Cross-sectional estimates of population health from the Survey of Health, Ageing and Retirement in Europe (SHARE) are biased due to health-related sample attrition

Magdalena Muszynska-Spielauer* Martin Spielauer**

*Institute of Applied Statistics, JKULinz; **Austrian Institute of Economic Research (WIFO)

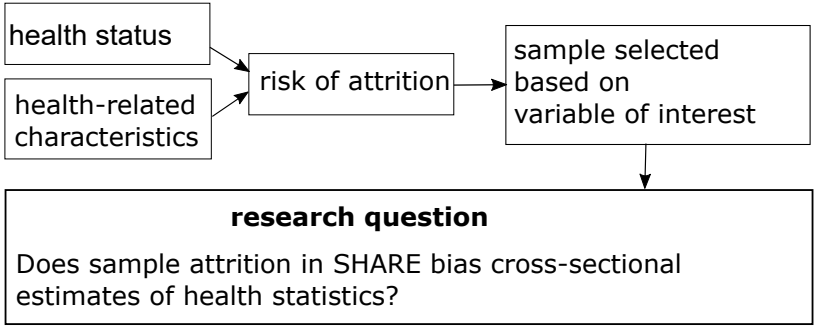
JKU Institute of Applied Statistics

WIFO AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH

longitudinal samples in SHARE $\rightarrow$ likely non-representative $\rightarrow$ biased estimates of population statistics from longitudinal and cross-sectional datasets

SHARE,Wave 7,Austria, 2017

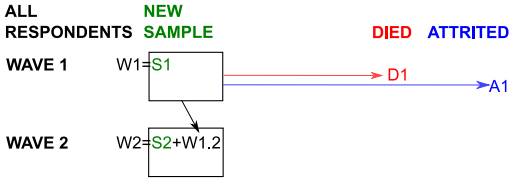
Samples Drawn	Longitudinal Sample			Attrition	Respondents in Wave 7	
Wave (Year)	All	Died	Attrited	Rate (in %)	Total	% All Samples
1 (2004)+4 (2011)	6221	587	2463	40	3219	51

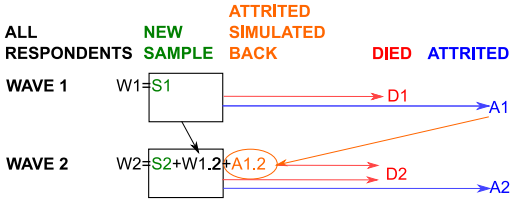

health status
health-related characteristics
risk of attrition
sample selected based on variable of interest
research question
Does sample attrition in SHARE bias cross-sectional estimates of health statistics?

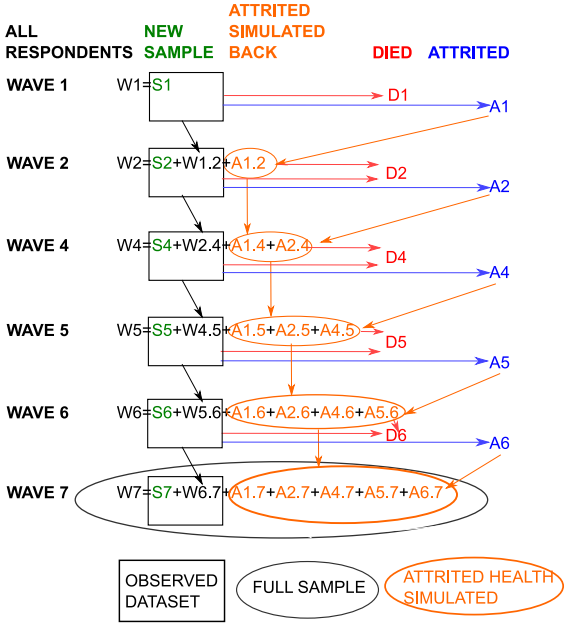
research question

Does sample attrition in SHARE bias cross-sectional estimates of health statistics?

1. Is there a pattern of health-related attrition?
2. If yes, can it be applied to replace missing data?
3. Replace missing data, compare health statistics in normal and full sample


ALL
RESPONDENTS
NEW
SAMPLE
DIED
ATTRITED
WAVE 1
W1=S1
D1
A1
WAVE 2
W2=S2+W1.2
WAVE 4
WAVE 5
WAVE 6
WAVE 7


ALL
RESPONDENTS
NEW
SAMPLE
ATTRITED
SIMULATED
BACK
DIED
ATTRITED
WAVE 1
W1=S1
D1
A1
WAVE 2
W2=S2+W1.2
A1.2
D2
A2
WAVE 4
WAVE 5
WAVE 6
WAVE 7


ALL RESPONDENTS
NEW SAMPLE
ATTRITED SIMULATED BACK
DIED
ATTRITED
WAVE 1
W1=S1
D1
A1
WAVE 2
W2=S2+W1.2+A1.2
D2
A2
WAVE 4
W4=S4+W2.4+A1.4+A2.4
D4
A4
WAVE 5
W5=S5+W4.5+A1.5+A2.5+A4.5
D5
A5
WAVE 6
W6=S6+W5.6+A1.6+A2.6+A4.6+A5.6
D6
A6
WAVE 7
W7=S7+W6.7+A1.7+A2.7+A4.7+A5.7+A6.7
OBSERVED DATASET
FULL SAMPLE
ATTRITED HEALTH SIMULATED

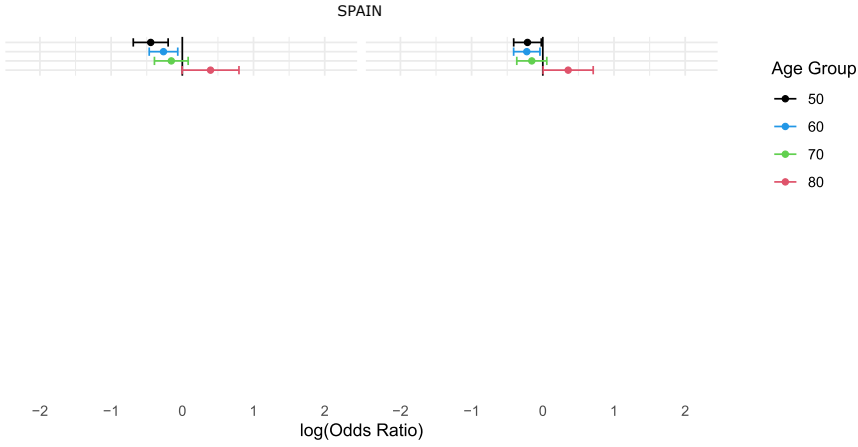
Country		Joined		Refreshment	Attrition
Code	Name	Wave	Year	Wave	Rate
HR	Croatia	6	2015	-	13
PL	Poland	2	2006-07	7	18
EE	Estonia	4	2010-11	-	23
LU	Luxemb.	5	2013	6	26
SI	Slovenia	4	2011	5,6	27
DK	Denmark	1	2004	5,6	28
ES	Spain	1	2004	2,4	33
PT	Portugal	4	2011	-	35
IT	Italy	1	2004	2-6	37
AT	Austria	1	2004	4	40
BE	Belgium	1	2004-06	2-6	40
HU	Hungary	4	2011	-	40
SE	Sweden	1	2004-05	2,5	40
FR	France	1	2004-05	2,4,6	41
CH	Switzerland	1	2004	2,4	41
CZ	Czech Rep.	2	2006-07	4,5	41
GR	Greece	1	2004-05	2,6	41
DE	Germany	1	2004	2,5	52

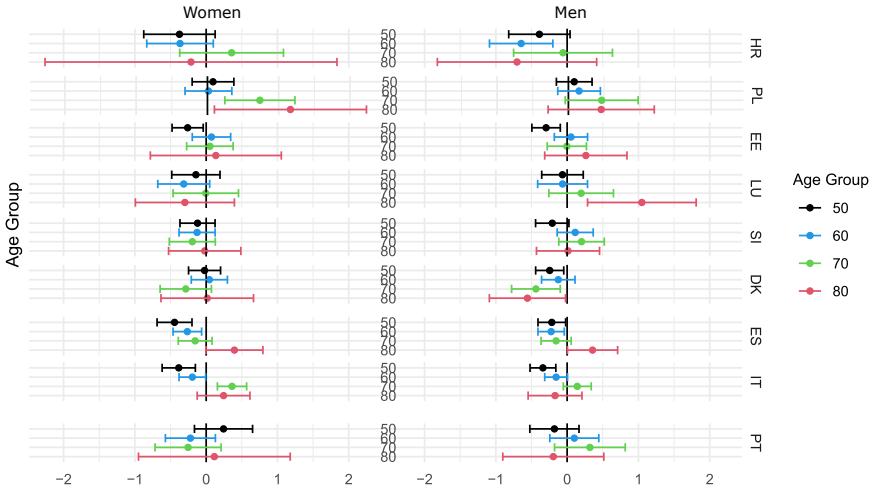
1.Pattern of health-related attrition

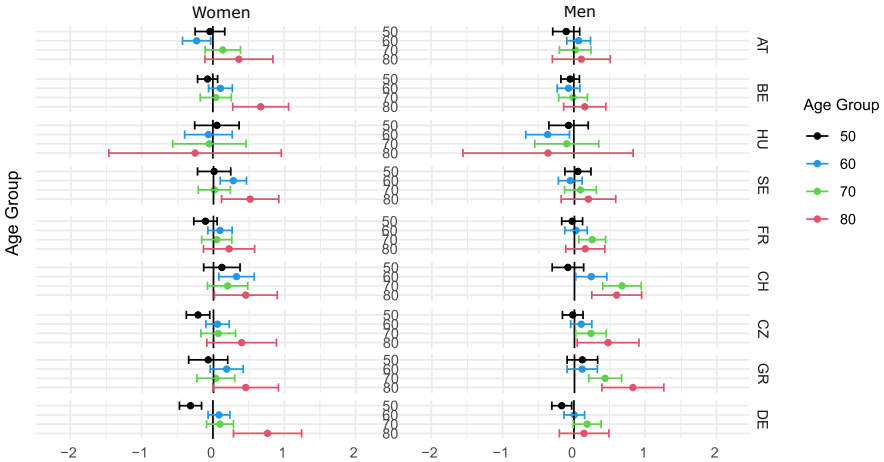
1. multinomial logistic regression - outcome status: attrited or dead vs. re-interviewed
2. explanatory variables: country, sex, 10-year age group, age group x health status at start wave
3. controls: educational attainment, marital status
4. model for pooled all pairs of waves: start wave + end wave

1. **Health dimension:** activity limitations measured with the Global Activity Limitation Indicator (GALI)
2. Interview Question: "For at least the past six months, to what extent have you been limited because of a health problem in activities people usually do? Would you say you have been ... ?"
3. Answers: "Not limited at all" = full health
   "Severely limited" + "Limited but not severely" = decreased health

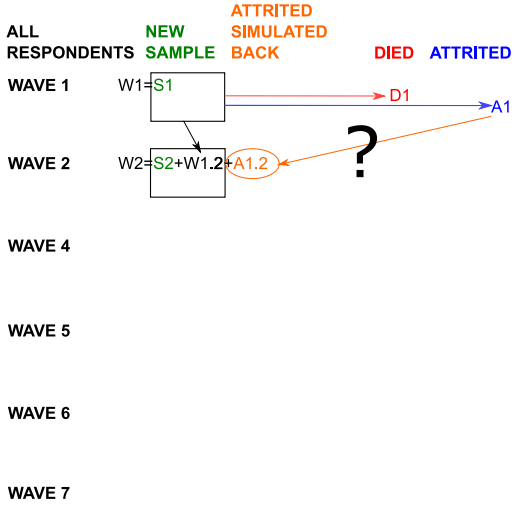
Log(odds) of attrition from the longitudinal samples vs. being re-interviewed due to decreased health




Women
Men
Age Group
50
60
70
80
HR
PL
EE
LU
SI
DK
ES
IT
PT
−2
−1
0
1
2


Women
Men
Age Group
50
60
70
80
AT
BE
HU
SE
FR
CH
CZ
GR
DE
−2
−1
0
1
2

2. Data-replacement strategy based on the observed attrition pattern

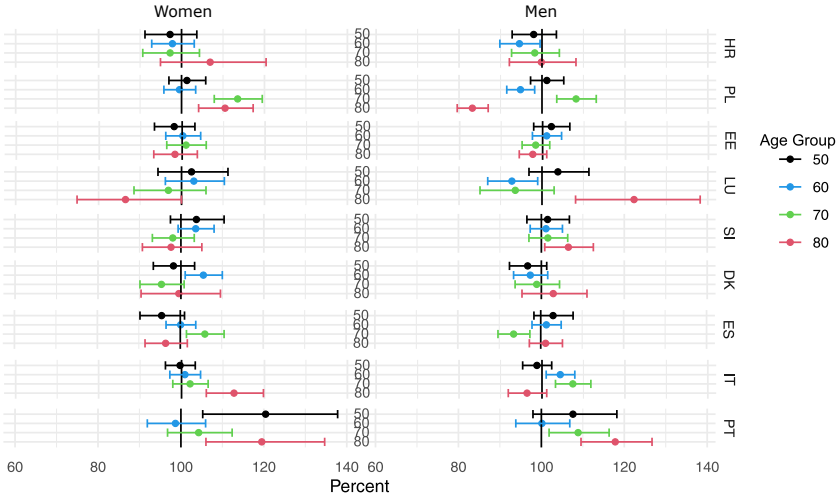


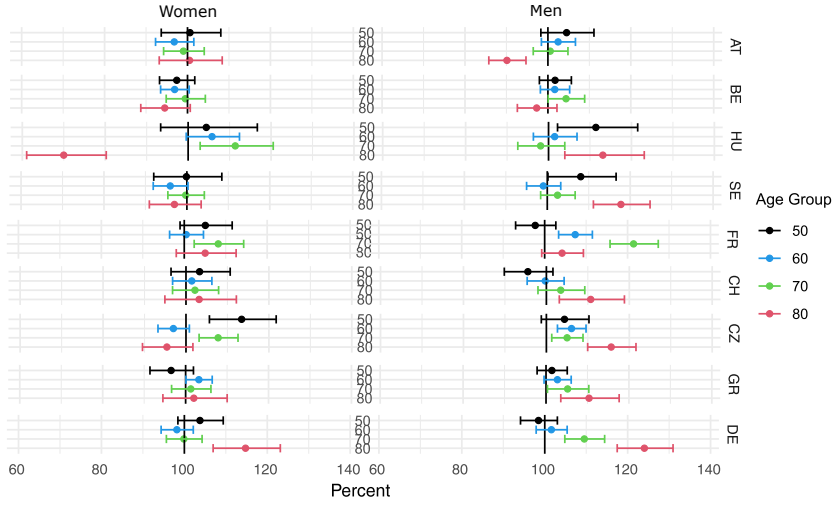
2. Data-replacement strategy based on the observed attrition pattern

- no clear-cut effect of health status on risk of attrition
- we make assumptions:
  1. attrition changes only the composition of the population by health status
     and hence
  2. transitions between health states of those who attrited would have been identical to
     those observed

1. **transitions probabilities** between health states with multinomial logistic model between each two waves.
2. **explanatory variables**: country, sex, health status, age, educational attainment, marital status
3. simulated **health status at Wave 7** according to last observed characteristics and transition probabilities, updated at each wave: age, health and marital status
4. **weights:** attrited: base weights at last observation, observed: base weights at Wave 7, raked to margin population by sex, age and NUTS1 in 2017 from Eurostat

Results: Prevalence of full health in data/ in full sample wave 7 (in%)




Women
Men
AT
BE
HU
SE
FR
CH
CZ
GR
DE
50
60
70
80
60
80
100
120
140
Percent
Age Group
50
60
70
80

Country		Healthy Life Years (HLY)					
		Women			Men		
Code	Name	Observed	Full	Difference	Observed	Full	Difference
HR	Croatia	12.7	13.0	-0.2	13.2	13.7	-0.4
PL	Poland	11.1	10.8	0.2	12.6	12.8	-0.2
EE	Estonia	10.9	11.1	-0.1	13.1	13.0	0.1
LU	Luxembourg	16.3	16.3	-0.0	16.5	16.6	-0.1
SI	Slovenia	14.5	14.3	0.3	16.2	15.9	0.3
DK	Denmark	18.2	18.2	0.0	18.6	18.9	-0.3
ES	Spain	19.3	19.7	-0.4	19.9	19.9	0.0
PT	Portugal	15.9	15.4	0.5	13.3	12.4	0.9
IT	Italy	20.5	20.1	0.4	20.0	19.6	0.4
AT	Austria	15.7	15.7	0.0	15.4	15.3	0.1
BE	Belgium	15.8	16.1	-0.3	16.1	15.9	0.3
HU	Hungary	14.3	14.0	0.4	16.1	15.6	0.5
SE	Sweden	18.9	19.4	-0.5	18.1	17.1	1.0**
FR	France	16.9	16.3	0.6	19.0	18.0	1.1***
CH	Switzerland	21.7	21.8	-0.1	21.7	21.6	0.1
CZ	Czech Rep.	13.4	12.2	1.2**	14.0	13.3	0.7*
GR	Greece	22.6	22.5	0.1	24.5	23.7	0.8**
GE	Germany	13.1	12.9	0.3	13.4	12.9	0.5

***p¡0.01, **p<0.05, *p<0.1

Small changes in rankings of countries based on HLY

	Women		Men	
Rank	Observed	Full	Observed	Full
1	Greece	Greece	Greece	Greece
2	Switzerland	Switzerland	Switzerland	Switzerland
3	Italy	Italy	Italy	Spain
4	Spain	Spain	Spain	Italy
5	Sweden	Sweden	France	Denmark
6	Denmark	Denmark	Denmark	France
7	France	Luxembourg	Sweden	Sweden
8	Luxembourg	France	Luxembourg	Luxembourg
9	Portugal	Belgium	Slovenia	Slovenia
10	Belgium	Austria	Belgium	Belgium
11	Austria	Portugal	Hungary	Hungary
12	Slovenia	Slovenia	Austria	Austria
13	Hungary	Hungary	Czech Rep.	Croatia
14	Czech Rep.	Croatia	Germany	Czech Rep.
15	Germany	Germany	Portugal	Estonia
16	Croatia	Czech Rep.	Croatia	Germany
17	Poland	Estonia	Estonia	Poland
18	Estonia	Poland	Poland	Portugal

1. at younger age healthy respondents are more likely to leave the longitudinal sample, at older age - those unhealthy
2. estimates of health statistics based on cross-sectional datasets are likely biased by health-related attrition
3. HLYs are more likely to be overestimated due to selective sample attrition, but it is not universal across the countries

SSM - Population Health 20 (2022) 101290

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Cross-sectional estimates of population health from the survey of health and retirement in Europe (SHARE) are biased due to health-related sample attrition☆

Magdalena Muszyńska-Spielauer [a,c,d,*], Martin Spielauer [b]

[a] Vienna Institute of Demography, Austrian Academy of Sciences, Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, UniVie), Austria
[b] Austrian Institute of Economic Research (WIFO), Austria
[c] Wittgenstein Centre for Demography and Global Human Capital (IIASA, OeAW, University of Vienna)
[d] Department of Public Finance and Infrastructure Policy, Institute of Spatial Planning, Vienna University of Technology, Vienna, Austria