Use it too much and lose everything? The effects of hours of work on health

Shinya Kajitani a, Colin McKenzie b,*, Kei Sakata c

a Faculty of Economics, Kyoto Sangyo University, Kyoto, Japan
b Faculty of Economics, Keio University, 2-15-45 Minato-ku, Tokyo, 108-8345, Japan
c Australian Institute of Family Studies, Southbank, Australia

ARTICLE INFO

JEL classification:
Noc 110 J2
Keywords:
Health Working hours Endogeneity Pensions Retirement

ABSTRACT

In this paper, we extend the ‘use it or lose it’ hypothesis to analyse whether the negative effects of working hours eventually dominate the positive effects of work as the hours of work increase. Using panel data from the HILDA survey, we estimate the optimal hours of work for the health status of middle age and elderly workers. We deal with the potential endogeneity of working hours by using the instrumental variable estimation technique with instruments based on the age for pension eligibility. For males working relatively moderate hours (up to around 24–27 h a week), an increase in working hours has a positive impact on their health outcomes, but thereafter an increase in working hours has a negative impact on health outcomes. When weekly working hours exceed 50 h, an individual’s health status is worse off than when he is not working at all.

1. Introduction

How does work affect health? Is work bad for health? Does work have any health benefits? There is an extremely large literature in epidemiology, occupational psychology, and health economics that examines these issues (see, for example, Bannai & Tamakoshi, 2014; Bassanini & Caroli, 2015). Some papers examine the extensive margin of work (working or not working), for example, by examining the impact of unemployment and job loss on health outcomes. Using fixed-effect models for Australian, Canadian, and UK panel data, Llena-Nozal (2009) shows that the shift from being employed to being unemployed has adverse effects on mental health. On the other hand, using German data, Schmitz (2011) finds no significant effect of plant closures on various health outcomes. Heller-Sahlgren (2017) finds no immediate effect, but rather a delayed impact of retirement on mental health. Another stream of research on the extensive margin examines whether retirement has any impact on cognitive functioning and health. Such studies test the so-called ‘use it or lose it’ hypothesis (Atalay et al. 2019; Blake & Garrantue, 2019; Bonsang et al. 2012; Carrino et al. 2020; Coe & Zamarro, 2011; De Grip et al. 2012; Kajitani, Sakata, & McKenzie, 2017; Mazzonna and Peracchi 2012, 2017; Rohweder & Willis, 2010). For Australian women, using the first 11 waves of the Household, Income and Labour Dynamics in Australia (HILDA) survey, Zhu (2016) provides evidence that retirement status has a positive impact on both physical and mental health outcomes. Overall, these studies tend to suggest that retirement has a negative impact on cognitive functioning, but positive impacts on health outcomes.

Other papers examine the intensive margin of work, that is, the number of hours worked. The main focus of these analyses is on the effects of working long hours on various health outcomes. Such studies reveal that working long hours has adverse effects on health (Bannai et al. 2015; Frijters et al. 2009; Nie et al. 2015; Sparks et al. 1997; Spurgeon et al. 1997). In a paper that emphasises the importance of employment policy in Europe in the context of the relationship between work and health, Barnay (2016) argues that while there is a large literature in epidemiology on the intensive margin of work on health outcomes, there is next to no literature in economics (see also Bannai & Tamakoshi, 2014). Lee and Lee (2016) examine the impacts of working hours on injury rates in the worker’s workplace which is regarded as one of the risk factors for their health, exploiting a quasi-natural experiment in South Korea. Assuming that the injury risk is a quadratic function of working hours, they find that the function is convex, which indicates that shortening very long working hours could be effective in reducing the injury rate. However, previous studies do not examine the effects of moderate working hours on health, or the optimal number of hours worked. Robone et al. (2011) indicate that having a part-time job, as

* Corresponding author.
E-mail address: mckenzie@keio.jp (C. McKenzie).
Table 1
Australian age pension eligibility ages and the sample distribution.

| Birthday (Age) | Pension eligibility age before April 2009 | Sample (%) | Pension eligibility age after May 2009 | Sample (%) |
|---------------|------------------------------------------|------------|--------------------------------------|------------|
|               | Wave 1 | Wave 8 | Wave 1 | Wave 8 |
| Up to 30 June | 65     | 68.7%  | 51.1% | 65 | 68.7% | 51.1% |
| 1952 (Aged 49 and older in Wave 1) | 65 | 5.0% | 3.9% | 65 | 5.0% | 3.9% |
| 1 July 1952 to 31 December 1953 (Aged 47–48 in Wave 1) | 65 | 3.7% | 2.6% | 66 | 2.8% |
| 1 January 1954 to 30 June 1955 (Aged 46 in Wave 1) | 65 | 4.9% | 3.7% | 66 | 3.6% |
| 1 July 1955 to 31 December 1956 (Aged 44–45 in Wave 1) | 65 | 17.6% | 38.8% | 67 | 41.8% |
| From 1 January 1957 (Aged 43 and younger in Wave 1) | Total | 100.0% | 100.0% | 100.0% |
| Number of Observations | 2,881 | 2,644 | 2,643 |

Source: For pension eligibility ages: Commonwealth of Australia (2009), p. 9. Sample proportions in each group are authors’ calculations using data from the HILDA Survey.

Table 2
Descriptive statistics.

| Variable                  | Mean  | Std. Dev. | Min | Max |
|---------------------------|-------|-----------|-----|-----|
| Physical functioning      | 79.63 | 23.79     | 0   | 100 |
| Bodily pain               | 70.07 | 24.64     | 0   | 100 |
| General health            | 65.16 | 21.69     | 0   | 100 |
| Vitality                  | 61.59 | 19.72     | 0   | 100 |
| Mental health             | 76.36 | 16.59     | 0   | 100 |
| Working hours             | 27.48 | 23.40     | 0   | 100 |
| Working hours-squared     | 1302.75 | 1361.62 | 0   | 7056 |
| Age                       | 57.09 | 11.94     | 40  | 86 |
| School years 7–10         | 0.49  | 0.50      | 0   | 1 |
| School years 11 and over  | 0.48  | 0.50      | 0   | 1 |
| Married                   | 0.79  | 0.41      | 0   | 1 |
| Number of dependent children | 0.63   | 1.07    | 0   | 9 |
| Ownhouse                  | 0.83  | 0.38      | 0   | 1 |
| Inner regional            | 0.27  | 0.44      | 0   | 1 |
| Outer regional            | 0.13  | 0.33      | 0   | 1 |
| Remote/Very remote        | 0.02  | 0.14      | 0   | 1 |
| Vacancy rate              | 1.41  | 0.42      | 0.56| 3.26 |
| Age difference 1          | 2.18  | 4.59      | 0   | 21 |
| Age difference 2          | −10.47 | 9.15     | −27 | 0 |
| White-collar              | 0.58  | 0.49      | 0   | 1 |

Source: Authors’ calculations using data from Waves 1–12 of the HILDA Survey. Note: 34,835 observations from 5,824 individuals are used to compute the descriptive statistics in each case.

that the relationship between work and health may not be linear. Work can be a double-edged sword in that it can have both positive and negative effects. Interactions with people at work may help maintain an individual’s cognitive functions and his/her mental health. Moreover, working individuals have more incentive to invest in health repair activities to be ‘fit’ in the labour market. On the other hand, long working hours can cause fatigue and stress on both physical and mental levels which potentially damage an individual’s overall health and reduce the amount of time that can be invested in health repair activities. Most of the previous studies treat long working hours as a 0–1 dummy variable which defines long working hours as working more than 50 or 60 h per week. This means that they implicitly assume that long working hours have a constant shift effect on health status. They do not deal with the potential non-linear effects of working hours on health.

For health outcomes, the literature on the extensive margin suggests that working may be better than not working, while the literature on the intensive margin suggests that working extremely long hours is worse than working a normal working week. Combining these two observations suggests there is a non-linear relationship between work hours and health outcomes which we seek to capture using a quadratic form in work hours. Although not directly focusing on health outcomes, one explanation for the results generated in Pencavel’s (2014, 2016) analysis of the nonlinear relationship between working hours and productivity in munitions factories in Britain during the First World War is stress and fatigue generated by long working hours that then affects productivity (see also Collewet & Sauermann, 2017).

We contribute to the existing literature in two ways. First, we focus on not only labor market participation (the extensive margin), but also working hours (the intensive margin). Secondly, the literature examining the impact of retirement on cognitive function examines the ‘use it or lose it’ hypothesis, namely that not working (not using your brain) leads to losses of cognitive functioning. Here, we also examine the relevance of this hypothesis for a broader set of health outcomes. In addition, we focus on the ‘use it too much and lose everything’ hypothesis which refers to the situation where working too much can lead to not just a loss of cognitive functioning (see Kajitani, McKenzie, & Sakata, 2017), but also declines in health status across the board.

We examine the causal impact of working hours on the health outcomes of middle-aged and older male adults (men aged 40 years and

---

1 In contrast, for women, Llena-Nozal et al. (2004) find that compared to working full-time, moving to part-time work or overtime leads to improvements in mental health outcomes.
over) in Australia using Wave 1 to Wave 12 of the HILDA Survey. The health outcomes are measured using five self-assessed health score components computed from the SF-36 (the 36-Item Short Form Health Survey) which is one of the most widely used self-assessed measures of health status. These score components cover both an individual’s physical and mental health (see Ware, 2000). One of the issues in estimating the causal relationship between hours worked and health is what is called the ‘healthy worker effect’ (Bassanini & Caroli, 2015); that is, healthy workers are more likely to be employed and work longer. Thus, the presence of the healthy worker effect implies the existence of reverse causality. We deal with the potential endogeneity of decisions relating to working hours by using the instrumental variable estimation technique. One advantage of using a sample of middle-aged and older adults is that it enables us to use information related to the eligibility age for pension benefits as instruments for variables related to working hours.

Our empirical evidence shows that there is non-linearity in the effects of working hours on self-assessed health status. To be more specific, there is an inverted U-shaped relationship. When working hours are less than around 24–27 h a week, increases in working hours have a positive impact on health. However, when working hours are greater than this threshold, increases in working hours have negative impacts on health. Compared to males who do not work, working hours slightly over 50 h per week will lead to worse health outcomes depending on the measure of health. One advantage of using a sample of middle-aged and older adults is that it enables us to use information related to the eligibility age for pension benefits as instruments for variables related to working hours.

Table 3
The impacts of working hours on health.

| VARIABLES                                                | 1st Stage (FE) | 2nd stage (FEIV) |
|----------------------------------------------------------|----------------|------------------|
|                                                          | (3.1A)         | (3.1B)           |
|                                                          | (3.2A)         | (3.2B)           |
|                                                          | (3.2C)         | (3.2D)           |
|                                                          | (3.2E)         | (3.2F)           |
|                                                          | (3.2G)         | (3.2H)           |
|                                                          | (3.2I)         | (3.2J)           |
|                                                          | (3.2K)         | (3.2L)           |
|                                                          | (3.2M)         | (3.2N)           |
|                                                          | (3.2O)         | (3.2P)           |
|                                                          | (3.2Q)         | (3.2R)           |
|                                                          | (3.2S)         | (3.2T)           |
|                                                          | (3.2U)         | (3.2V)           |
|                                                          | (3.2W)         | (3.2X)           |
|                                                          | (3.2Y)         | (3.2Z)           |
|                                                          | (3.3A)         | (3.3B)           |
|                                                          | (3.3C)         | (3.3D)           |
|                                                          | (3.3E)         | (3.3F)           |
|                                                          | (3.3G)         | (3.3H)           |
|                                                          | (3.3I)         | (3.3J)           |
|                                                          | (3.3K)         | (3.3L)           |
|                                                          | (3.3M)         | (3.3N)           |
|                                                          | (3.3O)         | (3.3P)           |
|                                                          | (3.3Q)         | (3.3R)           |
|                                                          | (3.3S)         | (3.3T)           |
|                                                          | (3.3U)         | (3.3V)           |
|                                                          | (3.3W)         | (3.3X)           |
|                                                          | (3.3Y)         | (3.3Z)           |
|                                                          | (3.4A)         | (3.4B)           |
|                                                          | (3.4C)         | (3.4D)           |
|                                                          | (3.4E)         | (3.4F)           |
|                                                          | (3.4G)         | (3.4H)           |
|                                                          | (3.4I)         | (3.4J)           |
|                                                          | (3.4K)         | (3.4L)           |

Notes.

1) The first stage equations have been estimated using an individual fixed effects (FE) estimator and the second stage equations have been estimated using an individual fixed effect instrumental variable (FEIV) estimator.

2) *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

3) Figures reported in square brackets are robust standard errors adjusted for clustering.

4) *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

5) Notes.

6) The Hausman test for fixed effects reports a Hausman test for the null hypothesis that the coefficients on Age difference 1 and Age difference 2 are jointly zero.

7) The test of zero restrictions reports a test of the null hypothesis that all the coefficients except the constant are jointly zero. This test is computed as an F-test in Stage 1.

8) The Cragg-Donald Wald F statistic reported is computed using the Cragg-Donald Wald F statistic for weak instruments.

9) The Endogeneity test (F-statistic) is based on 500 bootstrap repetitions using the Cragg-Donald Wald F statistic for weak instruments.

10) The Hausman test for fixed effects reports a Hausman test for the null pooled model estimated by instrumental variables against the fixed effect model estimated by instrumental variables. The test is based on 500 bootstrap repetitions using the “hausman” command in STATA 17.

11) The Endogeneity test tests the null hypothesis that working hours and Working hours-squared in the second stage can be treated as being exogenous. Following Wooldridge (2010, pages 352–354), this test is implemented as an F-test that tests the joint significance of the residuals from the two first stage equations when added to each of the stage 2 models and these second stage models are estimated by a fixed effects estimator.

12) The Cragg-Donald Wald F statistic reported is computed using the “xtivreg2” command in STATA 17.
Using the parameter estimates from Table 3 for All and Table 5 for the white-collar and blue-collar groups, this Table presents point estimates of $\alpha_1$, $\beta_0$, their standard errors, and their confidence intervals computed using the “nlcom” command in STATA 17. The estimates of the standard errors are based on 500 bootstrap replications.

| Panel A: All | Coef. | Std. Err. | 95% Conf. Interval |
|-------------|-------|-----------|--------------------|
| Physical functioning | 26.394 | 0.898 | 24.725 - 27.997 |
| Bodily pain | 27.196 | 1.076 | 25.001 - 28.987 |
| General health | 25.986 | 1.044 | 23.945 - 27.838 |
| Vitality | 25.928 | 0.949 | 24.019 - 27.609 |
| Mental health | 23.949 | 1.435 | 20.543 - 26.134 |

| Panel B: White-collar |
|-----------------------|-------|-----------|--------------------|
| Physical functioning | 27.269 | 0.976 | 25.511 - 28.086 |
| Bodily pain | 28.017 | 1.209 | 25.657 - 30.332 |
| General health | 27.151 | 1.149 | 24.373 - 29.848 |
| Vitality | 26.609 | 1.029 | 24.606 - 28.549 |
| Mental health | 24.848 | 1.964 | 21.110 - 28.248 |

| Panel C: Blue-collar |
|----------------------|-------|-----------|--------------------|
| Physical functioning | 23.430 | 2.247 | 18.677 - 27.102 |
| Bodily pain | 24.193 | 2.293 | 18.481 - 27.910 |
| General health | 22.586 | 2.657 | 17.233 - 27.012 |
| Vitality | 23.147 | 2.315 | 18.505 - 27.183 |
| Mental health | 20.839 | 3.322 | 13.175 - 25.514 |

Note. Using the parameter estimates from Table 3 for All and Table 5 for the white-collar and blue-collar groups, this Table presents point estimates of $\alpha_1$, $\beta_0$, their standard errors, and their confidence intervals computed using the “nlcom” command in STATA 17. The estimates of the standard errors are based on 500 bootstrap replications.

The rest of this paper is organized as follows: Section 2 provides a brief description of the Australian pension system. Section 3 describes the data and Section 4 presents the empirical framework used in this paper. Section 5 reports the results of the estimation and discusses their implications. The last section concludes this paper.

2. Australian pension system

In Australia, the timing of retirement is closely related to the pension system. Retirement income consists of three sources: a means tested public pension, a mandatory employer-contributed private retirement savings account, and voluntary private retirement savings. Since there is no mandatory retirement age in Australia, an elderly Australian can also continue to work to supplement his/her pension.

The maximum benefit payment from the public age pension is set at 25% of male total average earnings. Since the introduction of the good service tax in 2000, a supplement for compensation has been added to the payment. The maximum basic rates of the public pension were A$10,262.20 per annum for the standard rate and A$8,569.60 per annum to the payment. The maximum basic rates of the public pension were A$10,262.20 per annum for the standard rate and A$8,569.60 per annum to the payment. The maximum basic rates of the public pension were A$10,262.20 per annum for the standard rate and A$8,569.60 per annum to the payment.

The rest of this paper is organized as follows: Section 2 provides a brief description of the Australian pension system. Section 3 describes the data and Section 4 presents the empirical framework used in this paper. Section 5 reports the results of the estimation and discusses their implications. The last section concludes this paper.

3. Data: Overview of the HILDA survey

Our data are drawn from the first 12 waves of the “Household, Income and Labour Dynamics in Australia (HILDA) Survey,” from Wave 1 conducted in 2001 to Wave 12 conducted in 2012. The HILDA Survey, which is conducted by the Melbourne Institute: Applied Economics and Social Research, the University of Melbourne, is a broad social and economic longitudinal survey. Since 2001, the HILDA Survey has asked Australian respondents about their economic and subjective well-being, family structures, and labor market dynamics. Households included in the survey were selected using a three-stage approach. First, a sample of 488 Census Collection Districts (CDs) were randomly selected from across Australia. Second, within each of these CDs, a sample of dwellings was selected based on expected response rates and occupancy rates. Finally, within each dwelling, up to three households were selected to be part of the sample. In addition, the sample was replenished in Wave 11. One aim of this replenishment was to provide better coverage of migrants for inclusion in the HILDA Survey.

The HILDA survey contains the SF-36 (the 36-Item Short Form Health Survey) which is one of the most widely used self-assessed measures of health status. It consists of eight scaled self-assessed health scores: physical functioning; role physical; bodily pain; general health; vitality; social functioning; role emotional and mental health. The eight categories are scaled by the weighted sums of their questions, and are converted to a 0–100 scale. 0 is equivalent to the highest disability, and 100 is equivalent to the lowest disability. Of the eight health scores, we use physical functioning, bodily pain, general health; vitality; social functioning; role emotional and mental health. The eight categories are scaled by the weighted sums of their questions, and are converted to a 0–100 scale. 0 is equivalent to the highest disability, and 100 is equivalent to the lowest disability. Of the eight health scores, we use physical functioning, bodily pain, general health; vitality; social functioning; role emotional and mental health. The other three scores, role physical; social functioning; and role emotional, are eliminated from the analysis because they display little variation. For data in Wave 1 of HILDA, Butterworth and Crosier (2004) examine the internal consistency of the health measures for each domain, and conclude that their analysis supports the internal consistency of the measures.

The general release HILDA data sets that we are using do not contain information on the day or month of birth of the respondent. In each wave, for each individual, two ages are reported: the age as at 30 June of the year of the wave, and the age at the time of the survey. Using the information on the age of respondents as at 30 June in Wave 1, it is possible to determine the pension eligibility ages of all individuals.
except those whose age is reported to be 44 and 47 in this wave. Using both ages and the information on when the respondent was interviewed, it is possible to determine the pension eligibility age for 25% and 32% of those respondents whose ages are reported to be 44 and 47 at 30 June 2001 in wave 1, respectively. Those respondents whose pension eligibility age could not be determined were excluded from the analysis. Table O1 in the online supplementary material indicates the impact of this exclusion on the sample size available in each wave.

As stated in section 2, the 2009 pension reform was announced on 12 May 2009. As the interview periods for HILDA’s waves 8 and 9 are from 20 August 2008 to 27 February 2009 and from 20 August 2009 to 11 March 2010, respectively, the 2009 pension reform falls right between waves 8 and 9.

The exact definitions of all the variables used in the analysis in this paper are summarized in Appendix I. The sample is restricted to individuals who meet the following five criteria: (i) males aged 40 and over in Wave 1 of the survey or males who turn 40 after Wave 1 but are only included for the Waves where they are aged 40 and over; (ii) all five scores relating to health status are available; (iii) age and working hours are less than age and working hours in the top 1% percentile, respectively; (iv) respondents who report they are unemployed are excluded; and (v) information on all the relevant variables is available. We target this age group as people start experiencing some health declines. For example, the Australian Heart Foundation recommends that people over 45 (over 35 for Aboriginal and Torres Strait Islanders) have a heart health check. The National Bowel Cancer Screening Program is offered to all Australians aged 50–74 in 2020. In addition, this sample selection is applied for our identification strategy as younger age groups are thought to be indifferent to changes in their retirement age. Even if we did not have attrition, the second part of criterion (i) means we will not have a balanced panel data set. After imposing these restrictions, a sample of 34,385 observations on 5,824 individuals remain. Table O1 in the online supplementary material indicates the effect of these criteria on the sample size available in each wave.

Table 2 displays descriptive statistics on all the variables used in the analysis. In this table, Working hours is the respondent’s usual hours of working per week. As a result, the mean value of Working hours for males is 27.48 h.

4. Estimation model and identification strategy

Our identification strategy exploits the variation in working hours, while controlling for time-invariant individual characteristics. In order to capture the possible non-linear effects of working hours on health
### Table 5
The impacts of working hours on health (FEIV) for White- and Blue-Collar Workers.

#### Panel A: White-collar

| VARIABLES          | 1st stage (FE) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) |
|--------------------|---------------|------------------|------------------|------------------|------------------|------------------|------------------|
|                    | (5.A1)        | (5.A2)           | (5.B1)           | (5.B2)           | (5.B3)           | (5.B4)           | (5.B5)           |
| Working hours      |               |                  |                  |                  |                  |                  |                  |
| Working hours-squared |             |                  |                  |                  |                  |                  |                  |
| Physical functioning |               |                  |                  |                  |                  |                  |                  |
| Bodily pain        |               |                  |                  |                  |                  |                  |                  |
| General health     |               |                  |                  |                  |                  |                  |                  |
| Vitality           |               |                  |                  |                  |                  |                  |                  |
| Mental health      |               |                  |                  |                  |                  |                  |                  |

#### Age difference 1
-2.585***
[0.243] -127.806***
[16,954]

#### Age difference 2
-3.167***
[0.278] -171.299***
[18,901]

Total sample size: 2,572

#### Test of the exclusion restrictions
17.63***
10.96***

#### Number of individuals
2,572

#### Sample size
14,444

#### Test of the exclusion restrictions
10% critical values

#### Hausman test for fixed effects
0.67
4.50
10.01
0.83
0.77

#### Test of the zero restrictions
53.82***
45.76***
54.47***
99.64***
108.50***
33.22***
25.09***

#### Test of the exclusion restrictions
65.16***
43.66***

#### Number of individuals
3,252

#### Sample size
19,941

#### Test of the exclusion restrictions
90.678***

#### Cragg-Donald Wald F statistic for weak instruments
22.84

#### Stock-Yogo weak ID test 10% critical values
7.03

#### Panel B: Blue-collar

| VARIABLES          | 1st stage (FE) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) | 2nd stage (FEIV) |
|--------------------|---------------|------------------|------------------|------------------|------------------|------------------|------------------|
|                    | (5.A1)        | (5.A2)           | (5.B1)           | (5.B2)           | (5.B3)           | (5.B4)           | (5.B5)           |
| Working hours      |               |                  |                  |                  |                  |                  |                  |
| Working hours-squared |             |                  |                  |                  |                  |                  |                  |
| Physical functioning |               |                  |                  |                  |                  |                  |                  |
| Bodily pain        |               |                  |                  |                  |                  |                  |                  |
| General health     |               |                  |                  |                  |                  |                  |                  |
| Vitality           |               |                  |                  |                  |                  |                  |                  |
| Mental health      |               |                  |                  |                  |                  |                  |                  |

#### Age difference 1
-1.409***
[0.276] -57.827***
[19,654]

#### Age difference 2
-1.803***
[0.307] -82.949***
[21,159]

Total sample size: 3,252

#### Test of the exclusion restrictions
65.16***

#### Test of the zero restrictions
53.82***
45.76***
54.47***
99.64***
108.50***
33.22***
25.09***

#### Test of the exclusion restrictions
65.16***
43.66***

#### Number of individuals
19,941

#### Sample size
19,941

#### Test of the exclusion restrictions
82.468***
84.238***

#### Cragg-Donald Wald F statistic for weak instruments
22.84

#### Stock-Yogo weak ID test 10% critical values
7.03

(continued on next page)
Table 5 (continued)

Panel A: White-collar

| VARIABLES          | 1st stage (FE) | 2nd stage (FE IV) |
|--------------------|----------------|-------------------|
|                    | (5.A1)         | (5.A2)            |
| Working hours      | 22.59***       | 16.88***          |
| Working hours-
  squared          |                |                   |
| Physical functioning| 45.35***      | 39.04***          |
| Bodily pain        | 7.11           | 12.10             |
| General health     | 13.31          | 3.98              |
| Vitality           | 10.08***       | 32.64***          |
| Mental health      | 13.28***       |                   |

Notes.
1) The first stage equations have been estimated using an individual fixed effects (FE) estimator and the second stage equations have been estimated using an individual fixed effect instrumental variable (FE IV) estimator. This Table differs from Table 3 because Panel A (Panel B) excludes all respondents whose job at the first interview of the survey was a blue-collar job (a white-collar job). For those people who were unemployed or retired at the time of the first interview, we based the division on their last reported job.
2) Notes 2-8 for Table 3.

status, we consider the following model for health outcomes:

\[ y_{it} = \alpha_1 + \alpha_2 W_{i,t-1} + X_{it}\beta + u_{it}, i = 1, \ldots, N, t = 1, \ldots, T; \]

(1a)

\[ u_{it} = \mu_i + \epsilon_{it}, \]

(1b)

where \( y_{it} \) denotes various health outcomes (physical functioning, bodily pain, general health, vitality, and mental health) for individual \( i \) at the time of the survey \( t \), and \( W_{i,t-1} \) is working hours. In estimating equation (1a), we include those individuals whose working hours are zero. \( X_{it} \) denotes a vector of time variant control variables: a spouse 0–1 dummy variable, \( Married \), which takes the value one if the respondent has a spouse and zero otherwise; the number of dependent children, \( Number \ of \ dependent \ children \); the respondent’s age, \( Age \), which controls for age-related effects; and a house ownership 0–1 dummy variable, \( Own\-house \), which indicates whether the respondent owns or is in the process of owning his house as a proxy for assets. The variables related to the respondent’s marital status and the number of dependent children are included because communication and interaction with other family members may prevent declines in health, particularly in mental health. In addition, the number of dependent children is included since it can be argued that people with dependent children may be likely to invest more in their health capital. The house ownership dummy is included to control for the effects of assets holdings on health. In addition, \( X_{it} \) includes three 0–1 regional dummies. \( N \) is the number of individuals and \( T \) is the number of observations available for individual \( i \) indicating that we have an unbalanced panel. As equation (1b) indicates, \( u_{it} \) is an error term which consists of a time invariant individual fixed effect, \( \mu_i \), and an idiosyncratic error, \( \epsilon_{it} \), so that we allow for some degree of individual heterogeneity. The coefficients \( \alpha_1 \) and \( \alpha_2 \) in equation (1a) capture the non-linear effect of working hours on a health outcome. Given the discussion in section 1 that some work is better than no work, and that too much work may be worse than some work, it is expected that \( \alpha_2 < 0 \) and \( \alpha_1 > 0 \). Holding everything else constant, it is easy to see that the value of a health score is maximized when \( WH_{i,t} = -\alpha_1/(2\alpha_2) \), and that for \( WH_{i,t} = -\alpha_1/\alpha_2 \) the level of health is the same as it would be if the respondent is not working.

The possibility of the endogeneity of the respondents’ working hours in equation (1a) is a major obstacle to estimate the causal impact of working hours on health. As discussed in section 1, this particular identification problem is called the ‘healthy worker effect’, that is healthy individuals are more likely to be employed and to work longer whereas unhealthy workers may decide to leave the workforce or work short hours. Individuals, who are healthier and, therefore, tend to earn a relatively higher wage, could decide to reduce their hours of work. The same logic can be applied to health outcomes.

Since one of the purposes of this paper is to make causal statements about the impact of working hours on health outcomes and to rule out the possibility of reverse causality from health outcomes to working hours, in estimating the relationship between working hours and health outcomes, our modelling approach has attempted to take account of both the potential endogeneity of working hours and the presence of individual effects, and this leads us to choose a fixed effect instrumental variable estimator.

In order to account for the potential endogeneity of working hours and its square in equation (1a), we use an instrumental variable estimator using two instruments, \( Age \ difference 1 \) and \( Age \ difference 2 \), that are based on pension eligibility ages. These instruments measure the distance from an individual’s retirement age. More specifically, \( Age \ difference 1 \) is the difference between a respondent’s age and the respondent’s pension eligibility age that the respondent has reached the eligibility age and zero otherwise, while \( Age \ difference 2 \) is the difference between a respondent’s age and the respondent’s pension eligibility age provided the respondent has not reached the eligibility age and zero otherwise. That is, we assume with the following models:

\[ WH_{i,1} = \mu_1 Age \ difference \ 1_{it} + \mu_2 Age \ difference 2_{it} + X_{it}\mu\ + w_{it}, \]

(2)

\[ WH_{i,2} = \mu_3 Age \ difference \ 1_{it} + \mu_4 Age \ difference 2_{it} + X_{it}\mu\ + w_{it}, \]

(3)

where the error terms, \( w_{it} \) and \( w_{it0} \), include individual fixed effects.

Given the endogeneity of working hours, to ensure the consistency of the instrumental variable estimator of equation (1), the two instruments, \( Age \ difference 1 \) and \( Age \ difference 2 \), must be uncorrelated with the error term in equation (1) and related to the endogenous variables in equation (1). The idea here is to treat the policy changes in the age of pension eligibility discussed in section 2 as a ‘natural experiment.’ Variation in these two instruments comes from variation in the age of respondents and policy changes in the pension eligibility age. Since the pension eligibility age is policy determined, these instruments are not likely to be correlated with the error term in equation (1). We assume that the policy change only affects health outcomes through changes in working hours.

As can be seen from Table 1, in our sample, a reasonable amount of variation in the eligibility age is observed. Although the pension
Table 6
The impacts of working hours on health (including Time dummies).

| VARIABLES                      | 1st Stage (FE) | 2nd stage (FEIV) |
|--------------------------------|---------------|-----------------|
|                                | (6.A1)        | (6.A2)          | (6.B1) | (6.B2) | (6.B3) | (6.B4) | (6.B5) |
| Working hours                  |               |                 | 5.557*** | 3.095*** | 2.479*** | 4.000*** | 1.549*** |
| Age difference 1               | -3.115***     | [0.237]         | [1.102]  | [0.741]  | [0.622]  | [0.840]  | [0.473]  |
|                                | [15.816]      |                 | [0.014]  | [0.011]  | [0.015]  | [0.009]  |           |
| Working hours-squared          | -0.106***     | [0.020]         | -0.034*** | -0.048*** | -0.079*** | -0.168*** | -0.127*** |
| Age                            | 3.786***      | [0.460]         | 2.475*   | 1.372    | 1.500**  | 2.094**  | 0.901    |
|                                | [32.244]      | [1.397]         | [0.888]  | [0.745]  | [1.065]  | [0.569]  |           |
| Married                        | 0.991*        | [0.995]         | 1.668    | -0.149   | -0.098   | 1.329    | 3.073**  |
|                                | [71.385]      | [1.815]         | [1.180]  | [0.996]  | [1.382]  | [0.791]  |           |
| Number of dependent children   | -0.513**      | [0.249]         | 1.912*   | 1.080*   | 0.802    | 0.940    | 0.224    |
|                                | [69.020]      | [17.489]        | [0.948]  | [0.508]  | [0.731]  | [0.394]  |           |
| Ownhouse                       | -0.470        | [0.514]         | -1.196   | -0.767   | -0.988   | 1.255    | -0.297   |
|                                | [34.598]      | [1.560]         | [0.794]  | [0.830]  | [1.167]  | [0.500]  |           |
| Remote/Very remote             | -0.345**      | [1.153]         | -0.736   | 0.179    | -0.153   | -0.022   | 0.335    |
|                                | [26.711]      | [69.020]        | [1.663]  | [1.421]  | [2.061]  | [1.083]  |           |
| wave2                          | -1.217**      | [0.423]         | -3.956** | -2.403** | -2.984*** | -3.575*** | -1.333**  |
|                                | [30.657]      | [1.677]         | [1.103]  | [0.901]  | [1.265]  | [0.690]  |           |
| wave3                          | -2.278**      | [0.761]         | -7.881***| -5.474***| -5.646***| -6.713***| -2.965**  |
|                                | [54.713]      | [2.995]         | [1.931]  | [1.666]  | [2.272]  | [1.215]  |           |
| wave4                          | -3.591**      | [1.119]         | -11.349**| -7.488** | -7.931***| -9.540***| -4.187**  |
|                                | [194.129]     | [80.153]        | [4.429]  | [2.836]  | [2.379]  | [3.350]  | [1.791]  |
| wave5                          | -4.290**      | [1.483]         | -13.908**| -9.403** | -9.619** | -11.569**| -4.862**  |
|                                | [105.901]     | [5.683]         | [3.632]  | [3.048]  | [4.308]  | [2.303]  |           |
| wave6                          | -5.513***     | [1.838]         | -17.550**| -11.596**| -11.675**| -14.176**| -6.157**  |
|                                | [130.826]     | [7.068]         | [4.526]  | [3.788]  | [5.366]  | [2.879]  |           |
| wave7                          | -6.688**      | [2.068]         | -20.972**| -13.586**| -13.993**| -16.869**| -7.071**  |
|                                | [157.341]     | [8.482]         | [5.445]  | [4.549]  | [6.432]  | [3.438]  |           |
| wave8                          | -7.650**      | [2.564]         | -24.007**| -15.333**| -16.192**| -19.786**| -8.314**  |
|                                | [192.995]     | [9.861]         | [6.338]  | [5.087]  | [7.473]  | [4.904]  |           |
| wave9                          | -12.302***    | [2.950]         | -27.696**| -17.103**| -18.011**| -22.325**| -9.323**  |
|                                | [209.992]     | [11.229]        | [7.200]  | [6.027]  | [8.514]  | [4.570]  |           |
| wave10                         | -13.062**     | [3.299]         | -30.226**| -19.573**| -20.714**| -24.313**| -10.363** |
|                                | [235.027]     | [12.483]        | [7.990]  | [6.686]  | [9.452]  | [5.073]  |           |
| wave11                         | -14.337**     | [3.647]         | -33.004**| -21.106**| -22.949**| -26.649**| -11.517** |
|                                | [260.115]     | [13.837]        | [8.856]  | [7.420]  | [10.477] | [5.625]  |           |
| wave12                         | -15.612**     | [4.011]         | -36.452**| -23.217**| -24.929**| -28.695**| -12.460** |
|                                | [286.590]     | [15.248]        | [9.760]  | [8.178]  | [11.555] | [6.194]  |           |
| Constant                       | -57.986       | [73.982]        | -4.773   | -12.164  | -50.383  | 29.486   |           |
|                                | [47.032]      | [39.405]        | [5.648]  | [50.188] |           |           |           |

Note. 1) Notes 1-8 for Table 3.

Eligibility age for men remained at the age of 65 for individuals born before or on 30 June 1952, the proportion of this group in Wave 9 (just after the 2009 pension reform) is just 48.1%. The proportion of men whose pension eligibility age is 67 in Wave 9 consists of 41.8%.

The two instruments in equations (2) and (3) are closely related to one of the standard instruments used for the analysis of the causal relationship between retirement and cognitive functioning (Atalay et al. 2019; Bonsang et al. 2012; Coe & Tamaro, 2011; Mazzonna and Peracchi 2012, 2017; Rohwedder & Willis, 2010). It is worth pointing out that, in the literature that tries to determine the causal impacts of retirement on health outcomes, these pension related instruments are used to explain the retirement decision, that is, the extensive margin, not the number of hours worked.

Here we rely on Neumark and Powers (2000, 2003/2004, 2005, 2006), Kudrna and Woodland (2011a, 2011b) and Vere (2011) to justify using these pension eligibility age related variables as instruments for also explaining the intensive margin of working hours as well, that is, the instruments are related to the endogenous variables. Neumark and Powers (2000, 2003/2004, 2005, 2006) provide a theoretical explanation of the connection between social security in the
Table 7
Robustness check for FEIV (Using the vacancy rate as an additional instrument).

| VARIABLES                  | 1st stage (FE) (7.A1) | 1st stage (FE) (7.A2) | 2nd stage (FEIV) (7.B1) | 2nd stage (FEIV) (7.B2) | 2nd stage (FEIV) (7.B3) | 2nd stage (FEIV) (7.B4) | 2nd stage (FEIV) (7.B5) |
|----------------------------|-----------------------|-----------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|                            | Working hours         | Working hours-squared | Physical functioning    | Bodily pain             | General health          | Vitality                | Mental health           |
| Age difference 1           | -2.271***             | -108.394***           | 5.698***                | 3.298***                | 2.592***                | 4.216***                | 1.726***                |
|                           | [0.190]               | [13.262]              | [1.156]                 | [0.779]                 | [0.655]                 | [0.887]                 | [0.497]                 |
| Age difference 2           | -2.811***             | -146.499***           | 0.108***                | -0.061***               | -0.050***               | -0.081***               | -0.035***               |
|                           | [0.217]               | [14.704]              | [0.021]                 | [0.014]                 | [0.012]                 | [0.016]                 | [0.009]                 |
| Vacancy rate               | 1.162***              | 57.457***             | -0.747***               | -0.612***               | -0.680***               | -0.407***               | -0.155                  |
|                           | [0.207]               | [17.755]              | [0.175]                 | [0.152]                 | [0.128]                 | [0.185]                 | [0.101]                 |
| Working hours              |                      |                      | 5.698***                | 3.298***                | 2.592***                | 4.216***                | 1.726***                |
|                           |                      |                      | [1.156]                 | [0.779]                 | [0.655]                 | [0.887]                 | [0.497]                 |
| Working hours-squared      |                      |                      | -0.108***               | -0.061***               | -0.050***               | -0.081***               | -0.035***               |
|                           |                      |                      | [0.021]                 | [0.014]                 | [0.012]                 | [0.016]                 | [0.009]                 |
| Age                       | 1.396***              | 86.064***             | 0.747***                | -0.612***               | -0.680***               | -0.407***               | -0.155                  |
|                           | [0.175]               | [12.586]              | [0.239]                 | [0.152]                 | [0.128]                 | [0.185]                 | [0.101]                 |
| Married                    | 0.913                 | 67.148***             | 0.576                   | 0.521                   | 0.291                   | 0.598                   | 0.598                   |
|                           | [0.383]               | [39.103]              | [0.239]                 | [0.152]                 | [0.128]                 | [0.185]                 | [0.101]                 |
| Number of dependent children| -0.568***             | -13.920               | 1.994**                 | 1.185*                  | 0.867*                  | 1.057                   | 0.315                   |
|                           | [0.250]               | [17.526]              | [0.968]                 | [0.610]                 | [0.522]                 | [0.749]                 | [0.405]                 |
| Ownhouse                   | -0.426                | -37.539               | -1.167                  | -0.747                  | -0.998                  | -1.237                  | -0.308                  |
|                           | [0.517]               | [34.671]              | [1.568]                 | [0.982]                 | [0.839]                 | [1.179]                 | [0.649]                 |
| Inner regional             | -4.969***             | -216.720***           | -0.576                  | 0.521                   | 0.291                   | 0.598                   | 0.598                   |
|                           | [1.168]               | [69.631]              | [2.767]                 | [1.717]                 | [1.462]                 | [2.123]                 | [1.117]                 |
| Outer regional             | -2.677                | -101.792              | 0.312                   | 2.526                   | 0.491                   | 2.711                   | 2.038                   |
|                           | [1.644]               | [99.596]              | [3.887]                 | [2.418]                 | [2.108]                 | [2.857]                 | [1.608]                 |
| Remote/Very remote         | 4.365                 | 338.165***            | 10.863                  | 7.066                   | 3.870                   | 7.297                   | 4.540                   |
|                           | [3.289]               | [194.474]             | [8.063]                 | [4.855]                 | [4.252]                 | [6.347]                 | [3.264]                 |
| Constant                  | -76.943***            | -3999.273***          | 104.401***              | 92.883***               | 97.546***               | 73.531***               | 80.804***               |
|                           | [11.898]              | [846.115]             | [37.539]                | [21.118]                | [11.182]                | [16.205]                | [8.809]                 |
| Sample size               | 34.385                | 34.385                | 34.385                  | 34.385                  | 34.385                  | 34.385                  | 34.385                  |
| Number of individuals      | 5,824                 | 5,824                 | 5,824                   | 5,824                   | 5,824                   | 5,824                   | 5,824                   |
| Test of the exclusion restrictions | 56.26***         | 37.45***              | 92.28***                | 126.66***               | 163.28***               | 54.08***                | 33.41***                |
| Hansen J test             | 0.01                  | 0.01                  | 0.01                    | 0.31                    | 0.01                    | 2.05                    | 18.41                   |
| Cragg-Donald Wald F statistic for weak instruments | 13.43 | 13.43 | 13.43 | 13.43 | 13.43 | 13.43 | 13.43 |

Notes.
1) The first stage equations have been estimated using an individual fixed effects (FE) estimator and the second stage equations have been estimated using an individual fixed effect instrumental variable (FEIV) estimator. This Table differs from Table 3 as it uses an additional instrument, the vacancy rate.
2) The Hansen J test is an overidentification test, and in this case has one degree of freedom.
3) Notes 2-8 for Table 3.

United States and hours worked prior to retirement, while Vere (2011) provides some empirical evidence for the US on the connection between pension payments and hours worked. In the context of a dynamic general equilibrium model, Kudrna and Woodland (2011a, 2011b) provide evidence for Australia on the effect of the major reform of the Australian pension system that was announced by the Australian Government in the 2009/2010 national budget that we are considering on working hours. Furthermore, in the data set we are using for individuals who report their working hours are positive, Fig. 1 presents the relationship between the average of working hours and the difference between a respondent’s age and pension eligibility age. For values of this difference less than 10, there would appear to be a negative relationship between this variable and the average of working hours. One reason for limiting our sample to males aged 40 and over is to ensure that there is a sufficient connection between the changes in pension eligibility ages and working hours.

In both equations (2) and (3), the dependent variable is non-negative, but Wooldridge (2010, p. 90) makes the important point that the application of the instrumental variable (IV) estimator is not limited to the case where the dependent variable(s) in the first stage is (are) continuous, so even though working hours are non-negative this is not an impediment to the application of the IV technique.

5. Estimation results
All regression results reported in this section are estimated using STATA version 17 (StataCorp., 2021).

5.1. Estimation using a fixed effect instrumental variable (FEIV) estimator
Table 3 presents the estimates for the equations in both the first and second stages of the fixed effect instrumental variable (FEIV) estimator. In stage 2 (columns (3.B1)–(3.B5)), where the fixed effect instrumental variable estimates of equation (1) are presented, it is observed that for each health variable both working hours and its square are individually statistically significant with the estimated coefficients have signs that are consistent with the a priori expectations suggested in section 3, so that there is an inverted U shape relationship between health and working hours. The endogeneity tests in Table 3 which test the joint significance of two residuals added to the model estimated by FE (see Wooldridge, 2010, pp. 352–354) clearly reject the null hypothesis that Working hours and Working hours squared are exogenously determined. The Cragg and Donald (1993) test indicates we do not have a problem of “weak” instruments. Finally, the Hausman (1978) tests for choosing between the pooled IV model and FEIV indicate clearly that the FEIV
The impacts of working hours on health (FEIV) (Only respondents whose Age minus Pensionable Age is less than or equal to 18 in absolute value).

| VARIABLES | 1st stage (FE) |  | 2nd stage (FEIV) |  |  |  |  |  |
|-----------|---------------|---|-----------------|---|---|---|---|---|
|           | (8.A1)        | (8.A2) | Physical functioning | (8.B1) | (8.B2) | (8.B3) | (8.B4) | (8.B5) |
|           | Working hours | Working hours-squared | | | | | | |
| Age difference 1 | -1.736*** | -83.419*** | | | | | | |
| | [0.207] | [18.870] | | | | | | |
| Age difference 2 | -2.575*** | -137.589*** | | | | | | |
| | [0.289] | [20.095] | | | | | | |
| Working hours | | | | 5.748*** | 3.474** | 2.828** | 5.006*** | 1.506* |
| | | | | [2.091] | [1.428] | [1.191] | [1.780] | [0.838] |
| Working hours-squared | | | | -0.107*** | -0.063*** | -0.053*** | -0.090*** | -0.031*** |
| | | | | [0.034] | [0.024] | [0.020] | [0.029] | [0.014] |
| Age | 0.908*** | 45.583** | | | | | | |
| | [0.259] | [18.327] | | | | | | |
| Married | 0.942 | 43.898 | | | | | | |
| | [0.797] | [50.900] | | | | | | |
| Number of dependent children | -0.811** | -15.077 | | | | | | |
| | [0.392] | [26.697] | | | | | | |
| Ownhouse | -1.097 | -77.320* | | | | | | |
| | [0.712] | [45.922] | | | | | | |
| Inner regional | -4.969*** | -245.396*** | | | | | | |
| | [1.140] | [81.142] | | | | | | |
| Outer regional | -2.449 | -75.423 | | | | | | |
| | [2.600] | [120.494] | | | | | | |
| Remote/Very remote | 3.819 | 331.999 | | | | | | |
| | [4.446] | [260.920] | | | | | | |
| Constant | -42.680** | -2278.548* | | | | | | |
| | [17.004] | [1217.442] | | | | | | |
| Number of individuals | 24.731 | 24.731 | | | | | | |
| | 4.231 | 4.231 | | | | | | |
| Test of the exclusion restrictions | 54.249*** | 50.16*** | | | | | | |
| Test of the zero restrictions | 77.90*** | 63.13*** | | | | | | |
| Hausman test for fixed effects | 15.42* | 47.45** | 67.41*** | 14.74* | 32.94*** |
| | [0.712] | [125.494] | [11.176] | [6.883] | [6.031] | [10.079] | [6.945] |
| Endogeneity test (F-statistic) | 10.72 | | | | | | | |
| Cragg-Donald Wald F statistic for weak instruments | 7.03 | | | | | | | |
| Stock-Yogo weak ID test 10% critical values | | | | | | | | |
| | | | | | | | | |
| Notes. | 1) Notes 1–8 for Table 3. |

10

S. Kajitani et al.

SSM - Population Health 20 (2022) 101245

Table 8

In this sub-section, we present the estimated results of heterogeneity and robustness checks. The heterogeneity check examines whether there is any heterogeneity between blue-collar and white-collar workers. Moreover, we conduct three types of robustness checks: 1) adding time dummies to the estimated equations; 2) adding another instrument to conduct an overidentification test; and 3) restricting the sample to respondents whose (Age – Aged Pension Eligibility Age) is less than or equal to 18 in absolute value.
equal to 18 in absolute value.

Kajitani, Sakata, and McKenzie (2017) show that the job tasks of a worker’s longest served job affects cognitive functioning after retirement. This may also be the case for the effects of working hours on health. The effects of working hours on health vary depending on types of occupation. In order to check for occupational heterogeneity in the response of health outcomes to working hours, we divide the sample into “blue-collar” workers and “white-collar” workers depending on their reported job types in Wave 1. For those people who were unemployed or retired at the time of Wave 1, we based the division on their last reported job. We define white-collar jobs as jobs that fall into the following categories: Manager; professional; community and personal service work; clerical and administrative worker; or sales worker. We define blue-collar jobs as jobs that fall into the following categories: Technicians and trades workers; machinery operators and drivers; or labourers. Panels A and B in Table 5 report the results for “white-collar” and “blue-collar” workers, respectively. Both Working hour and Working hour squared are statistically significant in all cases. In order to interpret these estimated results, we plot how the effects of working hours on health change according to the number of working hours in Fig. 2. The peaks appear to arrive slightly earlier and to be slightly higher for blue-collar workers than white-collar workers. In Panels B and C of Table 4, estimates of the number of hours worked when each of the health measures peak are presented for white-collar and blue-collar workers, respectively. However, we could not reject the null hypothesis that the hours at which a peak occurs for blue-collar and for white-collar workers are the same for each health measure. This suggests that the impact of working hours on health outcomes for blue-collar and white-collar workers are similar. Since health outcomes may improve over time due to improvements in the health system, the first robustness check adds time dummies for Waves 2–13 so that Wave 1 is the base year. The results are reported in Table 6. For all the health measures, the estimated coefficients on all the time dummies are negative and significant, while the estimated coefficients on Age are positive and sometimes significant. This suggests it is difficult to separate the effects of aging from the impacts of improvements in the health system. Working hours have an inverted U shape for health outcomes. The peaks appear to arrive slightly earlier and to be slightly higher for blue-collar workers than white-collar workers. In Panels B and C of Table 4, estimates of the number of hours worked when each of the health measures peak are presented for white-collar and blue-collar workers, respectively. However, we could not reject the null hypothesis that the hours at which a peak occurs for blue-collar and for white-collar workers are the same for each health measure. This suggests that the impact of working hours on health outcomes for blue-collar and white-collar workers are similar. The results reported to date are based on exactly identified models, that is the number of explanatory variables and the number of instruments are the same. Using an additional instrument enables us to check the model’s specification through an overidentification test. The results of this second robustness check are reported in Table 7, where in addition to Age difference 1 and Age difference 2 the vacancy rate is also used as an instrument when estimating equation (1a) by FEIV. The vacancy rate shows both geographical (differing across each state) and temporal variation, and since it provides a measure of macroeconomic conditions at the state level is likely to affect the hours worked. On the other hand, it is unlikely to directly affect any of the health scores of individuals and so it is an appropriate instrument. As can be seen in columns (7.B1)–(7.B5), Working hours and Working hours squared are individually highly significant in all the equations and have estimated coefficients with the expected signs. All five health score equations pass the overidentifying test, so we can have some confidence in the specification of the model. In the first stage (columns (7.A1)–(7.A2)), the vacancy rate is highly significant in both equations. The peaks of the health measures occur at around 26 h for physical functioning, general health and vitality, 27 h for role physical, and 25 h for mental health.

The final robustness check involves restricting the sample to respondents whose (Age – Aged Pension Eligibility Age) of more than 18 years is likely to exclude people who have already stopped working due to retirement and whose average health status is lower than the rest of the sample. Men with an (Age – Aged Pension Eligibility Age) of less than -18 years, individuals aged 40–44 in 2001, are a long way from their pension eligibility age, so changes in their pension eligibility age may have little impact on their working hours. The results in Table 8 are again consistent with the results in Table 3, so that making these exclusions do not undermine our results. The peaks of the health measures occur at around 27 h for physical functioning, general health and vitality, 28 h for role physical, and 24 h for mental health.

Two additional robustness checks restricting the sample to men aged between 45 and 75 and estimating the model using a three-step IV estimator are reported in Tables O2 and O3 of the Online Supplementary Material, respectively. The results for these two cases are qualitatively consistent with those reported in Table 3.

6. Concluding remarks

We examine the causal impact of working hours on the self-assessed health status of middle-aged and elderly males living in Australia using longitudinal data from the Household Income and Labour Dynamics in Australia (HILDA) Survey. The literature in this area is very limited in that it does not consider a non-linearity in the effect of working hours on health. Many previous studies examine the ‘use it or lose it’ hypothesis, which tests whether or not retirement (not using your brain) leads to losses of cognitive functioning. On the other hand, we also examine the relevance of this hypothesis for a broader set of health outcomes. In addition, we focus on the ‘use it too much and lose everything’ hypothesis, which refers to the situation where working too much can lead to not just a loss of cognitive functioning (see Kajitani, McKenzie, & Sakata, 2017), but also declines in health status across the board. This study is unique in that we focus on not only labor market participation (the extensive margin) but also the intensive margin of work (working hours) and that we determine the optimal working hours for middle aged and elderly male workers in terms of maximizing their health status.

Using five measures of self-assessed health status in the SF-36, it is found that for working hours up to 24–27 h per week increases in working hours have a positive impact on cognition for males depending on the health measure. After that, working hours have a negative impact on health status. Compared to males who do not work, working hours over 48–54 h will lead to worse health outcomes depending on the measure. This indicates that the differences in working hours are an important factor in explaining differences in the health outcomes of middle aged and elderly male adults.

Thus, in middle and old age, adopting part-time work as a pattern of work could be effective in maintaining/improving the health status of individuals compared to when they do not work. Previous studies on retirement and cognitive functioning indicate that increasing the qualifying age for a pension can not only reduce the government social security expenditures but can potentially reduce the risk of cognitive deterioration. However, our study highlights that raising the qualifying age for a pension can reduce the risk of health deterioration, but that too much work can have quite adverse effects on health status.

It is worth noting that the health measures peak at levels of working hours that are approximately 10 h lower than the current minimum level of 35 h set for full-time work; that is, full time workers must work more than is optimal for their health outcomes. This paper does not provide any information on how individuals might trade off health outcomes for higher wages or whether firms might offer higher wages to induce individuals to work more than the hours that optimize their health outcomes. This additional information is needed to formulate any policy responses here.

Two important areas for future research are whether these results reported here for Australian males can be found for females and workers.
in other countries. Throughout this paper, health is assumed to be a quadratic function of working hours. It is, of course, possible that a more complicated functional form is appropriate. In a parametric context, estimating a more complicated function would require additional instruments, while assuming a non-parametric form for the relationship would probably require assuming that working hours are determined exogenously.

**Funding**

This research was supported financially by three grants from the Japan Society for the Promotion of Science (JSPS) (Grants in Aid for Scientific Research (B) No. JP24330093, No. 16H03607 and 21H00725) and a grant from Kyoto Sangyo University (Grant No. E1901).

**Code availability**

The STATA code we have used to analyse the HILDA data can be obtained by contacting the Corresponding Author.

**Approval of ethics review board**

The data used in this paper has been collected by the Melbourne Institute: Applied and Social Economic Research within Melbourne University. Approval for collecting the data been obtained from the Human Research Ethics Committee of the University of Melbourne.

**Author statement**

All three authors have been involved in the development and writing of this paper. All the empirical results in this paper have been obtained by Shinya Kajitani.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability**

The authors do not have permission to share data.

**Acknowledgements**

The authors would like to thank three anonymous referees for their extremely helpful and constructive comments on an earlier version of the paper. The authors would also like to thank Daiji Kawaguchi and participants at the following workshops and conferences, the Ageing and Longevity in the UK and Japan: Perspectives from the Health and Social Sciences Workshop, the Globalization and the Building of a High Quality Economic System Conference, the 10th International Conference of the Thailand Econometric Society 2017, the 22nd Euroasia and Business and Economics Society Conference, the 2019 Gerontological Society of America Annual Scientific Meeting, and the Spring Conference of the Japanese Economic Association for their helpful and constructive comments on earlier versions of this paper. This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute: Applied Economic and Social Research, the University of Melbourne (the Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to DSS, the Melbourne Institute, or the Australian Institute of Family Studies. The authors wish to acknowledge that this research was supported by the Japan Society for the Promotion of Science (JSPS) Grant in Aid for Scientific Research (B) No. JP24330093 for a project on “Retirement Behaviour of the Aged and their Cognitive Ability and Health” (P.I.: Kei Sakata), a Japan Society for the Promotion of Science (JSPS) Grant in Aid for Scientific Research (B) No. 16H03607 for a project “An Empirical Analysis of Parental Employment Status, Time Allocation and Way of Thinking, and their Children’s Human Capital Formation” (Project Leader: Midori Wakabayashi), a Japan Society for the Promotion of Science (JSPS) Grant in Aid for Scientific Research (B) No. 21H00725 for a project on “Time Allocation, Intra-Household Productive Activities and Health Capital Formation” (P.I.: Shinya Kajitani), and Kyoto Sangyo University Research Grant No. E1901 (P.I.: Shinya Kajitani).

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at [https://doi.org/10.1016/j.ssmph.2022.101245](https://doi.org/10.1016/j.ssmph.2022.101245).

**Appendix I. Definitions of variables**

| Name                          | Definition                                                                 |
|-------------------------------|---------------------------------------------------------------------------|
| Physical functioning          | The SF-36 physical functioning score (0–100)                             |
| Bodily pain                   | The SF-36 bodily score (0–100)                                           |
| General health                | The SF-36 general health score (0–100)                                  |
| Vitality                      | The SF-36 vitality score (0–100)                                         |
| Mental health                 | The SF-36 mental health score (0–100)                                   |
| Working hours                 | The number of usual or average working hours per week the respondent works. |
| Working hours-squared         | (Working hours)                                                          |
| Age                           | Respondent’s age in years at the time of the survey                      |
| School years 7–10 (benchmark: the respondent’s highest years of school completed are under 7) | 0-1 dummy variable taking the value of unity if the respondent’s highest years of school completed are between 7 and 10, and 0 otherwise. |
| School years 11 and over (benchmark: the respondent’s highest years of school completed are under 7) | 0-1 dummy variable taking the value of unity if the respondent’s highest years of school completed are 11 and over, and 0 otherwise. |
| Married                       | The number of respondents’ children who reside with him/her and are either are aged under 15 years or are aged 16–24 years and are enrolled in full-time education. |
| Number of dependent children  | 0-1 dummy variable taking the value of unity if the respondent owns his/her own house or is currently paying off a mortgage, and 0 otherwise. |
| Ownhouse                      | 0-1 dummy variable taking the value unity if the respondent lives in inner regional Australia, and 0 otherwise. |
| Inner regional                | 0-1 dummy variable taking the value unity if the respondent lives in outer regional Australia, and 0 otherwise. |
| Outer regional                | (continued on next page)
Remote/Very remote
0-1 dummy variable taking the value unity if the respondent lives in remote or very remote Australia, and 0 otherwise.

Age difference 1
= (Respondent’s age in years at the time of the survey) - (Aged pension eligibility age) if the respondent has not reached Aged pension eligibility age at the time of the survey, = 0 otherwise

Age difference 2
= (Respondent’s age in years at the time of the survey) - (Aged pension eligibility age) if the respondent has not yet reached Aged pension eligibility age at the time of the survey, = 0 otherwise

Vacancy rate
(Job vacancyEmployed) 100 where Job vacancy denotes the number of job vacancies in the state where the respondent lives in each wave which are reported by the Australian Bureau of Statistics (ABS) and Employed denotes the number of total employed persons in the relevant state in November at each waves which are reported by the ABS. Noted that we use the values for May 2008 because of a lack of values for November 2008.

wave
0-1 dummy variable taking the value unity if the observation is for Wave j, and 0 otherwise (j = 2,...,13)

Blue-collar
0-1 dummy variable taking the value unity if the respondents job at the first interview of the survey for those unemployed or retired, their previous last job was a blue-collar job, and 0 otherwise. We define “Technicians and Trades Workers,” “Machinery Operators and Drivers,” or “Labourers” as a blue-collar job.

White-collar
0-1 dummy variable taking the value unity if the respondents job at the first interview of the survey for those unemployed or retired, their previous last job was a white-collar job, and 0 otherwise. We define “Manager,” “Professional,” “Community and Personal Service Work,” “Clerical and Administrative Worker,” or “Sales Worker” as a white-collar job.

References
Atalay, K., & Barrett, G. F. (2015). The impact of age pension eligibility age on retirement and program dependence: Evidence from an Australian experiment. The Review of Economics and Statistics, 97(1), 71-87. https://doi.org/10.1162/REST_a_00443
Atalay, K., Barrett, G. F., & Staneva, A. (2019). The effect of retirement on elderly cognitive functioning. Journal of Health Economics, 66, 37-53. https://doi.org/10.1016/j.jhealeco.2019.04.006
Australian Bureau of Statistics (ABS). (2011). $30.0 - household expenditure survey, Australia: Summary of results, 2009-10.9 September 2011. Available from http://www.abs.gov.au/ausstats/abs@.nsf/mf/6056.0.65.001.
Bassanini, A., & Caroli, E. (2015). Is work bad for health? The role of constraint vs. Freedom. Journal of Labor Economics, 33(1), 44-86. https://doi.org/10.1086/681430
Coe, N. B., & Zamarro, G. (2011). Retirement effects on health in Europe. Journal of Health Economics, 30(3), 644-655. https://doi.org/10.1016/j.jhealeco.2010.08.005
Gutman, S. Kajitani et al. continued
Hausman, J. A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251-1271. https://doi.org/10.2307/1913827
Heller-Sahlgren, G. (2017). Retirement blue. Journal of Health Economics, 54(1), 66-78. https://doi.org/10.1016/j.jhealeco.2017.01.007
Kajitani, S., McKenzie, C., & Sakata, K. (2017). Use it too much and lose it? The effect of working hours on cognitive ability. Panel Data Research Center at Keio University. Discussion Paper no 2017-05
Kajitani, S., Sakata, K., & McKenzie, C. (2017). Occupation, retirement and cognitive functioning. Ageing and Society, 37(8), 1568-1596. https://doi.org/10.1017/S0144666716004653
Kudrna, G., & Woodland, A. D. (2011a). Implications of the 2009 age pension reform in Australia: A dynamic general equilibrium analysis. The Economic Record, 87, 183-201. https://doi.org/10.1111/j.1475-4932.2010.00703.x
Kudrna, G., & Woodland, A. D. (2011b). An inter-temporal general equilibrium analysis of the Australian age pension means test. Journal of Macroeconomics, 33(1), 61-79. https://doi.org/10.1016/j.jmacro.2010.09.006
Lee, J., & Lee, Y. K. (2016). Can working hour reduction save workers? Labour Economics, 40, 25-36. https://doi.org/10.1016/j.labeco.2016.02.004
Llena-Nozal, A. (2009). The effect of work status and working conditions on mental health in four OECD countries. National Institute Economic Review, 209(1), 72-87. https://doi.org/10.1177/0027950109345234
Llena-Nozal, A., Lindeboom, M., & Portrait, F. (2004). The effect of work on mental health: Does occupation matter? Health Economics, 13, 1045-1062. https://doi.org/10.1002/hec.929
Mazzonna, F., & Peracchi, F. (2012). Ageing, cognitive abilities and retirement. European Economic Review, 56(4), 691-710. https://doi.org/10.1016/j.euroec rev.2012.03.009
Mazzonna, F., & Peracchi, F. (2017). Unhealthy retirement? Journal of Human Resources, 52(1), 128-151. https://doi.org/10.3368/jhr.52.1.128-151
Neumark, D., & Powers, E. (2000). Welfare for the elderly: The effects of SSI on pre-retirement labor supply. Journal of Population Economics, 13(1), 44-79. https://doi.org/10.1007/s10198-000-0074-y
Neumark, D., & Powers, E. (2006). Supplemental security income, labor supply, and migration. Journal of Population Economics, 19(3), 447-479. https://doi.org/10.1007/s10198-006-0074-y
Nie, P., Otterbach, S., & Sousa-Poza, A. (2015). Long hours and health in China. China Economic Review, 35, 212-229. https://doi.org/10.1016/j.chieco.2015.02.004
Pencavel, J. (2014). The productivity of working hours. Economic Journal, 125, 2052-2076. 10.1111/ej.12166.
Pencavel, J. (2016). Recovery from work and the productivity of working hours. Economic, 62, 545-563. https://doi.org/10.1111/ece.12206
Robone, S., Jones, A., & Rice, N. (2011). Contractual conditions, working conditions and their impact on health and well-being. The European Journal of Health Economics, 12 (5), 429-444. https://doi.org/10.1007/s10198-010-0256-0
Rohwedder, S., & Willis, R. J. (2010). Mental retirement. The Journal of Economic Perspectives, 24(1), 119-138. https://doi.org/10.1257/jep.24.1.119
Schmitz, H. (2011). Why are the unemployed in worse health? The causal effect of unemployment on health. Labour Economics, 18(1), 71-79. https://doi.org/10.1016/j.labeco.2010.08.005
Sparks, K., Cooper, C., Fried, Y., & Shimon, A. (1997). The effects of hours of work on health: A meta-analytic review. Journal of Occupational and Organizational Psychology, 70(4), 391-408.
Spurgeon, A., Harrington, J. M., & Cooper, C. (1997). Health and safety problems associated with long working hours: A review of the current position. *Occupational and Environmental Medicine, 54*(6), 367–375. https://doi.org/10.1136/oem.54.6.367

StataCorp. (2021). *Stata statistical software: Release, 17*. College Station, TX: StataCorp LLC.

Vere, J. P. (2011). Social security and elderly labor supply: Evidence from the health and retirement study. *Labour Economics, 18*, 676–686. https://doi.org/10.1016/j.labeco.2011.02.001

Ware, J. D. (2000). SF-36 health survey update. *Spine, 25*(24), 3130–3139.

Watson, N., & Wooden, M. (2013). Adding a top-up sample to the household, income and labour dynamics in Australia survey. *The Australian Economic Review, 46*(4), 489–498. https://doi.org/10.1111/1467-8462.12027

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data (2nd ed.)*. Cambridge: The MIT Press.

Zhu, R. (2016). Retirement and its consequences for women’s health in Australia. *Social Science & Medicine, 163*, 117–125. https://doi.org/10.1016/j.socscimed.2016.04.003