Network for Studies on Pensions, Aging and Retirement

The causal effect of retirement on health

A meta-analysis

Jim Been, Marike Knoef, Casper van Mourik

NETSPAR ACADEMIC SERIES

DP 06/2022-019

The causal effect of retirement on health: A meta-analysis

Jim Been[1] Marike Knoef[2] Casper van Mourik

June 2022

Abstract

Despite the many papers analyzing the causal effect of retirement on health, there is no consensus in the literature as to whether retirement deteriorates or improves health. This study is the first attempt to a large-scale quantitative meta-analysis. Using 576 results from 61 causal studies, we find that 15% of the results indicate that retirement deteriorates health, while 36% of the results indicate that retirement improves health. Our results indicate that the variation in effects found in the literature is primarily due to variation in health measures and most prominently due to the use of mortality versus other health measures.

JEL codes: C21, C26, I10, J26

Keywords: Health, Retirement, Causal inference, Meta-analysis

[1] Department of Economics at Leiden University, and Netspar. Corresponding author. E-mail: j.been@law.leidenuniv.nl
[2] Department of Economics at Leiden University, Department of Econometrics & OR at Tilburg University, and Netspar. E-mail: m.g.knoef@law.leidenuniv.nl

1 Introduction

A vast amount of literature has studied the relationship between retirement and health. A more recent strand within this literature has tried to estimate the causal effect of retirement on health. Despite the numerous studies, researchers have found no consensus on the sign of the effect of retirement on health thus far. Due to a lack of consent on the effects, scholars have remained interested in the topic (Kolodziej & García-Gómez, 2019). A likely explanation for the different findings is the variety of different countries, causal methods, and health indicators used in the different papers. Although a meta-analysis on existing studies is highly common in health-related issues in order to find consensus in the literature (Gallet & List, 2002; Gemmill et al., 2007; Bellavance et al., 2009; Trapero-Bertran et al., 2012; Gallet, 2013; Cornelsen et al., 2014; Nelson, 2014), there have been no large-scale attempts to identify the causal effect of retirement on health.

The current study is the first attempt to synthesize the findings on causal effects of retirement on health using a large-scale meta-analysis. Van der Heide et al. (2013), Bassanini & Caroli (2015), Zulka et al. (2018), and Nishimura et al. (2018) provide a recent literature review on the relation between retirement and health. Contrasting these studies, Barnay (2016) more explicitly categorize the literature that uses causal identification but does not perform a large scale quantitative meta-analysis. Our large scale quantitative meta-analysis with 576 causal results from 61 different studies contributes to the contemporary literature by including only results that deal with the endogenous relationship of retirement and health using valid methods to infer causality, such as regression discontinuity (e.g. Eibich, 2015; Fé & Hollingsworth, 2016 and Zhang et al., 2018), instrumental variables (e.g. Bingley & Martinello, 2013; Kolodziej & García-Gómez, 2019 and Chung et al., 2009), and difference-in-differences (e.g. Shai, 2018).

2 Methodology

2.1 Data collection

In the process of collecting causal studies of retirement on health, we include all articles that were reviewed by Van der Heide et al. (2013), Zulka et al. (2018) and Nishimura et al. (2018). The bibliography in these articles are used to find new articles. To find recent (working) papers that are not yet published by an international peer reviewed journal, we used Google Scholar and the SHARE-data website. A combined total of more than 170 studies are more closely reviewed.

Articles estimating an association without addressing the endogeneity of retirement and health are excluded. Focusing on the causality in studies is important as there is empirical evidence for reversed effects of poor health that induces early retirement (Currie & Madrian, 1999). If causality from retirement to health is not explicitly taken into account, reversed causation creates a downward bias in the estimated effect of retirement on health given the results from Currie & Madrian (1999).

We define retirement by using the studies' definition, but exclude papers explicitly focusing on early retirement (e.g. Coe & Lindeboom, 2008; De Grip et al., 2012). Also, articles addressing health inputs (e.g. smoking, drinking, and exercising) instead of health outputs (e.g. general, mental, physical, health care, and mortality) are excluded from the analysis. Papers investigating measures beyond health, such as subjective well-being and life satisfaction (e.g. Bonsang & Klein, 2012), are not considered either.

From the 170 reviewed articles we keep those 61 articles that focus on the causal estimation of retirement on health. We keep those studies that use the causal inference techniques of Instrumental Variables (IV), Difference-in-Differences (Diff-in-Diff), or Regression Discontinuity (RD). We only include studies with unquestionable identification strategies. For example, we include IV studies using policy variation as an instrument, while we exclude studies using gender and lifestyle variables as instruments. These instruments are unlikely to be valid. Although we allow for studies from all fields, our sample mostly consists of studies in the field of economics due to the exclusion of non-causal studies. As observed by Barnay (2016), economists have traditionally been very cautious regarding the causal interpretation of empirical analyses.

The oldest article we use in the analysis is Charles (2004) and the most recent article is Feng et al. (2020). From the 61 articles we use, multiple results from a single paper are included which leads to a total of 576 estimates from the literature. Many studies look at multiple health measures and publish multiple heterogeneous results. All those results are included, even if the study is mostly focused on the significant results. In Table 1, we present the summary statistics of the 576 estimates from the literature we use in the meta-analysis. The table shows that about 15% of all estimates indicate a negative effect of retirement on health, while 36% of the estimates indicate a positive effect. The remainder of the estimation results are not significant.

Table 1. Summary statistics of the data.

		Negative (%)	Not significant (%)	Positive (%)	Total
All results		15	50	36	576
Health measure					
General		12	41	47	121
	Health index	24	38	38	29
	Self-reported health	8	42	50	92
Mental		15	51	34	216
	Mental health index	18	39	42	33
	Depression	11	44	45	93
	Cognition	19	62	19	90
Physical		16	46	38	141
	Physical health index	0	29	71	21

Grip strength	35	30	35	20
BMI	27	57	16	44
Activities of Daily Living	7	50	43	56
Health care utilization	18	51	31	67
Drug prescription	0	0	100	3
Doctor visits	35	43	22	23
Hospitalization	10	67	23	30
Chronic conditions	9	36	55	11
Mortality	10	87	3	31
Sex				
Both	18	44	38	228
Male	15	53	32	189
Female	9	53	37	159
Short-term/long-term[a]				
No focus	11	54	36	424
Short-term	13	59	28	75
Long-term	36	19	44	77
Blue/white-collar				
No distinction	15	46	39	446
Blue-collar	24	51	25	71
White-collar	3	78	19	59
Country				
USA	19	33	48	88
UK	18	51	31	83
Europe (one country)	2	73	25	133
Europe (multiple countries)	19	50	31	193
Other (Aus/Chi/Kor/Jpn)	17	33	50	127
Data				
SHARE (Europe)	15	50	24	213
HRS (US)	25	25	50	52
ELSA (UK)	14	54	32	37
Other (Aus/Chi/Kor/Jpn)	13	46	41	241
Panel data	14	50	35	562
Method				
IV	14	48	38	498
Diff-in-Diff	58	42	0	12
RD	14	65	21	66
Publication type				
Unpublished working paper	9	46	45	188
Published	18	51	31	388
Health Economics	18	40	42	40
J Health Economics	26	58	15	65

[a] Short/long-term depends on whether the paper mentions to particularly focus on the short-run, long-run or both.

2.2 Estimation

Based on the categorization presented in Table 1, we estimate which features of past studies explain finding negative, positive, or no significant effects of retirement on health. Due to the large variety of health measures and their measurement scales we discretize the dependent variable into: deteriorating health (value: -1), no significant effect (value: 0), and improving health (value: 1). We make sure that all the different health measures go into the same direction. For example, significantly positive effects on the number of doctor visits reported in the literature is denoted as a negative effect in our study. Similarly, a significant increase in the number of depression symptoms is denoted as a negative effect.

The significance of the result depends on the significance level chosen and reported by the particular study. The features of the studies are included as independent variables and categorized as dummies. Our baseline model estimates the effects of studies' features on outcome by OLS. Since we include multiple results per study, we cluster standard errors at the study-level. Results are robust to different model specifications. To further increase our understanding, we estimate non-linear models, such as an Ordered Probit model and a Multinomial Probit Model, in Online Appendix B. These models provide additional information on how studies' features are associated with the probabilities to find a negative, non-significant, or positive effect.

3 Estimation results

Table 2 presents the estimation results of the meta-analysis using an OLS estimator with a dependent variable that is either -1 (negative effect), 0 (no effect), or 1 (positive effect). In Online Appendix B, we present estimation results using an Ordered Probit Model and a Multinomial Probit

Model, which provide conclusions consistent with the OLS estimator. The different model specifications in columns (1)-(10) indicate that the results are highly robust to the inclusion and exclusion of several independent variables. In column (1) we distinguish several types of health measures. In columns (2)-(7) we sequentially add specific groups of variables. Column (8) contains our baseline specification, and in (9)-(10) we test robustness for alternative specifications.

The estimation results indicate that the type of health measure has a substantial impact on the estimated causal result of retirement on health. Especially, we find that measures based on physical, mental, healthcare, and general measures show significantly more often a positive results compared to mortality (the reference category). The estimates suggest that a positive effect of retirement on health is most often found in studies using a general health measure, which is more prone to potential reporting bias than other measures. It could be that individuals report a better general health in retirement because they are better able to deal with health issues in retirement, although their true health is not better in retirement compared to working.

The coefficients of physical, mental, healthcare utilization, and general health measures do not significantly differ except for physical and general measures (p-value = 0.04), and mental and general measures (p-value = 0.03). This implies that much of the different conclusions from using different health measures comes from using mortality compared to other health measures, where mortality produced relatively negative effects and general health measures produced relatively positive effects.

The country (or countries) of analysis matters very little in explaining estimated effects. We find no significant effects for either the US or European countries relative to the UK. In (8) we do find a positive effect of "other countries" at the 10%-level, suggesting that retirement leads more often to improved health in these countries. "Other countries" primarily include other developed countries with developed pension systems (Mercer, 2018) including Australia, (urban) China, Japan, and Korea. China might be different compared to the other developed countries in the study. However, excluding results based on China (36 observations) does not change the conclusions (not reported here). The estimated effects for these countries do not significantly differ with respect to estimated effects for US and Europe. Conclusions are robust to analyzing the role of different (country-specific) datasets, as can be observed in (9).

Column (8) shows that studies with causal identification from Diff-in-Diff and RD are associated with more negative results than IV estimates. We should note, however, that only 12 of the 576 estimates exploit a Diff-in-Diff, and an RD is used in 66 of the cases (IV in all other 498 cases). Excluding Diff-in-Diff and/or RD estimates does not change the conclusions (not reported here). We do not find an effect of the size of the longitudinal dimension of the datasets used.

Choices in types of people (gender, type of work) analyzed and whether the estimates are for the short- or long-run matter very little. Studies using a less homogenous group of blue- and white-collar workers tend to find positive results more often, but for white-collar workers effects are not significantly different from the reference group of blue-collar workers.

The publication bias is a potentially important threat in meta-analyses. We find that published papers show negative effects significantly more often than working papers (column 8). To investigate publication bias more thoroughly, the model needs to be flexible enough to allow the n.s. category to have a different sign and magnitude compared to categories that contain significantly positive and significantly negative results. Therefore, Table B1 in the Appendix presents the results of a Multinomial Probit Model. This model shows no evidence for a publication bias, as non-significant effects are not significantly more often found in working papers compared to published papers. Table B1 shows that it is primarily positive estimation results that are less often published (15%-points).

In column (10), we distinguish between studies published in top-tier journals in health economics (about 18% of all 576 results), namely *Journal of Health Economics* (65/576 results) and *Health Economics* (40/576 results), compared to studies published in other academic journals. The sign of a published paper in *Health Economics* is not significantly different from the sign in unpublished working papers. In contrast, results published in the *Journal of Health Economics* have similar signs compared to other published results, but are significantly more negative than unpublished effects (i.e. the linear combination of published and this journal, p-value = 0.03). In column (10), significantly positive and significantly negative effects can counterbalance each other. Therefore, we also analyze the results of a Multinomial Probit model in Table B2 of the appendix. These estimates confirm that there is no evidence that insignificant results are more often found in unpublished working papers compared to articles in top-tier health economics journals. *Journal of Health Economics* even published non-significant results more often than is found in working papers (weakly significant). Hence, although economics is arguably more focused on causal identification than other social disciplines, we find no important bias in the publication in top-tier health economics journals.

Finally, estimation results are highly robust to excluding studies that cover a high number of the in total 576 observations, such as Nishimura et al. (103), Leimer (50), Rose (42), Mavromaras et al. (36), and Delugas & Balia (30) (not reported here). Hence, our results are not driven by the results from a particular study.

Table 2. Estimation results on factors explaining the causal outcome of retirement on health.

	1	2	3	4	5	6	7	8	9	10
Health measure										
(ref: Mortality)										
Healthcare utilization	0.199	0.209	0.253	0.245	0.336*	0.319**	0.228	0.562***	0.617***	0.510***
	(0.182)	(0.177)	(0.194)	(0.196)	(0.181)	(0.148)	(0.185)	(0.164)	(0.180)	(0.178)
Physical	0.277**	0.309**	0.341***	0.411**	0.412***	0.265**	0.299***	0.489***	0.630***	0.453**
	(0.107)	(0.125)	(0.124)	(0.166)	(0.147)	(0.107)	(0.105)	(0.162)	(0.199)	(0.183)
Mental	0.250**	0.256*	0.347***	0.417**	0.393**	0.233**	0.281***	0.497***	0.648***	0.485***
	(0.116)	(0.140)	(0.114)	(0.170)	(0.149)	(0.115)	(0.094)	(0.158)	(0.189)	(0.182)
General	0.420***	0.433***	0.486***	0.547***	0.556***	0.451***	0.466***	0.676***	0.824***	0.668***
	(0.139)	(0.155)	(0.104)	(0.182)	(0.166)	(0.132)	(0.115)	(0.152)	(0.187)	(0.175)
Publication type										
(ref: Working paper)										
Published		-0.228						-0.251*	-0.175	-0.256*
		(0.168)						(0.131)	(0.127)	(0.134)
Country										
(ref: UK)										
USA			0.127					0.108		0.120
			(0.190)					(0.217)		(0.215)
EU - single			0.134					0.218		0.287
			(0.157)					(0.192)		(0.217)
EU - multiple			-0.091					0.044		0.104
			(0.138)					(0.217)		(0.240)
Other			0.154					0.403*		0.489**
			(0.250)					(0.233)		(0.238)

Data							
(ref: SHARE)							
HRS	0.212					0.052	
	(0.194)					(0.186)	
ELSA	0.045					-0.045	
	(0.163)					(0.178)	
Other	0.267					0.335**	
	(0.163)					(0.145)	
Panel data		0.014			0.015	0.014	0.0167*
		(0.009)			(0.009)	(0.010)	(0.010)
Method							
(ref: IV)							
Diff-in-Diff			-0.931***		-1.084***	-1.030***	-0.921***
			(0.091)		(0.166)	(0.123)	(0.229)
RD			-0.171		-0.219	-0.341**	-0.119
			(0.155)		(0.197)	(0.158)	(0.182)
(Sub)samples (ref: blue collar, long term, males)							
Blue- & White-collar				0.245**	0.193**	0.231***	0.202**
				(0.101)	(0.086)	(0.079)	(0.089)
White-collar				0.137	0.135	0.160	0.136
				(0.103)	(0.108)	(0.105)	(0.109)
No short/long-term				0.233	0.267	0.267	0.325
				(0.267)	(0.206)	(0.207)	(0.204)
Short-term				0.110	0.334	0.291	0.326
				(0.316)	(0.277)	(0.273)	(0.274)
Males & females				0.083	0.031	0.055	0.043
				(0.097)	(0.087)	(0.077)	(0.086)
Females				0.115	0.056	0.060	0.053
				(0.080)	(0.070)	(0.068)	(0.072)

(ref: Other journal)										
Health Econ										0.298*
										(0.178)
J Health Econ										-0.182
										(0.189)
Constant	-0.065	0.075	-0.189	-0.332*	-0.354*	-0.037	-0.549*	-0.857**	-1.024***	-0.975***
	(0.078)	(0.157)	(0.162)	(0.176)	(0.181)	(0.080)	(0.281)	(0.332)	(0.344)	(0.364)
adj. R-sq	0.014	0.037	0.028	0.037	0.03	0.05	0.034	0.134	0.138	0.147

* denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. All regressions are estimated with OLS using N = 576. Standard errors are clustered at the study-level and are presented within parentheses.

4 Conclusion

Despite many papers analyzing the causal effect of retirement on health, there is no consensus in the literature as to whether retirement deteriorates or improves health. Based on a meta-analysis on 576 causal results, we find that the causal effect of retirement on health is insignificant in almost half of the results, positive for 36% of the results, and negative for 15% of the results. Our results suggest that the variation in estimated effects is primarily due to different health measures and most prominently driven by mortality versus other health measures. Effects of choices in countries, data sets, gender, educational level, length of the panel dimension, and type of causal inference seem to be less important in explaining the variation in effects. Conclusions are highly robust to model specifications, non-linear models, and excluding particular studies, countries, or journals. We do not find evidence for a publication bias; non-significant effects are not significantly less often published in (economic) journals than in working papers. Significantly positive effects, however, are underrepresented in published articles compared to unpublished working papers.

References

Barnay, T. (2016) Health, work and working conditions: A review of the European economic literature. The European Journal of Health Economics, 17, 693-709.

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.

Bellavance, F., Dionne, G. & Lebeau, M. (2009) The value of a statistical life: A meta-analysis with a mixed effects regression model. Journal of Health Economics, 28(2), 444-464.

Bingley, P. & Martinello, A. (2013) Mental retirement and schooling. European Economic Review, 63, 292-298.

Bonsang, E. & Klein, T.J. (2012) Retirement and subjective well-being. Journal of Economic Behavior & Organization, 83(3), 311-329.

Charles, K.K. (2004) Is retirement depressing? Labor force inactivity and psychological well-being in later life. Research in Labor Economics, 23, 269-299.

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

Coe, N., & Lindeboom, M. (2008). Does retirement kill you? Evidence from early retirement windows. IZA Discussion Papers, No. 3817.

Cornelsen, L., Green, R., Turner, R., Dangour, A.D., Shankar, B., Mazzocchi, M. & Smith, R.D. (2015) What happens to patterns of food consumption when food prices change? Evidence from a systematic review and meta-analysis of food price elasticities globally. Health Economics, 24(12), 1548-1559.

Currie, J. & Madrian, B. (1999) Health, health insurance, and the labor market. In: Handbook of Labor Economics. Ed. by O. Ashenfelter & David Card, Ch. 3, part C. Oxford: Elsevier, 3309-3415.

De Grip, A., 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.

Delugas, E., & Balia, S. (2019). A life change for the better? The health consequence of retirement. Mimeo.

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

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.

Gallet, C.A. & List, J.A. (2002) Cigarette demand: A meta-analysis of elasticities. Health Economics, 12(10), 821-835.

Gallet, C.A. (2013) Can price get the monkey off our back? A meta-analysis of illicit drug demand. Health Economics, 23(1), 55-68.

Gemmill, M.C., Costa-Font, J. & McGuire, A. (2007) In search of a corrected prescription drug elasticity estimate: A meta-regression approach. Health Economics, 16(6), 627-643.

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

Leimer, B. (2017). No 'Honeymoon Phase'-Whose Health Benefits from Retirement and When. Gutenberg School of Management and Economics & Research Unit "Interdisciplinary Public Policy" Discussion Paper Series, (1718).

Mavromaras, K., Richardson, S., & Zhu, R. (2013). Age pension, age eligibility, retirement and health outcomes in Australia. National Institute of Labour Studies WP, (201).

Nelson, J.P. (2014) Estimating the price elasticity of beer: Meta-analysis of data with heterogeneity, dependence, and publication bias. Journal of Health Economics, 33, 180-187.

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.

Rose, L. (2018). Healthcare for the Elderly, Doctoral dissertation, UC Santa Cruz.

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.

Trapero-Bertran, M., Mistry, H., Shen, J. & Fox-Rushby, J. (2013) A systematic review and meta-analysis of willingness-to-pay values: The case of malaria control interventions. Health Economics, 22(4), 428-450.

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), 1180.

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.

Zulka, L., Hansson, I. & Hassing, L. (2019). Impact of Retirement on Cognitive Function. GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry, 32 (4), 187-203.

Online Appendix A - Studies included in the meta-analysis

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.

Barnay, T. & Defebvre, É. (2018). Retired, at last? The short-term impact of retirement on health status in France. Mimeo.

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

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

Bertoni, M., Maggi, S. & Weber, G. (2018). Work, retirement, and muscle strength loss in old age. Health Economics, 27 (1), 115-128.

Bianchini, L. & Borella, M. (2015). Retirement and memory in Europe. Ageing and Society, 36, 1434–1458.

Bingley, P. & Martinello, A. (2013), Mental retirement and schooling. European Economic Review, 63, 292-298.

Binh Tran, D. & Zikos, V. (2019). The Causal Effect of Retirement on Health: Understanding the Mechanisms. Australian Economic Review, 52 (4), 427-446.

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), 1–14.

Bonsang, E., Adam, S. & Perelman, S. (2012), Does Retirement Affect Cognitive Functioning?. Journal of Health Economics, 31, 490-501.

Bozio, A., Garrouste, C., & Perdrix, E. (2019). Impact of later retirement on mortality: Evidence from France. Mimeo.

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

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.

Charles, K.K. (2004) Is retirement depressing? Labor force inactivity and psychological well-being in later life. Research in Labor Economics, 23, 269-299.

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

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

Clouston, S. A., & Denier, N. (2017). Mental retirement and health selection: analyses from the US Health and Retirement Study. Social Science & Medicine, 178, 78-86.

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

Coe, N.B. & Zamarro. G. (2015a). Does retirement impact health care utilization? CESR-Schaeffer Working Paper No. 2015-032.

Coe, N.B. & Zamarro. G. (2015b). How does retirement impact health behaviors? An international comparison. CESR-Schaeffer Working Paper No. 2015–033.

Delugas, E., & Balia, S. (2019). A life change for the better? The health consequence of retirement. Mimeo.

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

Etgeton, S. (2019). The Impact of Pension Reforms on Income Inequality, Savings, and Health, Doctoral dissertation Berlin University.

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.

Fonseca, R., Kapteyn, A., Lee, J., Zamarro, G. & Feeney, K. (2015). A longitudinal study of well-being of older Europeans: does retirement matter?. Journal of Population Ageing 7 (1), 21-41.

Godard, M. (2016). Gaining weight through retirement? Results from the SHARE survey. Journal of Health Economics, 45 (C), 27-46.

Gorry, A., Gorry, D., & Slavov, S. N. (2018). Does retirement improve health and life satisfaction?. Health Economics, 27 (12), 2067-2086.

Grøtting, M. W. & Lillebø, O. S. (2019). Health Effects of Retirement: Evidence from Survey and Register Data. Empirical Essays on Health and Aging, 71, 1-34.

Habets, M. (2018). Retirement and cognitive functioning: A longitudinal analysis. University of Liège.

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.

Hanemann, F. (2017). Working Conditions, Retirement and Health–Longitudinal Evidence from Europe and the US. MEA Discussion Papers No. 07-2017.

Heller-Sahlgren, G. (2017). Retirement blues. Journal of Health Economics, 54, 66-78.

Hernaes, E., Markussen, S., Piggott, J. & Vestad, O.L. (2013). Does retirement age impact mortality?. Journal of Health Economics, 32, 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.

Insler, M. (2014). The health consequences of retirement. Journal of Human Resources, 49 (1), 195-233.

Horner, E. M. & Cullen, M. R. (2016). The impact of retirement on health: Quasi-experimental methods using administrative data. BMC Health Services Research 19, 16–68.

Johnston, D. & Lee, W.-S. (2009). Retiring to the good life? The short-term effects of retirement on health. Economic Letters 103, 8–11.

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

Kuhn, A., Staubli, S., Wuellrich, J. P. & Zweimüller, J. (2018). Fatal Attraction? Extended Unemployment Benefits, Labor Force Exits, and Mortality, NBER Working Paper, 1–43.

Latif, E. (2013). The impact of retirement on mental health in Canada. Health Policy and Economics, 16 (1), 35–46.

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.

Leimer, B. (2017). No 'Honeymoon Phase'-Whose Health Benefits from Retirement and When. Gutenberg School of Management and Economics & Research Unit "Interdisciplinary Public Policy" Discussion Paper Series, (1718).

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

Mavromaras, K., Richardson, S., & Zhu, R. (2013). Age pension, age eligibility, retirement and health outcomes in Australia. National Institute of Labour Studies WP, (201).

Mazzonna, F., & Peracchi, F. (2012). Ageing, Cognitive Abilities and Retirement. European Economic Review 56 (4), 691–710.

Mazzonna, F., & Peracchi, F. (2017). Unhealthy Retirement? The Journal of Human Resources, 52 (1), 128.

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.

Oi, K. (2019). Does gender differentiate the effects of retirement on cognitive health?. Research on Aging, 41, 575–601.

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.

Rohwedder, S. & Willis, R.J. (2010). Mental retirement. Journal of Economic Perspectives 24 (1), 1–20.

Rose, L. (2018). Healthcare for the Elderly, Doctoral dissertation, UC Santa Cruz.

Saxena, A. (2018). Effect of Retirement on Health and Mortality (Doctoral dissertation).

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.

Tumino, A. (2016). Retirement and cognitive abilities, ISER Working Paper Series (No. 2016-06).

Vo, T. T., & Tran, D. T. (2019). The Retirement Effect on Mental Health in Europe During 2004-2015: Evidence of Ashenfelter's Dip from SHARE Panel Data. Available at SSRN 3357265.

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.

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

Online Appendix B – Nonlinear models

Table B1 presents estimation results of the meta-analysis using an Ordered Probit Model and a Multinomial Probit Model, with a dependent variable that is either -1 (negative effect), 0 (no significant effect), or 1 (positive effect). Compared to the OLS estimator in Section 3, this provides us with additional information on how the independent variables marginally increase/decrease the probability of finding a negative, no, or positive effect. In Table B1, we use the same variables as in Column 8 in Table 2 (the baseline specification). In the Multinomial Probit Model we take no significant effect (n.s.) as the reference category. The table reports marginal effects. For example, the Multinomial Probit Model in Table B1 shows that studies that use a mental health measure show 55%-points more often a significantly positive effect of retirement on health compared to studies that use mortality (at means of the covariates).

One of the main conclusions from Table 2 is that the different outcomes are strongly driven by different health measures. Since the coefficients of physical, mental, healthcare, and general measures do not significantly differ, except for physical and general measures and mental and general measures, much of the different results stem from using mortality as a health measure compared to other measures. Table B1 shows that these conclusions also hold for non-linear estimators. Moreover, the results in Table B1 indicate that these effects of different health measures (compared to mortality) primarily imply substantially higher probabilities of finding a significantly positive effect at the cost of non-significant but especially at the cost of significantly negative effects.

Furthermore, we find that significantly positive effects of retirement on health are less often found in published papers compared to working papers (15-16%-points). Results from countries other than Europe or the US show more often a positive effect of retirement on health (about 23-26%-

points more often compared to the UK, only significant in the ordered probit model). Compared to using IV for causal inference, Diff-in-Diff gave more often negative or non-significant results. Finally, our results suggest that using both blue-collar and white-collar workers in the sample tends to increase the likelihood of finding a positive effect of retirement on health (12-13%-points) while decreasing the probability of negative and non-significant effects.

Table B1. Estimation results using non-linear estimators.

	Order Probit			Multinomial Probit		
	neg.	n.s.	pos.	neg.	n.s.	pos.
Health measure (ref: Mortality)						
Healthcare utilization	-0.202***	-0.131***	0.333***	-0.365***	-0.252	0.616***
	(0.066)	(0.038)	(0.094)	(0.106)	(0.182)	(0.222)
Physical	-0.176***	-0.114***	0.289***	-0.294***	-0.299	0.594**
	(0.064)	(0.032)	(0.087)	(0.099)	(0.193)	(0.235)
Mental	-0.179***	-0.115***	0.294***	-0.346***	-0.207	0.554**
	(0.062)	(0.034)	(0.088)	(0.099)	(0.201)	(0.237)
General	-0.248***	-0.160***	0.408***	-0.400***	-0.281	0.681**
	(0.063)	(0.039)	(0.085)	(0.107)	(0.184)	(0.226)
Publication type (ref: Working paper)						
Published[a]	0.099*	0.064*	-0.163*	0.075	0.075	-0.150*
	(0.054)	(0.036)	(0.088)	(0.048)	(0.063)	(0.083)
Country (ref: UK)						
USA	-0.043	-0.028	0.071	-0.001	-0.108	0.107
	(0.078)	(0.052)	(0.130)	(0.079)	(0.097)	(0.120)
EU - single	-0.080	-0.052	0.132	-0.297***	0.241*	0.056
	(0.070)	(0.044)	(0.112)	(0.0689)	(0.128)	(0.125)
EU - multiple	-0.018	-0.011	0.029	-0.060	0.064	-0.003
	(0.079)	(0.051)	(0.130)	(0.066)	(0.096)	(0.126)
Other	-0.159*	-0.103	0.262*	-0.089	-0.137	0.227
	(0.093)	(0.064)	(0.153)	(0.062)	(0.117)	(0.143)
Data (ref: Cross-section)						
Panel waves	-0.006	-0.004*	0.009*	-0.005	-0.002	0.008
	(0.004)	(0.002)	(0.006)	(0.005)	(0.005)	(0.006)
Method (ref: IV)						
Diff-in-Diff	0.408***	0.264***	-0.672***	1.028***	1.770***	-2.798***
	(0.066)	(0.068)	(0.103)	(0.094)	(0.114)	(0.137)
RD	0.082	0.053	-0.136	0.034	0.142*	-0.177

	(0.073)	(0.044)	(0.116)	(0.067)	(0.088)	(0.119)
(Sub)samples (ref: blue collar, long-term, males)						
Blue- & White-collar	-0.074**	-0.048*	0.122**	-0.042*	-0.089	0.131*
	(0.031)	(0.025)	(0.054)	(0.025)	(0.068)	(0.070)
White-collar	-0.049	-0.032	0.081	-0.264***	0.304***	-0.041
	(0.040)	(0.027)	(0.066)	(0.091)	(0.068)	(0.102)
No short/long-term	-0.098	-0.063	0.161	-0.187***	0.210***	-0.022
	(0.080)	(0.048)	(0.127)	(0.062)	(0.066)	(0.092)
Short-term	-0.126	-0.081	0.207	-0.200**	0.230**	-0.030
	(0.106)	(0.060)	(0.164)	(0.092)	(0.103)	(0.145)
Males & females	-0.014	-0.009	0.024	0.055	-0.147***	0.093*
	(0.031)	(0.022)	(0.053)	(0.036)	(0.046)	(0.053)
Females	-0.022	-0.015	0.037	-0.025	0.002	0.023
	(0.025)	(0.017)	(0.043)	(0.036)	(0.038)	(0.046)

* denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. N=576. This table presents marginal effects. Standard errors are clustered at the study-level and presented within parentheses.

[a] A Chi2-test shows that the estimated coefficients for neg. and n.s. are not significantly different with p-value = 0.221 in the Ordered Probit (p = 0.994 in the Multinomial Probit).

In Table B2, we re-estimate the Multinomial Probit Model of Table B1, but extend the specification by including dummies for the studies that are published in *Health Economics* or *Journal of Health Economics*. In this way, we can test whether these journals published relatively often studies with either negative, non-significant, or positive effects. We report marginal effects of 'Published', 'Health Economics' and 'J Health Economics'. Other marginal effects are similar to those presented in Table B1. According the Multinomial Probit Model, published studies are significantly more often reporting negative effects and significantly less often report positive effects compared to working papers. Published studies are not more or less likely to present non-significant studies than non-published studies. Studies published in *Journal of Health Economics* do not significantly deviate from other published studies. Compared to working papers however, this journal more often presents non-significant results (weakly significant, p-value = 0.07). Contrasting other journals, we find that studies published in *Health Economics* do not present significantly negative effects more often than is found in working papers.

Table B2. Estimation results of the Multinomial Probit Model with top-tier health economics journals.

	Multinomial Probit		
	neg.	n.s.	pos.
Publication type (ref: Working paper)			
Published[a]	0.090*	0.063	-0.153*
	(0.056)	(0.067)	(0.083)
(ref: Other)			
Health Econ[a]	-0.139*	-0.038	0.177*
	(0.080)	(0.107)	(0.102)
J Health Econ[b]	0.036	0.160	-0.196
	(0.054)	(0.116)	(0.147)

* denotes significance at the 10% level, ** at the 5% level and *** at the 1% level. N=576. Presented coefficients are marginal effects. Standard errors are clustered at the study-level and presented within parentheses.
[a] Published + Health Econ gives p-values of 0.61, 0.83, and 0.06 for negative, non-significant, and positive effects respectively.
[b] Published + J Health Econ gives p-values of 0.05, 0.07, and 0.43 for negative, non-significant, and positive effects respectively.