

Christian Hunkler

Stefan Gruber

Agnes Orban

Stephanie Stuck

with support by

Martina Brandt

Dimitris Christelis

Danilo Cavapozzi

Giuseppe De Luca

George Papadoudis

Morten Wahrendorf

Mauricio Avendano Pabon

Radim Bohacek

Thorsten Kneip

Giacomo Pasini



Federal Ministry
of Education
and Research

National
Institute
on Aging
■ ◆ ★ ✱

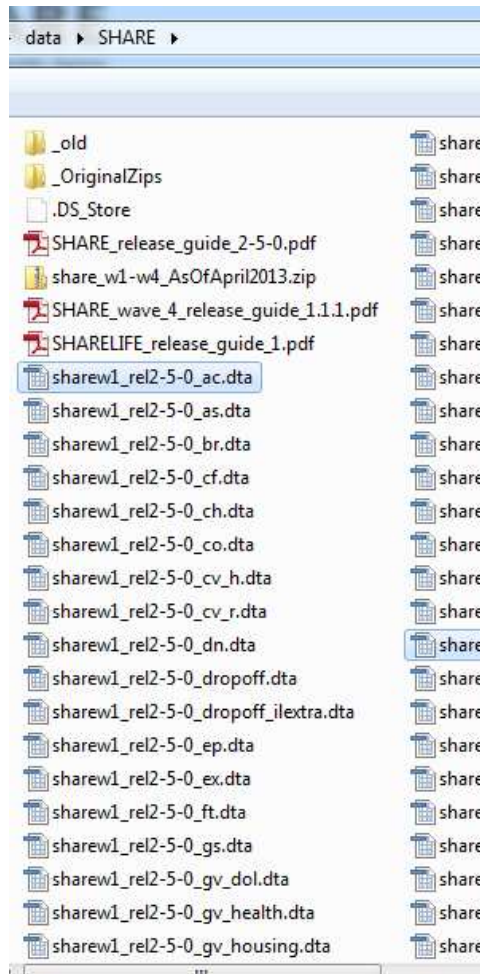
Liège, November 29, 2013



mea

MUNICH CENTER FOR THE ECONOMICS OF AGING
AT THE MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY

SHARE is complex



113 single data files

In addition:

- * Individual-, couple- and household-level data (family-/financial-/household-respondents)
- * Complex skip patterns for stable information
- * Complex sampling leading to variation in available subsamples ("vignette", "main", "drop-off"-samples ...)
- * Deviations across waves (e.g. questions "moving" from e.g. drop-off to CAPI)

* Idea

- * Easy structured, ready to use *single* file (no merging)
- * Selection of variables showing the central potentials of SHARE
- * Stimulate students/users to later transfer to SHARE main release

* Purpose

- * (Student-)training
- * Get students & researchers hooked on SHARE

* Main features

- * *easySHARE* includes *all respondents* of the SHARE main release (non-responding partners & deceased respondents excluded)
- * Long panel dataset covering *waves 1, 2, 3/SHARELIFE, and wave 4*

1. Inclusion of new indices and variables
(Age, Couple-ID, health scores, etc.)
2. Comparability with RAND HRS
(Adaption of existing variables to RAND HRS)
3. Integration of information
(assign household-/ financial-/family-respondent information, forwarding of stable information from previous waves, etc.)
4. Recoded and simplified variables
(extended missing code scheme, etc.)
5. Additional features:
Short documentation, "cookbook"-styled analysis examples, release of Stata do-file generating *easy*SHARE (easy adaption to SHARE main release)

- * Guidelines for choosing variables
 - * Cross-country comparability
 - * Maximum cross-wave comparability, i.e. longitudinal measures (exception: retrospective life histories collected in SHARELIFE)
 - * Low rate of missing information
 - * Indices instead of variables
 - * No complex filtering or documentation of filters in target variable

- * **Demographics:** age, gender, country of birth, citizenship, education, religion, marital status, age & gender of partner
- * **Household composition:** living with partner in the same household, household size, children living in the household
- * **Social Support & Network:** mother/father alive, number of children, residential proximity of children, number of grandchildren, number of living siblings, social activities, received and given social support

- * **Childhood conditions:** number of books at age ten, occupation of main breadwinner at age ten, relative mathematical skills when ten, relative language skills at age ten
- * **Health and health behaviour:** self-perceived health, number of chronic diseases, mental health variables, depression scale EURO-D, CASP-12 index for quality of life and well-being, health care utilization, grip strength, body mass index, smoking and drinking behavior, vigorous activities/sports

- * **Functional limitation indices:** mobility index, large muscle index, activities of daily living index, instrumental activities of daily living index, gross motor skills index, fine motor skills index, cognitive functions
- * **Work & Money:** current job situation, term of main job, working hours per week, satisfaction with main job, early retirement plans, able to make ends meet, monthly expenditure on food, household income percentiles


```
*>> Transfer information collected once (in baseline interviews)
*   to other waves of the respondent
*   (done for isced_r, edueyears, )
foreach var in isced_r edueyears {
    di "`var'"
    // first we check for deviations within person and set the
    // variable to missing if we have contradictory information
    // this occurs rarely and there is no way to know which
    // information is correct
    gen `var'c = `var'
    replace `var'c = . if `var'c < 0
    egen `var'c_sd = sd(`var'c), by(mergeid)
    replace `var'c = . if `var'c_sd > 0 & `var'c_sd < .

    // here we start to copy the information to all lines of
    // the respondent; i.e. isced_r from wave 1 is written into
    // all lines of the same respondent
    sort mergeid
    foreach i in $w {
        di "`var'"
        replace `var'c = `var'c[_n+`i'] if mergeid==mergeid[_n+`i'] & ///
            `var'c==. & `var'c[_n+`i'] !=.
        replace `var'c = `var'c[_n-`i'] if mergeid==mergeid[_n-`i'] & ///
            `var'c==. & `var'c[_n-`i'] !=.
    }
    // here we copy back the missing codes, only into the line of the wave
    // the missing code occurred, if we did not find a valid answer in any
    // other wave
    replace `var'c = `var' if `var'c == . & `var' < 0
    rename `var' `var'_orig // to preserve the label stored in `var'
    rename `var'c `var'
}
lab val isced_r isced
```



Working with *easySHARE*

Variables

Name	Label
mergeid	person identifier (fix across modules and waves)
hhid	household identifier in respective wave - see...
coupleid	couple identifier in respective wave - see var....
wave	wave
wavepart	wave participation pattern
int_year	interview year
int_month	interview month
country	country identifier
country_mod	country identifier (ISO coded)
language	language of questionnaire
female	gender: female=1, male=0
dn002_mod	month of birth
dn003_mod	year of birth
age	age at interview (in years)
birth_country	country of birth (ISO coded)
citizenship	citizenship of respondent (ISO coded)
q34_re	what religion do you belong or feel attached t...
iscd_r	education of respondent in ISCED-97 Coding
edueyears_mod	years of education
mar_stat	marital status
hhsz	household size
partnerinhh	living with spouse/partner
int_partner	interview of partner available
age_partner	age at interview of partner
gender_partner	gender of partner: female=1, male=0
mother_alive	is natural parent still alive: mother
father_alive	is natural parent still alive: father
siblings_alive	number of siblings alive (based on: dn036_d...
ch001	number of children

Results

```

. use "/Users/ChristianHunkler/Desktop/test easySHARE_rell1-0-0_stata/
> easySHARE_rell1-0-0.dta"
(easySHARE release 1.0.0 waves 1 2 3 4 doi 10.6103/SHARE.easy.100 )

. list mergeid wave birth_country wavepart if substr(mergeid,1,9)=="A
> T-986403" | substr(mergeid,1,9)=="SE-209636" , sepby(mergeid) noob
> s

```

mergeid	wave	birth_country	wavepart
AT-986403-01	1	40. Austria	1234
AT-986403-01	2	40. Austria	1234
AT-986403-01	3	40. Austria	1234
AT-986403-01	4	40. Austria	1234
AT-986403-02	1	40. Austria	1234
AT-986403-02	2	40. Austria	1234
AT-986403-02	3	40. Austria	1234
AT-986403-02	4	40. Austria	1234
SE-209636-01	1	752. Sweden	14
SE-209636-01	4	752. Sweden	14

Working with *easy*SHARE

. ta br010_,m

days a week consumed alcohol |
last 3 months | Freq.

-----+-----	
-15. no information	1,251
-13. not asked in this wave	26,836
-12. don't know / refusal	97
1. not at all	41,113
2. less than once a month	12,400
3. once or twice a month	14,177
4. once or twice a week	21,597
5. three or four days a week	8,534
6. five or six days a week	3,313
7. almost every day	23,553
-----+-----	
Total	152,871

. ta ep026_, m

satisfied with (main) job | Freq.

-----+-----	
-15. no information	1,599
-13. not asked in this wave	26,836
-12. don't know / refusal	320
-9. filtered: not worked	87,800
1. strongly agree	16,022
2. agree	17,470
3. disagree	2,272
4. strongly disagree	552
-----+-----	
Total	152,871

Working with *easy*SHARE

- * Set the missing codes to system missing:

```
mvdecode ep026_ br010_, mv(-1/-15)
```

- * Recode the variables:

```
recode ep026 (1=1) (2/4=0), gen(HappyJob)
```

```
recode br010 (1/6=0)(7=1) , gen(DrinkEveryDay)
```

- * Cross-sectional analysis: `logit Happy Drink if wave==4`

Logistic regression

Number of obs = 15634

LR chi2(1) = 4.76

Prob > chi2 = 0.0292

Log likelihood = -10709.583

Pseudo R2 = 0.0001

	Coef.	Std. Err.	z	P> z	[95% Conf. Inter
HappyJob					
DrinkEveryDay	.0949133	.0434681	2.18	0.029	.0097173 .1801093
_cons	-.2691981	.0176461	-15.26	0.000	-.3037838 -.2346124

Positive significant coefficient: people reporting to drink every day, are much more satisfied with their job.

Working with *easy*SHARE

- * To fully exploit SHARE, i.e. estimating a panel model, we first need to `xtset` the data.

- * Unfortunately Stata does not accept numerical person IDs, but there is a simple work-around:

```
bysort mergeid : gen mergeid_n = _n == 1  
replace                mergeid_n = sum(mergeid_n)
```

- * Then we can declare the data to be a panel

```
xtset mergeid_n wave
```

- * Then we can estimate a fixed effects panel model:

```
xtlogit Happy Drink, fe
```

Working with *easy*SHARE

Conditional fixed-effects logistic regression Number of obs
Group variable: mergeid_n Number of groups

Obs per group: min = 2
avg = 2.3
max = 3

Log likelihood = -2042.4895 LR chi2(1) = 0.57
Prob > chi2 = 0.4501

	HappyJob	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
+						
DrinkEveryDay	.0795758	.105378	0.76	0.450	-.1269612	.2861128

Unfortunately, the cross-sectional correlation we found, is gone. Respondents who increased their alcohol intake between waves, are not significantly more satisfied with their jobs.



easy access to *easy*SHARE

- * Registered SHARE users can download easySHARE from the SHARE Research Data Center.

→ Please note that the SHARE data access rules for the SHARE main release also apply to the easySHARE data set.

- * *easy* procedure for teachers/instructors:

1. You fill and sign the *easy*SHARE teacher statement (if you are already a registered SHARE user) which allows to share *easy*SHARE with a specified group/class.
2. Your students sign the (regular) SHARE user statement. You collect them and send both the teacher and the students' user statements to: jjanssen@uvt.nl
3. In due course, all your students are registered as regular SHARE users and will be provided with Login/passwords for the main distribution of SHARE.

Thank you for your attention!

Questions?

For more information on *easy*SHARE visit our website:

<http://www.share-project.org> → Data Access & Documentation → *easy*SHARE

Contact: info@share-project.org

BACKUP

Working with *easy*SHARE

To reshape the data from long into wide format :

```
reshape wide int_year ep005_ maxgrip, i(mergeid) j(wave)
```

Data	long	->	wide
Number of obs.	152871	->	86290
Number of variables	5	->	13
j variable (4 values)	wave	->	(dropped)
xij variables:			
	int_year	->	int_year1 int_year2 ... int_year4
	ep005_	->	ep005_1 ep005_2 ... ep005_4
	sphus	->	sphus1 sphus2 ... sphus4

Working with *easy*SHARE

Structure of the data:

mergeid	wave	birth_country	wavepart
AT-986403-01	1	Austria	1234
AT-986403-01	2	Austria	1234
AT-986403-01	3	Austria	1234
AT-986403-01	4	Austria	1234
SE-209636-01	1	Sweden	14
SE-209636-01	4	Sweden	14

Released today:

- * *easy*SHARE
- * General calibrated weights
 - * includes everything to generate your own:
 - * calibrated cross-sectional weights
 - * longitudinal weights
 - * both at the individual and at the household level
- * Release 2.6.0 of wave 1 & 2
 - * including minor corrections
 - * and Israel wave 2