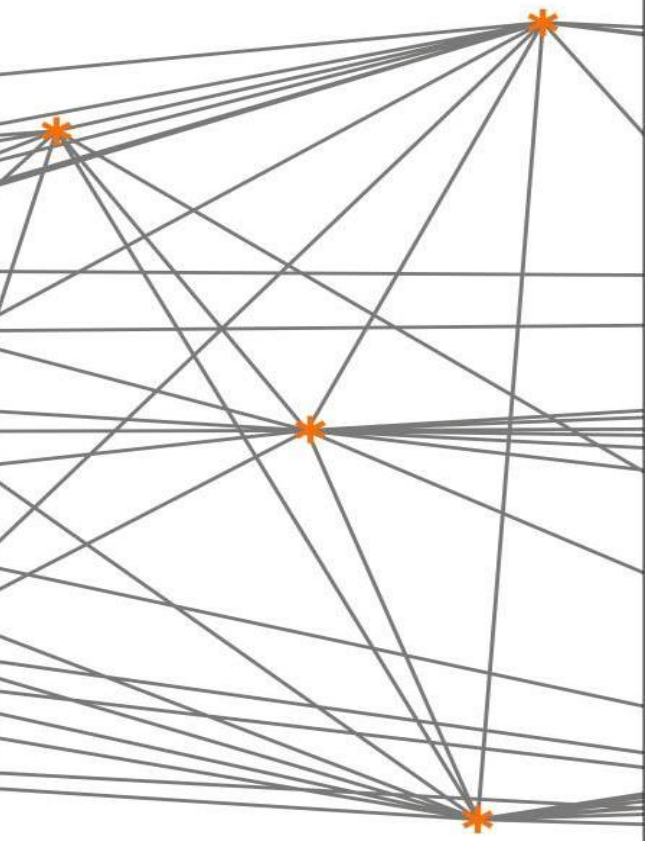
SHARE
Survey of Health, Ageing and Retirement in Europe



SHARELIFE database & usage

Giacomo Pasini

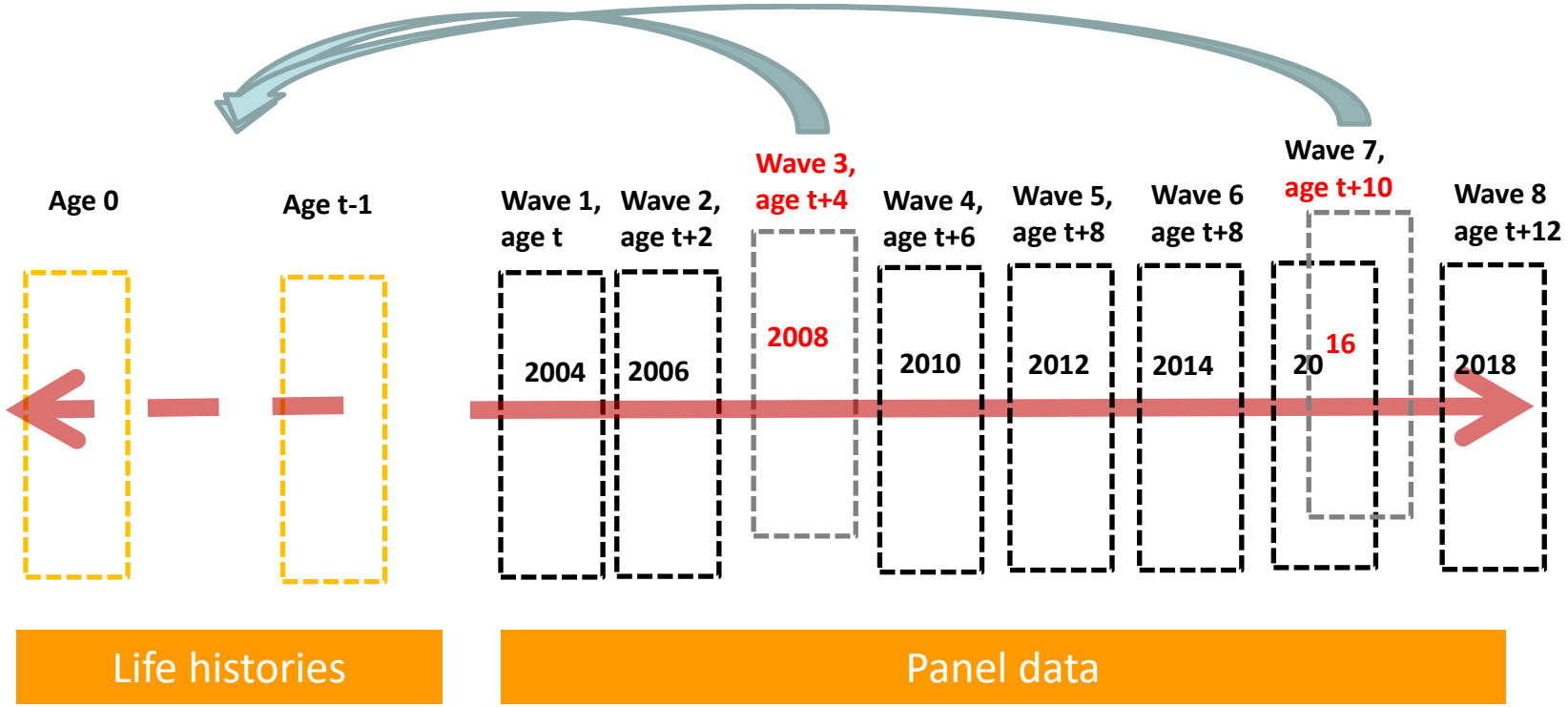
Ca' Foscari University of Venice

SHARE Users workshop

IER, Lubljana November 21st 2019

UNIVERSITÀ CA' FOSCARI VENEZIA

Data: SHARELIFE



- SHARE is a rich but complex source of information.
- Focus today: retrospective interviews in w3 and w7

2

SHARELIFE

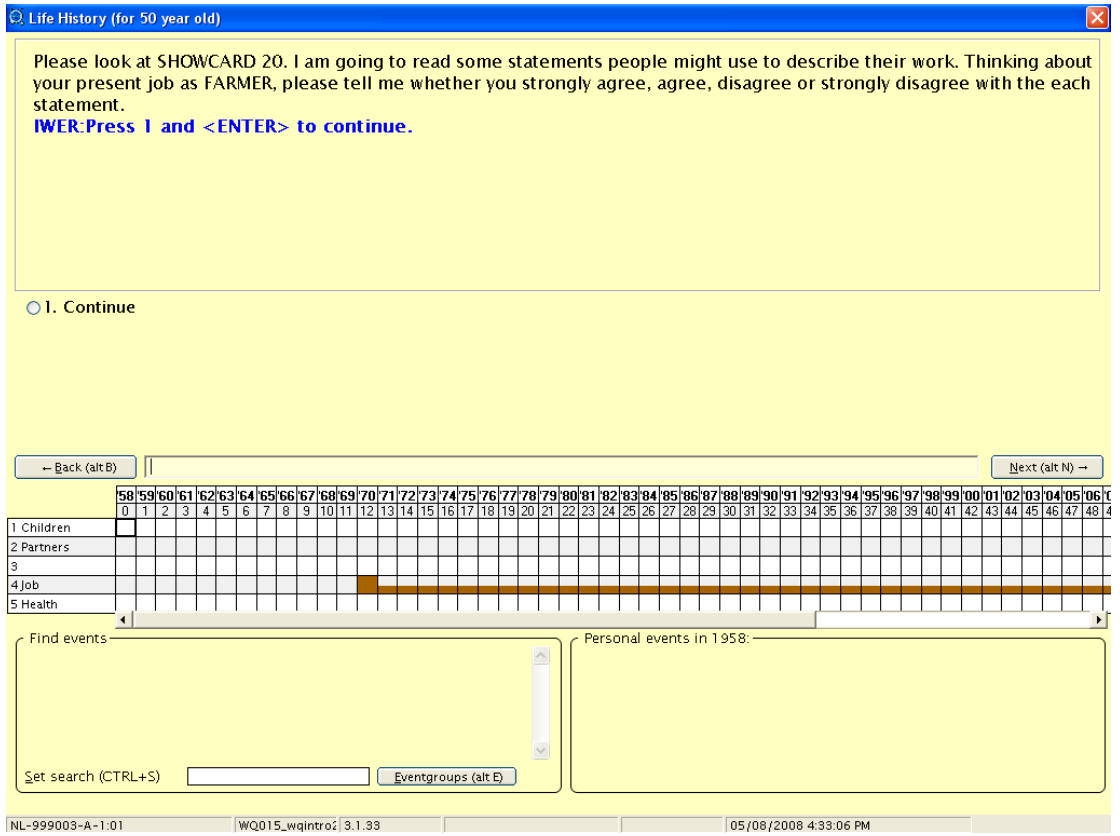
- The third wave of SHARE, named SHARELIFE, has been implemented to collect the retrospective histories of the SHARE respondents in order to obtain information about the respondents' lives before the baseline year of the survey (2004).
- The same questionnaire was offered in wave 7 to all countries that did not take part in w3 and to refreshers
  - The panel respondents in wave 7 were asked to answer the standard questionnaire

- The aim was to obtain data covering entire life cycles, overcoming the problems typical of long panels:
  - High cost
  - Selective attrition

ARE THE DATA GOOD ENOUGH?

A history of histories

- Selected recall questions in large general surveys,
  - e.g. earnings and other histories in BHPS/PSID etc.
- Detailed `autobiographical' life-histories in small scale studies,
  - e.g. age 65 year follow up of Boyd Orr cohort (Blane, 2005 IJE)
- Combination of two methods into CAPI life-grid calendars
  - Belli (1998, Memory), Belli, Shay and Stafford (2001, Public Opinion Quarterly)



A history of histories

- ELSA team consulted heavily with Blane and Belli, and created the ELSA Life History Interview as a supplement to ELSA wave 3
  - Combination of life-grid calendar with retrospective recall questions on early life circumstances for a full life-history questionnaire
- ELSA ⇢ SHARE ⇢ CHARLS ⇢ HRS
- Datasets then supplemented with contextual policy variables given the respondent's date of birth and age

Pros and cons

Retrospective interviews are good because:

- Create full life-history data on many cohorts
- The panel has no attrition by construction so is easy to work with
- Low cost

Worries over

- Selection:
  - the sample is representative of 50+ population at the time of interview (2004/2018)
  - We miss those that died early. SHARELIFE not representative e.g. of 20+ in 1974/1988
  - Is this relevant? Depend on the application
- Recall bias
  - Easy to remember date of birth of first child
  - Harder for e.g. firt wage in second job thirty years in the past
- Colouring:
  - How was your health when 10 years old?
  - If in bad health, Resp may “colour” his/her memory and report bad health back in the past

A history of histories

- As we said, most retrospective questions come from older studies
- Lifegrid reduce substantially recall bias, at least for questions asking “when” events took place
- A number of papers have argued good data validity
  - Havari and Mazzonna, 2016 Eur J Pop
  - Bingley and Martinello, 2014 WP
  - Trevisan, Pasini and Rainato 2011 WP
- Others have shown retrospective and prospective data from the same individuals can differ, but there is still value in the retrospective
  - Brown, 2014 Longitudinal and Life Course Studies
  - Reuben et al, 2016, Journal of Child Psychology and Psychiatry
  - Newbury et al, 2018, Journal of Psychiatric Research
- Ongoing work continues to look into this
- Life-history data not perfect but no data are. Better to ask whether the pros outweigh the cons

HOW TO HANDLE LIFE-CYCLE DATA

How to use SHARELIFE

1. Use **SHARELIFE** as the **third and seventh wave of a standard panel:**
   - SHARE respondents are observed seven times (in 2004/5, 2006/7, in 2008/9 with SHARELIFE, in 2010/11, in 2012/13, 2014 and again as a retrospective in 2016)

2. Use other waves of **SHARE** and **merge retrospective information** to study long term effects of life cycle events and policies

3. Build a **retrospective panel**:
   - Each respondent contributes as many observations as there are years of age from birth to the age at the moment of the interview.
   - Very long panel, but with rare events (e.g. we do not have income each year)

SHARELIFE as a «Standard wave»

- Some key information can be merged smoothly to other waves:
  - Grip strength

- In general, most questions are asked in a different but compatible way in wave 3 & 7
  - Employment status
  - Current health
  - Cronic health conditions

- Some of them are harmonized and readily available in EasySHARE
- Warning: EasySHARE is meant for study, not for research.

Example: labour market status

- Meschi; Padula, Pasini (2013) *Economic crisis and pathways to retirement* in FRB wave 4

- we look at the effect of the economic crisis on pathways to retirement looking at transitions out of labour force between waves 1 and 2 and 3 and 4.

- To do that, we need to build a detailed measure for labour market status from w3 variables, comparable with ep005 of standard waves

- Same for other controls

EP005

| **EP005**_ CURRENT JOB SITUATION
| Please look at card 20. In general, which of the following best describes your current employment situation?
| 1. Retired
| 2. Employed or self-employed (including working for family business)
| 3. Unemployed and looking for work
| 4. Permanently sick or disabled
| 5. Homemaker
| 97. Other (Rentier, Living off own property, Student, Doing voluntary work)

Labour market status in SHARELIFE

RE002 AGE FINISHED FULLTIME EDUCATION
In which year did you finish continuous full-time education at school or college?

RE003/007 SITUATION AT AGE 15 IF NO EDUCATION
| Please look at SHOWCARD 12. Which of these best describes the situation you were in at age 15?/straight after finishing full time education

SHOWCARD 12

| 1. Employee or self-employed
| 2. Unemployed and searching for a job
| 3. Unemployed but not searching for a job
| 4. Short term job (less than 6 months)
| 5. Sick or disabled
| 6. Looking after home or family
| 7. Leisure, travelling or doing nothing
| 8. Retired from work
| 9. Training
| 10. Further full time education
| 11. Military services, war prisoner or equivalent
| 12. Managing your assets
| 13. Voluntary or community work
| 14. Forced labour or in jail
| 15. Exiled or banished
| 16. Labor camp
| 17. Concentration camp
| 97. Other

RE008 DID SITUATION EVER CHANGE
| | Has your situation ever changed since you were <<from re007>>
| | | **RE009** YEAR OF CHANGE OF SITUATION
| | | In which year did your situation change?
| | | **RE010** SITUATION CHANGED TO
| | | Please look at SHOWCARD 12. Which of these best describes the situation you changed to?

...in practice

- The way data are organized in SHARELIFE is not obvious
- SHARELIFE is released as an individual-level dataset organizing sequences of life events in a flat file format

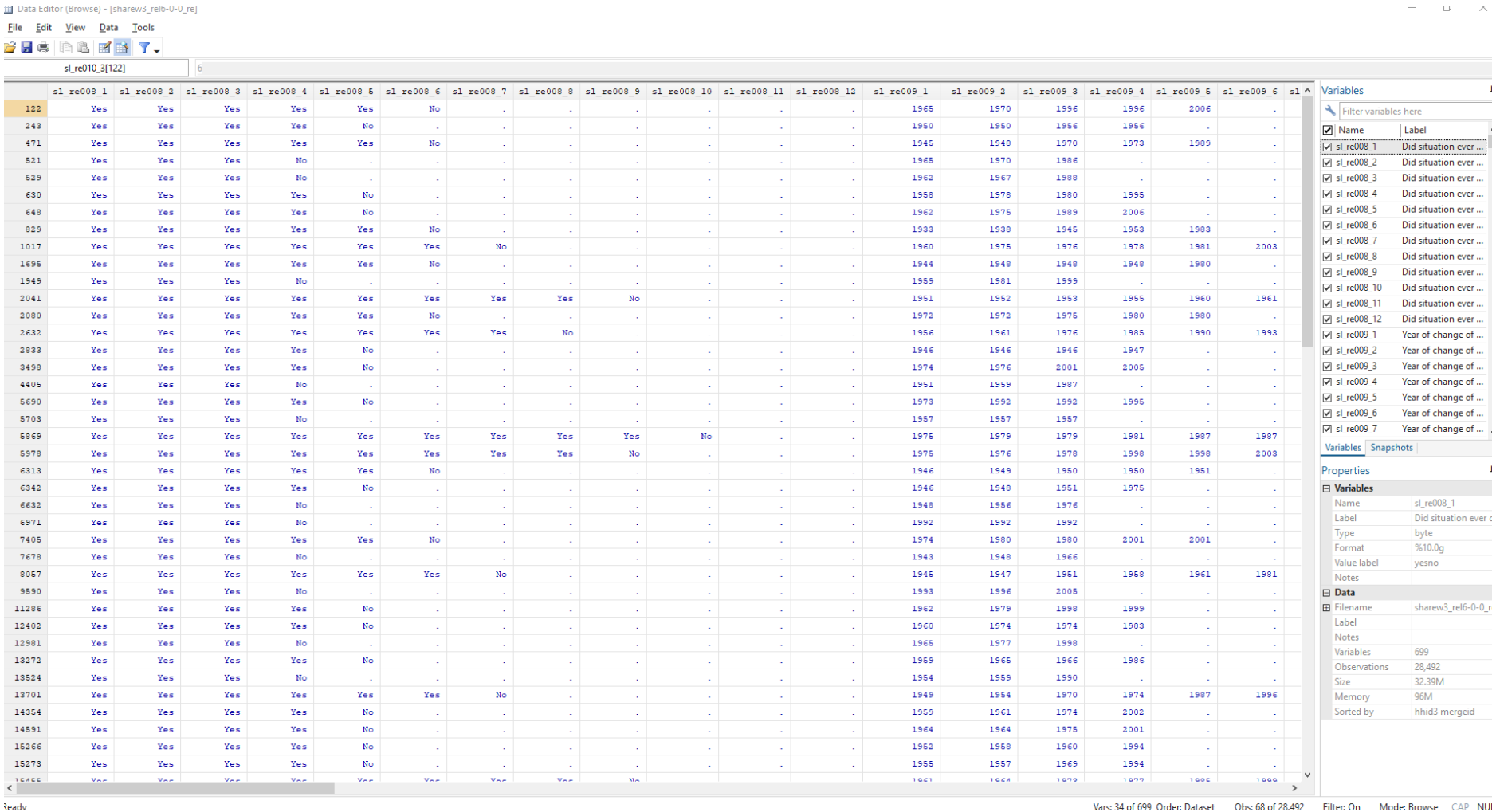
- Example: labour market status is looped over self reported spells and the information of Showcard 12 is stored as a set of numbered variables for each individual in the sample

..In practice

- If an individual had 4 labour market status spells:
  - sl_re008_1 to sl_re008_3 are equal to 1=Yes
  - sl_re008_4 is 5=No
  - sl_re008_5 onwards is set to missing
  - sl_re009_1 to sl_re009_3 report the year the spell started, sl_re009_4 onwards is missing
  - sl_re010_1 to sl_re010_3 report the status in the spell, sl_re010_4 onwards is missing

- In order to merge to ep005:
  - By looking at sl_re009 _# select the «current» spell
  - The corresponding sl_re010_# can be collapsed into ep005 categories

sl_re008 & sl_re009



STATA code

Economic crisis and pathways to retirement

Table X.2: ***Transition Matrix, Self Reported Labour Market Status in 2004 and 2006***

2004-06	LABOUR MARKET STATUS IN 2006			
LABOUR MARKET STATUS IN 2004	**Retired**	**In labour force**	**Disabled**	**Total**
Retired	2,366 (100)	0 *(0)*	0 *(0)*	2,366 *(100)*
In labour force	737 (*27.83*)	**1,831 (*69.15*)**	80 (*3.02*)	2,648 *(100)*
Disabled	127 (*41.64*)	27 (*8.85*)	151 (*49.51*)	305 (*100*)
Total	3,230 *(60.73)*	1,858 *(34.93)*	231 *(4.34)*	5,319 *(100)*

Table X.3: ***Transition Matrix, Self Reported Labour Market Status in 2008 and 2011***

2008-11	LABOUR MARKET STATUS IN 2011			
LABOUR MARKET STATUS IN 2008	**Retired**	**In labour force**	**Disabled**	**Total**
Retired	1,812 (100)	0 *(0)*	0 *(0)*	1,812 *(100)*
In labour force	653 (*26.21*)	**1,786 (*71.7*)**	52 (*2.09*)	2,491 *(100)*
Disabled	57 (*30.48*)	5 (*2.67*)	125 (*66.84*)	187 (*100*)
Total	2,522 *(56.17)*	1,791 *(39.89)*	177 *(3.94)*	4,490 *(100)*

- The fraction of workers who are still employed in the next wave is higher in 2011 than it was in 2006.
- This despite the longer time interval between waves!
- Is it a direct effect of the crisis or an indirect effect (via higher eligibility ages in some countries?)

easySHARE

• Some key variables from wave 3 & 7 are harmonized and readily available in easySHARE

•easySHARE includes the same number of observations as the main release of SHARE but it is restricted to a subset of variables. It contains the regular panel waves of SHARE (wave 1, 2, 4, 5, 6) and some information collected in the retrospective interviews of wave 3 and 7.

• It also incorporates variables from the “generated variables datasets” and create additional “easy to use”-variables.

20

Advantages of easySHARE

Reduce the complexity of data handling and data preparation

- One single data file without the need for complex merging of waves and modules
- Cross-wave comparability: the majority of those variables and indices included in the dataset have been collected in all waves of SHARE
- Inclusion of new variables and indices that are helpful for users but have not been made available in the SHARE data (such as age, wave specific couple identifier, health indices, etc).
- Code is provided (R, Stata and SPSS). Can be used as starting point to merge other SHARELIFE information

Data structure

mergeid	wave	birth_country	wavepart
AT-986403-01	1	Austria	1234567
AT-986403-01	2	Austria	1234567
AT-986403-01	3	Austria	1234567
AT-986403-01	4	Austria	1234567
AT-986403-01	5	Austria	1234567
AT-986403-01	6	Austria	1234567
AT-986403-01	7	Austria	1234567
SE-209636-01	1	Sweden	145
SE-209636-01	4	Sweden	145
SE-209636-01	5	Sweden	145

22

Disadvantages of easySHARE

EasySHARE is meant to have a quick access to data and data analysis. But

- In standard waves there is much more information, variables, etc
- The way some generated variables are done is non neutral and involve a number of assumptions, the researcher may want to have control on them

Please note

*easy*SHARE is intended for student training and teaching purposes. For scientific publications we recommend using the main data set of SHARE, or to carefully study this documentation and the Stata program that extracts and generates *easy*SHARE from the main release of SHARE.

merging retrospective data to Standard SHARE waves

- Many ‘HRS family’ papers on long term effect of early-life or early-adulthood circumstances and exposures
  - Health
  - SES
  - Unemployment
  - Single motherhood
  - War, recessions, …
- …on all kinds of late-life outcomes, e.g.
  - Lifetime earnings
  - Physical health
  - Mental health
  - Muscle strength
  - Physical activity
  - Chewing ability…

Early life circumstances and late-life outcomes

- The best of these exploit exogenous or quasi-exogenous variation in early life exposures
  - Kesternich, Siflinger, Smith and Winter, REStat 2014, 'The effects of World War II on economic and health outcomes across Europe'
  - Kesternich, Siflinger, Smith and Winter, EJ Features 2015, 'Individual behavior as a pathway between early-life shocks and adult health: Evidence from hunger episodes in post-war Germany'
  - van den Berg, Pinger and Schoch, EJ 2016, 'Instrumental variable estimation of the causal effect of hunger early in life on health later in life'
  - Bharadwaj, Graff Zavin, Mullins, Neidell, Am J Resp Critical Care Med, 2016, Early-life exposure to the Great Smog of 1952 and the development of Asthma
  - Antonova, Bucher-Konen, Mazzonna, Soc Sci Med 2015 'Long-term health consequences of recessions during working years'
  - Avendano, Brugiavini, Berkman, Pasini, Soc Sci Med, 2015, 'The long-run effect of maternity leave benefits on mental health: Evidence from European countries'
  - Schaan, Soc Sci Med, 2014 'The interaction of family background and personal education on depressive symptoms in later life.'

Example 1: maternity and late life depression

- Avendano, Berkman, Brugiavini, Pasini (2015): examine whether **paid maternity leave** policies have long-lasting effects on mother's late-life mental health
  - From SHARE waves 1 and 2: outcome variable (euro-D depression score) and covariates
  - From SHARELIFE: Year of first maternity, country of residence at childbirth, labour market status at childbirth, exact length of interruption
  - From external sources: policy data on maternity leave matched by year & country of residence at time of childbirth

Depression score by Full Wage Weeks of Maternity leave

	Full-wage week of maternity leave benefits		Difference high-low	% change
	Low	High		
working	2,64	2,51	-0,13	
not working	2,52	2,82	0,30	
		DiD	**-0,43**	**-16,17%**

Interpretation: Moving from a maternity leave with limited coverage to one with comprehensive coverage around the birth of a first child reduces late life depression scores by 16%

Example 2: Childhood conditions

- Increasing interest on early life conditions effects

- SHARELIFE extremely useful in this respect:
  - Individual-specific conditions early in life
  - Contextual bad conditions (hunger, bad economic conditions)
  - Insititutional variables can be merged by country & year

- Examples:
  - Kesternich et al (2014, REStat) effect of WW2 on later health and economic conditions
  - Schaan (2014, SS&M) looks at effect of childhood conditions on mental health

SHARELIFE – Childhood health

Information about Childhood Health in SHARELIFE:

- SRHS (five categories: excellent, very good, good, fair, poor);
- whether the respondent has missed school because of health problems
- whether or not individuals experienced any of the following diseases from birth until age 15:
  - infectious diseases, Polio, Asthma or other respiratory problems, allergies, Severe diarrhea, Meningitis, Chronic ear problems, speech impairment, difficulty seeing, Severe epilepsy/seizures, emotional, nervous, or psychiatric problems, fractures, appendicitis, diabetes, heart problems, leukemia, cancer or other not listed

SHARELIFE – Childhood SES

Childhood SES: it contains information on living conditions and family characteristics when respondents were 10 years old, for instance:

- number of rooms and number of people in the house;
- features of accomodation (fixed bath, cold and hot running water supply, inside toilet and central heating);
- number of books at home (from none to 2 or more bookcases);
- occupation of the main breadwinner (10 categories);
- assessments of respondents' relative position in Math and Language at age 10 (with respect to their class-mates)

Exploiting life cycles

- Probably the most promising feature of SHARELIFE is to study life cycle **trajectories**, their origins and consequences.

- A standard panel is likely to be too short to observe a sufficient longitudinal variation at individual level in low frequency/slowly changing phenomena like e.g. marital status or accomodation

- SHARELIFE contains detailed information on
  - fertility histories,
  - relationship and family composition,
  - accomodation
  - labour market histories

SHARELIFE variables (1)

• Children history (RC)
Own children, maternity leave, etc.

• Partnership history (RP)
Marriages, cohabiting partners, other important partners

• Accommodation history (AC)
Moves, locations, surroundings, buying/selling

• Childhood (CS)
Number of people, number of rooms, accommodation features, number of books, school performance

• Job history (RE)
Characteristics of the job, first salary, reason for leaving
- Includes also Work quality and Disability sub-modules

SHARELIFE variables (2)

- Financial history (FS)

First activities in financial market if any

- Health history

Childhood health and diseases, start/end menstrual period, insurgence cronic/severe health conditions, support during bad health

- Health care

Vaccination, when no regular care, dentist care, gynaecological care, blood tests, mammograms.
Healthy behavior

- General Life (GL)

Happiness, stress, poor health, hunger, financial hardship, persecuted

- Grip Strength (GS)

How to handle life cycle information

- As we already said, SHARELIFE is released as an individual-level dataset organizing sequences of life events in a flat file format
- Example: all job characteristics (type of job, wage, FT/PT, etc) are looped over all job episodes and the information is stored as sets of variables for each individual in the sample

- **Retrospective panel**: reorganize the data in such a way each respondent contributes as many observations as there are years of age from birth to the age at which they are observed at the moment of the interview.

SHARELIFE – Job History (1)

For each job spell (up to 20 jobs) we know:

- When job episode start (year)
- When job episode end (year)
- Job characteristics:
  - Title of job and job description
  - Industry
  - Type of job (employee, civil servant or self-employed)
- First monthly net wage or income
- If job was full or part time
  - Reason for working part-time
- Change between full to part-time or from part-time to full time:
  - When change happen
  - Why change happen (only from full to part time)

SHARELIFE – Job History (2)

In addition, we know:

- Age finished full time education
- Individual status between job spells (e.g. unemployment)
- Information on retirement
  - Last wage/labour income
  - First benefit received

- assessment of the psychosocial work environment of the last main job of the working career - lasting longer than 5 years- (work is uncomfortable, physically or emotionally demanding, involves conflict, health has suffered at work, etc.)

The retrospective panel

Job histories have been re-organized into a retrospective job episodes panel (JEP) and are available to users

- Documentation available in three SHARE working papers:
  - Brugiavini, Cavapozzi, Pasini, Trevisan (2013)
  - Antonova, Aranda, Pasini, Trevisan (2014)
  - Brugiavini, Orso, Genie, Naci, Pasini (2019)

- The starting point are the 28,492+62,561 individuals interviewed in w3 and in w7 retrospective

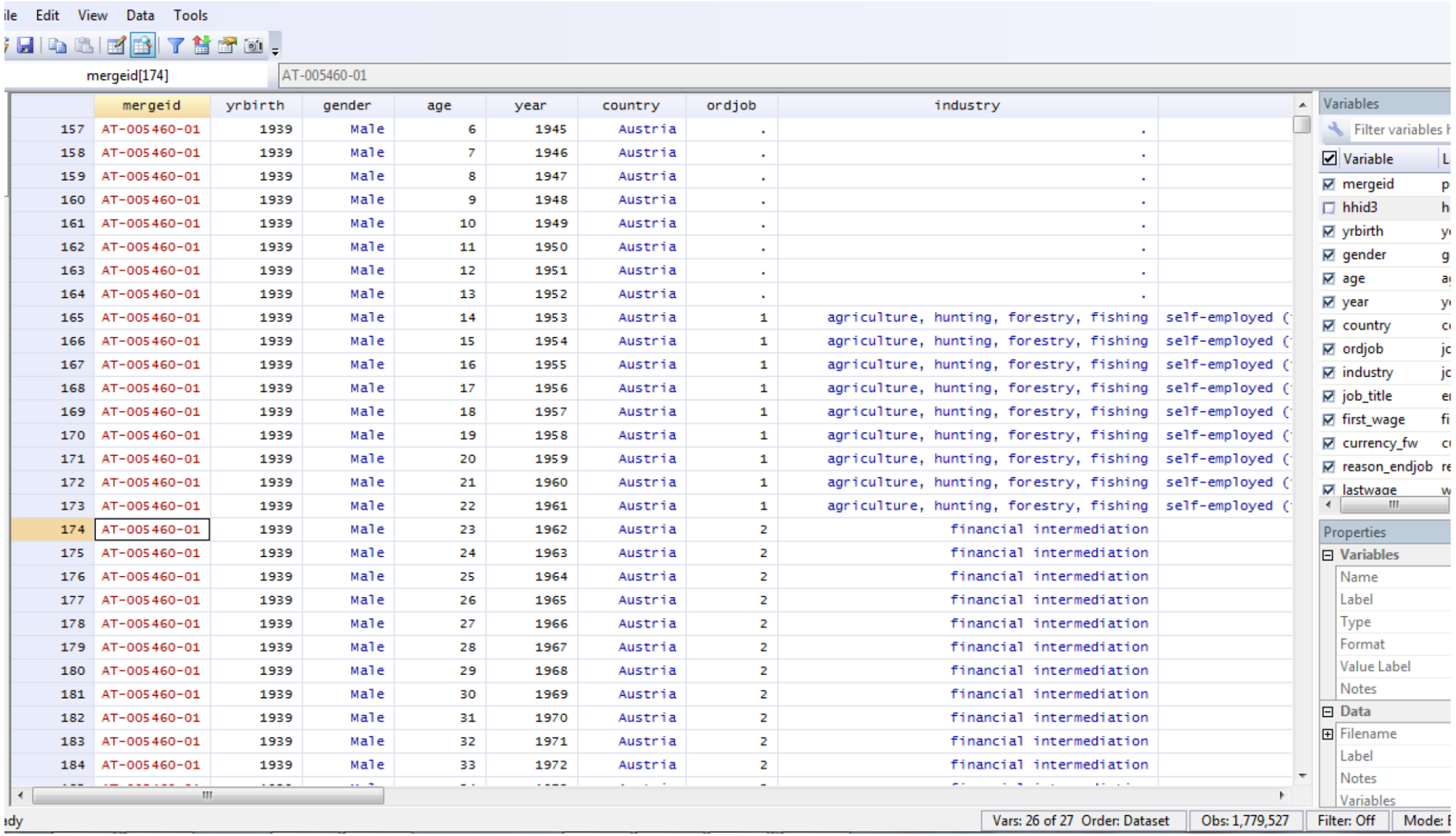
- SHARE release 7.0.0 includes JEP based oin wave 3 + wave 7

- The 6.0.0 release of JEP contains **6,116,307** person-year observations

Number of individuals and number of person-year observations by country

Country	Number of individuals	N. of person-year observations
Austria	3,722	258,703
Germany	4,902	326,451
Sweden	4,091	284,913
Netherlands	2,258	148,848
Spain	5,702	395,682
Italy	5,529	369,415
France	4,686	315,451
Denmark	4,105	266,062
Greece	4,252	280,749
Switzerland	2,972	200,956
Belgium	6,200	411,025
Israel	2,131	150,830
Czech Republic	5,115	348,781
Poland	5,499	353,903
Ireland	855	57,009
Luxembourg	1,254	83,508
Hungary	1,538	106,180
Portugal	508	34,724
Slovenia	3,692	255,422
Estonia	5,117	356,127
Croatia	2,407	159,908
Lithuania	2,032	134,562
Bulgaria	2,002	132,991
Cyprus	1,233	85,081
Finland	2,007	132,795
Latvia	1,754	116,932
Malta	1,260	84,246
Romania	2,112	137,727
Slovakia	2,064	127,326
Total	90,999	6,116,307

The database



	mergeid	yrbirth	gender	age	year	country	ordjob	industry	
157	AT-005460-01	1939	Male	6	1945	Austria	.	.	
158	AT-005460-01	1939	Male	7	1946	Austria	.	.	
159	AT-005460-01	1939	Male	8	1947	Austria	.	.	
160	AT-005460-01	1939	Male	9	1948	Austria	.	.	
161	AT-005460-01	1939	Male	10	1949	Austria	.	.	
162	AT-005460-01	1939	Male	11	1950	Austria	.	.	
163	AT-005460-01	1939	Male	12	1951	Austria	.	.	
164	AT-005460-01	1939	Male	13	1952	Austria	.	.	
165	AT-005460-01	1939	Male	14	1953	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
166	AT-005460-01	1939	Male	15	1954	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
167	AT-005460-01	1939	Male	16	1955	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
168	AT-005460-01	1939	Male	17	1956	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
169	AT-005460-01	1939	Male	18	1957	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
170	AT-005460-01	1939	Male	19	1958	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
171	AT-005460-01	1939	Male	20	1959	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
172	AT-005460-01	1939	Male	21	1960	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
173	AT-005460-01	1939	Male	22	1961	Austria	1	agriculture, hunting, forestry, fishing	self-employed (
174	AT-005460-01	1939	Male	23	1962	Austria	2	financial intermediation	
175	AT-005460-01	1939	Male	24	1963	Austria	2	financial intermediation	
176	AT-005460-01	1939	Male	25	1964	Austria	2	financial intermediation	
177	AT-005460-01	1939	Male	26	1965	Austria	2	financial intermediation	
178	AT-005460-01	1939	Male	27	1966	Austria	2	financial intermediation	
179	AT-005460-01	1939	Male	28	1967	Austria	2	financial intermediation	
180	AT-005460-01	1939	Male	29	1968	Austria	2	financial intermediation	
181	AT-005460-01	1939	Male	30	1969	Austria	2	financial intermediation	
182	AT-005460-01	1939	Male	31	1970	Austria	2	financial intermediation	
183	AT-005460-01	1939	Male	32	1971	Austria	2	financial intermediation	
184	AT-005460-01	1939	Male	33	1972	Austria	2	financial intermediation	

The variables (1)

Variables	Description	Questionnaire variables
mergeid	Person identifier fix across modules and waves	
hhid7	Household identifier wave 7	
hhid3	Household identifier wave 3	
jep_w	Number of Wave	
gender	Gender respondent	gender (cv_r module),
yrbirth	Year of birth respondent	yrbirth (cv_r module)
Age	Age respondent	int_year, yrbirth (cv_r module)
Year	Year	age, yrbirth (cv_r module)
country	Country of residence at the time of interview	country (cv_r module)
situation	Situation	re003, re007, re0010_*, re033_* (re module)
working	Working spell	re011_*, re026_* (re module)
unemployed	Unemployment spell	re031_*, re033_*, re006_, re007_, re035_* (re module)
in_education	In full time education	re002_ (re module)
retired	Retirement spell	re031_*, re033_*, re039a_*(re module); Waves 1, 2,4,5,6: ep329 (ep module)
mainjob	Main job spell	re040_, re011_*, re026_*
ordjob	Job spell numbering	re011_* (re module)
industry	Job industry	re014_* (re module)
job_title	Employee, civil servant or self-employed	re015_* (re module)
working_hours	Full time or part time	re016_*, re018_*, re020_* (re module)
first_wage	First wage for each job – nominal currency	re021_* (re module)
currency_fw	Currency coded first wage	re022c_* (re module)
first_income	First job income self-employment – nominal currency	re023_* (re module)
currency_fi	Currency coded first job income	re024c_* (re module)
reason_endjob	Reason left job	re031_* (re module)
afterlast	Situation after last job	re035_* (re module)
lastwage	Monthly wage at the end of main job	re041_ (re module)
currency_lw	Currency of monthly wage at the end of main job	re042_ , re022c_ (re module)
lastincome	Monthly income at the end of main job	re043_ (re module)
currency_li	Currency of monthly income at the end of main job	re044_, re024c_ (re module)
first_pension	First monthly pension benefit when retired	re036_* (re module). Wave 1,2,4,5 and 6: ep213_ (ep module)
currency_fp	Currency of first monthly pension benefit when retired	re037c_* (re module)

situation

- *"situat"* defines the job market status or the self-defined status if not in the labour force for each year.
- Respondents are asked to report what they did in gaps between jobs (re033) if two job spells are not adjacent, moreover they are asked to report their activity on the same item list if they never worked.
- Those who never worked can report up to 8 non-working spells.
- This information is combined with those about working, retired and in full time education spells.
- If a given year is reported to be both a working and a non-working year, *situat* reports it as working. Note that those overlaps are not necessary an indicator of recall bias. Most of them occur in the year of transition between a working and a non-working spell or in the year of retirement.

situation

Situation	Freq.	Percent	Cum.
Refusal	21	0.00	0.00
Don't know	172	0.00	0.00
Employee or self-employed	2,937,039	55.07	55.08
Unemployed and searching for a job	31,209	0.59	55.66
Unemployed but not searching for a job	28,267	0.53	56.19
Short term job (less than 6 months)	1,767	0.03	56.23
Sick or disabled	13,607	0.26	56.48
Looking after home or family	340,843	6.39	62.87
Leisure, travelling or doing nothing	5,541	0.10	62.98
Retired from work	771,246	14.46	77.44
Training	5,769	0.11	77.55
In education	1,141,620	21.41	98.95
Military services, war prisoner or equi	14,137	0.27	99.22
Managing your assets	2,146	0.04	99.26
Voluntary or community work	4,153	0.08	99.34
Forced labour or in jail	643	0.01	99.35
Exiled or banished	337	0.01	99.36
Labor camp	325	0.01	99.36
Concentration camp	156	0.00	99.37
Other	33,855	0.63	100.00
Total	5,332,853	100.00	

Country of residence

- Resp are asked all the changes in accommodation they had throughout their lives. If they changed Country of residence, they are asked to report the new country.

- Resp in w3 had to choose from a drop-down menu that included
  - all SHARE countries
  - UK, USA, Russia, Finland, Norway, Slovakia, Russia
  - two broader categories: “other European country” or “other non-European country”.

- SHARE wave 7 incorporates a “country-coder”, i.e. a built-in program that recognizes text strings and code them into a country name.
  - 36,18% individual-year observations from w7 refer to countries that in w3 would have been classified as “other European country” or “other non-European country”

Family composition

- SHARE contain information on:
  - the date of birth and death of each child of respondents,
  - the year of adoption for the adopted children
  - Year of start/end any cohabitation with a partner
  - Year of marriage and divorce
- We use this information to generate

nchildren_nat	Number of natural children alive in a given year	rc023 rc024_* rc027_* rc028_* (rc module)
nchildren	Number of children alive (including adopted children) in a given year	rc023,rc024_*, rc027_*, rc028_*, rc038_ ,rc039_ , rc041*, rc043*, rc044*, rc045*(rc module)
age_youngest_nat	Age of the youngest natural child alive in a given year	rc023, rc024_*, rc027_*, rc028_* (rc module)
age_youngest	Age of the youngest child alive, including adopted children, in a given year	rc023, rc024_*, rc027_*, rc028_*, rc038_, rc039_, sl_rc041*, sl_rc043*, sl_rc044*, rc045*(rc module)
withpartner	Dummy, takes value 1 if Resp is cohabiting with a partner	rp003, rp004b_*, rp011_*, rp012_* (rc module)
married	Dummy, takes value 1 if Resp is married	rp008_*, rp014_* (rc module)

Institutional variables in JEP

- Variables describing the pension legislation the respondent faced throughout their lives.
- Information is collected from the Mutual Information System on Social Protection (MISSOC) website, and from the Social Security Administration (SSA) website for Israel.
- Both sources provide information starting from 2004, therefore all the variables in table 5 are set to missing before that date.
- For countries that joined the EU after 2004, information is available since the time they joined the Union.

Variable	Description
contrib_employee	Pension contribution rate by the employee
contrib_employer	Pension contribution rate by the employer
ret_age	Statutory retirement age
early_age	Early retirement age
early_ret_reduction	Early retirement reduction rate
currency_min_pension	Currency of minimum pension
currency_max_pension	Currency of maximum pension
min_pension	Minimum pension benefits
max_pension	Maximum pension benefits

What is this dataset useful for?

- Duration and transition analysis
  - (We have complete job spells)
- Build synthetic life-cycle measures
- Merge with other SHARELIFE time-varying sections to build a fully fledged retrospective panel
  - Maternity history
  - Health history
  - Health care use
  - General Life questions

synthetic life-cycle measures

- Life histories can be «collapsed» in several ways into syntetic measures to be used then as time invariant chacteristics
  - Use job spells and reported wages to construct a lifetime/permanent income measure
    - Brunello, Weber, Weiss, EJ 2015 'Books are forever: Early life conditions, education and lifetime earnings in Europe'
  - Use employment/unemplyment/retirement spells + pension legislation variables to construct measures of Social Security Wealth
    - Alessie, Angelini, Van Santen EER 2013 'Pension wealth and household savings in Europe: Evidence from SHARELIFE'
    - Belloni, Agar Brugiavini, Buia, Carrino, Cavapozzi, Orso, Pasini JPEF 2019 'What do we learn about redistribution effects of pension systems from internationally comparable measures of Social Security Wealth?' THE SSW measure used in this paper is available to users and part of the gv variables of wave 4
  - Use sequence analysis to create "types" of individuals based on the working life trajectories
    - Wahrendorf, A&S 2015 'Previous employment histories and quality of life in older ages: sequence analyses using SHARELIFE.'
    - Wahrendorf, Blane, Bartley, Dragano, Siegrist, ALCR 2013 'Working conditions in mid-life and mental health in older ages.'

Example: SSW

- SSW is the discounted sum of pension benefits an individual (or household) receives throughout the lifecycle
- Discounting accounts for survival probabilities and financial discount rates
- When computed before retirement, it may only account for *accrued* pension rights or additionally include expected pension rights at a given (expected) retirement age
- What SSW can be used for?
  - Displacement of private wealth (Alessie, Angelini, Van Santen, 2013, EER)
  - Redistribution of pension systems (Brugiavini et al, 2019 JPEF)

Social Security Wealth: definitions and methods

- Since the release 5.0.0., Wave 4 includes among the generated variables two measures of individual accrued social security wealth
- Those measures are the first attempt of computing and deliver to the scientific community a set of *internationally comparable* measures of pension wealth computed for a large number of countries
- The SSW of retirees has been computed based on wave 4 data, SSW of workers relies on a variety of retrospective information obtained from the Job Episode Panel.

Social Security Wealth: definitions and methods

- We define the SSW **for workers** as follows:

$$SSW_i = \sum_{j=R}^{\Omega} \hat{P}_{i,j}(R)\, \pi(j|a)(1+r)^{a-j}$$

  - $\hat{P}(R)$: computed public old age pension benefit assuming that
    - the individual will retire at current age *a*
    - will start receiving pension income from the **old age retirement age *R***
  - *R* is included in the Job Episode Panel (*ret_age*).
  - $\hat{P}(R)$ depends on
    - Employment history
    - Earning history
    - National legislation

Social Security Wealth: definitions and methods

- A key step for the computation of $\hat{P}(R)$ is the reconstruction of individuals' working career including lifetime wages.
  - SHARELIFE reports – for each job episode – the self-reported after taxes first monthly wage;
  - it also reports after taxes last monthly wage for the main job spell.
  - We fill in missing wages within each job spell assuming that wages are constant in real terms.
  - Pension amounts depend sometimes on residential histories. Migration history from SHARELIFE again

- Belloni, Carrino, Orso, Buia, Cavapozzi, Pasini, Brugiavini (2016) Internationally comparable measures of individual social security wealth in SHARE Wave, SHARE Working Paper

Example: sequence analysis

- Wahrendorf 2015 A&S
- Identify 10 Clusters that group individuals into «type of career»

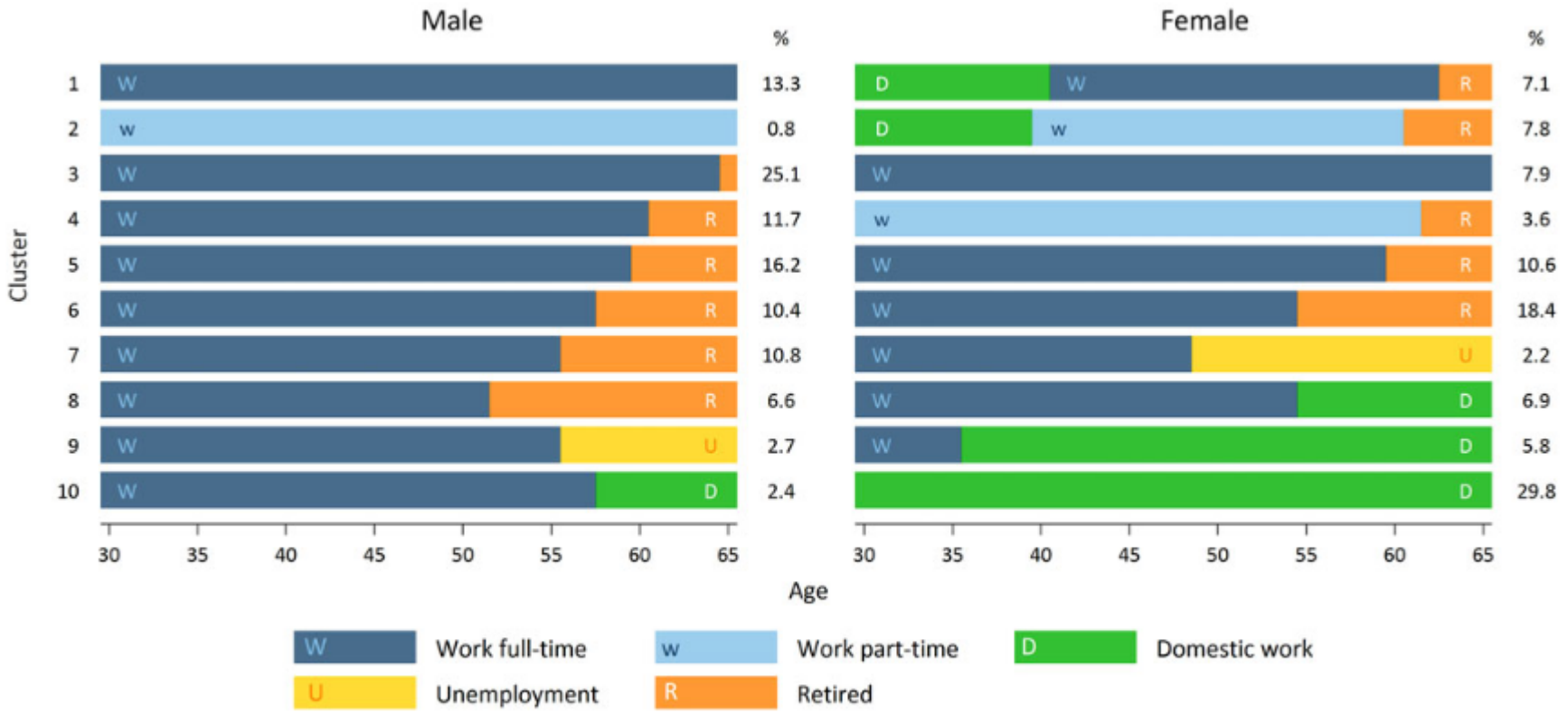


Figure 2. Prototypes of employment histories for men (N=4,808) and women (N=4,907) and frequencies (%). *Note*: Based on ten-cluster solutions calculated for men and women separately.

Quality of life and working trajectory

- Regress CASP index from standard waves on type dummies

TABLE 4. *Multi-level estimates for quality of life (CASP): regression coefficients and standard errors (SE) for men*

	Empty		Model 1		Model 2		Model 3	
			b	SE	*b*	SE	*b*	SE
Fixed parameters:								
Cluster:								
1			Ref.		Ref.		Ref.	
2			−1.33	0.98	−1.34	0.98	−1.24	0.97
3			−0.46	0.30	−0.46	0.30	−0.40	0.29
4			−0.08	0.35	−0.09	0.35	−0.11	0.35
5			−0.79*	0.33	−0.79*	0.33	−0.69*	0.33
6			−0.73*	0.37	−0.73*	0.37	−0.64	0.37
7			−0.81*	0.37	−0.78*	0.37	−0.77*	0.37
8			−0.46	0.43	−0.44	0.43	−0.45	0.43
9			−1.49*	0.59	−1.44*	0.59	−1.19*	0.59
10			−1.65**	0.62	−1.71**	0.62	−1.38*	0.62
Random parameters:								
Level 1: within country	5.48***	0.06	5.34***	0.06	5.32***	0.06	5.29***	0.06
Level 2: between country	2.59***	0.52	2.50***	0.50	2.51***	0.50	2.57***	0.51
Statistics:								
R^2_1 (level 1)			0.051		0.055		0.066	
R^2_2 (level 2)			0.073		0.063		0.018	
Log likelihood	−11,994.23		−11,892.85		−11,885.38		−11,863.85	
AIC	23,994.46		23,827.70		23,830.77		23,791.71	
BIC	24,013.22		23,959.00		24,018.33		23,991.78	

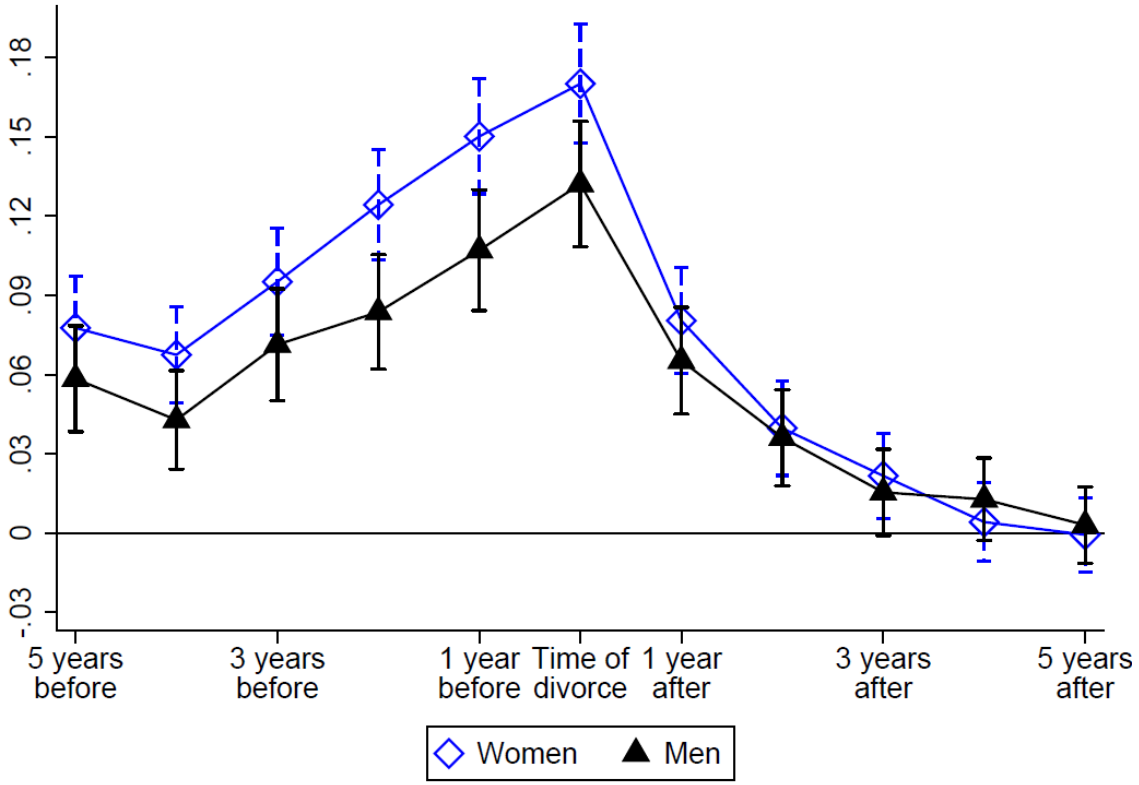
Example: augmented JEP

- Cavapozzi, Fiore, Pasini JEOA 2019 ‘Divorce and well-being. Disentangling the role of stress and socio economic status’
- Look at the effect of Divorce on General Life questions: Happiness, Financial Hardship, Stress
  - Add GL questions to JEP to build the dependent variable: dummies 1/0 whether happy/fin hard/stress in year t
  - Add Health conditions as controls

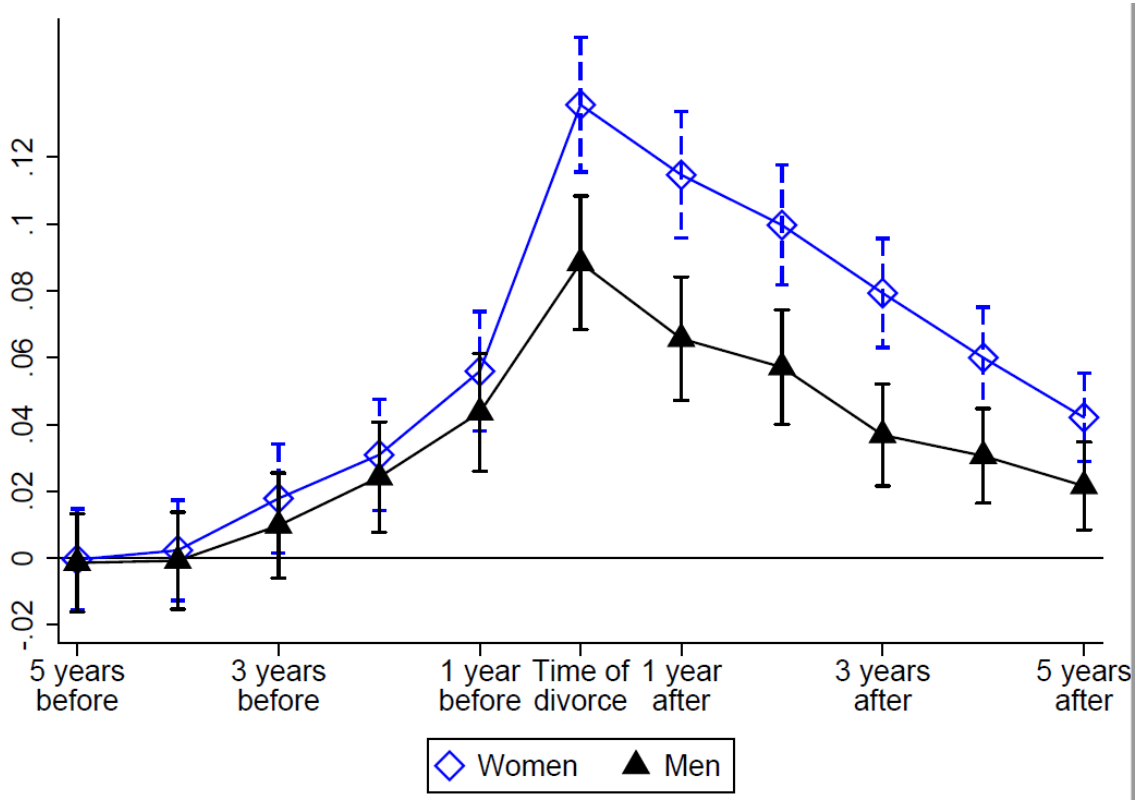
- Estimate effect of being in year of divorce on happiness
- Since the panel is long, look at the effect 5 years before and after

Example: augmented JEP

Stress around Divorce



Fianancial hardship around divorce



Merging time invariant variables

- CS section (early childhood), childhood illnesses
- Completed education and other variables from standard waves
- This is easy: merge m:1 by mergeid

Events in a given year

- Example: miscarriages
- | | **RC055** NUMBER OF PREGNANCIES NOT ALIVE CHILDREN
  | | How many such pregnancies did you have in all?
  | |
  | | _____ (0..20)
  | |
  | | *IF RC055 (NUMBER OF PREGNANCIES NOT ALIVE CHILDREN) > 0*
- | | | | **RC056** YEAR PREGNANCY ENDED
  | | | | [Let us begin with the first of these pregnancies.] In which year did [this/the] [1st/2nd/3rd/4th/5th/6th/7th/8th/9th/10th/11th/12th/13th/14th/15th/16th/17th/18th/19th/20th] pregnancy end?
  | | | |
  | | | | (1900..2009)

Stata code

```
use ${datain3}sharew3_rel1_rc.dta, clear

* cleaning miscarriages
keep mergeid sl_rc056_*                          /*flat format*/
reshape long sl_rc056_, i(mergeid) j(ordmiss) /*reshaped into panel format*/
drop if sl_rc056_==.
rename sl_rc056_ year                            /*generate an events panel*/

gen miscarriage=1

* Label variables
label var      ordmiss "Sequence number of miscarriage"
label var      miscarriage "Miscarriage in the year dummy"

save ${dataout}rc_section1.dta, replace

use ${dataout}longpanel0.dta, clear
merge 1:1 mergeid year using ${dataout}rc_section1.dta
drop if _merge==2
drop _merge

replace miscarriage =0 if miscarriage ==.
```

Definite periods

- Questions where we have just 1 period of time

- | **GL005** PERIOD OF STRESS
  | (Looking back on your life,) was there a distinct period during which you were under more stress compared to the rest of your life?
  |
  | 1. Yes
  | 5. No
  |
  | *IF GL005 (PERIOD OF STRESS) = 1. Yes*
  | |
  | | **GL006** WHEN STRESS PERIOD STARTED
  | | When did this stress period start?
  | |
  | | (1900..2009)
  | | **GL007** WHEN STRESS PERIOD STOPPED
  | | When did this period stop?
  | | IWER:Please code 9997 if this period is still ongoing
  | | (1900..9997)

Stata code

```
use ${dataout}gl_section.dta, clear                    /*Already in events database format*/
        rename sl_gl006 year
        keep mergeid year
        gen start=1
        keep if year<.
        merge 1:1 mergeid year using ${dataout}longgl.dta, keep(match using) nogenerate
        sort mergeid year
        save ${dataout}longgl.dta, replace

use ${dataout}gl_section.dta, clear
        rename sl_gl007 year
        keep mergeid year
        gen stop=1
        keep if year<.
        merge 1:1 mergeid year using ${dataout}longgl.dta, keep(match using) nogenerate
        sort mergeid year
        save ${dataout}longgl.dta, replace
```

Stata code

```
use ${dataout}longgl.dta, clear
gen stress=0
                gen c=year if start==1
                egen yrb=min(c), by(mergeid)
                drop c

                gen c=year if stop==1
                egen yrd=min(c), by(mergeid)
                drop c

                replace stress= 1 if yrb<=year & yrb<.
                replace stress= 0 if yrd<year & yrd<.

                drop yrb yrd start stop
                label var stress "in a stressful period"
save ${dataout}longgl.dta, replace

use ${datajep}sharew123_rel2_gv_job_episodes_panel.dta, clea
merge 1:1 mergeid year using ${dataout}longgl.dta, keep(match using) nogenerate
```

Illness period

- Individuals can report up to three periods of illness, and specify which type of illness in HS055 and HS056 (two code all that apply lists)
- Variables in the public release are
  - sl_hs055*i*_*j* and sl_hs056*i*_*j*, where
    - *i* is the ith illness from a showcard list
    - *j* goes from 1 to 3
- For each period, respondents report start and end date
  - Start: sl_hs059_1… sl_hs059_3
  - End: sl_hs060_1… sl_hs060_3

(simplified) Stata code

```
foreach y of numlist 1900/2009 {
foreach i of numlist 1/11 {
                qui gen hs055_illness`i'_`y'=.
                qui replace hs055_illness`i'_`y'=1 if ///
                (sl_hs055d`i'_1==1 & ((`y'>=sl_hs059_1 & `y'<=sl_hs060_1))) | ///
                (sl_hs055d`i'_2==1 & ((`y'>=sl_hs059_2 & `y'<=sl_hs060_2))) | ///
                (sl_hs055d`i'_3==1 & ((`y'>=sl_hs059_3 & `y'<=sl_hs060_3)))

                qui replace hs055_illness`i'_`y'=0 if ///
                (sl_hs055d`i'_1==1 & (`y'<sl_hs059_1 | `y'>sl_hs060_1) & sl_hs059_1!=.) | ///
                (sl_hs055d`i'_2==1 & (`y'<sl_hs059_2 | `y'>sl_hs060_2) & sl_hs059_2!=.) | ///
                (sl_hs055d`i'_3==1 & (`y'<sl_hs059_3 | `y'>sl_hs060_3) & sl_hs059_3!=.)                    }
}

qui reshape long hs055_illness1_ hs055_illness2_ hs055_illness3_ hs055_illness4_ ///
                 hs055_illness5_ hs055_illness6_ hs055_illness7_ hs055_illness8_ ///
                 hs055_illness9_ hs055_illness10_ hs055_illness11_ , i(mergeid) j(year) ///
save ${dataout}hs_section.dta, replace
merge 1:1 mergeid year using ${dataout}hs_section.dta
```

Note: There are some details to be added to deal with missings and those who report in hs054 they had more than 3 illness episodes

Period with attached information

- Some information have more details than simply start and end.
- Example: type of accomodation

Stata code

```
use ${datajep}sharew123_rel2_gv_job_episodes_panel.dta, clear
keep mergeid year
save ${dataout}longac.dta, replace

                                        *accomodation starts
forvalues i=1/29{
                        use ${dataout}ac_section.dta, clear
                        cap gen sl_ac011_`i'=.
                        sort mergeid
                        keep mergeid start`i' sl_ac008_`i' sl_ac011_`i'
                        rename start`i' year
                        gen start`i'=1 if sl_ac008_`i'==5
                        labe var start`i' "accomodation `i' starts and it is non private"
                        drop sl_ac008_`i'
                        keep if year<. & start`i'==1
                        merge 1:1 mergeid year using ${dataout}longac.dta, keep(match using) nogenerate
                        sort mergeid year
                        save ${dataout}longac.dta, replace
                        }

                                        *accomodation finishes
forvalues i=1/29{
                        use ${dataout}ac_section.dta, clear
                        sort mergeid
                        rename end`i' year
                        gen end`i'=1 if sl_ac008_`i'==5
                                        labe var start`i' "accomodation `i' ends and it is non private"
                        keep if year<.          & end`i'==1
                        keep mergeid year end`i'
                        merge 1:1 mergeid year using ${dataout}longac.dta, keep(match using) nogenerate
                        save ${dataout}longac.dta, replace
                        }
```

```
egen typeacc=rowmin(sl_ac011_29-sl_ac011_1)
sort id year
replace typeacc=l.typeacc if nonprivateacc==1 & l.typeacc<. & typeacc>=.
replace typeacc=0 if nonprivateacc==0
label val typeacc ac011

save ${dataout}longpanel7.dta, replace
erase ${dataout}longac.dta
```

References: «howto» SHARELIFE

- Job Episodes Panel
  - Brugiavini, Orso, Genie, Naci, Pasini (2019) "Combining the Retrospective Interviews of Wave 3 and Wave 7: The Third Release of the SHARE Job Episodes Panel". SHARE WP 36-2019
  - Antonova; Aranda; Pasini; Trevisan (2014) Migration, family history and pension: the second release of the SHARE Job Episodes Panel , SHARE WP 18-2014
  - Brugiavini; Cavapozzi; Pasini; Trevisan (2013) Working life histories from SHARELIFE: a retrospective panel , SHARE WP

- SSW
  - Belloni, Carrino, Orso, Buia, Cavapozzi, Pasini, Brugiavini (2016) Internationally comparable measures of individual social security wealth in SHARE Wave, SHARE Working Paper

- Validation of data
  - Trevisan, Pasini and Rainato (2011) Cross-country comparison of monetary values from SHARELIFE. SHARE WP
  - Havari, Mazzonna, (2015). Can we trust older people's statements on their childhood circumstances? Evidence from SHARELIFE. *European Journal of Population*, *31*(3), 233-257.
  - Bingley, Martinello (2014) Measurement error in the Survey of Health, Ageing and Retirement in Europe: A validation study with administrative data for education level, income and employment. SHARE WP 16-2014.