and Special-Care Homes Act,* and *The Saskatchewan Income Plan Act*.	Subsections 4(1) and 9(1) of the Special-Care Homes Rate Regulation of *The Housing and Special-Care Homes Act,* and *The Saskatchewan Income Plan Act*.	Maximum OAS+GIS (single)
AB	Subsection 3(1) of the Nursing Homes Operation Regulation of the *Nursing Homes Act*	Maximum OAS+GIS (single) less accommodation charge (net of long-term care supplement as prescribed in the *Seniors Benefit Act*.	Maximum OAS+GIS (single)
BC	Subsection 5(2) of the Continuing Care Fees Regulation of the *Continuing Care Act*	Maximum OAS+GIS (single) less minimum posted accommodation charge or 15% of maximum OAS+GIS (single).	Maximum OAS+GIS (single)

Table A2: Summary of Provincial Fee Assessment Policies

	Allowances			Services Included in Fee	
Province	Personal	Spousal	Room & Board	Medical	Asset Test
NL	Stated	Stated flat rate plus allowable household expenses	Yes	Yes	Yes
PEI	Stated	OAS/GIS[1]	Yes	Yes after 2006	Yes Before 2007
NS	Stated	Stated after 2005; OAS/GIS before 2005[2]	Yes	Yes after 2004	Yes Before 2005
NB	Stated	Stated	Yes	Yes after 2005	Yes Before 2006
ON	Stated	Stated	Yes	Yes	No
MA	OAS+GIS-Minimum Fee	Stated	Yes	Yes	No
SK	OAS+GIS+SK Benefit – Minimum Fee	None	Yes	Yes	No
AB	OAS+GIS+AB Seniors' Cash Benefit + AB Seniors' LTC Supplement - fee	OAS/GIS[3]	Yes	Yes	No
BC	Stated	OAS/GIS[4]	Yes	Yes	No

Notes: Income allowances and spousal allowances are either stated explicitly (i.e. in legislation or government documents) or pegged to the GIS/OAS rate for a single person. Specific annual values are reported in Table A2.

[1]Allows for variation in 50% division of joint spousal outcome when a spouse has financial need. Financial need is implicitly defined as an income below OAS/GIS for a single.

[2]No allowance stated prior to 2005; here we assume it is equal to OAS/GIS for a single.

[3]No stated allowance but policy designed to ensure spouse has "reasonable level of income," defined here as OAS/GIS.

[4]The spouse's retained income prior to 2010 is pegged to OAS/GIS. After 2010 the income test is based on individual after-tax income, but married residents can apply for a financial hardship waiver if their spouse requires more income. We assume here that financial hardship is defined as an income below OAS/GIS for a single senior.

Table A3: Minimum Annual Nursing Home Fees (Stated Minimums)

year	NL	PE	NS	NB	ON	MN	SK	AB	BC
2000	0	0	0	0	10,680	9,271	9,564	10,300	9,526
2001	0	0	0	0	11,038	9,417	9,866	10,300	9,746
2002	0	0	0	0	11,278	9,600	10,011	10,300	9,892
2003	0	0	0	0	11,720	9,636	10,278	10,300	9,892
2004	0	0	0	0	11,720	9,892	10,500	10,300	10,074
2005	0	0	0	0	11,720	10,038	10,676	10,300	10,293
2006	0	0	0	0	11,979	10,110	11,121	10,300	10,505
2007	0	0	0	0	12,220	10,512	11,534	10,789	10,727
2008	0	0	0	0	12,490	10,840	11,864	10,789	10,983
2009	0	0	0	0	12,603	11,169	11,864	11,570	11,016
2010	0	0	0	0	12,640	11,424	11,969	11,570	11,213

Notes: If minimum is not stated, then it is defaulted to \$0. Also, our simulations reduce the copay below the stated minimum if the income remaining after paying the minimum is below the personal allowance. The fee is gradually reduced until retained income is restored to the allowance. Thus, the effective minimum is always \$0, provided that incomes in the lower tail of the distribution are sufficiently low.

Table A4: Maximum Annual Nursing Home Fees

Year	NL	PE	NS	NB	ON	MN	SK	AB	BC
2000	34,065	36,135	39,164	39,811	16,034	21,644	18,000	10,300	18,250
2001	34,065	38,325	42,960	40,676	16,246	21,973	18,569	10,300	18,250
2002	34,065	42,705	46,392	42,957	17,348	22,411	18,842	10,300	18,250
2003	34,065	44,895	48,837	44,975	17,772	22,666	19,343	14,461	23,725
2004	34,065	48,180	48,472	47,016	17,772	23,250	19,762	14,461	24,163
2005	34,065	48,180	27,192	48,187	17,772	23,579	20,093	14,461	24,710
2006	34,065	49,640	27,558	25,550	18,162	23,761	20,931	14,461	25,196
2007	34,065	23,725	29,538	25,550	18,527	24,674	21,708	15,148	25,732
2008	34,065	24,455	31,663	25,550	18,936	25,440	22,329	15,148	26,342
2009	34,065	25,294	33,940	30,295	19,371	26,207	22,329	16,242	26,422
2010	34,065	25,294	36,135	30,295	19,429	26,791	22,527	16,242	26,893

Table A5: Annual Personal Allowance

year	NL	PE	NS	NB	ON	MN	SK	AB	BC
2000	1,800	996	1,260	1,056	1,344	1,990	1,967	3,780	1,689
2001	1,800	1,236	1,260	1,056	1,344	2,208	2,029	4,145	1,744
2002	1,800	1,236	1,260	1,056	1,344	2,200	2,059	4,319	1,770
2003	1,800	1,236	1,260	1,056	1,392	2,485	2,113	4,995	2,229
2004	1,800	1,236	1,260	1,056	1,392	2,498	2,159	5,263	2,315
2005	1,800	1,236	1,352	1,162	1,392	2,564	2,195	5,475	2,308
2006	1,800	1,236	1,440	1,220	1,428	3,028	2,287	3,180	2,633
2007	1,800	1,236	1,494	1,281	1,464	3,124	2,372	3,180	2,909
2008	1,800	1,236	1,546	1,288	1,500	3,193	2,440	3,180	3,051
2009	1,800	1,236	1,546	1,296	1,536	2,865	2,440	3,180	3,018
2010	1,800	1,236	1,560	1,296	1,560	2,736	2,461	3,180	3,300

Table A6: Annual Spousal Allowance

year	NL[1]	PE	NS	NB	ON	MN	SK	AB	BC
2000	n/a	11,261	11,261	9,129	14,265	10,222	0	11,261	11,261
2001	n/a	11,625	11,625	9,424	14,265	10,653	0	11,625	11,625
2002	n/a	11,800	11,800	9,566	14,265	11,101	0	11,800	11,800
2003	n/a	12,121	12,121	9,826	14,265	11,234	0	12,121	12,121
2004	n/a	12,389	12,389	10,044	14,775	11,674	0	12,389	12,389
2005	n/a	12,601	14,520	10,215	14,775	11,878	0	12,601	12,601
2006	n/a	13,138	15,045	10,650	15,032	11,840	0	13,138	13,138
2007	n/a	13,636	15,992	11,052	16,175	12,387	0	13,636	13,636
2008	n/a	14,034	16,165	11,374	16,457	13,075	0	14,034	14,034
2009	n/a	14,034	16,974	11,374	16,855	14,902	0	14,034	14,034
2010	n/a	14,160	16,994	11,477	16,855	15,440	0	14,160	0

[1]NL allows spouses to retain an annual spousal allowance of $10,800 (unchanged over the study period) plus additional amounts based on the spouse's actual household expenses. Allowable expenses include mortgage or rent payments, property taxes/insurance, property maintenance, utilities, cable, vehicle loan/lease payments and insurance, employment expenses, union dues, transportation (to the nursing home; $150-$300 per month), home care expenses, medical transportation and income taxes. Additional allowances are also permitted based on the discretion of the Department of Health and Community Services.

Table A7: GIS, OAS and CPI Values: 2000-2010

year	GIS (Married)	GIS (Single)	OAS	CPI
2000	3,983	6,115	5,145	95
2001	4,112	6,313	5,312	98
2002	4,174	6,408	5,392	100
2003	4,288	6,582	5,539	103
2004	4,383	6,728	5,661	105
2005	4,458	6,843	5,758	107
2006	4,746	7,235	5,903	109
2007	5,024	7,608	6,028	111
2008	5,171	7,830	6,204	114
2009	5,171	7,830	6,204	114
2010	5,217	7,901	6,259	117

Table A8: National Estimates of Expenditure-Income Ratios Based on Newfoundland and Labrador's Allowable Household Expenses for Spouses of Nursing Home Residents

	Gross Income Decile									
	1	2	3	4	5	6	7	8	9	10
Expenditure/ATI	0.52	0.41	0.34	0.36	0.34	0.33	0.32	0.29	0.27	0.22

Notes: The above table reports means and standard deviations of the ratio of allowable expenses income test to combined after-tax incomes of individuals and their spouses (allowable expenses are listed in Table D6). The ratios reported here are based on incomes and expenditures reported in the 1997-1999 waves of the Survey of Household Spending. Expenditures are set equal to the sum of observed household expenditures on rent, mortgage, property taxes, maintenance and insurance for a principal residence, telephone and cellular phone service, health care, taxes, union dues, personal income taxes, life insurance premiums and statutory employee contributions (e.g. CPP/QPP, EI). The sample includes households with respondents aged 80 or older and spouses aged 65 and older or with respondents 65 and older and spouses aged 80 and older. We also limit the sample to households who have a combined after-tax income of greater than $2,000. This is comparable to the SLID sample, in which individuals and spouses have after-tax incomes of at least $1,000. The resulting sample contains 693 observations.

Appendix B: Survey Questions for Well-Being, Online Communication and Social Media Use

B.1 Online Communication

GSS 2000: In the last month, how often did you communicate with your family or relatives... by E-mail?

GSS 2003: In the last month, how often did you communicate with relatives [friends] on the internet (including by e-mail)? Was it: everyday, a few times a week, a few times a month, once a month, not in the last month.

GSS 2008: In the past month, how often did you communicate with any of your relatives [friends] by e-mail or internet (outside of the people you live with)? Everyday, a few times a week, once a week, 2 or 3 times a month, once a month, not in the last month, did not use e-mail or internet in the past month.

GSS 2013: In the past month, how often did you communicate with any of your relatives [friends] by e-mail or internet (outside of the people you live with)? Everyday, a few times a week, once a week, 2 or 3 times a month, once a month, not in the last month. (Respondents instructed to include all forms of internet communication, including social networking websites, instant messaging and Skype).

B.2. Use of Social Networking Websites, 2013 GSS[86]

In the past 12 months, have you used the internet to access a social networking website (such as Facebook or Twitter)? Yes/No Answer.

[86] Use of social networking websites was added as a new question in the 2013 GSS.

How often do you access your social networking site(s)? Several times a day, once a day. 3-5 times a week, 1-2 times a week, a few times per month, less than once a month, never.

B.3 Life Satisfaction[87]

GSS 2003 and 2008: Using a scale of 1 to 10, where 1 means "Very dissatisfied" and 10 means "Very satisfied", how do you feel about your life as a whole right now?[88]

GSS 2013: Using a scale of 0 to 10, where 0 means "Very dissatisfied" and 10 means "Very satisfied", how do you feel about your life as a whole right now?"[89]

[87] The 2000 GSS survey did not ask about life satisfaction.
[88] Respondents who reported "no opinion" (11) are dropped from the sample.
[89] Zeros are recoded to 1 to conform with the scales from 2003 and 2008.