DOI: 10.1111/joes.12527

JOURNAL OF ECONOMIC SURVEYS WILEY

ARTICLE

Retirement and health outcomes in a meta-analytical framework

Mattia Filomena | **Matteo Picchio**

Department of Economics and Social Sciences, Marche Polytechnic University, Faculty of Economics, Ancona, Italy

Correspondence
Mattia Filomena, Department of Economics and Social Sciences, Marche Polytechnic University , Faculty of Economics, Piazzale Martelli 8, 60121 Ancona, Italy.
Email: m.filomena@pm.univpm.it

Abstract

This paper presents a meta-analysis on the effects of retirement on health. We selected academic papers published between 2000 and 2021 and studying the impact of retirement on physical and mental health, self-assessed general health, healthcare utilization and mortality. Our search resulted in a dataset consisting of 308 observations from 85 articles. Using meta-regression analysis and after checking for the presence of publication bias, we found that the average effect of retirement on health outcomes is very small and barely significant, under the assumption of a common true effect. We applied model averaging techniques to explore possible sources of heterogeneity of the true effect. Our findings suggest that effect heterogeneity across results is explained by the differences in both health measurements and retirement schemes. In particular, mandatory or involuntary retirement is associated with a negative impact of retirement on health, although it is small in magnitude.

KEYWORDS
health, meta-analysis, meta-regression, publication bias, retirement

JEL CLASSIFICATION
I10, J14, J26

1 INTRODUCTION

In recent years, interest in the effects of retirement on workers' physical and mental health has grown considerably, becoming a topic of interest not only in the medical or psychological field, but also among labour and health economists. For the financial sustainability of the pension systems, in most of the OECD countries the standard retirement age has indeed increased and will continue to increase in the future (OECD, 2019). Understanding the health consequences of retirement is of utmost importance to provide policy-makers with a clearer picture for the design of pension policies, labour market reforms, and healthcare investments that are welfare improving.

The identification of the causal health effects of retirement is the crux of this strand of research, and it involves methodological issues that are not easy to handle. Kuhn (2018) provides a clear non-technical summary of these methodological issues. First of all, estimation biases due to reverse causality may arise, because causality may not only run from retirement to health but is also likely to go from health to retirement decisions. Second, estimation biases may be due to measurement errors when researchers adopt subjective health measures as outcome variables. Indeed, the decision to retire early may influence the replies to the subjective answers of the interviewees, because they may assess their own health differently after retirement. This may happen for example because, when people retire, their reference group changes (Johnston & Lee, 2009). To deliver credible estimates of the causal impact of retirement on health, more recent studies have addressed endogeneity issues by means of different methodological strategies, especially using instrumental variables methods or regression discontinuity design (RDD).

Different identification strategies of the causal health effects of retirement may explain different estimates among studies. However, different findings are also explained by other reasons. For example, some recent reviews of the literature suggest that the heterogeneity in the estimated health effects of retirement depends also on the country or countries involved in the studies or the time span considered by the authors or covered by pension reforms. Furthermore, also the degree of freedom in choosing whether and when to retire matters: Bassanini and Caroli (2015), when reviewing the literature on the effect of working on health, found that both being forced to continue to work while one would like to retire and being forced to retire when one would prefer to continue working have similar adverse effects on health. They also found that voluntary retirement often has a positive effect on mental health. They concluded therefore that different findings among studies may be related to the voluntariness of the retirement decision.[1] Nishimura et al. (2018) investigated the source of differences among different studies by focusing on the methodological aspect and considering eight recent papers in the economic literature. They concluded that the key factors in explaining different results are the choice of the estimation method and the countries surveyed. They also found that their results were not sensitive to the definition of retirement. van der Heide et al. (2013) summarized 22 longitudinal studies on the health effects of retirement, describing differences in terms of voluntary, involuntary, and regulatory retirement and between blue-collar and white-collar workers. While they found strong evidence for retirement having a positive effect on mental health, their review also revealed that contradictory findings emerge when the studies use perceived general health and physical health as outcome variables. Picchio and van Ours (2020) presented a selection of most recent studies focusing on differences in set-up, identification strategy, dependent variables, and heterogeneity of the retirement effects. Pilipiec et al. (2020) investigated the empirical evidence on the effects of increasing the retirement age on the health, well-being, and labor force participation of older workers. Focusing on 19 studies, they found that the evidence that an increase of the retirement age impacts on health and well-being is scant and inconclusive, because of the heterogeneity of the retirement effect among different groups of workers, and between workers far from retirement and older workers closer to the retirement age. Finally, Zulka et al. (2019) focused on the impact of retirement on cognitive functioning by using a sample of 20 studies. They suggested that different effects may be due to different types of prior occupation.

Although detailed, the aforementioned literature reviews focus on single aspects of a multifaceted phenomenon (Kuhn, 2018) and their concluding summaries may be deceptive (Stanley et al., 2013). According to Kuhn (2018), a meta-analysis, i.e. a research methodology used to bring together in a systematic way and with a quantitative perspective all the findings from previous studies on a given issue, has the potential to yield significant insights into the factors that trigger various health effects of retirement. To the best of our knowledge, only van Mourik (2020) has taken up this challenge and proposed a meta-analysis on the effects of retirement on several measures of health by collecting 576 results from 61 manuscripts. However, this meta-analysis did not comply with the guidelines of the Meta-Analysis of Economics Research Network (MAER-Net) (Stanley et al., 2013; Havránek et al., 2020). The analysis, in fact, was built on a trinomial outcome instead of effect sizes, revealing that 15% of the studies reported negative health effects of retirement, 35% positive health effects, and 50% statistically insignificant results. Furthermore, it includes not only articles published in scientific journals, but also working papers and Ph.D. dissertations. Also Sewdas et al. (2020) have provided a meta-analysis, but with a focus limited to the link between mortality and early and on-time retirement. Using a sample of 25 studies, they estimated a random-effects meta-regression to identify the effects of retirement and to assess the influence of gender, prior health, and demographics. They concluded that early retirement, compared to continued working, is not associated with a higher risk of mortality. However, on-time retirement, compared to continued working, is associated with a higher mortality risk, which may reflect the healthy worker effect, that is, people in the group of those who work beyond the standard retirement age are on average healthier than those who retire on-time. Finally, both Pabón-Carrasco et al. (2020) and Li et al. (2021) only focus on depressive symptoms:[2] according to the former, the retirees with the highest prevalence of depression are those who retire in a mandatory fashion or due to illness; the latter show that the association of involuntary retirement with more depressive symptoms is stronger than voluntary or regulatory retirement, and it is more pronounced in Eastern developed countries.

A rigorous and extensive meta-analysis on the subject is lacking. The main contribution of our article is to fill this gap by means of a meta-analysis on the evidence of the health effects of retirement which (i) follows the MAER-Net guidelines (Stanley et al., 2013; Havránek et al., 2020); (ii) is based only on articles published in peer-reviewed journals, to reduce the probability that they contain mistakes (Xue et al., 2021), and in English, for the sake of accessibility (Vooren et al., 2019); (iii) does not focus on a particular measure of health but instead considers the ones most frequently used in the literature, such as self-reported general health, physical and mental health, healthcare utilization, and mortality; and (iv) focuses on studies published from 2000 onward in order to consider a more homogeneous labour market and pension policy background. Indeed, in most European countries the intensity of pension reforms has been particularly strong since the 2000s, with changes in eligibility criteria like the retirement age, the required contributory period, and the pension calculation scheme.[3] These changes have been implemented gradually and over long time periods. Thus, the increasing attention of policy-makers toward pension system reforms due to financial sustainability reasons and increasing workers' life expectancy after the mid-1990s has generated a large research interest among labour and health economists since the 2000s.

Our meta-analysis was carried out on 85 articles. It included the estimation of meta-regression models which enabled us to investigate the issue of publication bias and to look for patterns among different study characteristics after correcting the findings for it. We took into account all the main factors that might lead to different estimates of the effect sizes among studies, such as the institutional context, the research design, the causal effect identification strategy, and other study-related characteristics.

The rest of the paper is organized as follows. Section 2 focuses on the meta-analytical approach, describing the databases used, the research methods, and presenting preliminary and descriptive results of our meta-analysis. Section 3 assesses whether there is publication bias in this empirical literature. Section 4 provides heterogeneity analysis by using meta-regressions with the inclusion of covariates on the basis of Bayesian criteria for model selection. Section 5 concludes.

2 | META-DATASET

2.1 | Search strategy and study selection criteria

The empirical literature does not report clear-cut results on the health effect of retirement. Several reasons may explain different findings: different methodologies of analysis, different identification strategies of the causal effect, different countries, different time spans considered by the studies or covered by pension reforms. Hence, a simple comparison among the different studies and of their results may be misleading (Stanley et al., 2013). A rigorous meta-analysis enabled us to systematically review the literature by combining the results of multiple and different studies so as to identify patterns among diverse study results while taking into account the uncertainty behind each point estimate of the relation of interest and remove bias induced by publication biases. Publication bias (also named 'file drawer problem') is the bias arising from the tendency of editors to prefer to publish findings consistent with the conventional view or with statistically significant results, while studies that find small or no significant effects tend to remain unpublished (Card & Krueger, 1995).

Our search for studies followed the MAER-Net guidelines (Havránek et al., 2020). These guidelines are an attempt to create a shared subjectivity in conducting meta-analyses in economics and thereby improve the transparency, replicability and quality of the reported results. We searched studies from November 2020 to March 2021 in Ideas/EconPapers, Google Scholar, Scopus and Web of Science by using the following keywords: 'retirement', 'health' and one among 'mental health', 'physical health', 'psychological well-being', 'healthcare' and 'mortality'. We only considered articles published in peer-reviewed journals of health economics, labour economics, social sciences, psychology, and medicine and with the SCImago Journal Rank (SJR) indicator.[4] We excluded theoretical works and studies concerning only cross-partner retirement effects of retiring (Atalay & Zhu, 2018; Bloemen et al., 2019), or general life satisfaction as dependent variable (Abolhassani & Alessie, 2013; Bender, 2012; Horner, 2014; Kesavayuth et al., 2016), or only health behavior analysis (Evenson et al., 2002; Henkens et al., 2008; Zhao et al., 2017; Motegi et al., 2020).[5] Hence, we selected only micro-level studies on the health effects of retirement. We excluded 11 papers because they had not been published in peer-reviewed journals, that is, discussion papers (see e.g. Waldron, 2001; Bound & Waidmann, 2007; Coe & Lindeboom, 2008; Lalive & Staubli, 2015; Zulkarnain & Rutledge, 2018) and two book chapters (Charles, 2004; Börsch-Supan & Schuth, 2014). At this point we had 96 articles. Finally, we had to remove 11 articles because they do not contain sufficient information with which to compute the t-statistic of the estimated

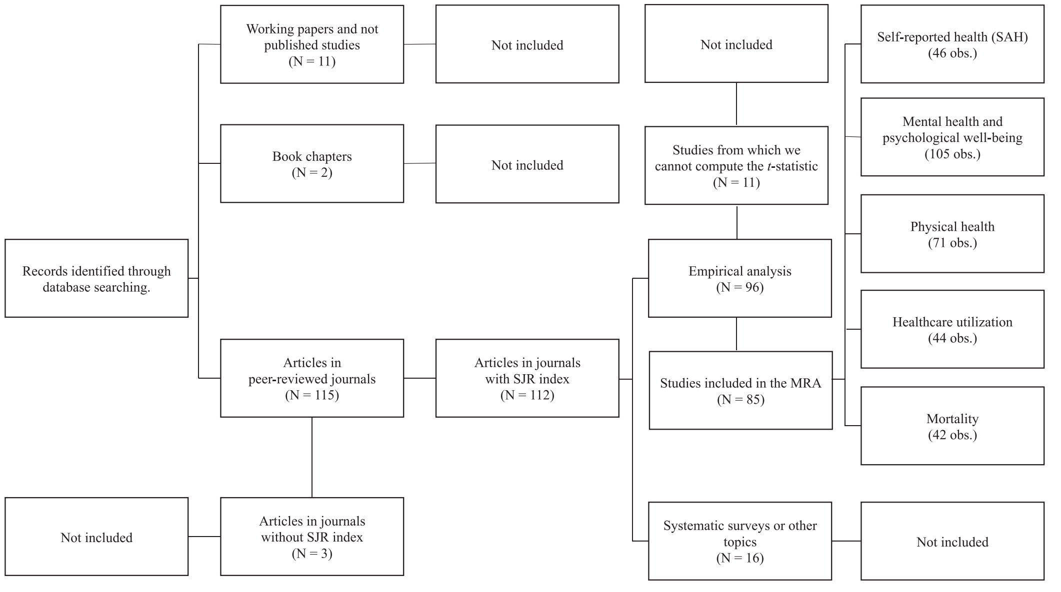


FIGURE 1 PRISMA flow diagram

retirement effect, on which we would build our meta-regressions.[6] Our final meta-analytic sample consisted of 85 articles, which are listed in Table A.1 in the appendix. Many studies dealt with the retirement effect on multiple health outcomes, and some others disaggregated the analysis by gender or by the type of previous occupation. In these cases, multiple data points were delivered and our final dataset consisted of 308 observations. Figure 1 is a PRISMA flow diagram (Moher et al., 2009): it graphically reports the rules we followed to include/exclude articles in our final sample.

From most of the articles, we directly extracted the estimated retirement effects ($\widehat{\beta}_i$) along which their standard errors ($SE_i(\widehat{\beta}_i)$) and computed the t-statistics as their ratio. In other cases, we could directly retrieve the t-statistics because they were reported among the study results. Finally, in some studies only the estimated effects and their 95% confidence intervals were displayed. In these cases, we approximated the standard errors in linear models (and then we computed the t-statistics) as follows: $SE_i = (ub - lb)/(2 \times 1.96)$, where ub and lb are the upper bound and the lower bound of the confidence interval, respectively. For studies with non-linear models, such as multinomial logit or Cox proportional hazard models, and reporting only the odds ratio (OR) and its 95% confidence interval, we calculated the standard error as $SE_i = [\ln(ub) - \ln(lb)]/(2 \times 1.96)$ and then the t-statistic as $t_i = [ln(\widehat{\beta}_{1i})\widehat{\beta}_{1i}]/SE_i$.

The health outcomes were quite different among, and sometimes within, studies. In some cases, when the sign of the coefficient of retirement was positive, this meant that there was a health improvement, like for general physical health indexes or self-assessed health. In some other cases, it was the negative sign that implied a health improvement, such as when mortality or depression were the health outcomes. We altered the sign of the t-statistics so that a ‘positive’ (‘negative’) sign means a health improvement (deterioration), and all the rest of our analysis is based on this modification of the t-statistics.

Graph (a) of Figure 2 shows the distribution of t-statistics, which is quite dispersed, with a minimum of -15.66, a maximum of 14.70, and a standard deviation of 3.13. Most of the findings (60.4%, 186 outcomes) are not significantly different from 0, having a t-statistic smaller than 1.96

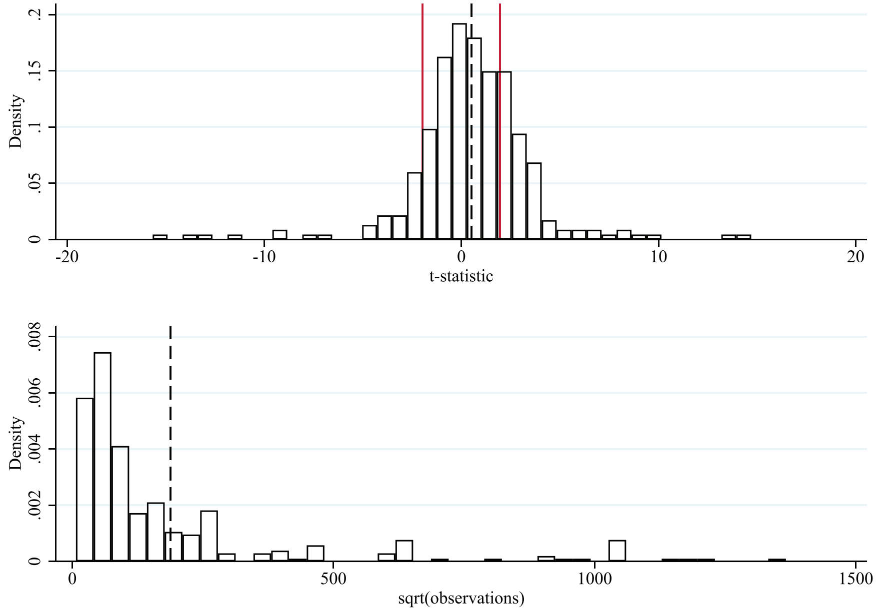


FIGURE 2 Distribution of t-statistics and observations of study outcomes
Notes: The number of study results is 308. The dashed vertical lines are the sample average of t-statistics in the upper graph (0.508) and of the square root of observations in the lower graph (188.23). The solid vertical lines in the upper graph denote the critical values for the 5% significance level in two-tailed tests (±1.96).
[Colour figure can be viewed at wileyonlinelibrary.com]

in absolute value; in 27.9% (11.7%) of the cases, 86 (36) results, the retirement effect on health is instead significantly positive (negative). Graph (b) of Figure 2 plots the distribution of the square root of the observations used to estimate the retirement effects. The number of observations is also very heterogeneous, with a minimum of 49 and a maximum of 1,866,974. Since in what follows the t-statistics and the number of observations would then be used to build a comparable measure of the estimated effect across different studies, the presence of extreme values in these two key variables raised concerns about outliers, especially because the linear models typically used in meta-regressions may be particularly sensitive to them (Viechtbauer & Cheung, 2010). As suggested by Xue et al. (2021), who had a similar problem when conducting a meta-analysis on the education effect on health, we moderated the problem by winsorization of t-statistics and number of observations at the top and bottom of their distribution: we replaced values that were lower (larger) than the 5th (95th) percentile with the value of the 5th (95th) percentile.[7]

2.2 | Descriptive statistics

We provide some basic descriptive statistics of our meta-analytic sample by research findings. Table 1 reports summary statistics by research outcomes[8] of those covariates that we used in the meta-regressions to capture the factors underlying the heterogeneous effects in the empirical literature: journal subject area, the number of citations on average per year (retrieved from Google Scholar on 05/04/2021), the journal SJR indicator at the time of publication, publication year, identification strategy, gender, institutional context, geographical area, type of previous occupa-

TABLE 1 Descriptive statistics of explanatory variables used in the meta-regressions

	Negative effect			Null effect			Positive effect		
	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.
Scimago subject areas									
Multi area (Reference category)	9	0.250	0.439	53	0.285	0.453	25	0.291	0.457
Economics/Business	12	0.333	0.478	50	0.269	0.445	26	0.302	0.462
Medicine/Psychology	15	0.417	0.500	83	0.446	0.498	35	0.407	0.494
Health outcomes									
Mortality (Reference category)	8	0.222	0.422	32	0.172	0.378	2	0.023	0.152
General and self-reported health	5	0.139	0.351	17	0.091	0.289	24	0.279	0.451
Physical health	10	0.278	0.454	47	0.253	0.436	14	0.163	0.371
Mental health	12	0.333	0.478	60	0.323	0.469	33	0.384	0.489
Healthcare utilization	1	0.028	0.167	30	0.161	0.369	13	0.151	0.360
Identification strategies									
Other methods (Reference category)	4	0.111	0.319	21	0.113	0.317	13	0.151	0.360
Regression discontinuity design (RDD)	6	0.167	0.378	33	0.177	0.383	20	0.233	0.425
Instrumental variables (IV)	16	0.444	0.504	89	0.478	0.501	47	0.547	0.501
Difference-in-differences (DiD)	–	–	–	18	0.097	0.296	5	0.058	0.235
Propensity score matching (PSM)	3	0.083	0.280	14	0.075	0.265	–	–	–
Fixed-effects/First-differences	7	0.194	0.401	11	0.059	0.237	1	0.012	0.108
Institutional contexts									
Statutory retirement (Reference category)	22	0.611	0.494	110	0.591	0.493	67	0.779	0.417
Mandatory or involuntary retirement	8	0.222	0.422	17	0.091	0.289	7	0.081	0.275
Early retirement	5	0.139	0.351	36	0.194	0.396	8	0.093	0.292
Postponed retirement	1	0.028	0.167	23	0.124	0.330	4	0.779	0.417

(Continues)

TABLE 1 (Continued)

	Negative effect			Null effect			Positive effect		
	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.
Geographical areas									
Multi-country analyses (Reference category)	4	0.111	0.319	25	0.134	0.342	12	0.140	0.349
Europe	11	0.306	0.467	92	0.495	0.501	40	0.465	0.502
Extra-European countries	21	0.583	0.500	69	0.371	0.484	34	0.395	0.492
Sex									
Males (Reference category)	12	0.333	0.478	59	0.317	0.467	31	0.360	0.483
Females	7	0.194	0.401	62	0.334	0.473	24	0.279	0.451
Males+Females	17	0.472	0.506	65	0.349	0.478	31	0.360	0.483
Calculation of t-statistic									
from 95% CI or from OR (Reference category)	4	0.111	0.319	20	0.108	0.311	16	0.186	0.391
t-statistic from $\widehat{\beta}_i/SE_i$	32	0.889	0.319	166	0.892	0.311	70	0.814	0.391
Birth cohorts									
Other (Reference category)	28	0.777	0.422	121	0.651	0.478	63	0.733	0.445
Only birth cohorts ≤ 1950	8	0.222	0.422	65	0.349	0.478	23	0.267	0.445
Type of previous occupation									
White collar (Reference category)	–	–	–	13	0.070	0.256	3	0.035	0.185
Blue collar	2	0.056	0.232	16	0.086	0.281	6	0.070	0.256
Not specified	34	0.944	0.232	157	0.844	0.364	77	0.895	0.308
Study-related characteristics									
Google Scholar citations per year	36	20.104	14.078	186	9.339	9.225	86	11.534	10.314
Scimago Journal Ranking	36	2.210	1.938	186	1.771	1.082	86	1.757	1.185
Year of publication	36	2012.861	5.117	186	2015.962	4.387	86	2015.174	4.671
Observations		36			186			86	

Notes: Females+Males = observations for which authors do not separate estimates for men and women. Other methods = OLS regressions and non-linear models (logit, multinomial logit, ordered probit and Cox proportional hazard models). The sign of the effect is based on the value of t-stat: "negative" means $t \leq -1.96$; "positive" is for $t \geq 1.96$; "null" when $-1.96 < t < 1.96$. When articles found that postponed retirement had a negative effect, we labeled the effect of retirement as "positive".

[(a)] At the time of publication, some journals did not yet have the SJR index, either because they had been published in too recent years or because the journal was not yet indexed in Scimago. In these cases, we assigned to the journal the available value of the SJR index which was chronologically closer.

tion, birth cohort, and the way in which the t-statistic was calculated. We considered three subject areas according to the Scimago classification: (i) Economics, Econometrics and Finance or Business, Accounting and Management (28.6% of our observations); (ii) Medicine or Psychology (43.2% of the observations); (iii) a residual category containing journals belonging to multiple subject areas (28.2% of the observations).[9]

The average number of yearly citations was the smallest (9.3) when the null hypothesis of no effect could not be rejected. It was instead the highest (20.1) when significant negative effects emerged and almost twice as large as the average number of yearly citations of findings supporting significant positive effects (11.5). Differences in the scientific influence of the journals where the articles had been published were smaller. In both cases, articles finding negative outcomes displayed a larger standard deviation. It is noteworthy that statistically insignificant results were not under-represented in journals of high scientific influence compared to those with more clear-cut findings; rather, they corresponded to almost 60% of our sample. This might suggest that, at a first and very descriptive level, publication bias is not an issue in this research strand.

Since health is a multidimensional concept, we referred to the main measures analysed in the empirical literature. Among the particular health measures evaluated, positive effect had the largest absolute frequency when we focused on general or self-assessed health. In all the other cases, no statistically significant effect was the prevailing outcome. These various health measures were physical health (23.1%), mental health (34.1%),[10] healthcare utilization, which included doctor visits and hospitalization (14.3%), and mortality (13.6%).

Identifying the causal effect of retirement on health is not straightforward because there are several sources of potential endogeneity of the retirement decision, such as reverse causality, negative self-selection, unobserved heterogeneity,[11] and measurement error.[12] These could affect not only the magnitude but also the sign of the estimated effect. Hence, we used a set of indicators to control for the methodology employed to identify and estimate the impact of retirement on health. The instrumental variables (IV) method was the one used most frequently (49.4%), followed by regression discontinuity design (RDD) (19.2%). The difference-in-differences (DiD) estimator was mostly used to evaluate policy reforms and represented 7.5% of our observations. In 12.3% of the study results, no particular methods was used to tackle the endogeneity of the retirement decision (e.g., linear model, multinomial logit or Cox proportional hazard models).

Some indicator variables were used to capture the institutional context and, in particular, the retirement scheme. The survey of the empirical literature provided by Bassanini and Caroli (2015) highlights the role played by choice vs. constraint in shaping the health impact of work and retirement. They focus on that strand of the literature which studies the voluntariness of retirement and from which evidence of adverse health effects arises when individuals are forced to stop working. In our analysis, we considered both the voluntariness of the retirement decisions and its timing: we distinguished among early (15.9%), postponed (9.1%), mandatory or involuntary (10.4%), and statutory retirement, that is, retiring at the standard retirement age (64.6%).

A further control variable is the gender associated with the estimated effect. The retirement effects may be different for men and women, for example because the career trajectory and the involvement in the labour market are typically different by gender. We also controlled for the geographical areas. In particular, we considered results for Europe (46.4%), for extra-European countries (40.3%), and from multi-country analyses (13.3%).

The health effects of retirement could be associated with the birth cohort because working conditions and the attention to occupational health changed during the 20th century, impacting on the physical and mental stress at work (Cullen, 1999; Harrison & Dawson, 2016) We coded the birth cohort using two dummy indicators: a dummy equal to one if the result came from individuals who had all been born before 1950 (31.2%); a dummy equal to one for results not specifying the birth cohort or covering both the period before and after 1950 (68.8%).

For similar reasons, the health effects of retirement may depend on the kind of occupation. Although very few studies provide separate estimates related to the type of previous occupation, we distinguished between blue-collar (7.8%) and white-collar workers (5.2%), and we grouped in a residual category all the other results which did not distinguish between the types of occupation (87%).

Finally, we also controlled for the method used to calculate the t-statistics. 87% of our observations were based on t-statistics derived from the ratio between $\widehat{\beta}_i$ and the corresponding standard error. The remaining 13% were derived from 95% confidence intervals or starting from odds ratios (OR).

2.3 | Comparable effect sizes

The estimated retirement effects on health $\widehat{\beta}_i$ are not easily comparable across the models used by the studies surveyed and the estimation techniques generating them. In this regard, we observed a large heterogeneity in the health measures used as outcome variables. For example, those most frequently used were self-reported general health, physical health indexes, like the body mass index (BMI) or the activities of daily living (ADL), mental health measures, like depression or the 5-item mental health inventory (MHI-5), healthcare utilization, and mortality. The units of measurement used by the studies were therefore not comparable. Moreover, even when a similar health outcome was used across studies, different model specifications and/or different estimation methods could alter their comparability. For example, although most of the estimated models were linear, in some cases nonlinear models, like multinomial logit or Cox proportional hazard models, were estimated.

To make the effect estimates comparable, we computed the partial correlation coefficient r_i, which has been commonly used in meta-analyses in economics, business and social sciences since Doucouliagos (1995). The partial correlation coefficient is a measure of the association between two variables, keeping other covariates constant. It is independent of the metrics with which the dependent and the independent variables are measured (Ugur, 2014). Very recent examples are Churchill and Mishra (2018) and Xue et al. (2021), who used the partial correlation coefficient in reviewing returns to education on the labour market and on health, respectively. Xue et al. (2020), in their meta-analysis on the health effects of social capital, used the partial correlation coefficient as a way to combine estimated effects that were not comparable because of different measures of health used and different types of econometric models estimated, as in our framework.[13]

The partial correlation coefficient is computed as

$$r_i = \frac{t_i}{\sqrt{t_i^2 + dk_i}}, \tag{1}$$

where dk_i is the degrees of freedom in the model from which the i-th t-statistic is derived. Keef and Roberts (2004) show that the estimate of r_i contains a small positive bias, since it increases as the number of independent variables in the regression model increases, that is, as the degrees of freedom decrease. However, asymptotically this bias disappears. Moreover, in our meta-dataset many studies did not provide precise information about the number of covariates. Consequently,

TABLE 2 Summary statistics of partial correlations, t-statistics, and number of observations by type of health outcome

Outcome variables used as health measures	Number of studies	Number of results	Relative frequency of results (%)	Average partial correlation (r)	Average t-statistic[(a)]	Average sample size[(a)]
Mental health	47	105	34.1	0.0095	0.8611	12,568
Physical health	30	71	23.1	0.0069	0.0334	47,394
General and self-reported health	32	46	14.9	0.0091	1.0967	17,178
Healthcare utilization	15	44	14.3	−0.0048	0.6342	289,704
Mortality	19	42	13.7	0.0004	−0.5500	290,393
Total	85[(b)]	308	100.0	0.0055	0.4807	98,761

[(a)] These averages are computed before the winsorization.
[(b)] This amount is not the sum of the absolute frequencies reported in this column, because the same study could have focused on multiple health dimensions and therefore could count in multiple lines of the same column.

we could not recover the degrees of freedom. When this was the case, we approximated dk_i with the number of observations (minus 2).[14] Because the smallest number of observations, after the aforementioned winsorization, was 523, this approximation generated a very mild upward bias which asymptotically disappeared. The standard error of the partial correlation coefficient is given by

$$SE(r_i) = \sqrt{\frac{1 - r_i^2}{dk_i}}. \tag{2}$$

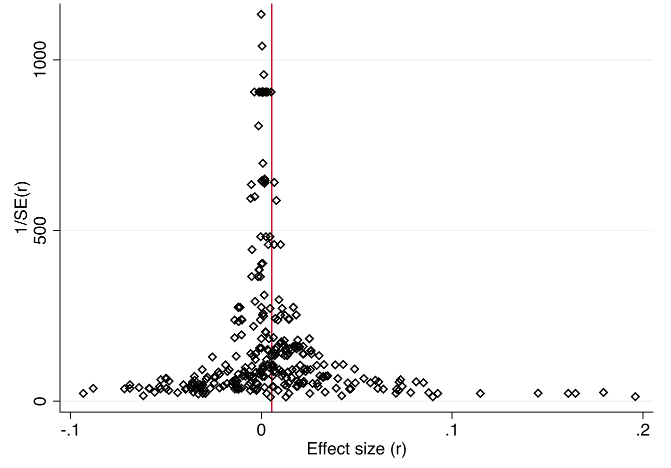
It can be shown that $r_i/SE(r_i) = t_i$.

The partial correlation coefficient r is a unitless measure, which takes a value between −1 and 1. It enables direct comparisons among the different ways to approach and measure health outcomes in the empirical literature and in the diverse literatures (Doucouliagos & Laroche, 2009). The partial correlation coefficient drops as the degrees of freedom or the sample size increase. This implies that nearly similar t-statistics will produce very different partial correlations if the sample sizes are diverse: the larger the sample size, the more the effect size measured by the partial correlation is scaled down.

Table 2 displays summary statistics of partial correlations, t-statistics, and number of observations of the full sample and of the results by the type of health measure. As in Xue et al. (2020) and Xue et al. (2021), we included in our meta-analysis different types of health measures, ranging from physical health, mental health, self-reported general health and healthcare service utilization. One may wonder whether we mixed together outcomes which measured too diverse phenomena. On the one hand, one of the aims of Section 4 is to understand if such heterogeneity is related to the findings, and this source of diversity was explicitly taken into account in the meta-regression analysis. On the other hand, we will do it partially, because in the specification of the meta-regression models we imposed that the impact of all the other covariates was not a function of the particular measure of health. Dividing the sample into as many subsamples as the five different measures of health would result in small sample size problems for some of them.

The graph in Figure 3, known as funnel plot (Light & Pillemer, 1984), shows the scatter plot of the partial correlation coefficient and its precision, measured by the inverse of its standard

FIGURE 3 Funnel plot of precision ($1/SE(r)$) versus effect size (r)
Notes: The number of observations is 308. The vertical line is the average of the partial correlation coefficients r (0.0055). [Colour figure can be viewed at wileyonlinelibrary.com]



error as defined in Equation (2). In the absence of publication bias, the partial correlation coefficient should vary randomly around its average, which is an estimate of the true effect. Hence, the symmetry of the funnel around the average effect is of help in graphically visualizing a possible publication bias (Stanley, 2005). The funnel plot shows a mild asymmetry, given the longer tail to the right of the average partial correlation coefficient. It is not easy to reach a conclusion about publication bias by means of this graphical approach. Indeed, it relies on the assumption that there is a single 'true' effect common to all empirical studies. Hence, if there is heterogeneity among articles due to different datasets, time spans, countries or methodologies, it may cause the funnel's skewness. In this case, the funnel plot seems to suggest that there is not an evident publication bias. However, in the next section, on the basis of Meta-Regression Analysis (MRA), we will formally test for the presence of publication bias.

3 | TESTING FOR PUBLICATION BIAS

To formally assess the relevance of publication bias and to eventually remove it from the estimate of the genuine retirement effect on health, we used the "Funnel Asymmetry Test – Precision Effect Test" (FAT-PET) (Egger et al., 1997; Stanley, 2005, 2008), which is a standard model to assess the presence of publication bias. Used since the end of the 1990s in the economic literature (Card & Krueger, 1995; Ashenfelter et al., 1999; Görg & Strobl, 2001), it is based on a simple regression of the i-th effect size on a constant and its standard error:

$$r_i = \gamma_1 + \gamma_0 SE(r_i) + \varepsilon_i, \tag{3}$$

where ε_i is the idiosyncratic error terms and γ_0 will be equal to zero when the effect size r_i varies randomly around the precision effect γ_1, meaning no publication bias. Publication bias is proportional to the inverse of the square root of the sample size, which in turn is proportional to the standard error (Begg & Berlin, 1988). The Funnel Asymmetry Test (FAT) tests the hypothesis of no publication bias (Egger et al., 1997), that is, $H_0 : \gamma_0 = 0$, and is therefore also a test of funnel asymmetry (Sutton et al., 2000). If the null hypothesis is rejected, a publication bias is affecting this strand of the literature, potentially posing a serious problem for interpretation of the scientific research (Begg & Berlin, 1988). The Precision Effect Test (PET) tests the null hypothesis

TABLE 3 FAT-PET and PEESE tests and corrections for publication bias

	FAT-PET								PEESE[(c)]	
	(1)		(2)		(3)		(4)		(5)	
	OLS		WLS-FE		WLS-FE[(a)]		FAIVE[(b)]		WLS-FE	
Publication bias	0.487	(0.384)	0.409*	(0.227)	0.414*	(0.233)	0.282	(0.269)	10.145	(7.768)
Precision effect	−0.002	(0.004)	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)	0.001**	(0.001)
R^2	0.034		0.023		0.023		0.028		0.009	

Standard errors robust heteroskedasticity and within-study correlation are in parentheses. The number of observations (studies) is 308 (85).

[(a)] The inverse of the square root of the sample size is used instead of $SE(r_i)$ as a precision measure.

[(b)] The F-statistic for the power of the excluded instrument is 32.97.

[(c)] PEESE gives a less biased correction for publication bias (Stanley & Doucouliagos, 2012, 2014).

$H_0 : \gamma_1 = 0$. The rejection of the null hypothesis can be interpreted as the presence of an authentic empirical effect, corrected for publication selection: when the sample size goes to infinity and the standard error goes to 0, the observed effects goes to γ_1 (Stanley, 2008).

Table 3 displays the results of different estimation and specifications of Equation (3). Model (1) reports the ordinary least squares (OLS) estimates of Equation (3), without taking advantage of the known form of heteroskedasticity affecting the distribution of r_i, as seen in Equation (2). This knowledge is instead exploited in Model (2), which displays the results when Equation (3) is estimated by Weighted Least Squares (WLS-FE) using $1/SE(r_i)^2$ as weights. Models (3) and (4) are robustness checks. In Model (3) we replicate our simple FAT-PET estimates by replacing $SE(r_i)$ with the inverse of the square root of the sample size as an alternative precision measure. Because the sample size is not subject to estimation error, it avoids an errors-in-variables bias that could instead affect $SE(r_i)$. If $SE(r_i)$ is endogenous in Models (1) and (2) because it is affected by measurement error, we may solve the problem by using an IV approach, instrumenting $SE(r_i)$ with the square root of the number of observations, which is strongly correlated to the standard error but should not be able to explain the estimated effect once we control for the standard error. This is called the Funnel Asymmetry Instrumental Variable Estimator (FAIVE) by Stanley (2005). Finally, in Model (5) we report the results if in Equation (3) we replace $SE(r_i)$ with its square to capture eventual non-linearities: this is the Precision Effect Estimate with Standard Error (PEESE) model, which is a meta-regression method to be preferred in correcting for publication bias when there is a genuine nonzero effect (Stanley & Doucouliagos, 2012, 2014).

From the five models reported in Table 3, we find weak evidence of publication bias only in the FAT-PET model estimated by WLS-FE. Furthermore, the FAT-PET point estimates of γ_0, ranging from 0.282 to 0.487, suggest that, if it exists, the publication bias is positive and small.

The precision coefficient is equal to 0.001 and significant only in the PEESE model. Hence, the mean effect of retirement on health is positive. However, it is extremely low, considering that, according to Cohen (1988), a partial correlation coefficient of 0.1 is to be considered as 'small', and in the analysis of Doucouliagos (2011), who focused on economic results, it should be at least 0.07 to be considered as 'small'.[15]

The recognition of publication bias as a threat to the reliability of the scientific knowledge has taken place at different times in different disciplines. For example, psychological and medical research has acknowledged it since the end of the 1950s (Sterling, 1959; Rosenthal, 1979; Begg & Berlin, 1988). The economic research has taken instead some more years, until the 1990s (see e.g. Card & Krueger, 1995; Ashenfelter et al., 1999). Therefore, one might wonder whether researchers and journal editors have different sensitivities to the problem across different disciplines, resulting in publication bias being limited only to some disciplines. To check whether this might be the case, we distinguished the study results into three broad subject areas: medicine/psychology, economics/business, and a residual category. Then, we generalized Equation (3) by having one constant per each subject area and the standard error interacted with the subject area indicator. We found that publication bias does not arise in any of the separate subject areas. We report the tests for publication bias by subject area in the online appendix.

In recent years, further techniques have been developed to detect publication bias. In the online appendix, we present the findings of the Endogenous Kink (EK) meta-regression model (Bom & Rachinger, 2019). This attempts to better fit the non-linearity of the relationship between the estimated effect and its standard error in the presence of publication bias using a piecewise linear model instead of a quadratic term. We also followed the suggestion by Andrews and Kasy (2019) to focus on the distribution of p-values or t-statistics across published studies. Indeed, if there is no publication bias, the distribution of the t-statistics and p-values should not display discontinuities, especially at critical values, like ± 1.96 for the former and 0.05 for the latter. From these further tests, which are displayed in the online appendix, we did not detect evidence of publication bias.

To sum up, after a battery of tests, we concluded that publication bias is not importantly affecting this strand of the literature, and that the mean effect of retirement on health is positive but very close to zero. The next meta-regressions reported were aimed at understanding possible heterogeneity among studies in the retirement effect on health. We kept the PEESE specification as the benchmark model, so as to correct for publication bias when multiple covariates were included in the model specification.

4 | MULTIPLE META-REGRESSIONS

To detect possible sources of heterogeneous effects of retirement on health, we included in the PEESE specification a series of covariates: measures of health, methods to identify the effect, institutional contexts, geographical areas, gender, year of publication, SJR index, the average number of Google scholar citations per year, type of previous occupation, birth cohorts, and the way in which we derived the t-statistics. We employed the PEESE specification because its quadratic form of the standard errors has been proven to be less biased and often more efficient than the FAT-PET specification when there is a nonzero genuine effect (Stanley & Doucouliagos, 2014).[16]

Formally, we estimated by WLS-FE the following equation for our effect size

$$r_i = \gamma_1 + \gamma_0 SE(r_i)^2 + \boldsymbol{\beta}_1 \mathbf{x}_i + \varepsilon_i, \tag{4}$$

which is equivalent to estimating by OLS the transformed model

$$\frac{r_i}{SE(r_i)} = \gamma_1 \frac{1}{SE(r_i)} + \gamma_0 SE(r_i) + \boldsymbol{\beta}_1 \frac{\mathbf{x}_i}{SE(r_i)} + \frac{\varepsilon_i}{SE(r_i)}, \tag{5}$$

where $\mathbf{x}_i$ is the vector of result characteristics.

A problem in estimating Equation (5) is related to the model uncertainty about which variables should be included. We overcame it by employing one of the most commonly used tools in meta-analysis, Bayesian Model Averaging (BMA). BMA takes into account all possible models by running many regressions with different subsets of control variables and computing the weighted averages of the estimated coefficients. The weights are Posterior Model Probabilities (PMP) and are related to the goodness of fit of each model. The sum of PMPs indicates the Posterior Inclusion Probability (PIP) for each regressor, which provides the information on the likelihood of the regressor belonging to the true specification. A PIP above 0.5 for a given regressor is usually used as a rule of thumb to include it in the final model (Eicher et al., 2011). For each covariate, BMA returns the posterior coefficient distribution, which yields the posterior mean (PM) of the regression coefficient and the posterior standard deviation (PSD).

We used the BMA estimator discussed by Magnus et al. (2010), who introduced the distinction between two subsets of explanatory variables. The first subset is the set of 'focus' regressors, which are those wanted in the model for theoretical (or other) reasons. In our case, the focus variables were those capturing the publication bias and the precision effect. The second subset is the set of "auxiliary" regressors, which are additional covariates that may be relevant to explaining the estimated effect, but this is not certain. Since we had 25 auxiliary covariates, the number of possible models to be considered was 2^{25}. BMA proceeds by applying conventional non-informative priors on the focus variables and the error variance σ^2, and an informative multivariate Gaussian prior on the auxiliary variables.

In a subsequent step, we performed a model-average procedure by using the Weighted Average Least Squares (WALS) (Magnus et al., 2010). WALS occupies an intermediate position between the Bayesian approach of BMA and the frequentist model-averaging procedure. In fact, it is a Bayesian combination of frequentist estimators (Magnus & De Luca, 2016). WALS uses conventional non-informative priors on the focus regressors and the error variance σ^2 and a distribution with zero mean for the independent and identically distributed elements of the t-ratios associated with linear combinations of the auxiliary regressors.[17] Unlike BMA, WALS relies on preliminary orthogonal transformations of the auxiliary regressors and their parameters, which reduce the computational burden from 2^{25} to 25. For this reason, WALS does not allow computation of the PIPs. An auxiliary covariate is considered to be robustly correlated with the outcome variable if the t-ratio of its coefficient is greater than 1 in absolute value or, equivalently, if the corresponding one-standard error band does not include zero (De Luca & Magnus, 2011). The advantage of WALS over BMA is that it does not impose an ad hoc assumption on the prior on the model space (in general BMA uses a uniform prior assigning equal probability to each model), but it is theoretically based (Magnus & De Luca, 2016).

Finally, like Havranek et al. (2015) and Xue et al. (2021), we conducted a frequentist check by estimating Equation (5) by OLS after restricting the set of regressors to those with PIP > 0.5 according to BMA. We ran the same frequentist check after the WALS estimates.

Table 4 reports the estimation results. For the BMA, we show the estimated posterior means, the posterior standard deviations, and the posterior inclusion probabilities of each regressor. For the WALS, we include the results deriving from two different assumptions about the model prior distributions. In the last columns of Table 4, we present the findings of the frequentist checks.

As regards the focus regressors, whilst for these variables the Posterior Inclusion Probabilities from BMA model are not informative, OLS estimates reveal no publication bias, even after controlling for a set of covariates. According to BMA results, there are six auxiliary covariates which are significant in explaining the heterogeneous effects of retirement on health (PIP > 0.5): two measures of health outcomes, fixed-effects/first-difference estimator, mandatory or involuntary retirement, year of publication and the dummy for the birth cohort. WALS results are quite similar, although some further covariates seem to be important: physical health and healthcare utilization, postponed retirement, the SJR indicator, estimates not distinguishing between males and females, RD design, and PSM estimator.

TABLE 4 Heterogeneity in the estimated effects of retirement on health

	Bayesian Model Averaging[(a)]			Weighted-Average Least Square				OLS check after BMA[(c)]		OLS check after WALS[(d)]	
				$(q = 1)$[(b)]		$(q = 0.5)$[(b)]					
	PM	**PSD**	**PIP**	**Coeff.**	**Std. Error**	**Coeff.**	**Std. Error**	**Coeff.**	**Std. Error**	**Coeff.**	**Std. Error**
Focus Regressors											
Publication bias	**8.635**	**5.761**	**1.000**	**8.274**	**5.657**	**8.131**	**5.682**	9.536	8.962	8.809	8.246
Precision effect	**0.004**	**0.003**	**1.000**	**0.013**	**0.004**	**0.013**	**0.004**	0.002***	0.001	0.008**	0.004
Auxiliary Regressors											
Google Scholar citations per year	0.000	0.000	0.220	0.000	0.000	0.000	0.000	–	–	–	–
Scimago Journal Ranking	−0.000	0.001	0.320	**−0.002**	**0.001**	**−0.002**	**0.001**	–	–	−0.002***	0.001
Year of publication	**0.001**	**0.000**	**0.550**	**0.001**	**0.000**	**0.001**	**0.000**	0.000***	0.000	0.001***	0.000
Scimago subject areas (Reference category: Multi-area journals)											
Economics/Business	−0.000	0.000	0.060	0.000	0.002	0.000	0.002	–	–	–	–
Medicine/Psychology	−0.000	0.000	0.070	0.001	0.002	0.000	0.002	–	–	–	–
Health outcomes (Reference category: Mortality)											
General and self-reported health	**0.012**	**0.002**	**1.000**	**0.010**	**0.002**	**0.010**	**0.002**	0.011***	0.002	0.012***	0.002
Physical health	0.001	0.001	0.160	**0.002**	**0.001**	**0.001**	**0.001**	–	–	0.002*	0.001
Mental health	**0.009**	**0.002**	**1.000**	**0.007**	**0.002**	**0.007**	**0.002**	0.009***	0.002	0.008***	0.003
Healthcare utilization	0.001	0.002	0.400	**0.001**	**0.001**	**0.001**	**0.001**	–	–	0.002**	0.001
Identification strategies (Reference category: Other methods)											
Regression discontinuity design (RDD)	0.000	0.001	0.120	**−0.006**	**0.004**	**−0.005**	**0.004**	–	–	−0.002**	0.001
Instrumental variables (IV)	0.000	0.001	0.070	−0.004	0.005	−0.003	0.005	–	–	–	–
Difference-in-differences (DiD)	0.000	0.001	0.050	−0.001	0.004	−0.001	0.004	–	–	–	–
Propensity score matching (PSM)	−0.000	0.002	0.050	**−0.010**	**0.009**	**−0.010**	**0.009**	–	–	−0.008	0.009
Fixed-effects/First-differences	**−0.013**	**0.003**	**1.000**	**−0.014**	**0.005**	**−0.014**	**0.005**	−0.013***	0.004	−0.013***	0.003

(Continues)

TABLE 4 (Continued)

	Bayesian Model Averaging[a]			Weighted-Average Least Square				OLS check after BMA[c]		OLS check after WALS[d]	
				($q = 1$)[b]		($q = 0.5$)[b]					
	PM	PSD	PIP	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Institutional contexts (Reference category: Statutory retirement)											
Mandatory or involuntary retirement	**−0.025**	**0.009**	**0.960**	**−0.021**	**0.007**	**−0.022**	**0.007**	−0.026***	0.008	−0.026***	0.007
Early retirement	−0.000	0.001	0.110	−0.001	0.001	−0.001	0.001	–	–	–	–
Postponed retirement	−0.001	0.002	0.230	**−0.005**	**0.002**	**−0.006**	**0.002**	–	–	−0.006***	0.002
Geographical areas (Reference category: Multi-country analyses)											
Europe	−0.000	0.001	0.060	**−0.002**	**0.002**	**−0.002**	**0.002**	–	–	−0.002	0.003
Extra-European countries	−0.000	0.001	0.060	**−0.003**	**0.002**	**−0.003**	**0.002**	–	–	−0.003	0.003
Sex (Reference category: Males)											
Females	0.000	0.000	0.050	0.000	0.001	0.000	0.001	–	–	–	–
Males+Females	−0.000	0.001	0.070	**−0.002**	**0.002**	**−0.002**	**0.002**	–	–	−0.002	0.002
Calculation of t-statistic (Reference category: from 95% CI)											
t-statistic from $\widehat{\beta}_i/SE_i$	−0.001	0.002	0.230	−0.002	0.003	−0.002	0.003	–	–	–	–
Type of previous occupation (Reference category: White collars)											
Blue-collar	0.000	0.004	0.040	0.001	0.002	0.001	0.002	–	–	–	–
Not specified	0.000	0.000	0.050	0.001	0.001	0.001	0.002	–	–	–	–
Birth cohorts (Reference category: Others)											
Birth cohorts ≤ 1950	**−0.002**	**0.002**	**0.650**	−0.001	0.001	**−0.002**	**0.001**	−0.003***	0.001	–	–

Notes: The results are from the PEESE specification using the inverse of the SE_i^2 as weights. PM = Posterior Mean of the coefficient; PSD = Posterior Standard Deviation; PIP = Posterior Inclusion Probability. The number of observations (studies) is 308 (85). Auxiliary variables for which the PIP is above 0.5 in BMA or the corresponding one-standard error band does not include zero in WALS are in bold. *** Significant at 1%, ** significant at 5%, * significant at 10%.

[a] In the BMA, we use the uniform distribution for model priors, the Zellner's g prior for the distributions of the coefficients and a Markov Chain Monte Carlo algorithm to search over the model space, by distinguishing between focus and auxiliary regressors.

[b] $q = 1$ indicates the Laplace model prior distribution; $q = 0.5$ implies the Subbotin model prior distribution.

[c] The model specification under "OLS" includes those variables which have a PIP > 0.5 in BMA ($R^2 = 0.29$).

[d] The second model specification under "OLS" includes those variables which are relevant according to WALS ($R^2 = 0.35$).

All the models reveal that the studies which used general and self-reported health indicators or mental health measures were the ones most likely to report positive effects of retirement on health. The analyses focusing on physical health or healthcare utilization were more likely to find positive effects than those dealing with mortality, although the difference in terms of correlation points was negligible. These findings reflect the results of some earlier systematic surveys in this field: as pointed out by Bassanini and Caroli (2015) or suggested by Nishimura et al. (2018) after re-estimating previous analyses, most of the evidence concerning the health effects of retirement shifts towards a positive impact on physical and mental dimensions of health, better self-assessed health, and lower healthcare utilization.

The results for the identification strategy suggest that the heterogeneity across this dimension is not particularly important in explaining different findings. We find that only those studies using a fixed-effects or a first-differences approach are more likely to report negative effects on health. This finding contrasts with the one reported by Nishimura et al. (2018), who instead showed that the choice of the estimation strategy is one of the key factors in explaining why the estimated results of the retirement effect on health differ.

An important factor in explaining heterogeneous estimated effects of retirement on health is the institutional context and the retirement scheme: mandatory or involuntary retirement has a PIP close to 1 and the greatest negative effect in magnitude. In the WALS results and, although with a lower magnitude, studies focusing on postponed retirement are also associated with a lower chance of detecting positive retirement effects than are studies dealing with early or statutory retirement. These findings confirm the conclusions of Bassanini and Caroli (2015), who showed that being forced to work while preferring to retire and, symmetrically, being forced to stop working because employees have no control on the retirement and work decisions have a health damaging effect. Similar results are reported by Pabón-Carrasco et al. (2020) and Li et al. (2021), but only on the effects on depressive symptoms. Moreover, the negative impact of postponed retirement on health, compared to statutory retirement, may reflect the consequences of being stuck in employment while one had planned to retire, for example due to pension reforms which raise the retirement age or the length of the contribution period required for entitlement to a pension (see e.g. Blake & Garrouste, 2019; Shai, 2018).

Regarding the publication year, we find that the estimated effects of retirement on health tend to be more and more positive over time: the year of publication presents a PIP = 0.55 and a positive and significant coefficient. As regards study-quality measures, WALS estimates reveal that the SJR indicator is negatively correlated to the partial correlation coefficient, meaning that the more positive the detected relation between retirement and health, the lower the SJR index of the journal where the result was published. Finally, the health effects of retirement are independent of geographical area, gender, and the previous type of occupation. Concerning this last, it should be taken into account that the number of study results distinguishing between blue- and white-collar workers is fairly low. Hence, our meta-analysis is not endowed with the statistical power to shed light on this particular source of heterogeneity.

Finally, the coefficient of the dummy for study results coming from individuals who were born before 1950 is significant and negative. This means that when studies include more recent cohorts in their samples, the retirement effect on health is more likely to be positive. Nevertheless, the difference is very small.

The results presented in Table 4 suggest sources of heterogeneity in the study results. However, it is not easy to visualise from it if for particular combinations of study features the expected retirement effect is significantly positive or significantly negative. To be more informative from this point of view, we used the OLS estimates from the frequentist check after BMA and computed the expected partial correlation coefficients for all the combinations of the covariates, after fixing the publication year to the median and setting γ_0 to zero, so as to mimic the absence of publication bias.

Table 5 displays the expected partial correlation coefficients for all the combinations of the explanatory variables. We find that, for the most frequent combination involving mental health as outcome variable (third line of Table 5), retirement has a positive and highly significant impact, with a partial correlation coefficient equal to 0.010. The covariate profile with the largest positive predicted partial correlation coefficient (0.013) has general and self-reported health as outcome variable. According to the classifications in Cohen (1988) or Doucouliagos (2011), which set to 0.1 and 0.07, respectively, the size of the partial correlation coefficient to be considered as 'small', the detected magnitudes are very modest. In the case of physical health or healthcare utilization or mortality, the predicted average effect for the chosen combinations of covariates is even closer to zero. Finally, regardless the health outcome, when a study focuses on mandatory or involuntary retirement, we predict an expected negative effect between -0.029 and -0.013.

5 | CONCLUSIONS

We summarized the literature on the impact of retirement on health using meta-analytic techniques. Our meta-sample consisted of 308 observations from 85 articles published in peer-reviewed journals in the period 2000–2021. Among these findings, 28% supported the hypothesis according to which retirement improves health; 60% provided no statistically significant effects; and only 12% reported evidence in favour of a worsening health status after retirement.

In a first step, we checked for the presence of publication bias under the assumption of a common effect and by using a battery of meta-regression based techniques and by considering the distribution of the t-statistics and p-values at critical values. We did not find evidence for publication bias. The average retirement effect is extremely small, considering the figures suggested by Cohen (1988) or Doucouliagos (2011) to value the size of a partial correlation coefficient as 'small'.

We then used model averaging strategies to explore possible sources of effect heterogeneity across several study characteristics, like research design, estimation strategy, institutional context, and type of previous occupation. Our results suggest that the different reported estimates are linked to the differences in health outcomes used by studies. The identification/estimation strategy does not appear to be particularly important for explaining heterogeneous findings, although studies which opted for fixed-effects or first-differences tended to report more negative estimated effects. Finally, a further source of heterogeneity is the type of retirement scheme. Compared to standard retirement, mandatory/involuntary retirement and, to a lesser extent, postponed retirement are associated with more negative health outcomes.

These findings have important implications for public policy, especially because many OECD countries still adopt mandatory retirement ages (OECD, 2017, Section 2.4) and are rising further their retirement age (OECD, 2019). Although we find that the effect of retirement on health outcomes is in general very small in magnitude, the predicted precision effects for different combinations of covariates displayed in Table 5 suggest that having no choice about the timing of retirement and being involuntarily retired (i.e. the category that we called "mandatory or involuntary retirement") may have some health damaging implications. Policy-makers should consider not only the financial sustainability of the pension system, but also the raising healthcare spending due to the negative impact of mandatory or involuntary retirement. Optimal welfare pension

TABLE 5 Expected partial correlation coefficients of the health effect of retirement for all the combinations of covariates (covariates not mentioned in each line are fixed at the reference)

	Coeff.	Std. Err.	*p*-value	Frequencies Abs.	Frequencies Rel. (%)
Physical health or healthcare utilization or mortality	0.002***	0.001	0.001	73	23.70
Physical health or healthcare utilization or mortality + birth cohorts ≤ 1950	−0.001*	0.001	0.064	66	21.43
Mental health	0.010***	0.002	0.000	65	21.10
General and self-reported health	0.013***	0.002	0.001	28	9.09
Mental health + mandatory or involuntary retirement	−0.016**	0.008	0.045	17	5.52
Mental health + birth cohorts ≤ 1950	0.007***	0.002	0.001	15	4.87
General and self-reported health + birth cohorts ≤ 1950	0.010***	0.002	0.000	11	3.57
Physical health or healthcare utilization or mortality + fixed-effects/first-differences	−0.011***	0.003	0.001	10	3.25
Physical health or healthcare utilization or mortality + mandatory or involuntary retirement	−0.024***	0.008	0.002	6	1.95
Mental health + fixed-effects/first-differences	−0.003	0.004	0.528	5	1.62
General and self-reported health + mandatory or involuntary retirement	−0.013*	0.008	0.093	4	1.30
General and self-reported health + fixed-effects/first-differences	0.000	0.004	0.995	3	0.97
Mental health + mandatory or involuntary retirement + birth cohorts ≤ 1950	−0.019**	0.008	0.019	2	0.65
Physical health or healthcare utilization or mortality + mandatory or involuntary retirement + birth cohorts ≤ 1950	−0.027***	0.008	0.001	2	0.65
Mental health + mandatory or involuntary retirement + fixed-effects/first-differences≤ 1950	−0.029***	0.009	0.001	1	0.33

Note: *** Significant at 1%, ** significant at 5%, * significant at 10%. Year of publication is normalized at its median value and γ_0 is set to zero.

policies should ensure workers a greater degree of freedom in choosing whether to retire and the timing of their retirement.

Finally, as suggested by Kuhn (2018), there are reasons to suspect that the health effects of retirement are heterogeneous across dimensions, such as different types of prior occupation (e.g. blue- vs. white-collar workers), different types of physically/mentally demanding previous jobs, time horizons or health behaviours, which are only partially investigated in our paper. We have tried to shed light on whether retirement differently affects blue- and white-collar workers. However, only a very limited number of the studies surveyed distinguished between blue- and white-collar workers. Hence, our meta-analysis has very limited statistical capacity to provide answers on this issue. We therefore conclude with a research suggestion: future research should take these further dimensions into account to gain a clearer picture of the multifaceted nature of the effects of retirement on health.

ACKNOWLEDGMENTS

We thank Jan van Ours for his comments. We also thank two anonymous reviewers and the Associate Editor for their comments and suggestions. Mattia Filomena acknowledges financial support from the Cariverona Foundation Ph.D. research scholarship.

ORCID

Mattia Filomena https://orcid.org/0000-0002-4099-9168
Matteo Picchio https://orcid.org/0000-0001-6996-4957

ENDNOTES

[1] To study the health effects of retirement, Bassanini and Caroli (2015) refer to 14 studies: five of them report negative effects of retirement on health.

[2] Pabón-Carrasco et al. (2020) collect a total of 11 articles, while Li et al. (2021) have a sample of 25 longitudinal studies.

[3] Carone et al. (2016) report that the average number of pension measures per year in Europe was less than 10 during the late 1990s and rose to 44 between 2009 and 2014.

[4] See https://www.scimagojr.com/SCImagoJournalRank.pdffor details on the calculation of the SJR. The following studies were not included in the final sample because their journals are not indexed in SCImago: Lee and Smith (2009), Fonseca et al. (2014), and Sonet et al. (2020).

[5] Drinking, smoking and physical activity are examples of health behavior outcomes.

[6] These 11 articles are: Allen and Alpass (2020), Barban et al. (2020), Carlsson et al. (2012), Dufouil et al. (2014), Finkel et al. (2009), Fisher et al. (2014), Kühntopf and Tivig (2012), Mazzonna and Peracchi (2012), Nishimura et al. (2018), Olesen et al. (2014), Rohwedder and Willis (2010).

[7] We replicated the empirical analysis without winsorization as a sensitivity analysis. Our findings were unchanged. We report estimation results without winsorization in the Online appendix.

[8] In Table A.2 in the appendix, we report similar summary statistics by the sign of the relation between retirement and health.

[9] This category also comprises two observations by Kalwij et al. (2013), the only article in our sample published in a social-sciences journal.

[10] Physical health included chronic conditions, mobility, body mass index (BMI), activities of daily living (ADL) and a measure of general physical status. Mental health consisted of cognitive functioning, depression or anxiety, and a more general measure which comprised general mental health index and psychological well-being (in this case, it also comprised happiness as a proxy for well-being).

[11] Omitted variables biases might be induced by differences in unobserved individual characteristics that influence both health and retirement decisions (e.g., subjective life expectancy). Unobserved heterogeneity could be time-

constant but also time-varying. To control for unobserved time-constant individual heterogeneity, researchers typically use individual fixed-effects panel data models (Eibich, 2015).

[12] Self-reported health measures are at risk of two kinds of measurement error: (i) self-assessed health may not be comparable across individuals ("classical measurement error"); (ii) individuals who do not work may justify their labour market status by their ill health ("justification bias"). The latter refers to retirees' tendencies to exaggerate their poor health conditions in order to provide socially acceptable justification for their retirement and observed health would be understated for retirees (Behncke, 2012; Insler, 2014).

[13] See Reed (2020) and the meta-analyses cited therein for other examples of meta-analyses using the partial correlation coefficient as the effect size.

[14] See Table B.11 in Lipsey and Wilson (2001).

[15] In Doucouliagos (2011), 0.17 is the threshold for 'moderate' and 0.33 for 'large'.

[16] Table A.3 in the appendix displays the results of the FAT-PET specification. The results are very similar to the ones from the PEESE model.

[17] The prior distribution of the t ratios can be either a neutral Laplace prior (Magnus et al., 2010), or a neutral Subbotin prior distribution (Einmahl et al., 2011).

REFERENCES

Abolhassani, M., & Alessie, R. (2013). Subjective well-being around retirement. *De Economist*, *161*(3), 349–366.

Allen, J., & Alpass, F. M. (2020). Trajectories of material living standards, physical health and mental health under a universal pension. *Journal of Epidemiology and Community Health*, *74*(4), 362–368.

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8), 2766–2794.

Apouey, B. H., Guven, C., & Senik, C. (2019). Retirement and unexpected health shocks. *Economics & Human Biology*, *33*, 116–123.

Ardito, C., Leombruni, R., Blane, D., & d'Errico, A. (2020). To work or not to work? The effect of higher pension age on cardiovascular health. *Industrial Relations: A Journal of Economy and Society*, *59*(3), 399–434.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, *6*(4), 453–470.

Atalay, K., & Barrett, G. F. (2014). The causal effect of retirement on health: new evidence from Australian pension reform. *Economics Letters*, *125*(3), 392–395.

Atalay, K., Barrett, G. F., & Staneva, A. (2019). The effect of retirement on elderly cognitive functioning. *Journal of Health Economics*, *66*, 37–53.

Atalay, K., & Zhu, R. (2018). The effect of a wife's retirement on her husband's mental health. *Applied Economics*, *50*(43), 4606–4616.

Bamia, C., Trichopoulou, A., & Trichopoulos, D. (2008). Age at retirement and mortality in a general population sample: The Greek EPIC study. *American Journal of Epidemiology*, *167*(5), 561–569.

Barban, N., De Luna, X., Lundholm, E., Svensson, I., & Billari, F. C. (2020). Causal effects of the timing of life-course events: Age at retirement and subsequent health. *Sociological Methods & Research*, *49*(1), 216–249.

Barrett, G. F., & Kecmanovic, M. (2013). Changes in subjective well-being with retirement: Assessing savings adequacy. *Applied Economics*, *45*(35), 4883–4893.

Bassanini, A., & Caroli, E. (2015). Is work bad for health? The role of constraint versus choice. *Annals of Economics and Statistics*, *119/120*, 13–37.

Bauer, A. B., & Eichenberger, R. (2018). Worsening workers' health by lowering retirement age: The malign consequences of a benign reform. *Journal of the Economics of Ageing*, *18*, 100296.

Begg, C., & Berlin, J. (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *151*(3), 419–463.

Behncke, S. (2012). Does retirement trigger ill health? *Health Economics*, *21*(3), 282–300.

Belloni, M., Meschi, E., & Pasini, G. (2016). The effect on mental health of retiring during the economic crisis. *Health Economics*, *25*(S2), 126–140.

Bender, K. A. (2012). An analysis of well-being in retirement: The role of pensions, health, and 'voluntariness' of retirement. *The Journal of Socio-Economics*, *41*(4), 424–433.

Bertoni, M., & Brunello, G. (2017). Pappa ante portas: The effect of the husband's retirement on the wife's mental health in Japan. *Social Science & Medicine*, *175*, 135–142.

Bianchini, L., & Borella, M. (2016). Retirement and memory in Europe. *Ageing & Society*, *36*(7), 1434–1458.

Binh Tran, D., & Zikos, V. (2019). The causal effect of retirement on health: understanding the mechanisms. *Australian Economic Review*, *52*(4), 427–446.

Blake, H., & Garrouste, C. (2019). Collateral effects of a pension reform in France. *Annals of Economics and Statistics*, *133*, 57–86.

Bloemen, H., Hochguertel, S., & Zweerink, J. (2017). The causal effect of retirement on mortality: Evidence from targeted incentives to retire early. *Health Economics*, *26*(12), e204–e218.

Bloemen, H., Hochguertel, S., & Zweerink, J. (2019). The effect of incentive-induced retirement on spousal retirement rates: Evidence from a natural experiment. *Economic Inquiry*, *57*(2), 910–930.

Bom, P. R., & Rachinger, H. (2019). A kinked meta-regression model for publication bias correction. *Research synthesis methods*, *10*(4), 497–514.

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*(3), 311–329.

Börsch-Supan, A., & Schuth, M. (2014). Early retirement, mental health, and social networks. In *Discoveries in the Economics of Aging*, NBER Chapters (pp. 225–250). University of Chicago Press.

Bound, J., & Waidmann, T. (2007). Estimating the health effects of retirement. Working Papers wp168, University of Michigan, Michigan Retirement Research Center.

Bozio, A., Garrouste, C., & Perdrix, E. (2021). Impact of later retirement on mortality: Evidence from France. *Health Economics*, *30*(5), 1178–1199.

Brockmann, H., Müller, R., & Helmert, U. (2009). Time to retire–time to die? A prospective cohort study of the effects of early retirement on long-term survival. *Social Science & Medicine*, *69*(2), 160–164.

Butterworth, P., Gill, S. C., Rodgers, B., Anstey, K. J., Villamil, E., & Melzer, D. (2006). Retirement and mental health: Analysis of the Australian national survey of mental health and well-being. *Social Science & Medicine*, *62*(5), 1179–1191.

Calvo, E., Sarkisian, N., & Tamborini, C. R. (2013). Causal effects of retirement timing on subjective physical and emotional health. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(1), 73–84.

Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *The American Economic Review*, *85*(2), 238–243.

Carlsson, S., Andersson, T., Michaëlsson, K., Vågerö, D., & Ahlbom, A. (2012). Late retirement is not associated with increased mortality, results based on all Swedish retirements 1991–2007. *European Journal of Epidemiology*, *27*(6), 483–486.

Carone, G., Eckefeldt, P., Giamboni, L., Laine, V., & Pamies, S. (2016). Pension reforms in the EU since the early 2000's: Achievements and challenges ahead. European Economy Discussion Paper No. 42, Brussels.

Carrino, L., Glaser, K., & Avendano, M. (2020). Later retirement, job strain, and health: Evidence from the new state pension age in the United Kingdom. *Health Economics*, *29*(8), 891–912.

Celidoni, M., Dal Bianco, C., & Weber, G. (2017). Retirement and cognitive decline. A longitudinal analysis using SHARE data. *Journal of Health Economics*, *56*, 113–125.

Celidoni, M., & Rebba, V. (2017). Healthier lifestyles after retirement in Europe? Evidence from SHARE. *The European Journal of Health Economics*, *18*(7), 805–830.

Charles, K. K. (2004). Is retirement depressing?: Labor force inactivity and psychological well-being in later life. In S. E. Polachek (Ed.), *Accounting for worker well-being*, volume 23 of *Research in Labor Economics* (pp. 269–299). Emerald Group Publishing Limited.

Che, Y., & Li, X. (2018). Retirement and health: Evidence from China. *China Economic Review*, *49*, 84–95.

Chung, S., Domino, M. E., & Stearns, S. C. (2009). The effect of retirement on weight. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *64*(5), 656–665.

Churchill, A., & Mishra, V. (2018). Returns to education in China: A meta-analysis. *Applied Economics*, *50*(54), 5903–5919.

Coe, N. B., & Lindeboom, M. (2008). Does retirement kill you? Evidence from early retirement windows. IZA Discussion Paper No. 3817, Institute of Labor Economics (IZA), Bonn, Germany.

Coe, N. B., von Gaudecker, H.-M., Lindeboom, M., & Maurer, J. (2012). The effect of retirement on cognitive functioning. *Health Economics*, *21*(8), 913–927.

Coe, N. B., & Zamarro, G. (2011). Retirement effects on health in Europe. *Journal of Health Economics*, *30*(1), 77–86.

Cohen, J. (1988). *Statistical power analysis in the behavioral sciences*. New York: Lawrence Erlbaum Associates: Ne.

Cullen, M. (1999). Personal reflections on occupational health in the twentieth century: Spiraling to the future. *Annual Review of Public Health*, *20*, 1–13.

Dave, D., Rashad, R. I., & Spasojevic, J. (2008). The effects of retirement on physical and mental health outcomes. *Southern Economic Journal*, *75*(2), 497–523.

Dayaram, K., & McGuire, A. (2019). Retirement reforms: Occupational strain and health. *Industrial Relations: A Journal of Economy and Society*, *58*(3), 522–542.

De Luca, G., & Magnus, J. R. (2011). Bayesian model averaging and weighted-average least squares: Equivariance, stability, and numerical issues. *Stata Journal*, *11*(4), 518–544(27).

Doucouliagos, C. (1995). Worker participation and productivity in labor-managed and participatory capitalist firms: A meta-analysis. *ILR Review*, *49*(1), 58–77.

Doucouliagos, H. (2011). How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics. Working Papers SWP 2011/5, Deakin University, Department of Economics.

Doucouliagos, H., & Laroche, P. (2009). Unions and profits: A meta-regression analysis. *Industrial Relations: A Journal of Economy and Society*, *48*(1), 146–184.

Dufouil, C., Pereira, E., Chêne, G., Glymour, M. M., Alpérovitch, A., Saubusse, E., Risse-Fleury, M., Heuls, B., Salord, J.-C., Brieu, & M.-A. (2014). Older age at retirement is associated with decreased risk of dementia. *European Journal of Epidemiology*, *29*(5), 353–361.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634.

Eibich, P. (2015). Understanding the effect of retirement on health: Mchanisms and heterogeneity. *Journal of Health Economics*, *43*, 1–12.

Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, *26*(1), 30–55.

Einmahl, J., Magnus, J., & Kumar, K. (2011). On the choice of prior in Bayesian model averaging. Discussion Paper 2011-003, Tilburg University, Center for Economic Research.

Evenson, K. R., Rosamond, W. D., Cai, J., Diez-Roux, A. V., & Brancati for the Atherosclerosis Risk in Communities Study Investigators, F. L. (2002). Influence of retirement on leisure-time physical activity: The atherosclerosis risk in communities study. *American Journal of Epidemiology*, *155*(8), 692–699.

Eyjólfsdóttir, 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*, 77–86.

Fé, E., & Hollingsworth, B. (2016). Short-and long-run estimates of the local effects of retirement on health. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *179*(4), 1051–1067.

Feng, J., Li, Q., & Smith, J. P. (2020). Retirement effect on health status and health behaviors in urban China. *World Development*, *126*, 104702.

Finkel, D., Andel, R., Gatz, M., & Pedersen, N. L. (2009). The role of occupational complexity in trajectories of cognitive aging before and after retirement. *Psychology and Aging*, *24*(3), 563–573.

Fisher, G. G., Stachowski, A., Infurna, F. J., Faul, J. D., Grosch, J., & Tetrick, L. E. (2014). Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *Journal of Occupational Health Psychology*, *19*(2), 231–242.

Fitzpatrick, M. D., & Moore, T. J. (2018). The mortality effects of retirement: Evidence from social security eligibility at age 62. *Journal of Public Economics*, *157*, 121–137.

Fonseca, R., Kapteyn, A., Lee, J., Zamarro, G., & Feeney, K. (2014). A longitudinal study of well-being of older Europeans: Does retirement matter? *Journal of Population Ageing*, *7*(1), 21–41.

Frimmel, W., & Pruckner, G. J. (2020). Retirement and healthcare utilization. *Journal of Public Economics*, *184*, 104146.

Gill, S. C., Butterworth, P., Rodgers, B., Anstey, K. J., Villamil, E., & Melzer, D. (2006). Mental health and the timing of men's retirement. *Social Psychiatry and Psychiatric Epidemiology*, *41*(7), 515–522.

Godard, M. (2016). Gaining weight through retirement? Results from the SHARE survey. *Journal of Health Economics*, *45*, 27–46.
Gorry, A., Gorry, D., & Slavov, S. N. (2018). Does retirement improve health and life satisfaction? *Health Economics*, *27*(12), 2067–2086.
Görg, H., & Strobl, E. (2001). Multinational companies and productivity spillovers: A meta-analysis. *The Economic Journal*, *111*(475), F723–F739.
Grip, A. d., Lindeboom, M., & Montizaan, R. (2012). Shattered dreams: the effects of changing the pension system late in the game. The *Economic Journal*, *122*(559), 1–25.
Grøtting, M. W., & Lillebø, O. S. (2020). Health effects of retirement: Evidence from survey and register data. *Journal of Population Economics*, *33*(2), 671–704.
Hagen, J. (2018). The effects of increasing the normal retirement age on health care utilization and mortality. *Journal of Population Economics*, *31*(1), 193–234.
Hallberg, D., Johansson, P., & Josephson, M. (2015). Is an early retirement offer good for your health? Quasi-experimental evidence from the army. *Journal of Health Economics*, *44*, 274–285.
Harrison, J., & Dawson, L. (2016). Occupational health: Meeting the challenges of the next 20 years. *Safety and Health at Work*, *7*(2), 143–149.
Havranek, T., Horvath, R., Irsova, Z., & Rusnak, M. (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, *96*(1), 100–118.
Havránek, T., Stanley, T. D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W. R., Rost, K., & van Aert, R. C. M. (2020). Reporting guidelines for meta-analysis in economics. *Journal of Economic Surveys*, *34*(3), 469–475.
Heller-Sahlgren, G. (2017). Retirement blues. *Journal of Health Economics*, *54*, 66–78.
Henkens, K., van Solinge, H., & Gallo, W. T. (2008). Effects of retirement voluntariness on changes in smoking, drinking and physical activity among Dutch older workers. *The European Journal of Public Health*, *18*(6), 644–649.
Hernaes, E., Markussen, S., Piggott, J., & Vestad, O. L. (2013). Does retirement age impact mortality? *Journal of Health Economics*, *32*(3), 586–598.
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.
Horner, E. M., & Cullen, M. R. (2016). The impact of retirement on health: quasi-experimental methods using administrative data. *BMC Health Services Research*, *16*(1), 1–9.
Hult, C., Stattin, M., Janlert, U., & Järvholm, B. (2010). Timing of retirement and mortality—A cohort study of Swedish construction workers. *Social Science & Medicine*, *70*(10), 1480–1486.
Insler, M. (2014). The health consequences of retirement. *Journal of Human Resources*, *49*(1), 195–233.
Johnston, D. W., & Lee, W.-S. (2009). Retiring to the good life? The short-term effects of retirement on health. *Economics Letters*, *103*(1), 8–11.
Jokela, M., Ferrie, J. E., Gimeno, D., Chandola, T., Shipley, M. J., Head, J., Vahtera, J., Westerlund, H., Marmot, M. G., & Kivimäki, M. (2010). From midlife to early old age: Health trajectories associated with retirement. *Epidemiology (Cambridge, Mass.)*, *21*(3), 284–290.
Kajitani, S., Sakata, K., & McKenzie, C. (2017). Occupation, retirement and cognitive functioning. *Ageing and Society*, *37*(8), 1568–1596.
Kalwij, A., Alessie, R., & Knoef, M. (2013). Pathways to retirement and mortality risk in the Netherlands. *European Journal of Population/Revue Européenne de Démographie*, *29*(2), 221–238.
Keef, S. P., & Roberts, L. A. (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology*, *57*(1), 97–129.
Kesavayuth, D., Rosenman, R. E., & Zikos, V. (2016). Retirement, personality, and well-being. *Economic Inquiry*, *54*(2), 733–750.
Kim, J. E., & Moen, P. (2002). Retirement transitions, gender, and psychological well-being: A life-course, ecological model. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(3), P212–P222.
Kim, S., & Koh, K. (2020). Does early access to pension wealth improve health? *Economic Inquiry*, *58*(4), 1783–1794.

Kolodziej, I. W., & García-Gómez, P. (2019). Saved by retirement: Beyond the mean effect on mental health. *Social Science & Medicine*, *225*, 85–97.

Kuhn, A. (2018). The complex effects of retirement on health. *IZA World of Labor*, *430*.

Kuhn, A., Staubli, S., Wuellrich, J.-P., & Zweimüller, J. (2020). Fatal attraction? Extended unemployment benefits, labor force exits, and mortality. *Journal of Public Economics*, *191*, 104087.

Kühntopf, S., & Tivig, T. (2012). Early retirement and mortality in Germany. *European Journal of Epidemiology*, *27*(2), 85–89.

Kuusi, T., Martikainen, P., & Valkonen, T. (2020). The influence of old-age retirement on health: Causal evidence from the Finnish register data. *Journal of the Economics of Ageing*, *17*, 100257.

Lalive, R., & Staubli, S. (2015). How does raising women's full retirement age affect labor supply, income, and mortality? Evidence from Switzerland. Working Paper NB 14–09, NBER Retirement Research Center.

Latif, E. (2011). The impact of retirement on psychological well-being in Canada. *Journal of Socio-Economics*, *40*(4), 373–380.

Lee, J., & Smith, J. P. (2009). Work, retirement, and depression. *Journal of Population Ageing*, *2*(1-2), 57–71.

Lei, X., & Liu, H. (2018). Gender difference in the impact of retirement on cognitive abilities: Evidence from urban China. *Journal of Comparative Economics*, *46*(4), 1425–1446.

Li, W., Ye, X., Zhu, D., & He, P. (2021). The longitudinal association between retirement and depression: A systematic review and meta-analysis. *American Journal of Epidemiology, 190(10), 2220-2230*.

Light, R., & Pillemer, D. (1984). *Summing up. The science of reviewing research*. Harvard University Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London: SAGE publications, Inc.

Litwin, H. (2007). Does early retirement lead to longer life? *Ageing and Society*, *27*(5), 739.

Lucifora, C., & Vigani, D. (2018). Health care utilization at retirement: The role of the opportunity cost of time. *Health Economics*, *27*(12), 2030–2050.

Magnus, J. R., & De Luca, G. (2016). Weighted-average least squares (WALS): A survey. *Journal of Economic Surveys*, *30*(1), 117–148.

Magnus, J. R., Powell, O., & Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, *154*(2), 139–153.

Mandal, B., & Roe, B. E. (2007). Job loss, retirement and the mental health of older Americans. *Journal of Mental Health Policy and Economics*, *11*(4), 167–176.

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.

Mein, G., Martikainen, P., Hemingway, H., Stansfeld, S., & Marmot, M. (2003). Is retirement good or bad for mental and physical health functioning? Whitehall II longitudinal study of civil servants. *Journal of Epidemiology and Community Health*, *57*(1), 46–49.

Messe, P.-J., & Wolff, F.-C. (2019a). Healthier when retiring earlier? Evidence from France. *Applied Economics*, *51*(47), 5122–5143.

Messe, P.-J., & Wolff, F.-C. (2019b). The short-term effects of retirement on health within couples: Evidence from France. *Social Science & Medicine*, *221*, 27–39.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Medicine*, *6*(7), 1–6.

Mojon-Azzi, S., Sousa-Poza, A., & Widmer, R. (2007). The effect of retirement on health: a panel analysis using data from the Swiss Household Panel. *Swiss Medical Weekly*, *137*(4142), 581–585.

Mosca, I., & Barrett, A. (2016). The impact of voluntary and involuntary retirement on mental health: Evidence from older Irish adults. *Journal of Mental Health Policy and Economics*, *19*(1), 33–44.

Motegi, H., Nishimura, Y., & Oikawa, M. (2020). Retirement and health investment behaviors: an international comparison. *Journal of the Economics of Ageing*, *16*, 100267.

Müller, T., & Shaikh, M. (2018). Your retirement and my health behavior: evidence on retirement externalities from a fuzzy regression discontinuity design. *Journal of Health Economics*, *57*, 45–59.

Neuman, K. (2008). Quit your job and get healthier? The effect of retirement on health. *Journal of Labor Research*, *29*(2), 177–201.

Nielsen, N. F. (2019). Sick of retirement? *Journal of Health Economics*, *65*, 133–152.

Nishimura, Y., Oikawa, M., & Motegi, H. (2018). What explains the difference in the effect of retirement on health? Evidence from global aging data. *Journal of Economic Surveys*, *32*(3), 792–847.
OECD (2017). *Pensions at a Glance 2017: OECD and G20 Indicators*. Paris: OECD Publishing.
OECD (2019). *Pensions at a Glance 2019: OECD and G20 Indicators*. Paris: OECD Publishing.
Oksanen, T., Vahtera, J., Westerlund, H., Pentti, J., Sjösten, N., Virtanen, M., Kawachi, I., & Kivimäki, M. (2011). Is retirement beneficial for mental health? Antidepressant use before and after retirement. *Epidemiology*, *22*(4), 553–559.
Olesen, K., Rugulies, R., Rod, N. H., & Bonde, J. P. (2014). Does retirement reduce the risk of myocardial infarction? A prospective registry linkage study of 617 511 Danish workers. *International Journal of Epidemiology*, *43*(1), 160–167.
Oshio, T., & Kan, M. (2017). The dynamic impact of retirement on health: evidence from a nationwide ten-year panel survey in Japan. *Preventive Medicine*, *100*, 287–293.
Pabón-Carrasco, M., Ramirez-Baena, L., López Sánchez, R., Rodríguez-Gallego, I., Suleiman-Martos, N., & Gómez-Urquiza, J. L. (2020). Prevalence of depression in retirees: a meta-analysis. *Healthcare*, *8*(3).
Picchio, M., & van Ours, J. C. (2020). Mental health effects of retirement. *De Economist*, *168*(3), 419–452.
Pilipiec, P., Groot, W., & Pavlova, M. (2021). The effect of an increase of the retirement age on the health, well-being, and labor force participation of older workers: a systematic literature review. *Journal of Population Ageing*, 14, 271-315.
Quaade, T., Engholm, G., Johansen, A. M. T., & Møller, H. (2002). Mortality in relation to early retirement in Denmark: a population-based study. *Scandinavian Journal of Public Health*, *30*(3), 216–222.
Reed, W. (2020). A note on the use of partial correlation coefficients in meta-analyses. Working paper no. 8/2020, Department of Economics and Finance, University of Canterbury, New Zealand.
Rijs, K. J., Cozijnsen, R., & Deeg, D. J. (2012). The effect of retirement and age at retirement on self-perceived health after three years of follow-up in Dutch 55-64-year-olds. *Ageing and Society*, *32*(2), 281–306.
Roberts, B. A., Fuhrer, R., Marmot, M., & Richards, M. (2011). Does retirement influence cognitive performance? The Whitehall II study. *Journal of Epidemiology and Community Health*, *65*(11), 958–963.
Rohwedder, S., & Willis, R. J. (2010). Mental retirement. *Journal of Economic Perspectives*, *24*(1), 119–38.
Rose, L. (2020). Retirement and health: evidence from England. *Journal of Health Economics*, *73*, 102352.
Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.
Sewdas, R., De Wind, A., Stenholm, S., Coenen, P., Louwerse, I., Boot, C., & Van Der Beek, A. (2020). Association between retirement and mortality: working longer, living longer? A systematic review and meta-analysis. *Journal of Epidemiology and Community Health*, *74*(5), 473–480.
Shai, O. (2018). Is retirement good for men's health? Evidence using a change in the retirement age in Israel. *Journal of Health Economics*, *57*, 15–30.
Son, C.-H. (2020). The effects of retirement on health-related quality of life of retirement-aged adults for four marital status subgroups. *International Journal of Economic Sciences*, *9*(1), 179–201.
Stanley, T. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics*, *70*(1), 103–127.
Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, *19*(3), 309–345.
Stanley, T. D., & Doucouliagos, H. (2012). *Meta-regression analysis in economics and business*. Abingdon: Routledge.
Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78.
Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., & Pugh, G. (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, *27*(2), 390–394.
Sterling, T. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance–or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34.
Sutton, A. J., Abrams, K. R., Jones, D. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester: Wiley.
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.
Tsai, S. P., Wendt, J. K., Donnelly, R. P., De Jong, G., & Ahmed, F. S. (2005). Age at retirement and long term survival of an industrial population: prospective cohort study. *British Medical Journal*, *331*(7523), 995.

Ugur, M. (2014). Corruption's direct effects on per-capita income growth: a meta-analysis. *Journal of Economic Surveys*, *28*(3), 472–490.

van der Heide, I., van Rijn, R. M., Robroek, S. J., 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.

van Mourik, C. (2020). A meta-analysis into the causal effect of retirement on health. Netspar Academic Series, Thesis for the Master Public Administration, Leiden University.

Viechtbauer, W., & Cheung, M. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Method*, *1*(2), 112–125.

Vooren, M., Haelermans, C., Groot, W., & van den Brink, H. (2019). The effectiveness of active labor market policies: a meta-analysis. *Journal of Economic Surveys*, *33*(1), 125–149.

Waldron, H. (2001). Links between early retirement and mortality. ORES Working Paper Series No. 93, Social Security Administration, Office of Policy, Washington.

Westerlund, H., Kivimäki, M., Singh-Manoux, A., Melchior, M., Ferrie, J. E., Pentti, J., Jokela, M., Leineweber, C., Goldberg, M., & Zins, M. (2009). Self-rated health before and after retirement in France (GAZEL): a cohort study. *The Lancet*, *374*(9705), 1889–1896.

Wu, C., Odden, M. C., Fisher, G. G., & Stawski, R. S. (2016). Association of retirement age with mortality: a population-based longitudinal study among older adults in the USA. *Journal of Epidemiology and Community Health*, *70*(9), 917–923.

Xue, X., Cheng, M., & Zhang, W. (2021). Does education really improve health? A meta-analysis. *Journal of Economic Surveys*, *35*(1), 71–105.

Xue, X., Reed, R. R., & Menclova, A. (2020). Social capital and health: A meta-analysis. *Journal of Health Economics*, *72*, 102317.

Zhang, Y., Salm, M., & van Soest, A. (2018). The effect of retirement on healthcare utilization: evidence from China. *Journal of Health Economics*, *62*, 165–177.

Zhao, M., Konishi, Y., & Noguchi, H. (2017). Retiring for better health? Evidence from health investment behaviors in Japan. *Japan and the World Economy*, *42*, 56–63.

Zhu, R. (2016). Retirement and its consequences for women's health in Australia. *Social Science & Medicine*, *163*, 117–125.

Zulka, L. E., Hansson, I., & Hassing, L. B. (2019). Impact of retirement on cognitive function: a literature review. *Journal of Gerontopsychology and Geriatric Psychiatry*, *32*(4), 187–203.

Zulkarnain, A., & Rutledge, M. S. (2018). How does delayed retirement affect mortality and health? CRR Working Paper 2018-11, Center for Retirement Research at Boston College.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Filomena, M., & Picchio, M. (2023). Retirement and health outcomes in a meta-analytical framework. *Journal of Economic Surveys*, *37,* 1120–1155. https://doi.org/10.1111/joes.12527

APPENDIX A

TABLE A.1 Articles included in the meta-analysis ($N = 85$)

Authors	Citations	Outcome(s)	Country	Data	Time Span	Id. Strategy	Effects	Heterogeneity
Apouey, Guven, and Senik (2019)	7	SAH, PH, MH	Australia	HILDA	2001-2014	FE	+, 0	No
Ardito et al. (2020)	1	HC	Italy	WHIP	2001-2014	IV	+	O, Ph, I
Atalay and Barrett (2014)	36	SAH, PH, MH	Australia	NHSs	1995-2008	IV	0 (SAH), 0/+ (PH, MH)	G
Atalay et al. (2019)	12	MH	Australia	HILDA	2012-2016	FD-IV	0	No
Bamia et al. (2008)	92	M	Greece	EPIC	1994-2006	Other	–	No
Barrett and Kecmanovic (2013)	35	MH	Australia	HILDA	2007	Other	0	V
Bauer and Eichenberger (2018)	1	SAH, PH	Switzerland	Swiss LFS	2004-2015	DiD	0	No
Behncke (2012)	247	SAH, PH, MH	England	ELSA	2002-2007	PSM, IV	– (SAH, PH), 0 (MH)	No
Belloni et al. (2016)	31	MH	10 EU	SHARE	2004-2013	FE-IV	0	O
Bertoni and Brunello (2017)	23	MH	Japan	JPSS	2008-2013	IV	–	No
Bianchini and Borella (2016)	16	MH	10 EU	SHARE	2004-2012	FE-IV	0, + (RD)	RD
Binh Tran and Zikos (2019)	8	SAH, PH, MH	Australia	HILDA	2002-2015	FE-IV	+	No
Blake and Garrouste (2019)	9	SAH, PH, MH	France	Health Barometer	1994-2003	DiD	0	E
Bloemen et al. (2017)	55	M	Netherlands	Administrative Data	2000-2005	FE-IV	0	No
Bonsang et al. (2012)	444	MH	USA	HRS	1998-2008	FE-IV	–	No
Bonsang and Klein (2012)	134	SAH	Germany	GSOEP	1995-2010	FE-IV	+, 0 (MI)	V
Bozio et al. (2021)	2	M	France	Administrative Data	2004-2017	IV	0	No
Brockmann et al. (2009)	104	M	Germany	Gmunder Ersatzkasse	1990-2004	Other	+ (M), 0 (F)	G, Ph, I
Butterworth et al. (2006)	256	MH	Australia	NSMHWB	1997	Other	0	O, AC
Calvo et al. (2013)	152	PH, MH	USA	HRS	1992-2010	FE-IV	–	T
Carrino et al. (2020)	1	SAH, PH, MH	UK	Understanding Society	2009-2016	DiD	0	O

(Continues)

TABLE A.1 (Continued)

Authors	Citations	Outcome(s)	Country	Data	Time Span	Id. Strategy	Effects	Heterogeneity
Celidoni et al. (2017)	55	MH	10 EU	SHARE	2004-2012	IV	+	T
Celidoni and Rebba (2017)	51	HC	10 EU	SHARE	2004-2012	FE-IV	0	MS
Che and Li (2018)	10	SAH	China	CHNS	1991-2006	IV	+	No
Chung et al. (2009)	64	PH	USA	HRS	1992-2002	FE-IV	–	AC, I, O
Coe et al. (2012)	153	MH	USA	HRS	1996-2008	IV	+, 0	O
Coe and Zamarro (2011)	506	SAH, MH	11 EU	SHARE	2004-2007	IV	+ (SAH), 0 (MH)	No
Dave et al. (2008)	485	SAH, PH, MH	USA	HRS	1992-2005	FE	–	MS, V
Dayaram and McGuire (2019)	1	PH, MH	Australia	HILDA	2003-2015	PSM	0	No
Eibich (2015)	268	SAH, PH, MH, HC	Germany	GSOEP	2002-2009	RDD	+ (SAH, MH, HC), 0 (PH)	E
Eyjólfsdóttir et al. (2019)	6	PH, M	Sweden	LNU, LISA, SWEOLD	2004-2014	PSM	0	No
Fé and Hollingsworth (2016)	14	SAH, MH	UK	BHPS	1991-2005	RDD	– (SAH), + (MH)	No
Feng et al. (2020)	12	PH	China	CHARLS	2001-2015	RDD	– (M), 0 (F)	E, G
Fitzpatrick and Moore (2018)	84	M	USA	MCOD, SSDMF	1979-2012	RDD	– (M), 0 (F)	G, E
Frimmel and Pruckner (2020)	7	HC	Austria	ASSD	1998-2012	FE-IV	+, 0 (F)	O, G
Gill et al. (2006)	95	MH	Australia	HILDA	2002-2003	Other	0	No
Godard (2016)	90	PH	8 EU	SHARE	2004-2011	FE-IV	0	O
Gorry et al. (2018)	84	SAH, MH, PH, HC	USA	HRS	1992-2014	IV	+, 0 (MH)	No
Grip et al. (2012)	121	SAH, MH, HC	Netherlands	Administrative Data	1997-2006	RDD	+ (MH), 0 (HC, SAH)	O, I
Grøtting and Lillebø (2020)	5	PH, HC, M	Norway	NORLAG	2002-2012	RDD	0, + (M PH)	No
Hagen (2018)	40	HC, M	Sweden	LOUISE	1987-2010	DiD	0	No
Hallberg et al. (2015)	64	M	Sweden	Administrative Data	1985-2010	DiD	+, 0	No
Heller-Sahlgren (2017)	67	MH	10 EU	SHARE	2004-2012	FE-IV	– (M), 0 (F)	E, G, O
Hernaes et al. (2013)	111	M	Norway	Administrative Data	1992-2010	IV	0	No
Hessel (2016)	54	SAH, PH	12 EU	EU-SILC	2009-2012	RE-IV	+, 0 (Chronic)	No
Horner and Cullen (2016)	23	PH, MH, HC	USA	Administrative Data	1997-2009	IV	0, + (Chronic)	No
Hult et al. (2010)	57	M	Sweden	Health Monitoring	1971-1993	Other	0	Ph

(Continues)

TABLE A.1 (Continued)

Authors	Citations	Outcome(s)	Country	Data	Time Span	Id. Strategy	Effects	Heterogeneity
Insler (2014)	198	SAH	USA	HRS	1992-2010	FE-IV	+	No
Johnston and Lee (2009)	149	SAH, PH, MH	UK	HSE	1997-2005	RDD	+, 0 (PH)	No
Jokela et al. (2010)	161	PH, MH	UK	WhiteHall II Cohort Study	1991-2006	Other	+ (MH), 0 (PH)	T, RD
Kajitani et al. (2017)	35	MH	Japan	NSJE	1987-2002	IV	0	O
Kalwij et al. (2013)	11	M	Netherlands	IPO	1996-2010	Other	0	I
Kim and Koh (2020)	1	SAH	Singapore	SLP	2015-2019	RDD	+	No
Kim and Moen (2002)	364	MH	USA	Cornell Retirement Study	1994-1999	Other	0	G, MS, Ph
Kolodziej and García-Gómez (2019)	20	MH	11 EU	SHARE	2004-2013	IV	+ (F), 0 (M)	MS, Ph, G
Kuhn et al. (2020)	19	M	Austria	ASSD	1972-2017	IV	0 (F), – (M)	Ph, O, G
Kuusi et al. (2020)	4	MH, HC	Finland	Statistics Finland	2000-2012	FE-IV	+, 0 (HC)	G, O, I
Latif (2011)	59	MH	Canada	CNPHS	1994-2006	FE-IV	+	AC, MS
Lei and Liu (2018)	14	MH	China	CHARLS	2011-2015	FE-IV	+, 0 (M), 0, – (F)	G, O
Litwin (2007)	51	M	Israel	NHS	1997-2004	Other	0	No
Lucifora and Vigani (2018)	16	HC	10 EU	SHARE	2004-2006	FE-IV	0	G, O
Mandal and Roe (2007)	117	MH	USA	HRS	1992-2002	IV	+, –	No
Mazzonna and Peracchi (2017)	138	SAH, MH	10 EU	SHARE	2004-2006	FD-IV	–, 0 (F SAH)	O
Mein et al. (2003)	348	PH, MH	UK	WhiteHall II Cohort Study	1991-1995	Other	+, 0 (PH F)	No
Messe and Wolff (2019b)	9	SAH, PH	France	LFS	2013-2016	DiD	+ (M), 0 (F PH)	G, O
Messe and Wolff (2019a)	7	SAH, PH	France	INSEE	2012	IV	0	No
Mojon-Azzi et al. (2007)	80	SAH, PH, MH	Switzerland	SHP	1999-2003	Other	+	No
Mosca and Barrett (2016)	17	MH	Ireland	TILDA	2009-2013	FD	0	V
Müller and Shaikh (2018)	49	SAH	19 C.	SHARE	2004-2013	RDD	+	No
Neuman (2008)	235	SAH, PH, MH	USA	HRS	1992-2004	IV	0, + (SAH)	No
Nielsen (2019)	13	PH, HC, M	Denmark	Administrative Data	1980-2010	IV, RDD	0 (IV), 0, + (HC)	G, AC, T
Oksanen et al. (2011)	103	MH	Finland	National Records	1995-2004	Other	+	No
Oshio and Kan (2017)	28	SAH, MH	Japan	PSMOA	2005-2014	FE-IV	+, 0 (F SAH)	G
Picchio and van Ours (2020)	12	MH	Netherlands	LISS	2007-2018	RDD	+ (M), 0 (F)	G, MS

(Continues)

TABLE A.1 (Continued)

Authors	Citations	Outcome(s)	Country	Data	Time Span	Id. Strategy	Effects	Heterogeneity
Quaade et al. (2002)	63	M	Denmark	Administrative Data	1986-1996	Other	–	No
Rijs et al. (2012)	41	SAH	Netherlands	LASA	1995-2009	Other	0	E, AC
Roberts et al. (2011)	51	MH	UK	WhiteHall II Cohort Study	1985-1988	Other	0	No
Rose (2020)	4	SAH, PH, MH, HC, M	England	ELSA, BHPS	1990-2011	RDD, FE-IV	+, 0, – (M M)	G, E, MS
Shai (2018)	44	SAH, HC	Israel	IHS, SHARE	1997-2013	DiD	0 (HC), +	E
Syse et al. (2017)	35	PH, MH	Norway	NORLAG	2002-2007	Other	+, 0	No
Tsai et al. (2005)	116	M	USA	Shell Oil	1973-2003	Other	0	No
Westerlund et al. (2009)	340	SAH	France	GAZEL	1990-2006	Other	+	O
Wu et al. (2016)	49	M	USA	HRS	1992-2010	Other	–	Ph
Zhang et al. (2018)	27	SAH, HC	China	CHARLS	2011-2013	RDD	0, – (F)	G
Zhu (2016)	51	SAH, PH, MH	Australia	HILDA	2001-2011	FE-IV	+	No

The sign of the effect is based on the value of t-stat: "–" means $t \leq -1.96$; "+" is for $t \geq 1.96$; "0" when $-1.96 < t < 1.96$. Note that the sign is positive even when articles estimate a negative effect of postponed retirement;*Identification Strategy*: PSM = Propensity Score Matching; IV = Instrumental Variables; DiD = Difference-in-differences; FE = Fixed Effects; RDD = Regression Discontinuity Design;Other methods = Ordinary Least Squares, Duration models, Logit, Multinomial logit, Ordered probit.*Outcome*: MH = Mental health; SAH = Self-assessed / General health; PH = Physical health; HC = Healthcare utilization; M = Mortality.*Effects*: M = Males; F = Females; B = Both; MI = Mandatory/Involuntary; ER = Early retirement; PP = Postponed; SR = Short-run; LR = Long-run; RD = Retirement duration.*Heterogeneity*: G = Gender; O = Occupation; Ph = Previous health; MS = Marital status; E = Educational attainment; I = Income; V = Voluntariness; AC = Age cohort; T = Timing; C = Country; LS = Living standard.

TABLE A.2 Descriptive statistics of explanatory variables used in the meta-regressions by the sign of the partial correlation coefficient

	$r \leq 0$			$r > 0$		
	Absolute frequencies	Mean	Std. Dev.	Absolute frequencies	Mean	Std. Dev.
Scimago subject areas						
Multi area	32	0.248	0.434	55	0.307	0.463
Economics/Business	36	0.279	0.450	52	0.291	0.455
Medicine/Psychology	61	0.473	0.501	72	0.402	0.492
Health outcomes						
Mortality (Reference category)	21	0.163	0.371	21	0.117	0.323
General and self-reported health	18	0.140	0.348	28	0.156	0.364
Physical health	33	0.256	0.438	38	0.212	0.410
Mental health	38	0.295	0.438	67	0.374	0.485
Healthcare utilization	19	0.147	0.356	25	0.140	0.348
Identification strategies						
Other methods (Reference category)	12	0.093	0.292	26	0.145	0.353
Regression discontinuity design (RDD)	22	0.171	0.378	37	0.207	0.406
Instrumental variables (IV)	60	0.465	0.501	92	0.514	0.501
Difference-in-differences (DiD)	14	0.109	0.312	9	0.050	0.219
Propensity score matching (PSM)	8	0.062	0.242	9	0.050	0.219
Fixed-effects/First-differences	13	0.101	0.302	6	0.034	0.180
Institutional contexts						
Statutory retirement (Reference category)	67	0.519	0.502	132	0.737	0.441
Mandatory or involuntary retirement	21	0.163	0.371	11	0.061	0.241
Early retirement	25	0.194	0.397	24	0.134	0.342
Postponed retirement	16	0.124	0.331	12	0.067	0.251
Geographical areas						
Multi-country analyses (Reference category)	19	0.147	0.356	22	0.123	0.329
Europe	57	0.442	0.499	86	0.480	0.501
Extra-European countries	53	0.411	0.494	71	0.397	0.491
Sex						
Males (Reference category)	38	0.295	0.458	64	0.358	0.481
Females	31	0.240	0.429	62	0.346	0.477
Males+Females	60	0.465	0.501	53	0.296	0.458
Calculation of t-statistic						
from 95% CI or from OR (Reference category)	13	0.101	0.302	27	0.151	0.359
t-statistic from $\widehat{\beta}_i/SE_i$	116	0.899	0.302	152	0.849	0.359
Birth cohorts						
Other (Reference category)	86	0.667	0.473	126	0.704	0.458
Only birth cohorts $\leq$ 1950	43	0.333	0.473	53	0.296	0.458

(Continues)

TABLE A.2 (Continued)

	$r \leq 0$			$r > 0$		
	Absolute frequencies	**Mean**	**Std. Dev.**	**Absolute frequencies**	**Mean**	**Std. Dev.**
Type of previous occupation						
White collars (Reference category)	5	0.039	0.194	11	0.061	0.241
Blue collars	11	0.085	0.280	13	0.073	0.260
Not specified	113	0.876	0.331	155	0.866	0.342
Study-related characteristics						
Google scholar citations per year	129	12.826	10.943	179	10.046	10.418
Scimago Journal Ranking	129	1.895	1.449	179	1.763	1.069
Year of publication	129	2015.023	4.523	179	2015.637	4.728
Observations	129			179		

Notes: Females+Males = observations for which authors do not separate estimates for men and women. Other methods = OLS regressions and non-linear models (logit, multinomial logit, ordered probit and Cox proportional hazard models).
[(a)] At the time of publication, some journals did not have the SJR index yet, either because they were published in too recent years or because the journal was not indexed yet in Scimago. In these cases, we assigned to the journal the available value of the SJR index which was chronologically closer.

TABLE A.3 Heterogeneity in the estimated effects of retirement on health (FAT-PET specification)

	Bayesian Model Averaging[(a)]			Weighted-Average Least Square $(q = 1)^{(b)}$		Weighted-Average Least Square $(q = 0.5)^{(b)}$		OLS check after BMA[(c)]		OLS check after WALS[(d)]	
	PM	**PSD**	**PIP**	**Coeff.**	**Std. Error**	**Coeff.**	**Std. Error**	**Coeff.**	**Std. Error**	**Coeff.**	**Std. Error**
Focus Regressors											
Publication bias	**0.308**	**0.174**	**1.000**	**0.245**	**0.175**	**0.217**	**0.178**	0.345	0.258	0.376	0.250
Precision effect	**0.003**	**0.003**	**1.000**	**0.012**	**0.004**	**0.012**	**0.004**	0.001*	0.001	0.006**	0.003
Auxiliary Regressors											
Google scholar citations per year	0.000	0.000	0.230	0.000	0.000	0.000	0.000	–	–	–	–
Scimago Journal Ranking	0.000	0.001	0.300	**−0.002**	**0.001**	**−0.002**	**0.001**	–	–	−0.002***	0.001
Year of publication	**0.000**	**0.000**	**0.630**	**0.001**	**0.000**	**0.001**	**0.000**	0.000***	0.000	0.001***	0.000
Scimago subject areas (Reference category: Multi-area journals)											
Economics/Business	0.000	0.000	0.060	0.000	0.002	0.000	0.002	–	–	–	–
Medicine/Psychology	0.000	0.000	0.070	0.001	0.002	0.001	0.002	–	–	–	–
Health outcomes (Reference category: Mortality)											
General and self-reported health	**0.011**	**0.003**	**1.000**	**0.010**	**0.002**	**0.010**	**0.002**	0.010***	0.002	0.011***	0.002
Physical health	0.000	0.001	0.130	0.001	0.001	0.001	0.001	–	–	–	–
Mental health	**0.008**	**0.002**	**0.990**	**0.007**	**0.002**	**0.007**	**0.002**	0.008***	0.002	0.007***	0.002
Healthcare utilization	0.001	0.001	0.390	**0.001**	**0.001**	**0.001**	**0.001**	–	–	0.002*	0.001
Identification strategies (Reference category: Other methods)											
Regression discontinuity design (RDD)	0.000	0.001	0.100	**−0.006**	**0.004**	**−0.005**	**0.004**	–	–	−0.002*	0.001
Instrumental variables (IV)	0.000	0.001	0.070	−0.004	0.005	−0.004	0.005	–	–	–	–
Difference-in-differences (DiD)	0.000	0.001	0.050	−0.001	0.004	−0.001	0.004	–	–	–	–
Propensity score matching (PSM)	0.000	0.003	0.050	**−0.010**	**0.008**	**−0.010**	**0.009**	–	–	−0.009	0.009
Fixed-effects/First-differences	**−0.013**	**0.003**	**1.000**	**−0.014**	**0.005**	**−0.014**	**0.005**	−0.013***	0.003	−0.012***	0.002

(Continues)

TABLE A.3 (Continued)

	Bayesian Model Averaging[(a)]			Weighted-Average Least Square (q = 1)[(b)]		Weighted-Average Least Square (q = 0.5)[(b)]		OLS check after BMA[(c)]		OLS check after WALS[(d)]	
	PM	PSD	PIP	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
Institutional contexts (Reference category: Statutory retirement)											
Mandatory or involuntary retirement	**−0.026**	**0.008**	**0.970**	**−0.021**	**0.006**	**−0.022**	**0.007**	−0.027***	0.008	−0.027***	0.008
Early retirement	0.000	0.001	0.100	−0.001	0.001	−0.001	0.001	–	–	–	–
Postponed retirement	−0.001	0.002	0.230	**−0.005**	**0.002**	**−0.006**	**0.002**	–	–	−0.006***	0.002
Geographical areas (Reference category: Multi-country analyses)											
Europe	0.000	0.000	0.050	−0.002	0.002	−0.002	0.002	–	–	–	–
Extra-European countries	0.000	0.001	0.060	**−0.003**	**0.002**	**−0.003**	**0.002**	–	–	−0.001	0.001
Sex (Reference category: Males)											
Females	0.000	0.000	0.050	0.000	0.001	0.000	0.001	–	–	–	–
Males+Females	0.000	0.001	0.080	**−0.002**	**0.002**	**−0.003**	**0.002**	–	–	−0.003	0.002
Calculation of t-statistic (Reference category: from 95% CI)											
t-statistic from $\hat{\beta}_i/SE_i$	−0.001	0.002	0.260	−0.002	0.003	−0.002	0.003	–	–	–	–
Type of previous occupation (Reference category: White collars)											
Blue collars	0.000	0.000	0.040	0.001	0.002	0.001	0.002	–	–	–	–
Not specified	0.000	0.000	0.050	0.001	0.001	0.001	0.002	–	–	–	–
Birth cohorts (Reference category: Others)											
Birth cohorts ≤ 1950	**−0.002**	**0.002**	**0.650**	−0.001	0.001	−0.001	0.001	−0.003***	0.001	–	–

Notes: The results are from the FAT-PET specification by using the inverse of the SE_i^2 as weights. PM = Posterior Mean of the coefficient; PSD = Posterior Standard Deviation; PIP = Posterior Inclusion Probability. The number of observations (studies) is 308 (85). Auxiliary variables for which the PIP is above 0.5 in BMA or the corresponding one-standard error band does not include zero in WALS are in bold. *** Significant at 1%, ** significant at 5%, * significant at 10%.

[(a)] In the BMA, we use the uniform distribution for model priors, the Zellner's g prior for the distributions of the coefficients and a Markov Chain Monte Carlo algorithm to search over the model space, by distinguishing between focus and auxiliary regressors.

[(b)] $q = 1$ indicates the Laplace model prior distribution; $q = 0.5$ implies the Subbotin model prior distribution.

[(c)] The model specification under "OLS" includes those variables which have a PIP > 0.5 in BMA ($R^2 = 0.30$).

[(d)] The second model specification under "OLS" includes those variables which are relevant according to WALS ($R^2 = 0.34$).