

Determinants of the attrition process in the first two waves of SHARE

Giuseppe De Luca
ISFOL

The Royal Society
London, March 23, 2009

1 Raw attrition rates

Table 1 and 2 presents the **raw sample attrition rates** in the first two waves of SHARE by country, gender and age class. These are defined as the fraction of all respondents who are lost between the first and the second wave.

Table 1: Samples sizes by country, sample type and gender.

Country	Baseline			Longitudinal			Sample Attrition		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
AT	782	1111	1893	509	729	1238	0.35	0.34	0.35
BE	1739	2088	3827	1281	1527	2808	0.26	0.27	0.27
CH	462	542	1004	317	379	696	0.31	0.30	0.31
DE	1380	1628	3008	710	834	1544	0.49	0.49	0.49
DK	771	936	1707	564	685	1249	0.27	0.27	0.27
ES	994	1402	2396	576	799	1375	0.42	0.43	0.43
FR	1386	1807	3193	848	1150	1998	0.39	0.36	0.37
GR	1244	1654	2898	966	1314	2280	0.22	0.21	0.21
IT	1132	1427	2559	765	1001	1766	0.32	0.30	0.31
NL	1368	1611	2979	804	973	1777	0.41	0.40	0.40
SE	1414	1639	3053	930	1080	2010	0.34	0.34	0.34
Total	12672	15845	28517	8270	10471	18741	0.35	0.34	0.34

Table 2: Samples sizes by country, sample type and gender.

Country	Age class							Total
	(54]	[55-59]	[60-64]	[65-69]	[70-74]	[75-79]	[80)	
AT	0.34	0.31	0.34	0.33	0.38	0.32	0.47	0.35
BE	0.27	0.23	0.24	0.27	0.29	0.24	0.36	0.27
CH	0.30	0.32	0.24	0.22	0.36	0.29	0.46	0.31
DE	0.53	0.43	0.43	0.45	0.52	0.56	0.58	0.49
DK	0.26	0.28	0.20	0.22	0.28	0.30	0.37	0.27
ES	0.42	0.40	0.39	0.39	0.43	0.45	0.52	0.43
FR	0.39	0.35	0.36	0.30	0.40	0.37	0.45	0.37
GR	0.23	0.20	0.20	0.16	0.15	0.26	0.31	0.21
IT	0.28	0.30	0.30	0.30	0.33	0.28	0.45	0.31
NL	0.41	0.40	0.37	0.37	0.36	0.45	0.52	0.40
SE	0.35	0.31	0.30	0.32	0.38	0.34	0.46	0.34
Total	0.35	0.32	0.32	0.32	0.35	0.36	0.44	0.34

Sample attrition is a key issue.

- About 9,800 individuals (34% of baseline sample) are lost between the first and the second wave.
- Sample attrition rates vary substantially by country, ranging from about 21% in Greece to about 49% in Germany.
- In most countries, sample attrition rates are slightly higher for men than for women.
- In most countries, we also notice a U-shape relationship between sample attrition and respondents' age.

2 How SHARE deal with attrition problems

SHARE provides **calibrated longitudinal weights (CLW)** which are meant to make the distribution of the sample by gender and age class in each country the same as the distribution of the target population.

Following Deville and Särndal (1992), CLW are as close as possible to the original sampling design weights, while also respecting a set of calibration margins reflecting the size of the target population across gender and age groups (50-59, 60-69, 70-79 and 80+).

Calibrated longitudinal weights are computed

- individual and household level,
- 3 variants of the sample (main, vignette, and joint longitudinal samples).

In most countries, information about the calibration margins comes from external sources, such as national population census.

Calibration margins are adjusted for mortality in the target population of wave 1 by using estimates of mortality rates obtained from life tables.

3 Attrition mechanism

A key issue in studying attrition is the nature of the underlying missing-data mechanism. This may depend on both observable and unobservable factors.

This distinction is important because it leads to different assumptions on the attrition process and different ways of controlling for the potential attrition selectivity bias.

Following Rubin (1976), this mechanism can be:

- missing completely at random (MCAR);
- missing at random (MAR);
- not missing at random (NMAR).

Under:

- MCAR, attrition can be ignored because it is independent from the distribution of the key survey variables. There is only an efficiency loss.
- MAR, weighting strategies and imputation procedures may lead to valid inference on the target population. However, it is important to select a suitable set of conditioning variables.
- NMAR, the attrition probability may depend on both observable and unobservable factors. This is the typical case considered in the literature on sample selection models.

4 Determinants of the attrition process

Knowing determinants of the attrition process is crucial for the implementation of ex-post statistical procedure (weights, imputations sample selection models) which attempt to correct for selectivity bias generated by MAR and NMAR sample attrition.

I classify determinants of the raw attrition process in three broad categories:

- mortality;
- survey design features that may affect the probability of continuing panel participation;
- factors related to individual decisions of leaving the panel.

4.1 Mortality

Notice that, mortality and other reasons of ineligibility (moving abroad) which lead to exit from the target population do not enter the computation of the attrition rates.

The incidence of these events cannot be easily isolated because they are unknown for all individuals not contacted between the first and the second wave (Sample Management System - SMS - data are still uncleaned).

Table 3 compares the raw average attrition rate of each country, gender and age class (50-54, 55-59, 60-64, 65-69, 70-74, 75-79) with the corresponding average mortality rate obtained from the life tables provided by Eurostat.

Table 3: Raw sample attrition rates (top row) and mortality rates (bottom row) by country, sex and age class.

Male							
Country	50-54	55-59	60-64	65-69	70-74	75-79	Total
AT	0.368	0.323	0.308	0.356	0.404	0.333	0.344
	0.014	0.019	0.028	0.043	0.073	0.115	0.041
BE	0.259	0.216	0.263	0.257	0.281	0.259	0.253
	0.013	0.019	0.029	0.044	0.076	0.125	0.043
CH	0.276	0.364	0.247	0.258	0.306	0.317	0.293
	0.009	0.015	0.022	0.036	0.059	0.098	0.033
DE	0.531	0.448	0.455	0.419	0.458	0.612	0.477
	0.013	0.019	0.029	0.046	0.074	0.116	0.043
DK	0.249	0.297	0.221	0.152	0.271	0.323	0.250
	0.013	0.020	0.031	0.051	0.082	0.140	0.043
ES	0.413	0.390	0.418	0.380	0.415	0.417	0.405
	0.012	0.018	0.026	0.043	0.068	0.114	0.045
FR	0.414	0.365	0.407	0.343	0.389	0.338	0.380
	0.014	0.019	0.027	0.041	0.064	0.103	0.037
GR	0.249	0.200	0.218	0.193	0.149	0.292	0.213
	0.012	0.018	0.027	0.042	0.067	0.117	0.037
IT	0.281	0.367	0.282	0.303	0.368	0.292	0.320
	0.009	0.015	0.024	0.039	0.066	0.109	0.038
NL	0.404	0.424	0.393	0.385	0.375	0.430	0.403
	0.010	0.017	0.026	0.045	0.078	0.130	0.040
SE	0.338	0.326	0.312	0.302	0.344	0.350	0.326
	0.009	0.014	0.022	0.038	0.064	0.109	0.037
Total	0.349	0.331	0.329	0.318	0.342	0.369	0.337
	0.012	0.018	0.027	0.043	0.070	0.116	0.040

Table 2: Raw sample attrition rates (top row) and mortality rates (bottom row) by country, sex and age class (continued).

Female							
Country	50-54	55-59	60-64	65-69	70-74	75-79	Total
AT	0.305	0.292	0.357	0.305	0.364	0.315	0.324
	0.006	0.010	0.014	0.021	0.038	0.071	0.023
BE	0.249	0.243	0.223	0.279	0.291	0.226	0.252
	0.007	0.010	0.014	0.022	0.040	0.074	0.024
CH	0.255	0.276	0.239	0.189	0.415	0.271	0.267
	0.005	0.008	0.011	0.018	0.032	0.059	0.019
DE	0.497	0.411	0.407	0.493	0.576	0.514	0.474
	0.007	0.010	0.014	0.023	0.041	0.076	0.023
DK	0.257	0.273	0.185	0.295	0.289	0.286	0.260
	0.009	0.013	0.020	0.035	0.057	0.088	0.031
ES	0.429	0.408	0.374	0.392	0.443	0.477	0.419
	0.005	0.007	0.010	0.018	0.034	0.063	0.021
FR	0.373	0.346	0.323	0.259	0.414	0.391	0.351
	0.006	0.008	0.012	0.018	0.030	0.055	0.019
GR	0.208	0.204	0.178	0.126	0.158	0.241	0.186
	0.005	0.007	0.011	0.020	0.038	0.084	0.022
IT	0.302	0.236	0.306	0.303	0.298	0.274	0.286
	0.005	0.008	0.012	0.019	0.033	0.060	0.018
NL	0.365	0.380	0.349	0.358	0.356	0.474	0.373
	0.007	0.011	0.015	0.025	0.042	0.078	0.022
SE	0.362	0.295	0.281	0.329	0.408	0.338	0.329
	0.006	0.009	0.014	0.022	0.038	0.071	0.021
Total	0.335	0.307	0.303	0.316	0.361	0.350	0.325
	0.006	0.009	0.013	0.021	0.038	0.071	0.022

Main findings: The comparison of the two sets of rates shows that

- in the selected age groups, mortality can explain about 3% of raw sample attrition rates.
- mortality becomes a relevant source of attrition only at the very old ages. It can explain more than 7% of attrition only after 70 years of age for men and 75 years of age for women;
- mortality cannot explain the huge cross-country variability in the attrition rates;
- the positive differences in the mortality rates between men and women provide a plausible explanation for the higher attrition rates of men.

4.2 Survey design factors

Features of the survey design and the implementation of the fieldwork procedures may also be important determinants of the attrition probability.

According to the literature important factors are:

- target number of households to be approached in the field;
- length of the fieldwork period;
- number and socio-demographic characteristics of the interviewers;
- interviewers training;
- tracing rules and nonresponse conversion strategies;
- period of time elapsed between two successive waves of the panel.

Since these aspects of the survey design are not completely standardized across countries, they are likely to play an important role in explaining the cross-country variability of the attrition rates.

We exploit information from the SHARE Sample Management System (SMS) to investigate the relationship between attrition rates and the following indicator of fieldwork burden

$$F = \frac{N}{M I}$$

where:

- N is the total number of eligible households to be approached (target households);
- M is the number of months of the fieldwork period;
- I is the median number of interviewers available in each month.

Table 3: Fieldwork burden and attrition rates by country.

Country	Target households	Fieldwork length	Median # interviewers	Fieldwork burden	Attrition rate
AT	1326	11	28	4.38	0.35
BE	2912	13	41	5.46	0.27
CH	1507	11	38	3.61	0.31
DE	3508	10	45	7.88	0.49
DK	2436	12	45	4.56	0.27
ES	2446	12	43	4.74	0.43
FR	2180	4	72	7.57	0.37
GR	2931	9	64	5.09	0.21
IT	2777	12	40	5.86	0.31
NL	2977	9	59	5.65	0.40
SE	3246	15	34	6.36	0.34

Figure 1: Attrition rate and fieldwork burden by country.

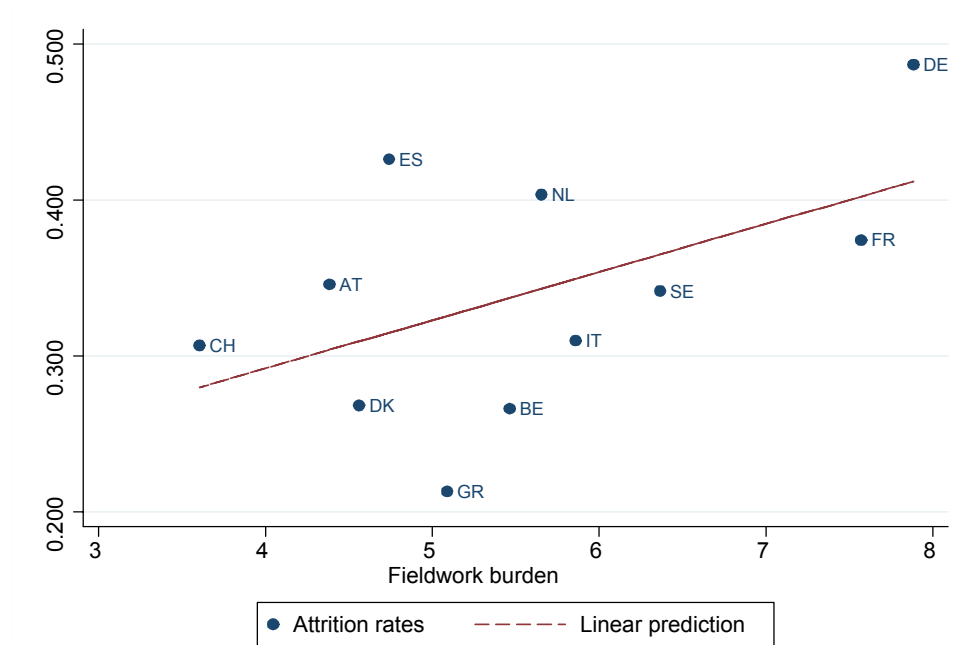


Figure 1 shows a positive correlation between attrition rates and our indicator of fieldwork burden ($\hat{\beta} = .03$ *s.d.* = .0173).

Given a target number of households, countries with a shorter fieldwork period and a lower number of interviewers in the field are likely to have higher attrition rates.

The effect is especially relevant for Germany and France where the indicator of fieldwork burden is much higher than the cross-country average.

4.3 Individual decisions of leaving the panel

Of the 4375 eligible households lost between the first and the second wave, 10% are due to non-contact and 90% are due to lack of cooperation. This suggests that reluctance to participate is likely to play an important role.

Lack of respondents' consent from wave 1

In Germany and Italy, a non negligible fraction of sample attrition (17 and 12 percent respectively) is due to the lack of respondents' consent to be re-approached in the next wave of the panel.

In these countries, privacy protection laws forbid survey agencies to contact again households where, at the end of the wave 1 interview, all respondents did not consent to be re-approached for the second wave.

Correlation with individual and household level characteristics

We estimate a simple probit model to investigate correlation between the attrition probability and several individual and household level characteristics.

We pool data from the various country and consider 7 model specifications:

- Specification A: country dummy, a 3rd-order polynomial in age, female dummy and its interaction with the linear term in age.
- Specification B: adds other socio-demographic characteristics (education, citizenship, location, marital status, household composition, employment status, eq. income).
- Specification C: adds cognitive functioning test scores (orientation in time, immediate and delayed recall, fluency, numeracy).
- Specification D: adds physical and mental health conditions (eurod, chronic diseases, gali, obese, mobility limitations, adl).
- Specification E: adds social activities (voluntary, cared sick, help family, sport, religious, political).
- Specification F: adds socio-demographic characteristics of the interviewer (age, gender and education).
- Specification G: adds features of the interview process in wave 1 (bed willingness to answer, often asked for clarifications, difficulties understanding questions, interview length).

Table 4: Probit estimates for the attrition probability. Sample size 25457.

Variable	A	B	C	D	E	F	G
constant	1.789**	1.921**	1.837**	1.469*	1.571**	1.646**	1.617**
AT	0.222**	0.219**	0.255**	0.241**	0.225**	0.206**	0.132**
CH	0.106*	0.091+	0.118*	0.105*	0.102*	0.071	0.066
DE	0.582**	0.567**	0.584**	0.580**	0.563**	0.563**	0.519**
DK	-0.044	-0.042	-0.021	-0.025	-0.021	-0.008	-0.003
ES	0.436**	0.413**	0.345**	0.351**	0.322**	0.306**	0.250**
FR	0.263**	0.248**	0.243**	0.241**	0.236**	0.241**	0.217**
GR	-0.191**	-0.226**	-0.270**	-0.278**	-0.291**	-0.369**	-0.367**
IT	0.152**	0.166**	0.094*	0.093*	0.068+	0.050	0.021
NL	0.379**	0.343**	0.360**	0.353**	0.366**	0.336**	0.329**
SE	0.216**	0.183**	0.240**	0.235**	0.247**	0.239**	0.220**
female	-0.049**	-0.058**	-0.047*	-0.057**	-0.058**	-0.058**	-0.059**
age	0.095**	0.102**	0.090**	0.071*	0.075*	0.077*	0.080*
age ²	0.001	0.001+	0.001	0.001	0.001	0.001	0.001
education med.		-0.032	0.016	0.009	0.017	0.018	0.025
education hig.		-0.124**	-0.038	-0.050+	-0.026	-0.023	-0.015
noncitizen		0.182**	0.141**	0.145**	0.134**	0.138**	0.132**
big city		0.170**	0.169**	0.170**	0.160**	0.159**	0.159**
single		-0.063**	-0.074**	-0.073**	-0.068**	-0.066**	-0.051*
hsize		-0.029**	-0.030**	-0.031**	-0.029**	-0.029**	-0.030**
children		-0.051	-0.051	-0.051	-0.050	-0.052	-0.028
working		0.006	0.017	0.013	0.002	0.001	-0.001
unemployed		-0.058	-0.066	-0.061	-0.065	-0.065	-0.067
sick		0.168**	0.135**	0.123*	0.109*	0.109*	0.106*
homemaker		0.008	0.005	0.006	0.009	0.010	-0.001
income		0.023*	0.031**	0.030**	0.033**	0.034**	0.034**
income ²		0.005+	0.005+	0.005+	0.005+	0.005+	0.005
orientation			-0.582**	-0.537**	-0.543**	-0.541**	-0.481**
recall imm.			-0.027**	-0.027**	-0.026**	-0.025**	-0.020**
numeracy			-0.009	-0.010	-0.009	-0.008	0.003
fluency			-0.012**	-0.012**	-0.011**	-0.011**	-0.010**
recall del.			-0.008	-0.008	-0.007	-0.008	-0.007
eurod				-0.008+	-0.008+	-0.009+	-0.010*
chronic				-0.028**	-0.027**	-0.027**	-0.019**
gali				-0.015	-0.018	-0.017	-0.022
obese				-0.125**	-0.124**	-0.127**	-0.115**
mobility				0.026**	0.025**	0.025**	0.024**
adl				0.012	0.009	0.009	0.007
voluntary					-0.121**	-0.120**	-0.108**
cared sick					-0.013	-0.013	0.004
help family					-0.087**	-0.085**	-0.060**
sport club					-0.042+	-0.041+	-0.033
political					-0.083+	-0.081+	-0.077+
iv. female						-0.023	-0.012
iv. age						0.003+	0.004**
iv. age ²						0.000**	0.000**
iv. edu. med.						-0.016	-0.028
iv. edu. hig.						-0.075*	-0.085**
bed will. answ.							0.373**
clarification							0.065+
understanding							-0.028
int. length							-0.003**
int. length ²							0.000**
Pseudo R ²	.026	.030	.036	.038	.040	.041	.049
AIC	31302	31203	31030	30979	30925	30914	30683
BIC	31433	31456	31340	31353	31349	31378	31196

Main findings:

- **Country dummies:** even after controlling for a large set of covariates, unobserved cross country heterogeneity still plays an important role;
- **Gender:** women are more likely to participate than men.
- **Age:** the attrition probability is positively related to respondent's age.
- **Other socio-demographic variables:** citizenship, living a big city, marital status, household size, income are strong predictors of sample attrition.
- **Cognitive functions:** people with poor cognitive abilities are more likely to drop out from the panel.
- **Health:** The evidence for health is less clear-cut. The attrition probability is positively related to self-reported health and mobility limitations, but negatively related to chronic diseases, depression and obesity.
- **Social activities:** people more involved in social activities are more likely to participate.
- **Interviewers:** attrition is positively associated with the interviewers' age and negatively associated with interviewers' education.
- **Interview process:** attrition is positively associated with the burden of the interview process in wave 1.

5 Conclusion

Our empirical evidence suggests that:

- the MCAR assumption is likely to be rejected because attrition is related to several time-invariant variables and other key survey variables collected in wave 1;
- the MAR assumption could be reasonable if we condition on a large set of covariates (weighting strategies and imputation procedures). As pointed out by Heckman and Navarro (2004), these procedures offer little guidance on how to pick the variables that should account for selection, and are not robust to the choice of the conditioning set.
- we also have a good setting to analyze NMAR attrition (sample selection models). Features of the data collection process and socio-demographic characteristics of the interviewers seem to provide a suitable set of exclusion restrictions (Fitzgerald *et al.* 1998, Nicoletti & Peracchi 2005).