



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DISIA

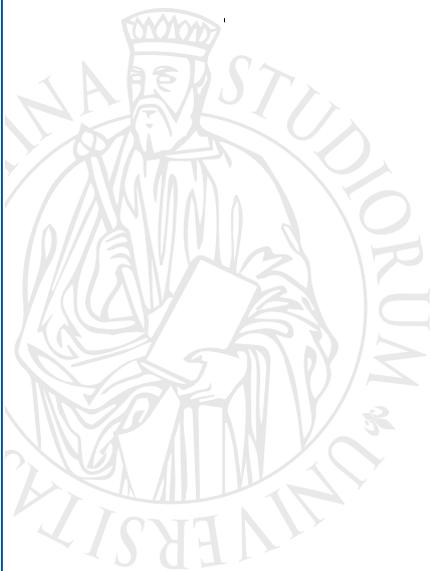
DIPARTIMENTO DI STATISTICA,
INFORMATICA, APPLICAZIONI
"GIUSEPPE PARENTI"

Retirement?

**Other ways out of the labour market are
far more worrying for health.**

Results from a matching approach study

Elena Pirani, Gustavo De Santis, Francesca Zanasi



**DISIA WORKING PAPER
2021/10**

© Copyright is held by the author(s).

Retirement?

Other ways out of the labour market are far more worrying for health.

Results from a matching approach study

Elena Pirani, elena.pirani@unifi.it

Dipartimento di Statistica, Informatica, Applicazioni – DiSIA University of Florence, Italy

Gustavo De Santis, gustavo.desantis@unifi.it

Dipartimento di Statistica, Informatica, Applicazioni – DiSIA, University of Florence, Italy

Francesca Zanasi, francesca.zanasi37@gmail.com

Dipartimento di Scienze Politiche, University of Bari Aldo Moro

Abstract

Objectives

Despite a growing body of research, the effects of retirement on health are not clear. The study explores the role played by the path out of the labour market (formal retirement vs. unemployment or family reasons), accounting for individual heterogeneity.

Methods

Propensity score matching approach is employed on longitudinal data from the Survey of Health, Ageing and Retirement in Europe (2004-2015).

Results

While health does not change significantly for those who formally retire, it worsens considerably for those who leave the labour market for other reasons. Moreover, health outcomes turn out to be highly heterogeneous, depending on individual socio-economic and job-related characteristics.

Discussion

Leaving the labour market in one's mature years is a complex transition. Future research should focus on understanding and combating the causes of premature exit from the labour market, a relevant concern both in economic terms and on health grounds, in the light of our results.

Keywords: Frailty index; Euro-D; self-rated health; propensity score; retirement-

Acknowledgments

This paper uses data from SHARE Waves 1, 2, 3, 4, 5, and 6 (DOIs: 10.6103/SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w3.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710), see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

We thank Alessandra Mattei for her advice on the use of the PSM technique.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

Funding

Financial support from the Italian MIUR, through the JPI MYBL / CREW Project is gratefully acknowledged (Joint Programme Initiative: More Years Better Life, 2016 Call. CREW: Care, retirement and wellbeing of older people across different welfare regimes. MIUR Decree: n. 3266/2018; Official Bulletin no. 32 7. Feb 2019).

Retirement?

Other ways out of the labour market are far more worrying for health.

Results from a matching approach study

Elena Pirani, Gustavo De Santis, Francesca Zanasi

Introduction

At first sight, the abundant literature about the effects of retirement on health seems inconclusive: the retirement-health relationship appears to be sometimes positive (retirement leading to better health), sometimes negative (retirement leading to worse health), and sometimes not significant (e.g., Van Der Heide et al. 2013).

Several reasons lie behind these contradictory findings. Some are substantive, because contrasting mechanisms may be at play. On the one hand, retirement may lead to a sudden break with support networks, colleagues, and friends, and may constitute a sort of shock. Besides, once out of the “productive chain”, individuals may lose the sense of belonging and usefulness they (probably) felt when they were employed, and they may also stop investing in intellectually engaging activities, with negative consequences on cognitive functions (Bonsang et al. 2012; Lee et al. 2019; Mazzonna and Peracchi 2012), self-perceived health and chronic conditions (Behncke 2012). On the other hand, retirement may instead prove beneficial: it is a relief from work-related stress, and it may encourage health-enhancing behaviours (Eibich 2015; Insler 2014) and lifestyles (Zantinge et al. 2013). This may improve physical and mental health (Syse et al. 2017), whether it is measured objectively (Coe and Zamarro 2011; Kolodziej and García-Gómez 2019), subjectively (e.g., self-perceived health; Hessel 2016), or indirectly, via questions on life satisfaction (Horner 2014).

The decision to stop working, and when, may imply a complex evaluation of costs and benefits, opportunities, and constraints (Whitley and Popham 2017): its consequences in terms of health and wellbeing may vary. This highlights the role of heterogeneity: individuals differ in their socio-demographic profile, behaviors and preferences, and they retire from different occupations. All these characteristics contribute to shape both retirement decision and post-retirement health – incidentally, this joint determination complicates the search for a true causal effect.

The contradictory findings of the literature depend also on the research design. For instance, the health outcome that one selects is not neutral, because health has several dimensions, and no single indicator suffices (Pirani 2017; Voss et al. 2020). Retirement may prove beneficial in some respects and detrimental in others (e.g., triggering depression while improving physical conditions), and in both cases, these effects work through the mediation of several other dimensions, family circumstances and institutional settings. The time since retirement is also very relevant: little is known on the mid- or long-term development of objective and subjective health after retirement, a path that may itself be causally linked to the change of status, for instance through an adaptation process. In empirical analyses, time since retirement is typically short, for lack of long panel data (a limitation that affects also our results, incidentally; for an exception, see Westerlund et al. 2009).

The contradictory findings of the literature depend also on definitions and analytical strategies. As for the former, the term “retirement” is often used loosely, and its meaning may range from acquiring the formal status of a retiree to the simple fact of leaving the labour market, past a certain age. In turn, this choice is linked to life histories – e.g., in terms of income, or individual’s family situation – probably the true underlying causes (Halleröd et al., 2013; De Preter, 2013). Data availability rarely permits researchers to keep these long-term factors under control, unless they reverberate in some observable variable around the time of retirement, or shortly before it. This connects to the thorny methodological issue of endogeneity (Nishimura et al., 2018). Retirement is not independent of previous health conditions (for instance, see Behncke 2012; Bonsang et al. 2012; Eibich 2015; Mazzonna and Peracchi 2012) and “reverse causality” (health affecting retirement, rather than vice versa), if not properly controlled for, may invalidate all conclusions. The purpose of this paper is to clarify three of the issues at stake in this field of study. As for the endogeneity problem, we adopt a propensity score matching (PSM) approach, a suitable way of estimating effects in observational contexts (e.g., Caliendo and Kopeinig 2008; Crown 2014). Relying on this technique, whose limits and potentialities we also discuss, we pursue our second aim, an analysis of the relevance of the path out of the labour market (see also Blomqvist et al. 2020; Halleröd et al. 2013; Whitley and Popham 2017). Finally, we deal with heterogeneity, stratifying our analyses by several individual characteristics.

While in this paper we focus on the individual level, the general picture should not be overlooked. Delaying retirement has become crucial for the survival of pension systems in most developed (and formerly developing) countries, which are ageing rapidly (OECD 2019). This strategy, in conjunction with others (such as adjusting contributions and benefits), is practically forced, and has in fact been pursued and periodically reinforced in all OECD countries. Besides, it seems consistent with the development of clerical occupations, the constant increase in life expectancies and the possible, although not so clear, expansion of healthy life expectancy (*ibidem*). Determining by how much retirement can be delayed without (major) negative consequences on people’s health appears to be of paramount importance: this paper aims to shed additional light on some of the relevant issues in this domain.

1. Between theoretical considerations and technical difficulties

1.1 On the definitions of retirement

Strictly speaking, retirement is a formal passage from employment to the status of retiree, receiving a pension benefit. Usually, this happens because people have reached a specific age, possibly in combination with a certain seniority (through old age retirement schemes), or because they have become ill or unable to work, and in this case one can usually count on invalidity or sickness pension schemes. However, there are also other reasons that may push people out of the labour market: people may lose their job, or may decide, or need, to do something else, such as taking care of sick relatives. Indeed, in recent decades, similarly to what happened in the labour market, also the phase of retirement has been characterized by an increasing level of complexity (Voss et al. 2020). For instance, more than half of the workforce in Europe leave their occupation through pathways other than formal retirement, with men generally exploiting disability or unemployment

benefits, while women become economically inactive (e.g. to take care of family members, OECD 2011). The boundaries between these situations are frequently blurred, and the transition may have more than just one cause: for example, unemployment frequently drives relatively old people out of the labour market, because re-employment perspectives are scarce.

The variety and indeterminacy of paths leading to labour market exits at later ages are recognized in the literature (e.g., Denton and Spencer 2009), but not always accounted for in empirical research, where several studies equate retirement with the end of the working life past a certain age (e.g., 60 or 65 years). For instance, a frequently used definition is the simple “not working for pay” (e.g., Coe and Zamarró 2011; Eibich 2015; Kolodziej and García-Gómez 2019; Oi 2019; Zhu 2016), sometimes refined with the intention of staying out permanently (e.g., Bonsang et al. 2012; Bonsang and Klein 2012). The downside of this broad definition is that the observation group becomes heterogeneous, because individuals who formally retire and receive a pension benefit get mixed up with those who leave the labour market for other reasons, such as unemployment, sickness or passage to economic inactivity – and health consequences may differ.

While transition from work to retirement (broadly defined) has led to ambiguous results (e.g., Van Der Heide et al. 2013), transition to unemployment, especially late in life, has consistently been found to generate depression and anxiety (Blomqvist et al. 2020), diminished social engagement and self-esteem (Whitley and Popham 2017) and lower psychological wellbeing and mental health (Voss et al. 2020). Studies on people – generally women – leaving job to take care of family and home are scarce, but a reduced mental engagement has generally emerged in these cases (Whitley and Popham 2017). These negative health consequences tend to be long lasting: returning to work, when possible, usually improves well-being, but fails to bring it back to its initial levels (Voss et al. 2020).

In short, leaving the labour market may have different consequences. To investigate which prevail, in what circumstances, we formed different groups of “mature” people (aged between 50 and 70 years) who were formerly employed, distinguishing those who formally retired from those who exited the labour market for unemployment or family reasons, and we measured the health effect of all of these paths.

1.2 Heterogeneity issues

Some authors claim that individual heterogeneity accounts for differences in research results (e.g., Blomqvist et al. 2020; Halleröd et al. 2013; Grøtting & Lillebø 2020). Retirement decisions result from the subjective evaluation of a number of push and pull factors, acting both at individual and institutional level (De Preter et al. 2013); on the one hand, socio-economic circumstances, family situation, and health all play a role, as well as job-related characteristics such as job satisfaction, labour market attachment, and income. On the other, push and pull factors can also be defined at institutional level (Whitley and Popham 2017), considering for example public pension schemes, with their penalties or incentives for early retirement, or other welfare programs (e.g. unemployment or invalidity schemes).

For example, education and socio-economic status may differ between individuals in different occupations, the highly educated and well-off typically occupying prestigious positions, less physically demanding, and less likely to deteriorate one’s physical or mental health. In turn,

education affect what may be labelled “cognitive investment” before and after retirement (Mazzonna and Peracchi 2012), and changes in lifestyle, such as smoking or physical activity (Celidoni and Rebba 2017). Also family networks have been found to be correlated to health in late life (Siegel et al. 2004; Pirani and Salvini 2012), although empirical evidence on the link with labour market exit is scarce. Considering job-related characteristics, dropping physically demanding or unpleasant jobs has been found to be a relief (Celidoni and Rebba 2017; Eibich 2015; Mazzonna and Peracchi 2017), and the same applies to jobs that create stress (Van Den Bogaard et al. 2016). Conversely, abandoning a prestigious and stimulating occupation may be experienced as a loss. Not surprisingly, the consequences on health differ in the two cases (e.g., Kolodziej and García-Gómez 2019; Westerlund et al. 2009).

In this paper, we take this heterogeneity into account, by estimating the effect of leaving the labour market across different subgroups of individuals.

1.3 Methodological perspectives

Another reason behind the contradictory findings of the literature on the health-retirement relation lies in the methodology of analysis. This is perhaps the main conclusion of Nishimura and colleagues (2018) in their systematic review of studies based on Global Aging data (SHARE, ELSA, HRS, HSE, JSTAR), where they also considered alternative definitions of retirement. The biggest challenge is to account for selection and for the possible endogeneity between retirement and health, which is one of the drivers of the decision to retire or just leave the labour market. To circumvent these problems, an instrumental variable (IV) approach is often adopted, combining age-, country-, and gender-specific incentives for retirement, together with the statutory retirement age (e.g. Behncke 2012; Coe and Zamarro 2011; Fe 2020; Kesavayuth et al. 2018; Kolodziej and García-Gómez 2019). Individual fixed-effects are sometimes also kept into account (e.g., Bonsang et al. 2012; Celidoni and Rebba 2017; Oi 2019; Zhu 2016).

However, a valid instrument, such as the frequently used “statutory retirement age”, must be correlated with the “cause” (retirement) but not with the “effect” (health), not even indirectly, for instance through some unobserved variables that affect health. Unfortunately, this is an assumption that cannot be empirically tested, and it turns out to be very difficult to find a conceptually valid instrument (Crown 2014). A tempting solution, in these cases, is to add as many control variables as possible, to avoid the omitted variable bias. However, all these controls should be unrelated to health and should not be themselves affected by retirement, which is rarely the case (Kolodziej and García-Gómez 2019; Mazzonna and Peracchi 2012). Finally, even studies that account for socio-economic heterogeneity in the relationship between retirement and health do not always consider that the statutory age at retirement employed as instrumental variable may vary by type of job and employment sector, as in Italy for instance, or because of other factors, such as (for women) the number of children, as in some Eastern European countries (OECD, 2019). In other cases, a statutory age at retirement simply does not exist (Eyjólfsdóttir et al. 2019).

In this paper, we opted for the Propensity Score Matching (PSM) approach. While rarely used in this field of study, and typically applied only to a reduced set of countries (i.e., Behncke 2012; Coppola and Spizzichino 2014; Eyjólfsdóttir et al. 2019; Radó and Boissonneault 2018), the PSM can produce an unbiased estimate of the causal effect using observational data sets, if the

underlying assumptions are satisfied (Crown 2014; Li 2012). The most important of these is that that assignment to treatment is strongly ignorable, or unconfounded: in other words, that there are no observed or unobserved variables that influence treatment selection (retirement). This risk can never be reduced to zero, but if the set of observed variables is large and appropriately selected, and if it can itself be interpreted as a proxy of other unobserved variables (think, for instance, of how telling education and income are of socio-economic status in general), the degree of confidence in the results of the analysis increases considerably.

Another reason to prefer the PSM approach to IV lies in the type of effect to be estimated. After the match based on the propensity score, one can compute the so-called ATT, or the treatment effect averaged over the subpopulation of treated people, a measure of the difference in the outcome (health) between those who stop working (the treated) and those who do not (the controls). Conversely, the IV approach lead to estimate the so-called LATE, or locale average treatment effect, which correspond to the average of the treatment effect over the individuals whose behavior would be different if the instrument (generally the statutory retirement age) were changed (Newey 2007). Our aim is to study whether and to what extent exiting the labour market, for whatever reason, triggers ill health, and this exit can take place not only for retirement strictly defined. We do not intend to evaluate the effect of a given policy measure (for instance, the increasing in statutory retirement age): instead, we assess the average health effect attributable to leaving the labor market, differentiating by possible routes of exit.

2. Data and Method

2.1 Data and sample

For our study, we used data from the longitudinal Survey of Health, Ageing and Retirement in Europe (SHARE) (Börsch-Supan et al. 2013), especially waves 1, 2, 4, 5 and 6, carried out respectively in 2004-2005, 2006-2007, 2010-2011, 2012-2013, and 2014-2015. We excluded waves 3 and 7, the retrospective SHARELIFE surveys. In each wave, SHARE data cover several key areas of life (health, socio-economic status, social and family networks, etc.) of more than 60,000 individuals aged 50 or over.

We focused on respondents aged between 50 and 70 years in the period between 2004 and 2015 and on a subset of the SHARE countries: those that participated in at least two consecutive waves – spanning two years from each other, except for waves 2 and 4 which span instead four years – and for which we had all the information that we needed for the analysis. Our transitions are those observed between waves 1 and 2; waves 2 and 4, waves 4 and 5; or waves 5 and 6. This means that each individual concurs to the sample with two consecutive observations: respondents must be employed at wave w and no longer employed (cases) or still in employment (controls) at wave $w+1$. We thus focus on the short-term effect of labour market exit on health (for a discussion, see section 4.1).

The countries included in the analysis are 15: Austria, Belgium, Czech Republic, Denmark, Estonia, France, Germany, Greece, Italy, the Netherland, Poland, Slovenia, Spain, Sweden, and Switzerland. Unfortunately, due to limited dimension of our sample, we could not perform country-specific analyses.

2.2 Analytical strategy

In our study, we applied a propensity score matching approach (paragraph 2.3), to identify the *causal effect* of leaving the labour market on various measures of health (paragraph 2.4), using a large set of individual information (paragraph 2.5). Very briefly, under a few assumptions, we estimate the effect of discontinuing one’s work (“treatment”) on health by comparing two groups of individuals who are as similar as possible in all observed respects, except for the fact that, between two consecutive waves of the SHARE survey, some leave the labour market (treated individuals, or cases) and some do not (matched individuals, or controls). Any systematic difference in the health outcome between the two groups can therefore be interpreted as statistically “caused” by the change in status.

In the first step of the analysis, we included all those who left the labour market, or *leavers*. These are those who stopped working between two consecutive waves, for *any* reason: they were employed at wave w , but not at wave $w+1$.

Then, to investigate possible differences in the routes to retirement, we broke this group down into three sub-groups, as follow (Table 1):

- i) *Retirees*: those who formally retired between two consecutive waves: they were employed at wave w , and formally retired at wave $w+1$;
- ii) *Drop-outs*: those who stopped working between two consecutive waves, but did not formally retire: they were employed at wave w , and either unemployed or looking after home/family, at wave $w+1$;
- iii) *Sick and disabled*: those who stopped working between two consecutive waves, but did not formally retire: they were employed at wave w , and permanently sick or disabled at wave $w+1$.

Table 1 – Cases (three types) and controls considered in this analysis, overall and by sex (aged 50 to 70 years, several EU countries, between 2004 and 2015)

		Total	Men	Women
Always employed individuals (control group, before matching)		7,350	3,624	3,726
Stopped working (treated group)	Leavers: stopped working for any reason (<i>i+ii + iii</i>)	4,800	2,485	2,315
	(i) Retirees: stopped working for formal retirement	3,794	2,048	1,746
	(ii) Drop-outs: unemployed, looking after home or family*	727	280	447
	(iii) Sick or disabled	279	157	122

* Although some differences may exist, due to the reduced sample size, we put together these two labour market exit patterns.
Source: SHARE data

As the last group – the *sick and disabled* (*iii*) – is small and highly selected, we decided to drop it. For the other groups instead we estimated the short-term health effect of leaving the labour market, keeping as a reference group those who remained in employment.

In the second step of our analysis, we estimated the health effect of “retirement” highlighting the role of a few in our opinion relevant stratification variables.

2.3 Method

To estimate the effect on people’s health of exiting the labour market, we followed the *potential outcome approach* (e.g., Imbens and Rubin 2015). For each individual i ($1, \dots, n$), we observed a

vector X_i of characteristics measured at wave w (pre-treatment characteristics). Let T_i denote the treatment variable indicator, equal to 1 if individual i leaves the labour market between two consecutive waves of the survey, zero otherwise. Under the Stable Unit Treatment Value Assumption (Rubin 1980, see Appendix), for each individual i observed in employment at wave w there are two potential outcomes at wave $w+1$, $Y_i(1)$ and $Y_i(0)$, if i does or does not leave the labour market, respectively.

A comparison of these potential outcomes, typically their difference, gives a measure of the causal effect of the transition out of the labour market. We consider the Average Treatment effect on the Treated (ATT), that is the average effect of this transition on those who actually leave the labour market:

$$ATT = E[Y_i(1) - Y_i(0) | T=1]$$

As we use observational (not experimental) data, we need two assumptions on the treatment-assignment mechanism. First, unconfoundedness, or “selection on observables”. Formally, this requires that, conditioning on *observed* covariates, the treatment assignment (leaving or not leaving the labour market) must be independent of the potential health outcomes – $T_i \perp (Y_i(0), Y_i(1)) | X_i$. Second, the overlap assumption requires that there is sufficient overlap in the distribution of the covariates between treated and control units, $0 < P(T_i=1 | X_i=x) < 1$ for each i , which basically means that both come from overall comparable groups and that for each treated individual, a “corresponding” control individual (with similar background characteristics) can be found.

This comparison is made through the propensity score, or the probability of leaving the labour market, estimated through a probit model in this case, conditional on the observed covariates. The propensity score is a balancing score: conditional on the propensity score, the distribution of the covariates is similar between treated and controls, and therefore independent of the treatment. If the unconfoundedness and the overlap assumptions hold, matching on the propensity score is sufficient to remove confounding, and what emerges (health effect) can be safely attributed to the treatment (exiting the labour market, in this case; Rosenbaum and Rubin 1983).

In our case, we considered a large set of covariates covering demographic, social, economic, family, and job-related aspects, thus capturing very many observed (and proxying also several unobserved) characteristics. We checked that the distributions proved comparable between treated and controls, and we also ran a sensitivity check (see Appendix). Finally, with PSM, we selected a sub-group of *controls*, that is a group of individuals who remained in the labour market between two subsequent waves and who were as similar as possible to the *treated*, those who left it. The variables used to calculate the propensity score are listed in paragraph 2.5 and in the Appendix.

In selecting controls, we opted for a 1-to-1 nearest neighbour matching (Abadie and Imbens 2011) with replacement. Among the potential controls (first row of Table 1), for each individual exiting the labour market we selected the closest control (still in the labour market at wave $w+1$, and with the closest propensity score), allowing a non-treated individual to be used more than once as a match. More details on the technical aspects of the matching procedure (i.e., caliper, common support and balancing before and after matching) are discussed in the Appendix.

2.4 Health outcomes

We evaluated the health differences between treated and control units – or between those who did and did not leave the labour market – considering three health outcomes measured at wave $w+1$. First, we used the frailty index, a 40-item index validated on the SHARE dataset, measuring the number of accumulated health deficits (Romero-Ortuno and Kenny 2012). It summarizes individual frailty both in objective (e.g., grip strength and body mass index) and subjective terms (e.g., self-reported health and mood). The count of deficits for each person is divided by 40 (total number of possible deficits considered), so that individuals in good health score 0, while individuals in the worst possible health conditions score 1.

Next, we used a composite indicator of depression constructed from the 12 basic items of the EURO-D scale (Prince et al. 1999): depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. This scale, ranging between 0 (not depressed) and 12 (severe depression), was developed to derive a common scale of depression symptoms, especially in late life, on different indicators in several European countries. We divided this count by 12, so that individuals with good mental health score 0, while individuals with severe depression score 1.

Finally, as is commonly done in this area of research (e.g., Voss et al. 2020), we considered also a general measure of health, self-perceived health, dichotomizing the original 5-point scale into 0=good (excellent, very good, or good) and 1=poor (fair or poor).

2.5 Propensity score matching variables

A critical issue in the design phase of an observational study is the choice of the pre-treatment variables conditional on which unconfoundedness and overlap assumptions are reasonable. To this end, a large and representative set of pre-treatment characteristics – all measured at wave w , prior to retirement to avoid the endogeneity problem – has been accurately selected. We based this selection on a priori substantive knowledge of the phenomenon (e.g. Caliendo and Kopeinig 2008) derived from the literature on retirement-health relationship. Below, we list the variables used for the probit estimation of the propensity score: more details about the coding of these variables and the balancing of the propensity score are reported in the Appendix.

First, in the specification of the propensity score we included various measures of health. Accounting for the health status *prior* to retirement is fundamental to control for potential selection, as bad health itself can push individuals out of the labour force. We included the same three health measures used as outcomes in the subsequent wave $w+1$: self-perceived health, frailty index, and depression. In addition, to account for other possible aspects linked to physical health, we also considered self-reported hearing and eyesight difficulties, limitation in activities of daily living (ADL) and in instrumental activities of daily living (IADL), mobility limitations, the global activity limitation indicator (GALI), the presence of chronic diseases, grip strength, and obesity. We added controls for smoking and physical activity, two important health-related behaviours.

Second, we considered several socio-demographic characteristics, potentially linked to both health and labour market exit: gender, age and age squared, education, subjective evaluation of one's own

economic situation. We kept into account the country of residence, to capture, at least roughly, unobserved characteristics differentiating individuals in our sample.

Third, we introduced the family network, controlling whether the respondent had a partner, if at least one parent was still alive, the number of siblings, children and grandchildren, and the household size.

Job-related characteristics form the fourth set of variables included in the estimation of the propensity score, as these are possible confounders in the link between labour market exit and health: type of contract and job satisfaction as declared by the respondent. To account for a potential anticipatory element, we also included the preference, expressed in wave w , to retire as early as possible from the current job.

Results

Table 2 shows the estimates of the ATT effects on various health outcomes according to the three different definitions of retirement used in this paper: *leavers* (those who stopped working *tout court*), *retirees* (those who formally retired), and *drop-outs* (those who stopped working without taking formal retirement). The health outcomes are coded so that the higher the score, the worse the health status (higher frailty, higher depression, worse perceived health). As previously described, the ATT effect represents the difference in the effect between the treated and the control group; in our case, a positive ATT effect means that treated individuals exhibit worse health conditions than controls, thus indicating a damaging effect of labour market exit on health.

First, let us consider the more comprehensive group of *leavers*. Leaving the labour market, at least between 50 and 70 years of age, *seems* to cause a general deterioration of health, especially in terms of frailty, although the magnitude of the effects is small.

This negative effect, however, depends on the route to labour market exit. Formal retirement (*i*) does not appear to affect health in our sample, in either way (a slightly worsening effect, significant at 10% level, can be detected only for the frailty index of men). Conversely, those who leave their job not because of formal retirement but for other reasons, the *drop-outs* (*ii*), later suffer from worse health conditions, with all the three health indicators considered here. In this case, the magnitude of these ATT effects is far from negligible, equalling 0.015, 0.029 and 0.068 respectively for frailty, depression and self-perceived health, with even higher values for men. Those who stopped working because of unemployment and family reasons, generally showed higher frailty levels, a higher number of depressive symptoms, and more frequently declared to be in poor health. This finding confirms what has generally been found in the literature: when discontinuing work is not a (totally) free decision, as it probably happens in these cases, distress, depression and a deterioration of both objective and subjective health may ensue (e.g., Daly and Delaney 2013).

These results bear, we believe, two relevant implications. The first is that a precise definition of “retirement” is necessary to determine the direction and magnitude of the health effects of an interruption in one’s labour activity, because of the different mechanisms behind this passage. The second is that remaining at work, even at relatively high ages, is not detrimental for health - and this is particularly important in an epoch when raising retirement ages seems to be one of the necessary tools to preserve the viability of pension systems.

Table 2 – Models results: estimates of the ATT effects on frailty index, depression, self-perceived health. (Overall and by gender, according to the three definitions of retirement defined in the text)

*Significance level: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$*

	N of treated individuals	Frailty Index ATT sig.	Depression ATT sig.	Bad self-perceived health ATT sig.
<i>Leavers (i+ii+iii)</i>				
Men and women	4,800	0.009***	0.003	0.029 +
Men	2,485	0.008*	0.001	0.037
Women	2,315	0.004	-0.002	0.000
<i>Retirees (i)</i>				
Men and women	3,794	0.002	-0.009	-0.016
Men	2,048	0.007+	-0.001	0.017
Women	1,746	-0.003	-0.013	-0.041
<i>Drop-outs (ii)</i>				
Men and women	727	0.015***	0.029**	0.068**
Men	280	0.016*	0.031*	0.140***
Women	447	0.013**	0.026	0.074*

Source: SHARE data, 15 European countries, between 2004 and 2015

Let us now move to the second step of our analysis, where we focus separately on the two subgroups of formal retirees and drop-outs. Table 3 presents ATT effect estimates of retirement on the *retirees* for our three health outcomes according to some selected individual and job-related characteristics (for the sake of brevity, results are shown here for men and women together, but they hold also for the two genders separately. Details available upon request).

Table 3 – Models results: estimates of the ATT effects computed for the group of retirees, by selected individual and job-related characteristics, on: frailty index, depression, self-perceived health.

*Significance level: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$*

	N of treated individuals	Frailty Index ATT sig.	Depression ATT sig.	Bad self-perceived health ATT sig.
<i>Educational Level</i>				
Low	2,397	0.008 +	0.003	0.039
High	1,397	0.001	-0.016	-0.050 +
<i>Psychologically demanding job</i>				
No	2,081	0.002	-0.008	-0.015
Yes	1,713	-0.002	-0.014	-0.055 *
<i>Presence of a partner</i>				
No	3,016	0.002	-0.12	0.009
Yes	778	0.007	0.005	-0.022
<i>Presence of grandchildren</i>				
No	1,452	0.002	-0.010	0.011
Yes	2,342	0.001	-0.008	-0.021
<i>Preference for early retirement</i>				
No	1,886	0.005	0.008	0.029
Yes	1,908	-0.005	-0.029 *	-0.080 **

Source: SHARE data, 15 European countries, between 2004 and 2015

Some differences emerge by educational level: lower education is associated with slightly worse health conditions after retirement (i.e., 0.008 for the frailty index), even if this is not significant for depression (0.003) or self-perceived health (0.039). Instead, a beneficial effect of retirement on perceived health emerges for the highly educated (-0.050), and for those who had a psychologically demanding job (-0.015). These results are slightly more robust for women (not shown). The presence of a partner or grandchildren has positive effects on the self-perceived health of women (not shown), but these effects disappear when considering simultaneously men and women. Finally, as expected, those who longed for retirement improve after reaching it, both in terms of depression (-0.029) and self-perceived health (-0.080).

As for the *drop-outs* – those who left the labour market for unemployment or family reasons, Table 4 – the negative effect of leaving the labour market emerges consistently in all the health dimensions considered here, and it is almost always significant, and occasionally even highly significant. Leaving the labour market deteriorates health especially when the previous occupation was psychologically demanding, and also for those with low education. Even those who were unsatisfied with their job are worse off once they discontinue it.

The results about the presence of a partner reported in Table 4 are interesting, but they hide remarkable gender differences. In particular, health deteriorates for the men who do not have a living partner, while for women it is the presence of a partner that is linked to worse health. The presence of grandchildren does not have relevant impact on health, for either sex.

Table 4 – Models results: estimates of the ATT effects computed for the group of drop-outs, by selected individual and job-related characteristics, on: frailty index, depression, self-perceived health.

*Significance level: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; + $p < 0.10$*

	N of treated individuals	Frailty Index ATT sig.	Depression ATT sig.	Bad self-perceived health ATT sig.
<i>Educational Level</i>				
Low	567	0.135 **	0.026 *	0.064 *
High	156	0.156 *	0.027	0.064
<i>Psychologically demanding job</i>				
No	407	0.014 **	0.027 *	0.086 *
Yes	320	0.017 **	0.027 +	0.104 **
<i>Job satisfaction</i>				
No	91	0.012	0.023	0.080
Yes	636	0.014 **	0.020 +	0.076 **
<i>Presence of a partner</i>				
No	585	0.010 *	0.019 +	0.066 *
Yes	142	0.030 ***	0.046 *	0.070
<i>Presence of grandchildren</i>				
No	364	0.009 +	0.034 *	0.083 *
Yes	363	0.013 *	0.011	0.067 +

Source: SHARE data, 15 European countries, between 2004 and 2015

Discussion

In this study, we analysed the health consequences of leaving the labour market in a group of European countries in recent times. Our purpose was to clarify three main issues. First, we applied a propensity score matching approach, a technique relatively rare in this field of study, which the rich set of pre-treatment variables available in the SHARE data made possible. These refer to health, life-style behaviours, demographic, social and family traits, and job-related features, and enabled us to minimize the possible influence of “confounders” on the relation of interest. Second, we considered different pathways out of the labour market because, as we suspected and as it turned out, results depend on the very definition of “retirement”. Third, we estimated the health effect of exiting the labour market differentiating by various individual and job-related characteristics.

We considered three different measures of health: *the frailty index* (a composite measure including objective and subjective aspects); a common and reliable measure of depression, *the Euro-D scale*; and *self-perceived health*, the quintessential indicator of subjective health.

Our results highlight the importance of considering the routes that lead a mature worker out of the labour market. Loose definitions of retirement, such as “all those who, past a certain age (e.g., 50 years), leave their occupation for whatever reason” may lead to wrong conclusions. In fact, formal retirement itself has virtually no effect on health, either good or bad. This is good news, in our opinion: health concerns cannot be used as an argument against postponing the retirement age in case of need, for instance for the safeguard of pension systems. Rather, increased effort should be put in keeping mature people at work, even when difficulties emerge, either in the labour market (unemployment) or in the family sphere: those who quit their job for these reasons are worse off – at least, shortly after retirement, the only effect that we could measure. Our findings are essentially independent of the health indicator considered: all of them signal a worsening of health when the labour market is abandoned for reasons other than formal retirement.

Our results also emphasise that individual and job-related characteristics may shape the relationship between labour market exit and health. As for self-perceived health, for instance, the idea that retirement may represent a relief for workers with psychologically demanding jobs found only limited support. A modest positive effect emerges also for highly educated individuals. Not surprisingly, those who expressed preferences to retire as early as possible report lower depression and improved self-perceived health after retirement. Focusing on the subgroup of drop-outs for unemployment or family reasons, the negative effect – which is generally present – is stronger for the low educated, possibly because their possibilities of re-employment are lower. The family context, especially the presence of a partner, matters as well, with gender differences. Health worsens for men without a partner and for women *with* a partner. A possible explanation lies in the caregiver role traditionally played by women, who may have to leave the labour market to look after family members in need (i.e., a sick or disabled relatives); conversely, men who leave the labour market before retirement suffer from the lack of a partner on whom to rely. Unfortunately, our data do not permit us to verify this conjecture.

Overall, we showed that it is not possible to identify a unique link between labour market exit and health. Individual heterogeneity is fundamental in shaping results.

4.1 Limitations and strengths

Our study has a number of limitations. First, the panel nature of our sample reduced the number of observations, which resulted in large standard errors: incidentally, this could explain why several parameters, while with the expected sign, did not prove significant. Moreover, due to the reduced sample size, we could not go into greater detail, and we were forced to ignore country differences, for instance, which may be relevant because of the specificity of their welfare and social protection systems (Richardson et al. 2019).

A second limitation of this study is its time frame: health consequences may emerge after a longer interval than we could cover here. We are aware that we focus only on the short-term health effects of retirement, and that we cannot measure the possible evolution of these effects, including, for example the so-called “honeymoon phase” for the newly retired, who are happy at first, but perhaps less so in the long run (Hershenson 2016). On the other hand, limiting our study to a two-wave span allows us to minimize attrition, and maximize our sample size. Attrition could potentially be problematic for studies on health and retirement, because of selection (people in worse health are more likely to drop out of the panel). In SHARE data, attrition is high, ranging between 30% (when two consecutive waves are considered, two years apart) and 50% (when the interval between waves reaches four years). However, if a bias were present, it would lead to an underestimation of the effects that we measure here, because people in poor health are more likely to exit the panel. Besides, the large number of variables considered in our PSM (also in terms of health) makes the two groups highly similar (cases and controls: see Figure A1 and Table A2 in the Appendix), which reinforces our confidence in our findings.

In our analysis we could not distinguish those who have literally just exited the labour market (a few days before the interview) from those who transitioned “long” before (up to two years; even up to four in the passage from wave 2 to wave 4 of the SHARE rounds). For the subgroup of the *retirees*, however, we have the exact date of retirement: in this case, we verified that the distribution of these retirement periods is at random, and that including it in the analysis leaves the results unchanged (not shown here but available upon request).

Another potential limitation of our study, typical of the PSM approach followed here, is due to unobserved factors, which could invalidate the unconfoundedness assumption. This is something that, by definition, cannot be verified. However, we controlled for a very large set of pre-treatment characteristics, including health, precisely to minimize the risk of selection and reverse causality. We also simulated the presence of potential unobserved confounders (see Appendix), and the results of this robustness check indirectly supports the claim that no unobserved confounders remain once controlling for the large set of observed characteristics.

Perhaps the most important lesson to be learnt from our study is the importance of distinguishing the routes out of the labour market: retirement proper *vs.* unemployment/family obligations. These groups differ in terms of characteristics, resources and health consequences, non-existing for formal retirees, but clearly visible (and negative) for the others. We deem that this study, by using PSM technique, further contributes to a better understanding of the inconsistency of previous findings on the influence of labour market exit on health.

4.2 Conclusion

Leaving the labour market in one's mature years (between 50 and 70 years of age) is a more complex transition than it may seem. Formal retirement is the most frequent case, but there are several other ways out and these alternatives (e.g., unemployment and family reasons) are associated with markedly worse health outcomes. Future research should focus on understanding and combating the causes of premature exit from the labour market, a reason of concern not only in economic terms but also on health grounds in the light of our results.

Our results suggest that labour market attachment, even at relatively high ages, does not damage health, thus reinforcing the idea that longer labour market participation is possible: all in all, this is good news in an era of rapid population ageing and endangered sustainability of pension systems. Between 50 and 70 years, individuals are exposed to age-related exclusionary forces, such as biological health decline and the shrinking of support network, and face limited possibility of re-entering the labour market (e.g. OECD 2011, 2019). Back in 2002, the World Health Organization adopted the term "active ageing" to define a policy framework for "*optimizing opportunities for health, participation and security in order to enhance quality of life as people age*" (WHO 2002, p. 12). Policy measures for improving the professional contribution and competitiveness of mature workers are possible, e.g. continuous re-training and up-skilling programs (Boudiny 2013). Our results suggest that they are worth pursuing.

Welfare schemes should also keep into account the needs of those who are forced out of the labour market for family reasons (e.g., to take care of sick relatives). Getting back to work is sometimes possible, even if difficult: when this happens, the wellbeing of these mature workers improves (Voss et al. 2020). Some categories of workers are especially problematic: they suffer from higher risks of dropping out of the labour market, have worse health consequences in these cases, and are the most difficult to re-employ. These are the cases that merit special attention.

References

- Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics*, 29(1), 1–11.
- Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal*, 2(4), 358–377.
- Behncke, S. (2012). Does retirement trigger ill health? *Health Economics*, 21, 282–300.
- Blomqvist, S., Westerlund, H., Alexanderson, K., & Magnusson Hanson, L. L. (2020). Labor market exit around retirement age in Sweden and trajectories of psychotropic drugs in a context of downsizing. *BMC Public Health*, 20(1), 1–12.
- Bonsang, E., Adam, S., & Perelman, S. (2012). Does retirement affect cognitive functioning? *Journal of Health Economics*, 31(3), 490–501.
- Bonsang, E., & Klein, T. J. (2012). Retirement and subjective well-being. *Journal of Economic Behavior & Organization*, 83, 311–329.
- Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., et al. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, 42(4), 992–1001.
- Boudiny, K. (2013). “Active ageing”: From empty rhetoric to effective policy tool. *Ageing and Society*, 33(6), 1077–1098.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72.
- Celidoni, M., & Rebba, V. (2017). Healthier lifestyles after retirement in Europe? Evidence from SHARE. *European Journal of Health Economics*, 18(7), 805–830.
- Coe, N. B., & Zamarro, G. (2011). Retirement effects on health in Europe. *Journal of Health Economics*, 30(1), 77–86.
- Coppola, L., & Spizzichino, D. (2014). The effect of retirement on self-reported health: a gender comparison in Italy. *Vienna Yearbook of Population Research*, 12, 53–82. <http://www.jstor.org/stable/43773386>
- Crown, W. H. (2014). Propensity-score matching in economic analyses: Comparison with regression models, instrumental variables, residual inclusion, differences-in-differences, and decomposition methods. *Applied Health Economics and Health Policy*, 12(1), 7–18.
- De Preter, H., Van Looy, D., & Mortelmans, D. (2013). Individual and institutional push and pull factors as predictors of retirement timing in Europe: A multilevel analysis. *Journal of Aging Studies*, 27(4), 299–307.
- Denton, F. T., & Spencer, B. G. (2009). What Is Retirement? A Review and Assessment of Alternative Concepts and Measures. *Canadian Journal on Aging / La Revue canadienne du vieillissement*, 28(1), 63–76.
- Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics*, 43, 1–12.

- Eyjólfssdóttir, H. S., Baumann, I., Agahi, N., Fritzell, J., & Lennartsson, C. (2019). Prolongation of working life and its effect on mortality and health in older adults: Propensity score matching. *Social Science & Medicine*, 226(November 2018), 77–86.
- Fe, E. (2020). Pension Eligibility Rules and the Local Causal Effect of Retirement on Cognitive Functioning. *SSRN Paper*, 1–42.
- Grötting, M. W., & Lillebø, O. S. (2020). Health effects of retirement: evidence from survey and register data. *Journal of Population Economics*, 33(2), 671–704.
- Halleröd, B., Örestig, J., & Stattin, M. (2013). Leaving the labour market: the impact of exit routes from employment to retirement on health and wellbeing in old age. *European Journal of Ageing*, 10, 25–35.
- Hershenson, D. B. (2016). Reconceptualizing retirement: A status-based approach. *Journal of Aging Studies*, 38, 1–5.
- Hessel, P. (2016). Does retirement (really) lead to worse health among European men and women across all educational levels?. *Social Science & Medicine*, 151, 19–26.
- Horner, E. M. (2014). Subjective Well-Being and Retirement: Analysis and Policy Recommendations. *Journal of Happiness Studies*, 15(1), 125–144.
- Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, 23, 305–327.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources*, 49(1), 195–233.
- Kesavayuth, D., Rosenman, R. E., & Zikos, V. (2018). Retirement and health behaviour. *Applied Economics*, 50(54), 5859–5876.
- Kolodziej, I. W. K., & García-Gómez, P. (2019). Saved by retirement: Beyond the mean effect on mental health. *Social Science & Medicine*, 225(August 2018), 85–97.
- Lee, Y., Chi, I., & Palinkas, L. A. (2019). Retirement, Leisure Activity Engagement, and Cognition Among Older Adults in the United States. *Journal of Aging and Health*, 31(7), 1212–1234.
- Leombruni, R., & Mosca, M. (2019). *PSMatching: A SAS Macro for Treatment Effect Estimation and Missing Data Imputation Based on Propensity Score Matching*.
- Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.
- Li, M. 2013. Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods*, 16(2), 188–226.
- Mazzonna, F., & Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4), 691–710.
- Mazzonna, F., & Peracchi, F. (2017). Unhealthy retirement? *Journal of Human Resources*, 52(1), 128–151.
- Nannicini, T. (2007). A Simulation-Based Sensitivity Analysis for Matching Estimators. *The Stata Journal*, (2005).

- Newey, W., 2007. Course materials for New Econometric Methods, Spring 2007. MIT OpenCourseWare (<http://ocw.mit.edu>), Massachusetts Institute of Technology.
- OECD, 2011. Trends in Retirement and in Working at Older Ages, in Pensions at a Glance 2011: Retirement-income Systems in OECD and G20 Countries, OECD Publishing, Paris.
- OECD. (2019). *Pensions at a Glance 2019*. OECD.
- Oi, K. (2019). Does Gender Differentiate the Effects of Retirement on Cognitive Health? *Research on Aging*, 41(6), 575–601.
- Pirani, E. (2017). On the Relationship Between Atypical Work(s) and Mental Health: New Insights from the Italian Case. *Social Indicators Research*, 130(1), 233–252.
- Pirani, E., Salvini, S. (2012) Socioeconomic Inequalities and Self-Rated Health: A Multilevel Study of Italian Elderly. *Population Research and Policy Review*, 31, 97–117.
- Prince, M., Reischies, F., Beekman, A. T. F., Fuhrer, R., Jonker, C., Kivela, S.-L., et al. (1999). Development of the EURO-D scale – a European Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*, 174(4), 330–338.
- Radó, M., & Boissonneault, M. (2018). Short and long-term change in subjective well-being among voluntary and involuntary retirees. *Journal of the Economics of Ageing*.
- Richardson, S., Carr, E., Netuveli, G., & Sacker, A. (2019). Country-level welfare-state measures and change in wellbeing following work exit in early old age: Evidence from 16 European countries. *International Journal of Epidemiology*, 48(2), 389–401.
- Romero-Ortuno, R., & Kenny, R. A. (2012). The frailty index in Europeans: association with age and mortality. *Age and Ageing*, 41, 684–689.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75(371), 591–593.
- Siegel, M. J., Bradley, E. H., Gallo, W. T., & Kasl, S. V. (2004). The effect of spousal mental and physical health on husbands' and wives' depressive symptoms, among older adults: longitudinal evidence from the health and retirement survey. *Journal of Aging and Health*, 16(3), 398–425.
- Syse, A., Veenstra, M., Furunes, T., Mykletun, R. J., & Solem, P. E. (2017). Changes in Health and Health Behavior Associated With Retirement. *Journal of Aging and Health*, 29(1), 99–127.
- Van Den Bogaard, L., Henkens, K., & Kalmijn, M. (2016). Retirement as a relief? The role of physical job demands and psychological job stress for effects of retirement on self-rated health. *European Sociological Review*, 32(2), 295–306.
- Van Der Heide, I., Van Rijn, R. M., Robroek, S. J. W., Burdorf, A., & Proper, K. I. (2013). Is retirement good for your health? A systematic review of longitudinal studies. *BMC Public Health*, 13(1), 1–11.
- Voss, M. W., Wadsworth, L. L., Birmingham, W., Merryman, M. B., Crabtree, L., Subasic, K., & Hung, M. (2020). Health Effects of Late-Career Unemployment. *Journal of Aging and Health*, 32(1), 106–116.

- Westerlund, H., Kivimäki, M., Singh-Manoux, A., Melchior, M., Ferrie, J. E., Pentti, J., et al. (2009). Self-rated health before and after retirement in France (GAZEL): a cohort study. *The Lancet*, *374*(9705), 1889–1896.
- Whitley, E., & Popham, F. (2017). Leaving the labour market later in life: How does it impact on mechanisms for health? *Occupational and Environmental Medicine*, *74*(12), 877–886.
- World Health Organization (2002). Active ageing: A policy framework, WHO/NMH/NPH/02.8, Geneva, World Health Organization.
- Zantinge, E. M., Van Den Berg, M., Smit, H. A., & Picavet, H. S. J. (2013). Retirement and a healthy lifestyle: Opportunity or pitfall? A narrative review of the literature. *European Journal of Public Health*, *24*(3), 433–439.
- Zhu, R. (2016). Retirement and its consequences for women's health in Australia. *Social Science & Medicine*, *163*(2016), 117–125.

Appendix

For the propensity score method to work properly, cases and controls must be comparable. In the following subsections we describe how we managed a few technical issues to ensure this.

SUTVA

The Stable Unit Treatment Value Assumption (SUTVA) requires that individuals' potential outcomes depend on their own treatment only and not on that of other individuals in the population. As the SHARE survey is at the household level, the assumption of no interference may be violated when there are couples in the relevant age range, because ego's decisions on retirement may be influenced by that of their partner. There is no simple way to circumvent this obstacle, but we also ran separate analysis by gender, obtaining perfectly comparable results (not shown here). Therefore, we believe that this potential violation does not undermine our conclusions.

Unconfoundedness assumption

The reliability of PSM estimates depends on the plausibility of the unconfoundedness assumption, which means that conditioning on observed covariates, selection into treatment is independent of the outcome. The unconfoundedness assumption does not hold if there are unobserved confounders, or relevant variables that were not considered in the estimation of the propensity score. This case cannot be verified directly, but sensitivity (or indirect) tests exist to signal possible violations of the unconfoundedness assumption. One of these, suggested by Ichino, Mealli and Nannicini (2008) and Nannicini (2007), is a sensitivity analysis (*sensatt* routine using the software Stata 14) based on the simulation of a potential unobserved (binary) confounder U .

The distribution of the simulated variable is obtained by choosing four parameters:

$$p_{ij} = \Pr(U = 1 | T = i, Y = j) = \Pr(U = 1 | T = i, Y = j, X)$$

with $i, j = \{0, 1\}$, which give the probability that $U = 1$ in each of the four groups defined by the combination of the treatment status (T) and the outcome value (Y). For example, p_{11} is the fraction of individuals who, having left the labour market ($T=1$) and poor health ($Y=1$), also have $U=1$.

The simulated U is then included in the set of covariates X used for the PSM, and the ATT is estimated as the average across several iterations (e.g. 1,000 times). The simulated ATT is then compared to the baseline ATT (where U is not included among the matching variables), that is to the result of our analysis (Table 2). This gives a feeling of how sizeable the confounding factor U must be to drive the ATT to zero. Moreover, the sensitivity analysis shows (in the odds ratio scale) the extent to which U affects Y_0 (outcome effect, Γ) and/or T (selection effect, Λ).

In our sensitivity analysis, we set the p_{ij} parameters to two different set of values, which we extracted from the simulation exercises proposed by Nannicini (2007) and Ichino, Mealli and Nannicini (2007). The former proposes to simulate a “killer” confounder, identified by the parameters $p_{11}= 0.8$, $p_{10}= 0.8$, $p_{01}= 0.6$, $p_{00}= 0.3$. The latter propose a “neutral” confounder, where all the p_{ij} parameters are set to 0.5. For each of the three groups (*leavers, retirees, drop-outs*), Table

A.2 shows the ATT effect obtained when adding U in the set of matching covariates, whereas baseline estimates are reported in Table A1.

Overall, when adding the simulated unobserved confounder, the estimates remain virtually the same, in terms of both size and statistical significance. Results hold also in the case of the “killer” confounder, when U is associated with large selection and outcome effects. In other words, the presence of such a strong unobserved confounder is hardly plausible given the vast set of covariates included in our matching set, but if it existed, it would not substantially alter our conclusions.

Table A.1 Sensitivity analyses on unobserved confounder, for the estimates of the ATT effects on frailty index, depression, self-perceived health

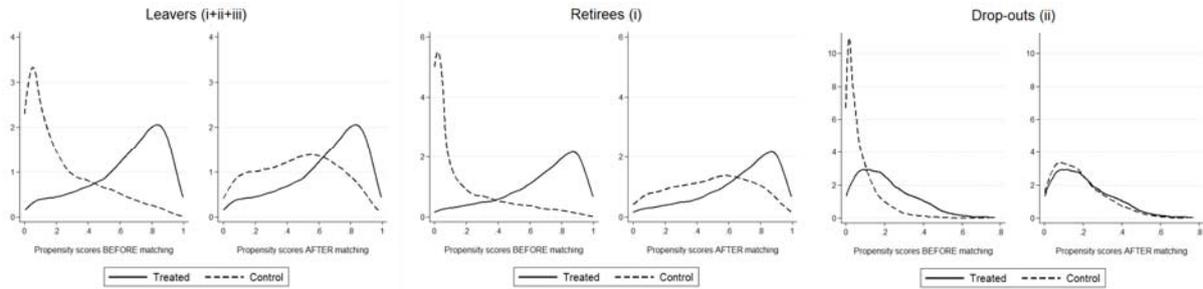
	Frailty Index				Euro-D				SRH			
	<i>Killer confounder</i>											
	ATT	SE	Γ	Λ	ATT	SE	Γ	Λ	ATT	SE	Γ	Λ
Leavers	0.009	0.003	3.508	5.710	0.003	0.006	3.508	5.138	0.030	0.016	3.501	7.390
Retirees	0.003	0.003	3.504	5.714	-0.010	0.007	3.504	5.150	-0.016	0.019	3.505	7.418
Drop-outs	0.016	0.004	3.500	5.754	0.032	0.010	3.508	5.165	0.072	0.026	3.498	7.434
	<i>Neutral confounder</i>											
	ATT	SE	Γ	Λ	ATT	SE	Γ	Λ	ATT	SE	Γ	Λ
Leavers	0.009	0.003	1.004	0.999	0.003	0.006	1.002	0.999	0.030	0.016	1.001	1.000
Retirees	0.003	0.003	1.003	1.002	-0.010	0.007	1.001	1.003	-0.016	0.019	1.000	1.001
Drop-outs	0.016	0.004	1.000	1.002	0.032	0.010	1.005	1.003	0.072	0.026	0.999	1.006

Note: the small differences between the ATT effects estimated here and those of Table 2 depend on the estimation procedure. *Sensatt* relies on the *att** package from Becker and Ichino (2002), whereas in the Table 2 we used the *psmatch2* procedure (Leuven and Sianesi 2003), which is in general allows the caliper setting (see e.g. Leombruni and Mosca, 2019 for a discussion of differences). Differences in the magnitude of the effects are negligible and not statistically significant.

Overlap assumption

The second assumption required for the propensity score matching procedure is that there is a sufficient overlap in the joint distribution of the covariates between treated and control groups, the so-called “common support”. Consequently, we excluded from our analysis the very few cases whose propensity score fell outside the range (smaller than the minimum or larger than the maximum) of the other group: one of the *leavers*, two among the *retirees*, and five among the *drop-outs*. On the remaining cases, we matched cases and controls to make sure that the estimation of treatment effects refers to comparable pairs, which is one of the strengths of the PSM approach (Crown 2014). As Figure A.1 shows, in all three cases (*leavers*, *retirees*, and *drop-outs*), the matching procedure consistently improved the similarity between cases and controls: after matching the distribution of the propensity score for the two groups reasonably overlaps, as the method requires.

Figure A.1 – Distribution of the estimated propensity scores according to the three definitions of retirement used, before and after matching



Caliper

Our analyses are based on a propensity score approach with 1-to-1 nearest neighbour matching. To avoid poor matches, pairs of treated and untreated subjects are formed whose propensity scores differ by at most a pre-specified amount, the so-called “caliper width”. As recommended for this type of applications, we set the caliper equal to 0.25 of the standard deviation of the estimated propensity score. Therefore, our caliper is 0.076 for the *leavers*, 0.082 for the *retirees*, and 0.024 for the *drop-outs*.

Balancing

In order to assess the quality of the matching procedure, a comparison between the covariates distributions across treated units and matched control units can be done. Table A2 presents the variables used to estimate the propensity score, together with their description and the coding used in our analysis, and their distribution according to the sub-groups considered in the analysis: first, we consider the entire group of controls (i.e. before matching: those always in employment between wave t and $t+1$). Then, for *leavers*, *retirees*, and *drop-outs* we show the distribution of the characteristics for the treated in each group (column T) and for the controls after matching (column C). The ASB, or Absolute Standardized Bias, contrasts the mean value of the variables in the treated and matched control groups, and it is standardized using the average standard deviation in the two samples:

$$ASB = \left| \frac{\bar{X}_T - \bar{X}_C}{\sqrt{0.5 (S_T^2 + S_C^2)}} 100 \right|$$

Typically, an ASB lower than 5% ensures a good balance (Caliendo and Kopeinig 2008), as it is the case for most of our variables. The coding of the variables itself was decided trading off between quality (which would require the maximum detail possible) and balance (which is achieved when the distribution is the same in the two groups). Even when the ABS is above 5%, comparison with the values before matching (the Always Employed column) shows that for all the variables considered, matching led to a sizeable improvement in the similarity between cases and controls.

Table A.2 – Distribution of the variables used for the estimation of the propensity score, across control individuals before matching (always employed), treated (T) and matched control (C) units, according to the three definitions of retirement used, and Absolute Standardized Bias (ASB) of the mean values of the two groups

Variable	Description	Coding	Always employed	Leavers			Retirees			Drop-outs		
				T	C	ASB	T	C	ASB	T	C	ASB
Health status												
Hearing	Self-reported hearing	% fair, poor (<i>alternative:</i> excellent, very good, good)	0.09	0.13	0.12	1.5	0.13	0.12	3.9	0.13	0.13	-0.4
Eyesight	Self-reported eyesight	% fair, poor (<i>alt.</i> excellent, very good, good)	0.17	0.19	0.18	0.9	0.17	0.18	-0.8	0.22	0.22	-1.4
Mobility limitations	Mobility limitations (mobility, arm function, and fine motor limitations)	% one or more limitations in mobility (<i>alt.</i> no limitation)	0.24	0.31	0.31	-0.2	0.3	0.3	-0.1	0.28	0.27	3.5
ADL	Activities of daily living (dressing, walking, bathing, eating, getting in or out of bed, using the toilet)	% one or more limitations with activities of daily living (<i>alt.</i> no limitation)	0.03	0.04	0.04	-3.9	0.03	0.05	-6.9	0.03	0.03	-3.5
iADL	Instrumental activities of daily living (using a map, preparing a hot meal, shopping for groceries, making telephone calls, taking medications, doing work around the house or garden)	% one or more limitations with instrumental activities of daily living (<i>alt.</i> no limitation)	0.03	0.05	0.05	-1.1	0.04	0.04	1.5	0.04	0.04	1.5
Chronic diseases	Chronic conditions (heart attack, high blood pressure and cholesterol, stroke, diabetes, lung disease, asthma, osteoporosis, cancer, ulcer, parkinson and alzheimer's disease, cataracts, fractures)	% one or more chronic diseases (<i>alt.</i> no limitation)	0.55	0.65	0.67	-3.6	0.65	0.67	-4.1	0.61	0.59	3.7
GALI	Global activity limitation indicator	% non-limited (<i>alt.</i> limited)	0.24	0.31	0.3	1.4	0.29	0.29	0.7	0.3	0.31	-4.4
Grip strenght	Maximum of grip strength measures	Mean (<i>range</i> : min 3 – max 92)	40.29	39.04	38.47	4.7	39.4	39.47	-0.6	36.9	36.23	5.6
Frailty index	Frailty Index, number of accumulated deficits	% 0.0241/0.05 (<i>alt.</i> 0/0.0240)	0.27	0.25	0.23	4.6	0.27	0.27	0.4	0.22	0.22	-1.3
		% 0.051/0.075	0.18	0.18	0.19	-2.8	0.18	0.2	-4.5	0.19	0.19	0.4
		% 0.0751/0.60	0.39	0.44	0.44	-0.6	0.41	0.4	4	0.46	0.47	-0.9
Self-perceived health	Self-perceived health - US version	% fair, poor (<i>alt.</i> excellent, very good, good)	0.15	0.21	0.24	-6.7	0.19	0.19	-1.1	0.22	0.25	-8.2
Obesity	BMI	% BMI >30 (<i>alt.</i> BMI ≤ 30)	0.18	0.19	0.19	0.8	0.19	0.18	1.2	0.19	0.19	0.0
Depression	EURO-D scale (depression, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment, tearfulness)	% 3 or more depressive symptoms on the EURO-D scale	0.3	0.3	0.32	-4.3	0.27	0.26	1.5	0.37	0.4	-7.1

(continued)

Table A.2 – (continued)

Variable	Description	Coding	Always Employed	Leavers			Retirees			Drop-outs		
				T	C	ASB	T	C	ASB	T	C	ASB
<i>Health-related behaviours</i>												
Physical activity	Engagement in physical activity	% inactive (<i>alt.</i> some physical activity)	0.02	0.03	0.03	-5.2	0.02	0.03	-5.6	0.03	0.03	0
Smoking	Ever smoked daily	% ever smoked daily (<i>alt.</i> never smoked)	0.56	0.6	0.59	0.5	0.59	0.58	1.9	0.6	0.59	1.1
<i>Socio-demographic characteristics</i>												
Gender		% female (<i>alt.</i> male)	1.51	0.48	0.49	-1.1	0.46	0.44	4.1	0.61	0.64	-5.6
Age		Mean (<i>range</i> : min 50; max 70)	54.42	59.03	59.2	-4.6	59.9	60.12	-6.4	55.86	55.49	10.7
Age squared			2974.61	3498.7	3521.1	-5.3	3599.6	3628.6	-7.2	3130.9	3088.8	10.8
Education		% high, ISCED 4, 5, 6 (<i>alt.</i> low, ISCED 0, 1, 2, 3)	0.4	0.34	0.3	7.6	0.37	0.39	-4.1	0.22	0.22	-1.2
Income evaluation	Whether the household is able to make ends meet	% with (great) difficulty (<i>alt.</i> (fairly) easily)	0.71	0.75	0.71	8.8	0.79	0.77	5.2	0.62	0.62	1.2
Country	Country of residence	% Germany (<i>alt.</i> Austria)	0.04	0.06	0.06	0.5	0.06	0.04	8.9	0.08	0.08	1.8
		% Sweden	0.07	0.11	0.1	2.4	0.12	0.1	7.9	0.03	0.03	-0.7
		% Netherlands	0.06	0.07	0.06	3	0.06	0.05	4.3	0.09	0.07	4.2
		% Spain	0.05	0.06	0.08	-7	0.05	0.05	-2.9	0.15	0.19	-13.6
		% Italy	0.04	0.06	0.07	-5.6	0.05	0.07	-6.6	0.09	0.08	5.5
		% France	0.11	0.11	0.11	0.9	0.11	0.12	-2.7	0.13	0.12	3.4
		% Denmark	0.1	0.09	0.07	6.6	0.09	0.1	-2.1	0.08	0.07	3.9
		% Greece	0.05	0.01	0.02	-3.9	0.01	0.02	-4.6	0.02	0.01	1.6
		% Switzerland	0.11	0.07	0.07	-1.1	0.07	0.06	3	0.09	0.1	-3.7
		% Belgium	0.1	0.11	0.11	0.3	0.12	0.14	-8	0.07	0.06	5
		% Czech Republic	0.07	0.09	0.08	5.1	0.1	0.08	4.8	0.06	0.05	5.6
		% Poland	0.01	0.02	0.03	-7.8	0.02	0.02	-0.8	0.03	0.04	-6
		% Slovenia	0.02	0.01	0.01	0.5	0.01	0.01	1.1	0.01	0.02	-3.2
		% Estonia	0.11	0.05	0.06	-4.0	0.05	0.05	-0.1	0.04	0.04	1.1

(continued)

Table A.2 – (continued)

			Always Employed	Leavers			Retirees			Drop-outs		
Variable	Description	Coding		T	C	ASB	T	C	ASB	T	C	ASB
Family network												
Household size	Number of persons in the household	Mean (range: min 0; max 3+)	2.31	2.13	2.12	1.4	2.09	2.1	-1.5	2.27	2.25	2.7
Partnership status		% single, divorced, widowed vs having a partner	0.23	0.21	0.22	-2.4	0.21	0.22	-3.3	0.2	0.22	-6.1
Living parents		% respondent's parents still alive (alt. no)	0.57	0.37	0.4	-5.6	0.34	0.33	0.9	0.48	0.51	-5.6
Grandchildren		Mean (range: min 0; max 3+)	0.74	1.32	1.34	-1.2	1.4	1.4	-0.2	1.01	0.94	6.7
Living siblings		Mean (range: min 0; max 3+)	1.82	1.82	1.83	-0.3	1.79	1.84	-4.3	1.96	1.87	8
Number of children		Mean (range: min 0; max 3+)	1.94	1.96	1.96	-0.7	1.96	1.92	4.6	1.93	1.91	2.4
Employment characteristics												
Preference for early retirement	Thinking about your present job, would you like to retire as early as you can from this job?	% yes (alt. no)	0.39	0.51	0.5	2.3	0.5	0.49	2.9	0.5	0.54	-8.7
Physically demanding job	Agreement with the statement: My job is physically demanding.	% disagree, strongly disagree (alt. strongly agree, agree)	0.44	0.46	0.49	-6	0.44	0.44	0.5	0.5	0.49	3.3
Psychologically demanding job	Agreement with the statement: I am under constant time pressure due to a heavy workload"	% disagree, strongly disagree (alt. strongly agree, agree)	0.49	0.45	0.47	-3.9	0.45	0.49	-7.9	0.44	0.47	-6.4
Job satisfaction	Agreement with the statement: All things considered I am satisfied with my job	% disagree, strongly disagree (alt. strongly agree, agree)	0.94	0.92	0.9	7.6	0.93	0.92	4.8	0.88	0.86	7.2
Type of job	Type of job	% Civil servant (alt. employee)	0.2	0.19	0.17	5.1	0.21	0.24	-7.7	0.07	0.09	-6.6
		% Self-employed	0.16	0.14	0.18	-9.8	0.13	0.17	-9.4	0.17	0.16	2.2

