

## **The effect of retirement on loneliness:**

### **A longitudinal comparative analysis across Australia, China, and the United States**

#### **Supplementary files**

##### **List of Supplementary Files**

Appendix 1 - STROBE Statement—Checklist of items that should be included in reports of cohort studies .....	4
Appendix 2 – recoding process and original variables .....	7
Retirement .....	7
Nature of retirement .....	7
Loneliness.....	8
Social Engagement:.....	8
Mental health:.....	9
Health status .....	10
Appendix 3 – Statistical analysis .....	18
References .....	21
Appendix 4 – Additional results .....	32

### List of Supplementary Tables

Table S1.1. STROBE Statement—Checklist of items that should be included in reports of cohort studies .....	4
Table S2.1. Surveys original questions, responses options and recoding.....	12
Table S3.1. Summary of pooled samples and missing data according to wave .....	23
Table S3.2. Summary of pooled samples according to retirement status and wave .....	24
Table S4.1. Lagged model of loneliness prevalence according to retirement status in Australia, China and US .....	32
Table S4.2. Fully lagged model of loneliness estimated adjusted prevalence according to retirement status in Australia, China and US.....	32
Table S4.3. loneliness estimated adjusted prevalence according to nature of retirement among Australia and the US .....	33
Table S4.4. The modifying effect of social engagement on the association between the estimated adjusted prevalence of loneliness and retirement according to social engagement status among the three surveys: Australia, China and the US .....	34
Table S4.5. Adjusted estimated adjusted prevalence of loneliness according to retirement status among urban participants in Australia, China and the US.....	35

### List of Supplementary Figures

Figure S3.1. Participants flow chart.....	25
Figure S3.2. Household, Income and Labour Dynamics in Australia (HILDA) primary analysis balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) .....	25
Figure S3.3. China Health and Retirement Study (CHARLS) primary analysis balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	26
Figure S3.4. United-States Health and Retirement Study (HRS) primary analysis balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	27
Figure S3.5. Household, Income and Labour Dynamics in Australia (HILDA) nature of retirement balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	27
Figure S3.6. United-States Health and Retirement Study (HRS) nature of retirement balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	28
Figure S3.7. Household, Income and Labour Dynamics in Australia (HILDA) urban sample balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	28

Figure S3.8. China Health and Retirement Study (CHARLS) urban sample balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	29
Figure S3.9. United-States Health and Retirement Study (HRS) urban sample balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	29
Figure S3.10. Household, Income and Labour Dynamics in Australia (HILDA) fully lagged model balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	30
Figure S3.11. China Health and Retirement Study (CHARLS) fully lagged model balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	30
Figure S3.12. United-States Health and Retirement Study (HRS) fully lagged model balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup> .....	31
Figure S4.1. Adjusted estimated prevalence of loneliness according to retirement status among the three countries: Australia, China and the US (fully lagged model).....	35

## Appendix 1 - STROBE Statement—Checklist of items that should be included in reports of cohort studies

**Table S1.1. STROBE Statement—Checklist of items that should be included in reports of cohort studies**

	<b>Item No</b>	<b>Recommendation</b>	<b>Page No</b>
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1-2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
<b>Introduction</b>			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
<b>Methods</b>			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	6
		(b) For matched studies, give matching criteria and number of exposed and unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-8
Bias	9	Describe any efforts to address potential sources of bias	8-9
Study size	10	Explain how the study size was arrived at	10, Tables S4.1-S4.2
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8, Appendix 2

Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	8-10, Appendix 3
		(b) Describe any methods used to examine subgroups and interactions	9
		(c) Explain how missing data were addressed	9-10, Appendix 4
		(d) If applicable, explain how loss to follow-up was addressed	9-10, Appendix 4
		(e) Describe any sensitivity analyses	9
<b>Results</b>			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	10, Tables S4.1-S4.2
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	Figure S4.1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10
		(b) Indicate number of participants with missing data for each variable of interest	Table S4.1
		(c) Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	Report numbers of outcome events or summary measures over time	11-12
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11-12
		(b) Report category boundaries when continuous variables were categorized	11-12
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12
<b>Discussion</b>			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	15-16

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-16
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-16
<b>Other information</b>			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	17

\*Give information separately for exposed and unexposed groups.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at <http://www.strobe-statement.org>.

## Appendix 2 – recoding process and original variables

### Retirement

- **Australia:** Retirement was measured by the item “retired completely from the workforce” (‘yes’, ‘no’, ‘never in workforce’) which was asked only among those who were not working and above the age of 45. Those who were not asked this question were recoded as having a working status. This question was available in all waves except from wave 11 which had a similar question: “Do you consider yourself to be completely retired from the paid workforce, partly retired or not retired at all?” with four optional answers: ‘completely retired’, ‘partly retired’, ‘not retired at all’, ‘not relevant - have never been in paid work’. Both items were recoded as 0 = ‘not retired’/ ‘partly retired’, 1 = ‘retired’, 2 = ‘never worked’, in the final analysis.
- **China:** Measured by ‘work status’ with possible answers including: ‘Work in agriculture’, ‘work in non-agriculture’, ‘retired’, ‘unemployed’, ‘never worked’. The item was recoded to the following: 0 = ‘work in agriculture’/ ‘work in non-agriculture’, 1 = ‘retired’, 2 = ‘unemployed’/ ‘never worked’.
- **US:** Participants were asked about their retirement: “do you consider yourself to be completely retired, partly retired, or not retired at all?” Answers were recoded to the following: 0 = ‘not retired’/ ‘partly retired’, 1 = ‘completely retired’, 2 = ‘not relevant’ (never worked).

### Nature of retirement

Was available only in HRS and HILDA.

- **HILDA:** Twenty-three items regarding different reason for retirement (e.g., “could not find another job”, “own ill health”, “reached compulsory retirement age”). These items were available only in waves 2011 and 2015. Data on nature of retirement was available in 2019 as well. However, since we used a lagged model analysis, the loneliness outcome for that wave should have been taken from the subsequent wave. Since we excluded waves after 2019 due to potential COVID-19 impact, we could not include the 2019 nature of retirement data). A new item was computed in which 1 = those who only indicated voluntary reasons for retirement, 2 = those who only indicated involuntary reasons for retirement, 3 = those who indicated both voluntary and involuntary reasons for retirement. Health problems, caring responsibilities and employment issues (e.g., made redundant/ dismisses/ had no choice) reflected

involuntary reasons, while age and lifestyle considerations (e.g., to spend more time with spouse/ partner) reflected voluntary reasons (1, 2).

- **HRS:** examined by an item which asked whether retirement was wanted or forced with possible answers (1 = Wanted, 2 = Forced, 3 = Part wanted, part forced).

### Loneliness

- **HILDA:** Loneliness was measured by a single item (“I often feel very lonely”) derived from a loneliness and social isolation scale constructed for the HILDA survey purposes (3). The item was on a 7-point Likert scale (1 = Strongly disagree, 7 = Strongly agree). To be consistent with the two other surveys, loneliness was dichotomised to: 0 = levels 1-4, 1 = levels 5-7 (indicating agreement with the statement), based on previous studies that used the same item (4, 5).
- **CHARLS:** Loneliness was measured by a single item from the Center for Epidemiologic Studies Depression Scale (CES-D) regarding experiencing loneliness during the past week (6). The item was on a 4-point Likert scale: 1. ‘Rarely or none of the time (less than 1 day)’, 2. ‘Some or a little of the time (1-2 days)’, 3. ‘Occasionally or a moderate amount of time (3-4 days)’, 4. ‘Most or all of the time (5-7 days)’. To be consistent with the HRS loneliness item, the item was recoded to 0 = levels 1-2 in the original scale, 1 = level 3-4 in the original scale.
- **HRS:** Similar to CHARLS, loneliness was measured by a single item from the Center for Epidemiologic Studies Depression Scale (CES-D) regarding experiencing loneliness during the past week (6). The original scale had Four levels: 1. ‘Rarely or none of the time (less than 1 day)’, 2. ‘Some or a little of the time (1-2 days)’, 3. ‘Occasionally or a moderate amount of time (3-4 days)’, 4. ‘Most or all of the time (5-7 days)’. The item was recoded by HRS as 0 = levels 1-2 in the original scale, 1 = level 3-4 in the original scale (7).

### Social Engagement:

- **Australia:** Measured by one item from the HILDA self-completion questionnaire: “In general, how often do you get together socially with friends or relatives not living with you?” The item was measured on a 7-point Likert scale (1 = ‘every day’, 7 = ‘less often than once every 3 months’). To be consistent with the other two surveys,



the item was recoded as follows: 0 = levels 6-7 (not socially active), 1 = levels 1-5 (socially active).

- **China:** Social Engagement was assessed by the item: “Have you interacted with friends in the past month?” (‘no’/‘yes’). This item was part of a list of activities for which participants were asked whether they participated in these activities. Since most of the activities were not necessarily indicative of social engagements, and since the social engagement items that were available in the two other surveys referred to friends or family engagements, we decided to use this item alone.
- **US:** An overall score was built from a series of questions assessing the number of social interactions and frequency of engagements. Four questions ask respondents if they have spouses/partners, children, family, and friends (‘no’/‘yes’), followed by four questions to assess the extent to which respondents are in contact with each of their social networks excluding spouses: meet up, speak on the phone, write or email and using social media (from 2014 onwards). The Four items were rated on a 6-point Likert scale (1 = Three or more times a week, 6 = Less than once a year or never) (8). For the current study, we used the questions evaluating relationships with family and friends, similar to the CHARLS and HILDA items. Communication by social media was excluded as it was not available for all of the waves (9). The Three items included were dichotomised to 0 = ‘three or more times a week’ – ‘once or twice a month’, 1 = ‘every few months’ – ‘less than once a year or never’. A count variable was built from the Six items (Three for each contact) which was then dichotomised to 0 = ‘no social engagements’, 1 = ‘at least one social engagements’.

#### Mental health:

- **Australia:** Mental health was measured by 5 items from the emotional wellbeing section of the 36-Item Short Form Health Survey (SF-36) (10). Participants were asked about the frequency of experiencing different emotions. Items were on a 6-point Likert scale (1 = 1 all of the time, 6 = none of the time). Negative items were recoded to reflect worse mental health with a higher score. Total items were summarised and ranged between 5-30.
- **China:** Mental health was measured by the 10-items CES-D-R10 scale regarding how participants felt and behaved during the past week (6). Items were on a 4-point Likert scale (0 = Rarely or none of the time, 3 = Most or all of the time). We excluded the

item: “I felt lonely” since it was already used as a measure of loneliness. Positive items were recorded, and the total items were summarised. Scores ranged from 0-30 with a higher score indicating worse mental health (11).

- **US:** The mental health scale included Eight items from the CES-D scale regarding how participants felt and behaved during the past week (6). The original scale was 4-point Likert scale (as described above). Answers were recoded by HRS as 0 = levels 1-3 in the original CES-D scale, 1 = level 4 in the original CES-D scale. A count score of negative items (i.e., “I had trouble keeping my mind on what I was doing”) minus the positive items (i.e., “I was happy”) was created (7). We excluded the item: “I felt lonely” since it was already used as a measure of loneliness. Scores ranged between -2 and 5 with a higher score indicating worse mental health.

#### Health status

In all three surveys, reported health status was measured by one item from the SF-36 questionnaire (10) questionnaire on a 5-point Likert scale (1 = Excellent, 5 = poor).

Summary of the recoding process is in **Table S2.1**.

## References

1. Ding D, Grunseit AC, Chau JY, Vo K, Byles J, Bauman AE. Retirement-A Transition to a Healthier Lifestyle?: Evidence From a Large Australian Study. *Am J Prev Med*. 2016;51(2):170-8.
2. Vo K, Forder PM, Tavenor M, Rodgers B, Banks E, Bauman A, et al. Retirement, age, gender and mental health: findings from the 45 and Up Study. *Aging Ment Health*. 2015;19(7):647-57.
3. Manera KE, Smith BJ, Owen KB, Phongsavan P, Lim MH. Psychometric assessment of scales for measuring loneliness and social isolation: an analysis of the household, income and labour dynamics in Australia (HILDA) survey. *Health and quality of life outcomes*. 2022;20(1):40.
4. Kung CSJ, Kunz JS, Shields MA. Economic Aspects of Loneliness in Australia. *Aust Econ Rev*. 2021;54(1):147-63.
5. Astell-Burt T, Hartig T, Eckermann S, Nieuwenhuijsen M, McMunn A, Frumkin H, et al. More green, less lonely? A longitudinal cohort study. *Int J Epidemiol*. 2022;51(1):99-110.
6. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Assessment*. 1977;1:385-401.
7. Health and Retirement Study. RAND HRS Longitudinal File (V2) public use dataset. . Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). 2018
8. Smith J, Ryan G, Fisher GG, Sonnega A, Weir W. HRS Psychosocial and Lifestyle Questionnaire 2006-2016. Documentation Report Core Section LB. Ann Arbor, Michigan: Survey Research Center Institute for Social Research University of Michigan; 2017.
9. Smith J, Ryan L, Fisher G, Sonnega A, Weir D. HRS Psychosocial and Lifestyle Questionnaire 2006-2016. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, University of Michigan; 2017.
10. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-83.
11. Phillips D, Green H, Petrosyan S, Shao K, Wilkens J, Lee J. Gateway to Global Aging Data. The Harmonised CHARLS dataset and Codebook, Version D June 2021 (R01 AG030153, RC2 AG036619, R03 AG043052). The National Institute on Aging 2021.

**Table S2.1. Surveys original questions, responses options and recoding**

		<b>Australia (1)</b>	<b>China (2)</b>	<b>US (3)</b>	<b>Final item recoding</b>
<b>Retirement</b>	<b>Question</b>	Retired completely from the workforce:	Work status	Do you consider yourself to be completely retired, partly retired, or not retired at all?	0 = not retired 1 = retired 2 = never worked
	<b>Responses</b>	1. Yes 2. No 3. Never in the workforce	1. Work in agriculture 2. Work in non-agriculture 3. Retired 4. Unemployed 5. Never worked	1. Completely retired 2. Partly retired 3. Not retired at all 4. Not relevant (never worked)	
	<b>Recode</b>	0 = level 2 1 = level 1 2 = level 3	0 = levels 1-2 1 = level 3 2 = levels 4-5	0 = levels 2-3 1 = level 1 2 = level 4	
<b>Nature of retirement</b>	<b>Question</b>	Twenty-three items regarding different possible reason for retirement (e.g., “could not find another job”, “own ill health”, “reached compulsory retirement age”).	No variable available	Thinking back to the time you [partly/completely] retired, was that something you wanted to do or something you felt you were forced into?	0 = not retired 1 = retired voluntary 2 = retired involuntary 3 = retired partly voluntary partly involuntary 4 = never worked
	<b>Responses</b>	0 = no 1 =yes		1. Wanted to 2. Forced to 3. Partly wanted partly forced	
	<b>Recode</b>	1 = those who only indicted voluntary reasons for retirement 2 = those who only indicated involuntary reasons for retirement 3 = those who indicated both voluntary and involuntary reasons for retirement		1 = level 1 2 = level 2 3 = level 3	
<b>Loneliness</b>	<b>Question</b>	The following statements have been used by many people to describe how much support they get from other people. How much do you	I felt lonely during the past week (CES-D)	I felt lonely during the past week (CES-D)	0 = not lonely 1 = lonely

		<b>Australia (1)</b>	<b>China (2)</b>	<b>US (3)</b>	<b>Final item recoding</b>
		agree or disagree with each? The more you agree, the higher the number of the box you should cross. The more you disagree, the lower the number of the box you should cross. I often feel very lonely (HILDA questionnaire (4)). 7-point Likert scale			
	<b>Responses</b>	1. Strongly disagree 7. Strongly agree	1. Rarely or none of the time (less than 1 day) 2. Some or a little of the time (1-2 days) 3. Occasionally or a moderate amount of time (3-4 days) 4. Most or all of the time (5-7 days)	1. Rarely or none of the time (less than 1 day) 2. Some or a little of the time (1-2 days) 3. Occasionally or a moderate amount of time (3-4 days) 4. Most or all of the time (5-7 days)	
	<b>Recode</b>	0 = levels 1-4 1 = levels 5-7	0 = levels 1-2 1 = levels 3-4	0 = levels 1-2 1 = levels 3-4	
<b>Social engagement</b>	<b>Question</b>	In general, how often do you get together socially with friends or relatives not living with you?	Have you interacted with friends in the past month?	1. Do you have any other immediate family, for example, any brothers or sisters, parents, cousins or grandchildren 2. Do you have any friends? 3. How often do you a) meet up, b) speak on the phone, c) write or email?	0 = not socially engaged 1 = socially engaged

		Australia (1)	China (2)	US (3)	Final item recoding
	<b>Responses</b>	1. Every day 2. Several times a week 3. About once a week 4. 2 or 3 times a month 5. About once a month 6. Once or twice every 3 months 7. Less often than once every 3 months	1. no 2. yes	1-2. yes/no 3(a-c). 6-point Likert scale (1 = Three or more times a week, 6 = Less than once a year or never)	
	<b>Recode</b>	0 = levels 6-7 (not socially active) 1 = levels 1-5 (socially active)	0 = level 1 1 = level 2	6 items recoded: 0 = 'three or more times a week' – 'once or twice a month' 1 = 'every few months' – 'less than once a year or never' A count variable was built from the Six items (Three for each contact) which was then dichotomised to 0 = 'no social engagements' 1 = 'at least one social engagements'.	
<b>Mental health</b>	<b>Question</b>	SF-36 (5): emotional wellbeing scale	Eight items from the CES-D scale (6) (excluding 'loneliness')		Summary score from each survey was used
	<b>Responses</b>	1. All of the time 2. Most of the time 3. A good bit of the time 4. Some of the time 5. A little of the time 6. None of the time	1. Rarely or none of the time (less than 1 day) 2. Some or a little of the time (1-2 days) 3. Occasionally or a moderate amount of time (3-4 days) 4. Most or all of the time (5-7 days)	1. no = 'none' – 'occasionally' in the original scale 2. yes = 'most or all of the time' in the original scale.	

		<b>Australia (1)</b>	<b>China (2)</b>	<b>US (3)</b>	<b>Final item recoding</b>
	<b>Recode</b>	Five items summary score	Summary score of 8 items	A count score of negative items minus the positive items.	
<b>Health status</b>		Self-assessed health- SF-36 (5) 1. Excellent 2. Very good 3. Good 4. Fair 5. Poor			
<b>Age</b>		Numerical values			
<b>Gender</b>		1. Male 2. Female			
<b>Education</b>	<b>Responses</b>	1. Postgrad - masters or doctorate 2. Grad diploma, grad certificate 3. Bachelor or honours 4. Adv diploma, diploma 5. Cert III or IV 6. Year 12 7. Year 11 and below	1. Did not finish primary school but capable of reading and/or writing 2. Elementary school 3. Middle school 4. No formal education (illiterate) 5. Sishu/home school 6. High school 7. Two-/Three-Year College/Associate degree 8. Vocational school 9. Doctoral degree/Ph.D. 10. Four-Year College/Bachelor's degree 11. Master's degree	1. Lt High-school 2. GED 3. High-school graduate 4. Some college 5. College and above	1. Below high school 2. High school 3. Tertiary-not university 4. College/university
	<b>Recode</b>	1 = level 7 2 = level 6 3 = levels 2, 4-5 4 = levels 1, 3	1 = levels 1-5 2 = level 6 3 = 7-8 4 = 9-11	1 = level 1 2 = levels 2-3 3 = level 4 4 = level 5	
<b>Marital status</b>	<b>Responses</b>	1. Legally married	1. Cohabitated	1. Married	

		<b>Australia (1)</b>	<b>China (2)</b>	<b>US (3)</b>	<b>Final item recoding</b>
		2. De facto 3. Separated 4. Divorced 5. Widowed 6. Never married and not de facto	2. Married with spouse present 3. Married but not living with spouse temporarily for reasons such as work 4. Divorced 5. Never married 6. Separated 7. Widowed	2. Married, spouse absent 3. Partnered 4. Separated 5. Divorced 6. Separated/ divorced 7. Widowed 8. Never married	1. Married/partnered 2. Not married 3. Widowed
	<b>Recode</b>	1 = levels 1-2 2 = levels 3-4, 6 3 = 5	1 = levels 1-2 2 = levels 3-6 3 = 7	1 = levels 1, 3 2 = levels 2, 4-6, 8 3 = levels 7	
<b>Location of residence</b>	<b>Responses</b>	1. Major City 2. Inner Regional Australia 3. Outer Regional Australia 4. Remote Australia 5. Very Remote Australia	1. Urban community 2. Rural village	1. Urban (Beale Rural-Urban Continuum code 1) 2. Suburban (Beale Rural-Urban Continuum code 2) 3. Ex-urban (Beale Rural-Urban Continuum codes 3,4,5,6,7,8,9) 9. No match for Rural-Urban code	1. Urban 2. Rural
	<b>Recode</b>	0 = level 1 1 = levels 2-3 2 = levels 4-5		1 = level 1 2 = level 2 3 = level 3	

**Abbreviations:** CES-D: Center for Epidemiologic Studies Depression Scale; SF-36: The Short Form survey 36 items



## References

1. Summerfield M, Garrard, B, Jin, Y, Kamath, R, Macalalad, N, Watson, N, Wilkins, R, Wooden, M. HILDA User Manual – Release 20. Melbourne Institute: Applied Economic and Social Research. University of Melbourne; 2021.
2. Phillips D, Green H, Petrosyan S, Shao K, Wilkens J, Lee J. Gateway to Global Aging Data. The Harmonised CHARLS dataset and Codebook, Version D June 2021 (R01 AG030153, RC2 AG036619, R03 AG043052). The National Institute on Aging 2021.
3. Health and Retirement Study. RAND HRS Longitudinal File (V2) public use dataset. . Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). 2018
4. Manera KE, Smith BJ, Owen KB, Phongsavan P, Lim MH. Psychometric assessment of scales for measuring loneliness and social isolation: an analysis of the household, income and labour dynamics in Australia (HILDA) survey. *Health and quality of life outcomes*. 2022;20(1):40.
5. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-83.
6. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Assessment*. 1977;1:385-401.

### Appendix 3 – Statistical analysis

We estimated the effect of retirement on loneliness using marginal structural models estimated using inverse probability of treatment weights (IPTW) (1).

IPTW marginal structural models are a method of estimating unbiased causal effects, even in the presence of complex confounding (2). In order to produce valid inference, MSMs requires four assumptions be met: 1) no interference; 2) consistency; 3) positivity; and 4) conditional exchangeability.

1. The no interference assumption states that the effect of the exposure on the outcome is independent of the exposure of other participants (3). In this case, because the samples are large population samples, it is unlikely that there would be interference between participants of the study.
2. The consistency assumption assumes that the potential (possibly counterfactual) outcome associated with a given exposure, is the same as the actual outcome observed if that exposure occurs (4). This assumption can be violated if, for example, the exposure is defined ambiguously, but is unlikely in this study.
3. Positivity (5) (or near-positivity), assumes that there are no participants for whom the probability of exposure is zero (or very close to zero). In this case, given the age of the samples included and the fact that most people retire at ages covered by the study, makes positivity violations highly unlikely.
4. Last, conditional exchangeability, also commonly called *no unmeasured confounding* (6), assumes that all common causes of the exposure and outcomes can be eliminated via modelling. While it is impossible to be certain that no residual confounding exists, there are no major sources of confounding identified in the literature beyond those

controlled for in the study, making it unlikely there are major sources of unmeasured confounding.

To estimate the IPTW-MSM, we first estimated the propensity score, the probability of being exposed, in each of the three cohorts, using the set of confounders described in Appendix B1. Because the exposure variable is categorical, we estimated the propensity using multinomial logistic regression, fit using vector generalised linear models (VGLMs) via the ‘VGAM’ package in R (7). To allow for possible non-linear effects of continuous variables (eg age and mental health), we fit continuous variables using natural splines, varying the degrees of freedom between 1 and 5 (8), and selecting the version with the best model fit based on Bayesian information criteria (BIC) (9). We then used the propensity score to estimate a stabilised inverse probability weight, using the raw probability of exposure as the numerator:

$$W_A = \frac{P(A = a)}{P(A = a|L = l)}$$

where A is the exposure variable and L is the set of confounding variables.

We then assessed the balance in the covariates between groups. Plots showing the between group balance are included in **Figures S3.2-S3.12**.

We then estimated the probability of being lost to follow-up using a similar procedure, but using logistic regression:

$$W_C = \frac{P(C)}{P(C|L)}$$

Both the propensity and loss to follow-up models incorporated the survey weights for the survey, to account for the fact the samples are not simple random samples.

We then created a final analysis weight, as the product of the three weights (propensity, loss, and sampling):

$$W_{final} = W_A \times W_C \times W_{sampling}$$

We then created a new data set including the outcome variable (loneliness), exposure variable (retirement), the final analysis weight, the social support variable, and a categorical indicator variable labelling each dataset. We then estimated the effect of exposure on the outcome (loneliness) using a robust log-Poisson model to estimate relative risks. We then used these models to calculate marginal predicted probabilities (prevalence), with standard errors estimated using the delta method (8). To be certain that the exposure preceded the outcome, the outcome variable was drawn from the wave subsequent to exposure.

#### *Missing data*

Since the surveys were used for prevalence estimates, they were reweighted each wave, so that the current sample is representative, regardless of who is lost to follow-up. Since retention was accounted for by the sample weights, there was no need to impute or adjust for retention in the sample itself [10-12].

Second, we used complete case analysis instead of imputing the ‘exposure’ wave (i.e., the wave in which we took the ‘retirement’ variable from, in each of our paired waves). When conditioning on variables in a model, complete case analysis is unbiased, provided that missingness is related to those variables. That means that imputing baseline variables does not reduce bias, it only increases power [13]. However, the current sample sizes were large enough not to warrant that.

Third, bias can occur when there is missingness in the outcome. Therefore, wherever the ‘lagged’ outcome was not available, a censoring weight was created to adjust for the missing outcome [14].

The pooled samples and missing data can be found in **Tables S3.1-S3.2. Participants flow chart is in Figure S3.1.**

## References

1. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;550-60.
2. Clare PJ, Dobbins T, Mattick RP. Causal models adjusting for time-varying confounding – a systematic review of the literature. *International Journal of Epidemiology*. 2018;48(1):254-65.
3. VanderWeele TJ, Tchetgen Tchetgen EJ, Halloran ME. Interference and sensitivity analysis. *Statistical Science*. 2014;29(4):687-706.
4. Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Current Epidemiology Reports*. 2016;3(1):63-71.
5. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*. 2012;21(1):31-54.
6. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48.
7. Yee TW. *Vector generalized linear and additive models: with an implementation in R*: Springer; 2015.
8. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. *American journal of epidemiology*. 2011;174(8):984-92.
9. Neath AA, Cavanaugh JE. *The Bayesian information criterion: background, derivation, and applications*. Wiley Interdisciplinary Reviews: Computational Statistics. 2012;4(2):199-203.
10. Watson N. HILDA Project Technical Paper Series No. 2/12. Longitudinal and Cross-sectional Weighting Methodology for the HILDA Survey. The Melbourne Institute, Faculty of Business and Economics, The University of Melbourne, December 2012.
11. Lee S, Nishimura R, Burton P, et al. HRS 2016 Sampling Weights. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan, 2021.
12. Zhao Y, Hu Y, Smith JP, et al. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). *Int J Epi* 2014;43(1):61-68. doi: 10.1093/ije/dys203.

13. Hughes RA, Heron J, Sterne JAC, et al. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int J Epi* 2019;48(4):1294-304. doi: 10.1093/ije/dyz032
14. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* 2013;22(3):278-95. doi: 10.1177/0962280210395740 [published Online First: 2011/01/12]

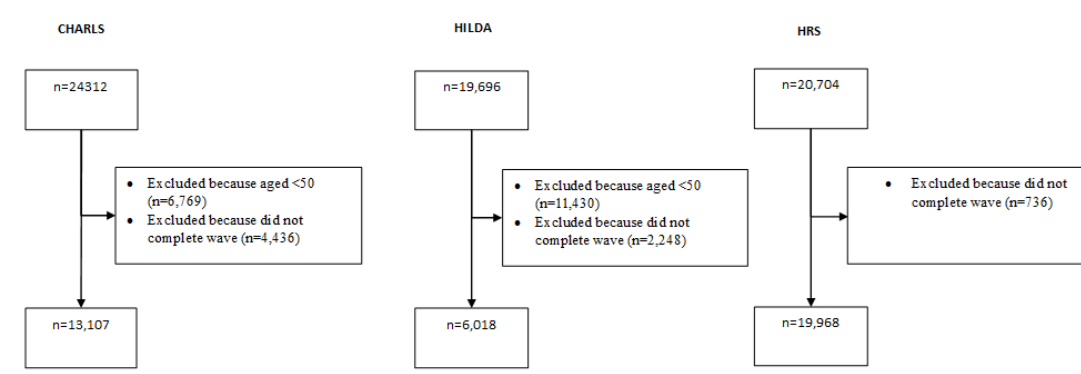
**Table S3.1. Summary of pooled samples and missing data according to wave**

Year	Australia (n = 6,018)		Unites States (n = 19,968)		China (n = 13,107)	
	Included in year	Missing lagged outcome	Included in year	Missing lagged outcome	Included in year	Missing lagged outcome
<b>2008</b>	3531 (58.7%)	119 (3.4%)	6280 (31.5%)	851 (13.6%)		
<b>2009</b>	3596 (59.8%)	137 (3.8%)				
<b>2010</b>	3817 (63.4%)	160 (4.2%)	7526 (37.7%)	609 (8.1%)		
<b>2011</b>	3831 (63.7%)	146 (3.8%)			8385 (64.0%)	1038 (12.4%)
<b>2012</b>	3915 (65.1%)	138 (3.5%)	6693 (33.5%)	712 (10.6%)		
<b>2013</b>	4018 (66.8%)	149 (3.7%)			6143 (46.9%)	855 (13.9%)
<b>2014</b>	4119 (68.4%)	129 (3.1%)	6714 (33.6%)	841 (12.5%)		
<b>2015</b>	4160 (69.1%)	153 (3.7%)				
<b>2016</b>	4244 (70.5%)	135 (3.2%)	5916 (29.6%)	900 (15.2%)		
<b>2017</b>	4234 (70.4%)	127 (3.0%)				

**Table S3.2. Summary of pooled samples according to retirement status and wave**

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
<b>Unites States</b>											
<b>Not retired</b>	2037		3336		2772		2603		2362		13110
<b>Fully retired</b>	2820		3310		3023		3143		2552		14848
<b>Never worked</b>	572		271		186		127		102		1258
<b>Total</b>	5429		6917		5981		5873		5016		29216
<b>Australia</b>											
<b>Not retired</b>	1653	1732	1834	1878	1894	1953	2048	2039	2059	2039	19129
<b>Fully retired</b>	1726	1703	1799	1761	1861	1891	1925	1936	2016	1950	18568
<b>Never worked</b>	33	24	24	46	22	25	17	32	34	30	287
<b>Total</b>	3412	3459	3657	3685	3777	3869	3990	4007	4109	4019	37984
<b>China</b>											
<b>Not retired</b>				4427		3444					7871
<b>Fully retired</b>				1087		300					1387
<b>Never worked</b>				1833		1544					3377
<b>Total</b>				7347		5288					12635

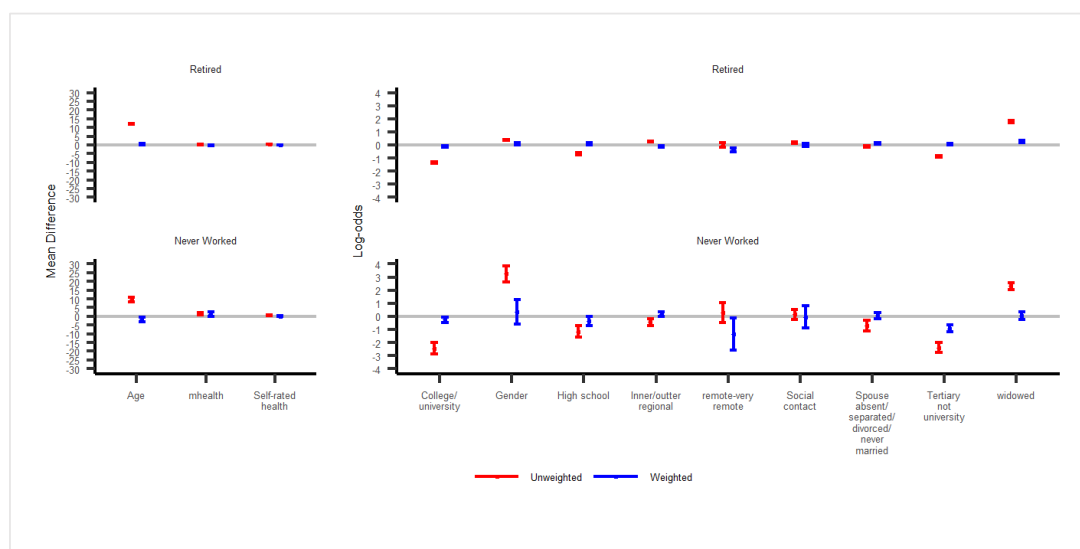


**Figure S3.1. Participants flow chart**

*Note.*

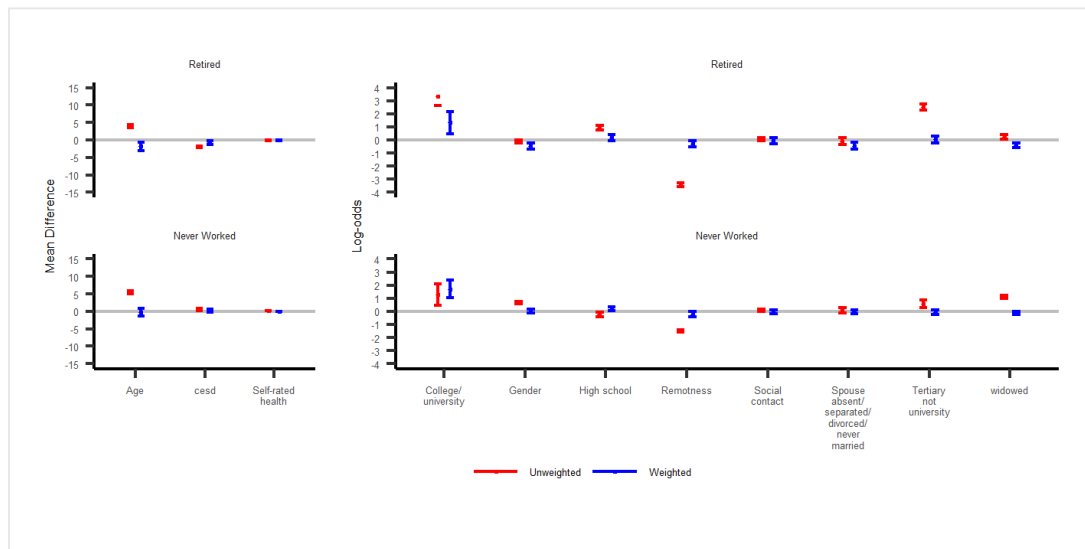
Number of incomplete cases in HRS is low due to higher number of missing/zero weights that were dropped.

**Figure S3.2. Household, Income and Labour Dynamics in Australia (HILDA) primary analysis balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**

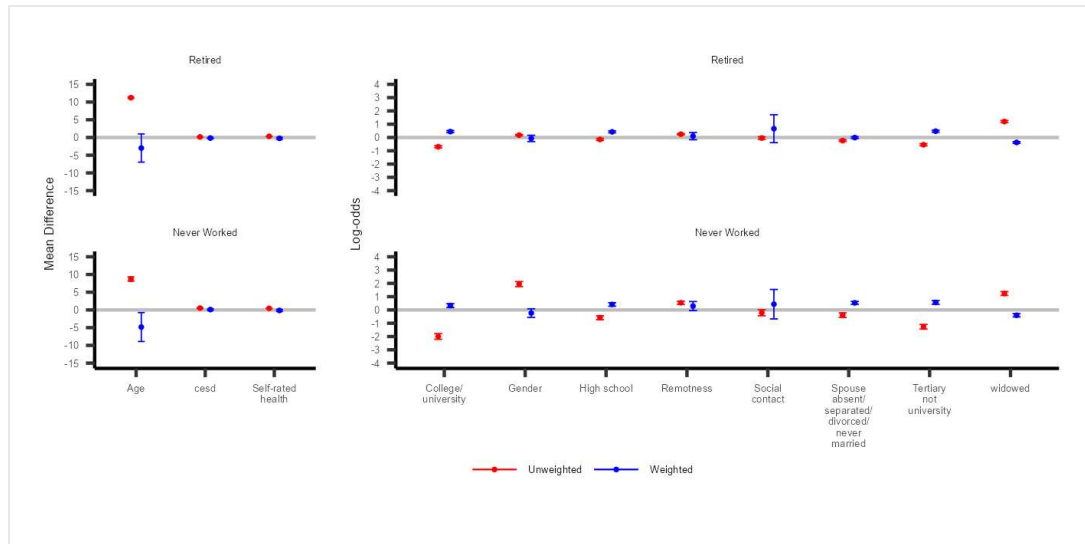


<sup>1</sup> Note: point estimates are the difference between exposed and unexposed groups. If the propensity model performs well the weighted differences should be smaller than the unweighted differences, and ideally close to zero.

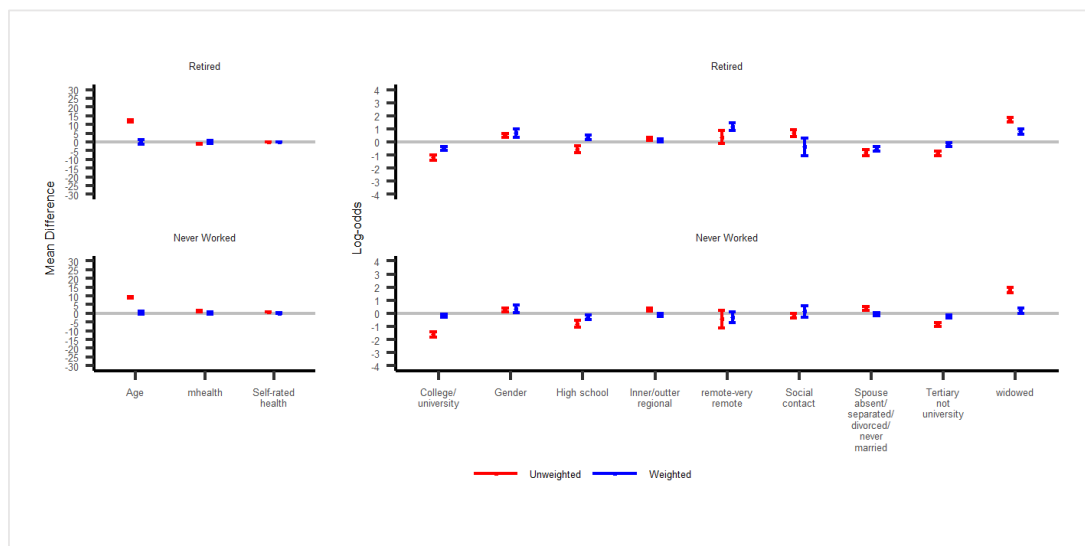
**Figure S3.3. China Health and Retirement Study (CHARLS) primary analysis balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



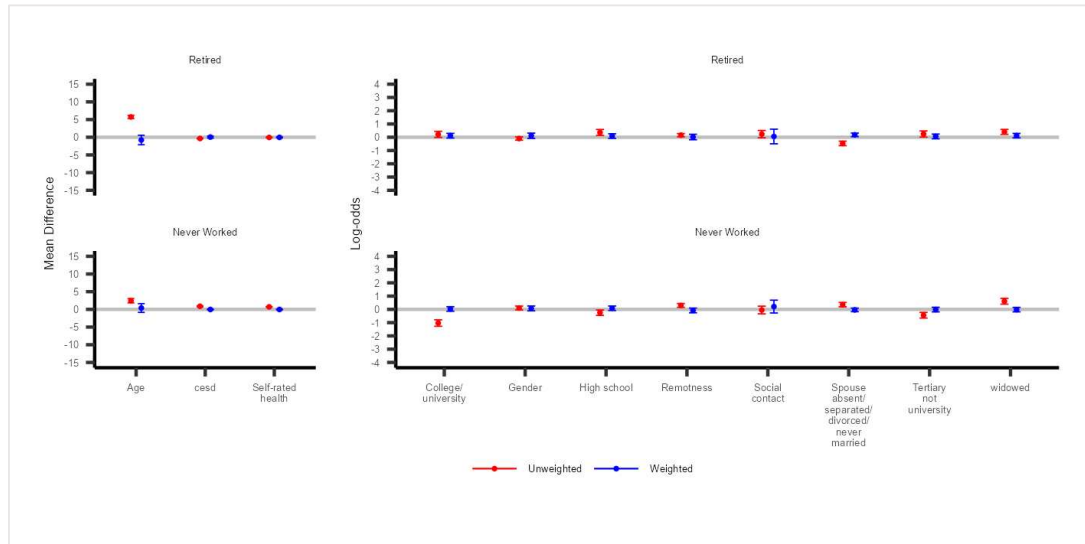
**Figure S3.4. United-States Health and Retirement Study (HRS) primary analysis balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



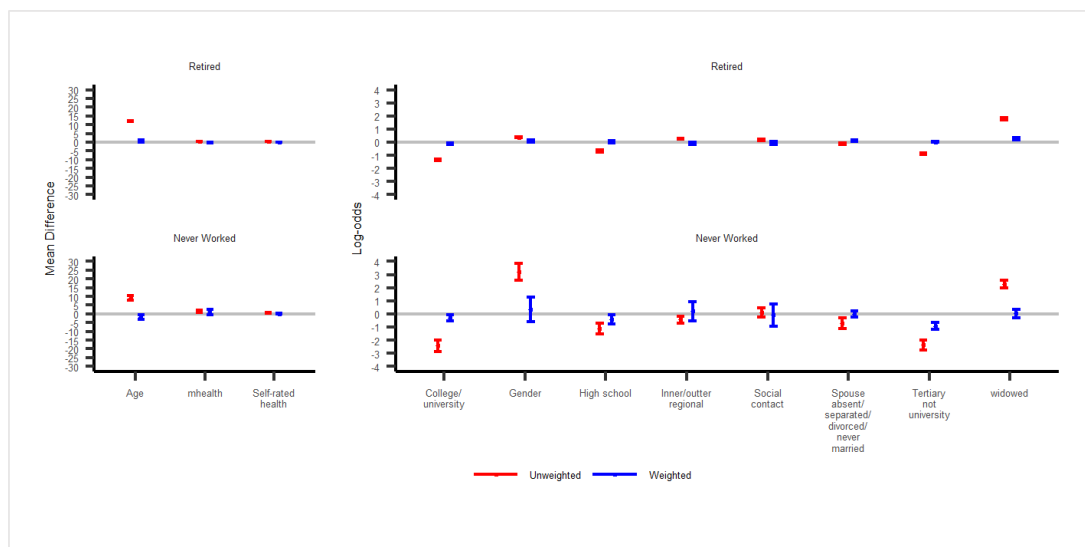
**Figure S3.5. Household, Income and Labour Dynamics in Australia (HILDA) nature of retirement balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



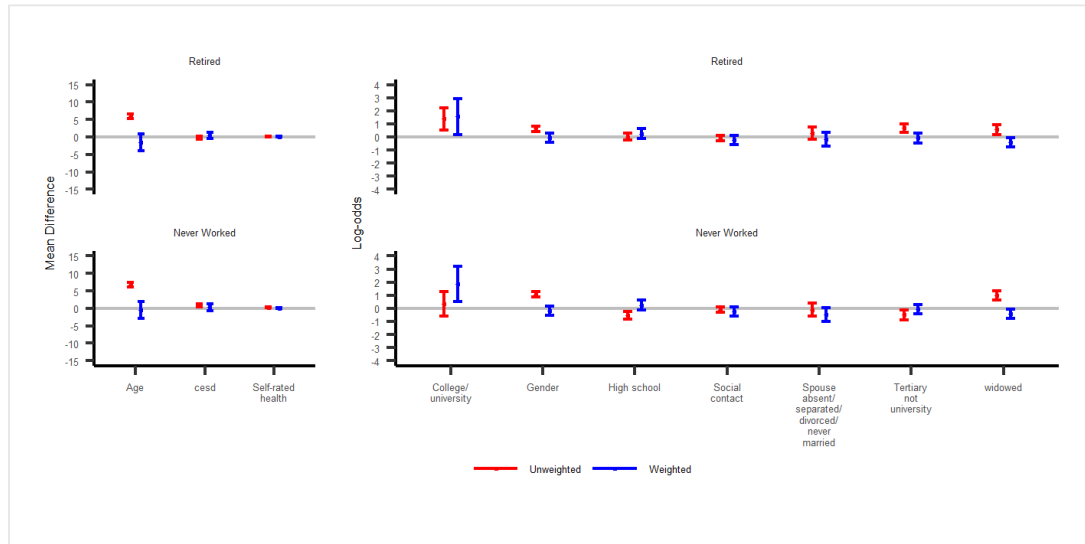
**Figure S3.6. United-States Health and Retirement Study (HRS) nature of retirement balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



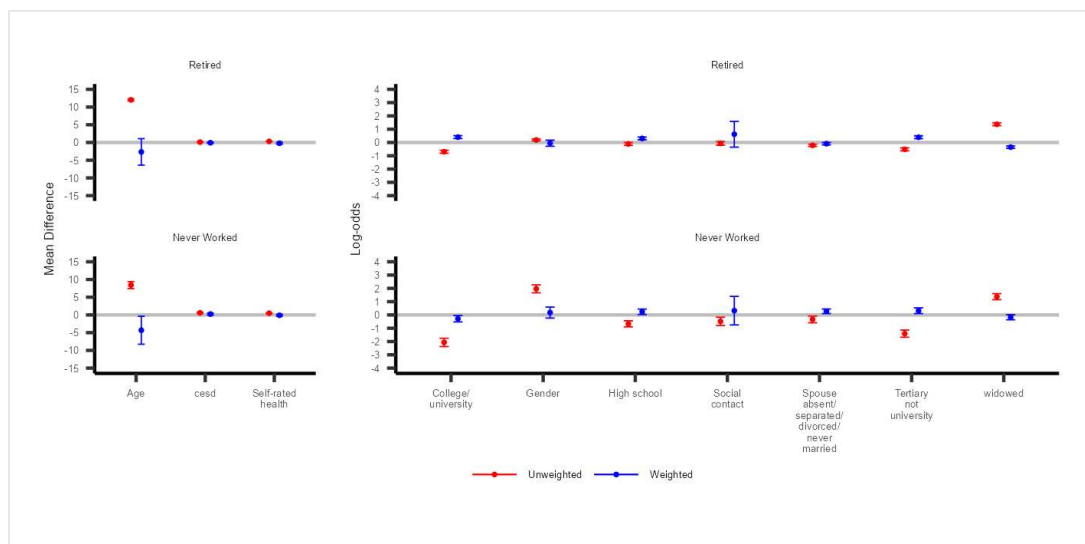
**Figure S3.7. Household, Income and Labour Dynamics in Australia (HILDA) urban sample balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



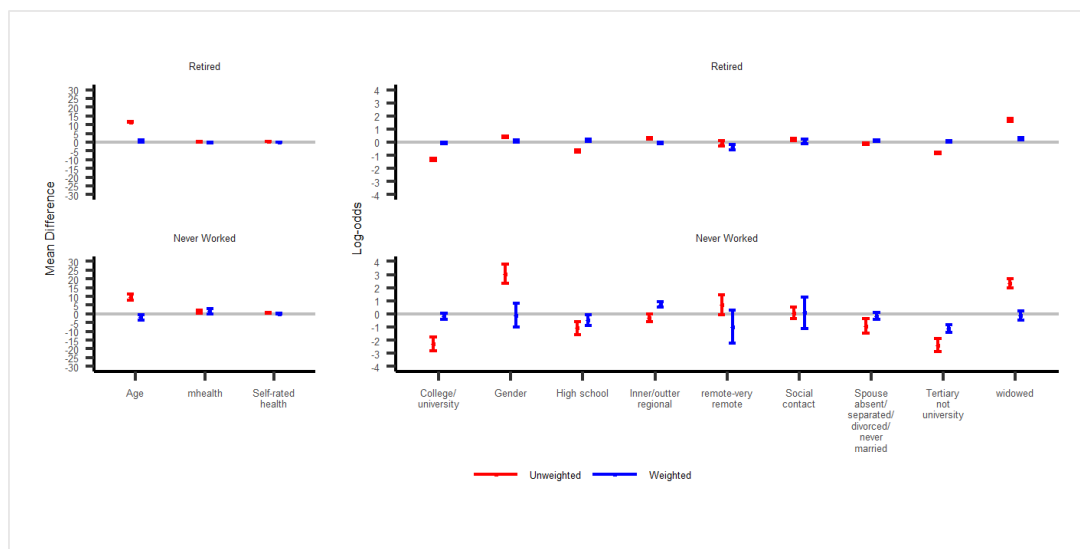
**Figure S3.8. China Health and Retirement Study (CHARLS) urban sample balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



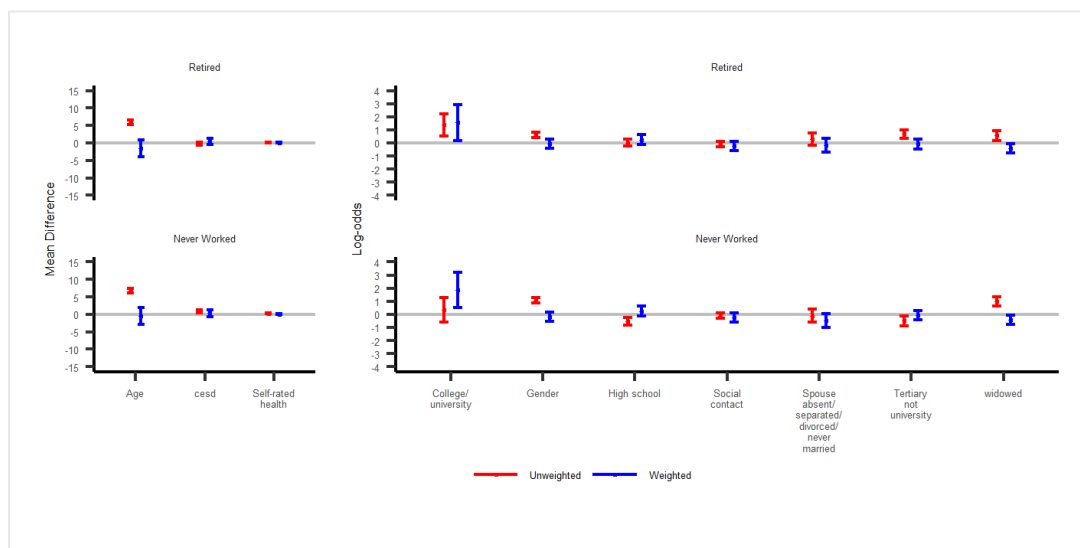
**Figure S3.9. United-States Health and Retirement Study (HRS) urban sample balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



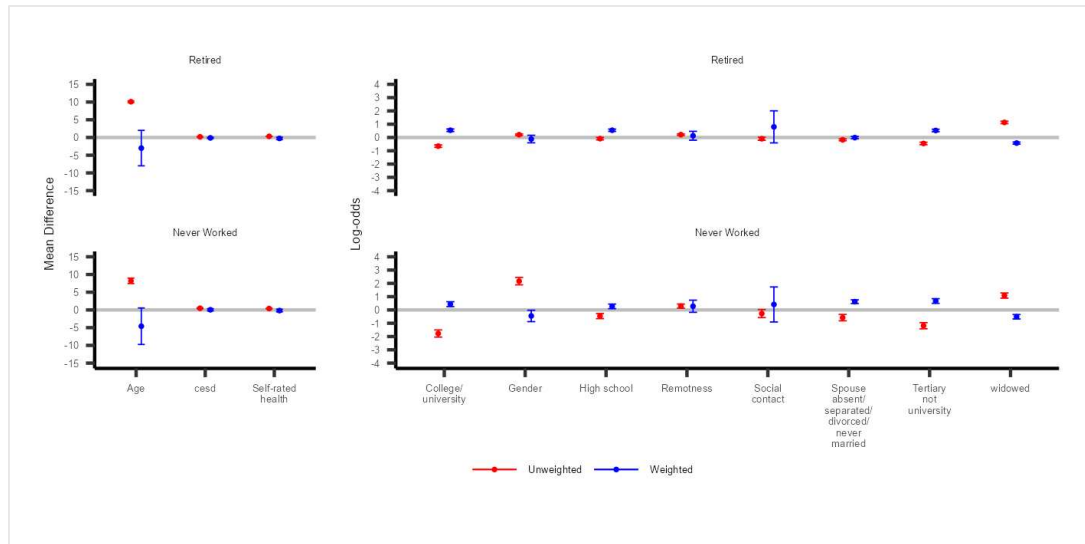
**Figure S3.10. Household, Income and Labour Dynamics in Australia (HILDA) fully lagged model balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



**Figure S3.11. China Health and Retirement Study (CHARLS) fully lagged model balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



**Figure S3.12. United-States Health and Retirement Study (HRS) fully lagged model balanced weights: mean differences of weighted outcomes compared to the unweighted outcomes (reference group = not retired) <sup>1</sup>**



**Appendix 4 – Additional results****Table S4.1. Lagged model of loneliness prevalence according to retirement status in Australia, China and US**

<b>Retirement status</b>	<b>Sample</b>	<b>Prevalence estimate</b>	<b>95% CI Lower</b>	<b>95% CI Upper</b>
Not retired	Australia	0.17	0.16	0.18
Not retired	China	0.17	0.16	0.19
Not retired	US	0.16	0.14	0.17
Retired	Australia	0.19	0.18	0.21
Retired	China	0.10	0.08	0.12
Retired	US	0.19	0.18	0.21
Never worked	Australia	0.21	0.14	0.32
Never worked	China	0.18	0.16	0.19
Never worked	US	0.22	0.19	0.27

**Abbreviations:** 95%CI: confidence interval.

**Table S4.2. Fully lagged model of loneliness estimated adjusted prevalence according to retirement status in Australia, China and US**

<b>Retirement status</b>	<b>Sample</b>	<b>Prevalence estimate</b>	<b>95% CI Lower</b>	<b>95% CI Upper</b>
Not retired	Australia	0.16	0.15	0.18
Not retired	China	0.22	0.20	0.24
Not retired	US	0.15	0.14	0.17
Retired	Australia	0.19	0.17	0.20
Retired	China	0.14	0.09	0.21
Retired	US	0.19	0.17	0.20
Never worked	Australia	0.16	0.10	0.25
Never worked	China	0.20	0.18	0.22
Never worked	US	0.20	0.17	0.24

**Abbreviations:** CI: confidence interval.



**Table S4.3. loneliness estimated adjusted prevalence according to nature of retirement among Australia and the US**

<b>Retirement status</b>	<b>Sample</b>	<b>Prevalence estimate</b>	<b>95% CI Lower</b>	<b>95% CI Upper</b>
Not retired	Australia	0.16	0.15	0.17
Not retired	US	0.14	0.13	0.15
Voluntary retired	Australia	0.10	0.07	0.12
Voluntary retired	US	0.11	0.09	0.13
Involuntary retired	Australia	0.17	0.14	0.21
Involuntary retired	US	0.18	0.15	0.21
Mixed reasons	Australia	0.12	0.09	0.15
Mixed reasons	US	0.11	0.06	0.19
Never worked	Australia	0.14	0.08	0.23
Never worked	US	0.17	0.14	0.20

**Abbreviations:** CI: confidence interval.

**Table S4.4. The modifying effect of social engagement on the association between the estimated adjusted prevalence of loneliness and retirement according to social engagement status among the three surveys: Australia, China and the US**

<b>Retirement status</b>	<b>Sample</b>	<b>Social engagement</b>	<b>Prevalence estimate</b>	<b>95% CI Lower</b>	<b>95% CI Upper</b>
Not retired	Australia	Not social	0.30	0.26	0.34
Not retired	Australia	Social	0.15	0.14	0.16
Not retired	China	Not social	0.17	0.15	0.18
Not retired	China	Social	0.18	0.16	0.20
Not retired	US	Not social	0.21	0.17	0.26
Not retired	US	Social	0.15	0.14	0.17
Retired	Australia	Not social	0.33	0.29	0.37
Retired	Australia	Social	0.18	0.16	0.19
Retired	China	Not social	0.09	0.07	0.12
Retired	China	Social	0.11	0.08	0.16
Retired	US	Not social	0.28	0.24	0.33
Retired	US	Social	0.18	0.17	0.20
Never worked	Australia	Not social	0.51	0.20	0.82
Never worked	Australia	Social	0.17	0.11	0.27
Never worked	China	Not social	0.17	0.15	0.20
Never worked	China	Social	0.18	0.15	0.20
Never worked	US	Not social	0.28	0.16	0.44
Never worked	US	Social	0.22	0.18	0.26

**Abbreviations:** CI: confidence interval.

**Table S4.5. Adjusted estimated adjusted prevalence of loneliness according to retirement status among urban participants in Australia, China and the US**

Retirement status	sample	Prevalence estimate	95% CI Lower	95% CI Upper
Not retired	Australia	0.16	0.15	0.17
Not retired	China	0.07	0.06	0.10
Not retired	US	0.14	0.12	0.16
Retired	Australia	0.18	0.17	0.20
Retired	China	0.09	0.07	0.11
Retired	US	0.18	0.16	0.20
Never worked	Australia	0.20	0.13	0.30
Never worked	China	0.10	0.08	0.12
Never worked	US	0.23	0.17	0.30

**Abbreviations:** CI: confidence interval.

**Figure S4.1. Adjusted estimated prevalence of loneliness according to retirement status among the three countries: Australia, China and the US (fully lagged model)**