

SOEP Survey Papers

Series C – Data Documentation

SOEP – The German Socio-Economic Panel study at DIW Berlin

2015

Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2014)

Martin Kroh, Simon Kühne, Rainer Siegers

Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at
<http://www.diw.de/soepsurveypapers>

Editors:

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Martin Kroh, Simon Kühne, Rainer Siegers. 2015. Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2014). SOEP Survey Papers 297: Series C. Berlin: DIW Berlin / SOEP



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). © 2015 by SOEP

ISSN: 2193-5580 (online)

German Socio-Economic Panel (SOEP) | DIW Berlin
Mohrenstr. 58
10117 Berlin, Germany
Contact: soeppapers@diw.de

Data Documentation:

Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2014)

Martin Kroh Simon Kühne Rainer Siegers

November 13, 2015

DIW Berlin – Deutsches Institut für Wirtschaftsforschung e.V.
Sozio-oekonomisches Panel (SOEP)
Mohrenstraße 58
10117 Berlin

Contents

1	Introduction	6
2	Developments in Sample Size	7
2.1	Development of the Number of Successful Interviews by Cross-Section	7
2.2	Continuance and Exit: The First Wave Gross Samples and their Participatory Behavior	20
2.3	New Entrants through Birth or Move into SOEP Households and Their Participation Behavior	27
2.4	The Risk of Survey-Related Panel Attrition	34
3	Panel Attrition Due to Unsuccessful Follow-Ups	39
3.1	The Frequency of Successful Follow-Ups	39
3.2	Predicting the Probability of Successful vs. Unsuccessful Follow-Ups in the Year 2014	41
4	Panel Attrition Due to Refusals	45
4.1	The Frequency of Participation	45
4.2	Predicting the Probability of Re-Interviewing versus Refusal in the Year 2014 .	47
5	Summary Statistics of the Derived Longitudinal and Cross-Sectional Weights	56

List of Tables

3.1	The Frequency of Households to be Re-Contacted and the Percentage of Successful Follow-Ups, Subsamples A to M by Year.	40
3.2	Definition of the Regressors of the Logit Model of Unsuccessful Follow-Ups	42
3.3	Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2014	43
3.3	Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2014	44
4.1	The Frequency of Re-Contacted Households and the Percentage of Participation, Subsamples A to M by Year.	46
4.2	Definition of the Regressors of the Logit Model of Refusal	48
4.2	Definition of the Regressors of the Logit Model of Refusal	49
4.2	Definition of the Regressors of the Logit Model of Refusal	50
4.2	Definition of the Regressors of the Logit Model of Refusal	51
4.3	Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2014	52
4.3	Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2014	53
4.3	Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2014	54
4.3	Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2014	55
5.1	Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples A through D (Percentiles of \$HBLEIB up to Wave BE).	57
5.2	Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples E through G (Percentiles of \$HBLEIB up to Wave BE).	58
5.3	Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples H, J, K and M (Percentiles of \$HBLEIB up to Wave BE).	59
5.4	Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples L1, L2 and L3 (Percentiles of \$HBLEIB up to Wave BE).	59
5.5	Summary Statistics of the Derived Cross-Sectional Weights at the Household Level (Percentiles of \$HHRF up to Wave 31).	60
5.6	Summary Statistics of the Derived Cross-Sectional Weights at the Person Level (Percentiles of \$PHRF up to Wave 31).	61

List of Figures

1	The Number of Successful Interviews with Persons by Subsamples A through M, Waves 1 to 31.	7
2	Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 31	8
3	Comparison of Successful Interviews with Persons and Households (Subsamples C), Waves 1 to 25	9
4	Comparison of Successful Interviews with Persons and Households (Subsamples D), Waves 1 to 20	10
5	Comparison of Successful Interviews with Persons and Households (Subsamples E), Waves 1 to 17	11
6	Comparison of Successful Interviews with Persons and Households (Subsamples F), Waves 1 to 15	12
7	Comparison of Successful Interviews with Persons and Households (Subsamples G), Waves 1 to 13	13
8	Comparison of Successful Interviews with Persons and Households (Subsamples H), Waves 1 to 9	14
9	Comparison of Successful Interviews with Persons and Households (Subsample J), Waves 1 to 4	15
10	Comparison of Successful Interviews with Persons and Households (Subsample K), Waves 1 to 3	16
11	Comparison of Successful Interviews with Persons and Households (Subsample L1), Waves 1 to 5	17
12	Comparison of Successful Interviews with Persons and Households (Subsample L2), Waves 1 to 5	18
13	Comparison of Successful Interviews with Persons and Households (Subsample L3), Waves 1 to 4	19
14	All First-Wave Persons (Gross Subsample A). Development up to Wave 31	20
15	All First-Wave Persons (Gross Subsample B). Development up to Wave 31	21
16	All First-Wave Persons (Gross Subsample C). Development up to Wave 25	21
17	All First-Wave Persons (Gross Subsample D). Development up to Wave 20	22
18	All First-Wave Persons (Gross Subsample E). Development up to Wave 17	22
19	All First-Wave Persons (Gross Subsample F). Development up to Wave 15	23
20	All First-Wave Persons (Gross Subsample G). Development up to Wave 13	23
21	All First-Wave Persons (Gross Subsample H). Development up to Wave 9	24
22	All First-Wave Persons (Gross Subsample J). Development up to Wave 4	24
23	All First-Wave Persons (Gross Subsample K). Development up to Wave 3	25
24	All First-Wave Persons (Gross Subsample L1). Development up to Wave 5	25
25	All First-Wave Persons (Gross Subsample L2). Development up to Wave 5	26
26	All First-Wave Persons (Gross Subsample L3). Development up to Wave 4	26
27	Entrants and their Participation Behavior (Subsample A)	27
28	Entrants and their Participation Behavior (Subsample B)	28
29	Entrants and their Participation Behavior (Subsample C)	28
30	Entrants and their Participation Behavior (Subsample D)	29
31	Entrants and their Participation Behavior (Subsample E)	29

32	Entrants and their Participation Behavior (Subsample F)	30
33	Entrants and their Participation Behavior (Subsample G)	30
34	Entrants and their Participation Behavior (Subsample H)	31
35	Entrants and their Participation Behavior (Subsample J)	31
36	Entrants and their Participation Behavior (Subsample K)	32
37	Entrants and their Participation Behavior (Subsample L1)	32
38	Entrants and their Participation Behavior (Subsample L2)	33
39	Entrants and their Participation Behavior (Subsample L3)	33
40	Successful Re-Interviewing of First-Wave Respondents by Subsamples A, B, C. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	34
41	Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	35
42	Successful Re-Interviewing of First-Wave Respondents by Subsamples G, H, J and K. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	35
43	Successful Re-Interviewing of First-Wave Respondents by Subsamples L1, L2 and L3. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	36
44	Successful Re-Interviewing of All First-Wave Respondents by Age Categories. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	36
45	Successful Re-Interviewing of All First-Wave Respondents by Occupation. Kaplan- Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	37
46	Successful Re-Interviewing of All First-Wave Respondents by Income Quin- tiles. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	37
47	Successful Re-Interviewing of All First-Wave Respondents by Education. Kaplan- Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad	38

1 Introduction

This data documentation¹ is meant to provide SOEP users with a general overview of the longitudinal development of the survey over the past 31 years and the derivation of weights that compensate for selective panel attrition. In the first section, we report the number of household and personal interviews by cross-section. We do so for the entire SOEP sample as a whole, as well as for sub-samples A through K individually, the migration sample M and the recently integrated samples L1, L2, and L3, boost samples of specific family types. For a general overview on the integration of refreshment samples into the SOEP see Kroh et al. (2015).

The SOEP study surveys not only the original sample from the first wave, but also households and persons that entered the survey at later points in time. They enter, for example, when SOEP households split (i.e., individuals move out and form their own households), when people move into SOEP households, and when an original sample member gives birth to a “new sample member”. For a detailed review of the SOEP inclusion rules for new sample units and their treatment within the weighting framework see Spiess et al. (2008) and Schonlau et al. (2010). The second section of the present paper on the longitudinal development of the SOEP reports descriptive figures of the participatory behavior of the original sample members and the entrance patterns of new sample members.

Households may leave the survey for several reasons. SOEP’s weighting strategy distinguishes between survey-related reasons and reasons unrelated to the survey (for a detailed description of the SOEP weighting strategy, see Rendtel (1995); Schonlau et al. (2013) and for a general overview, Haisken-DeNew and Frick (2001)). We ignore panel attrition of the latter form due to respondents moving abroad or dying, since these cases technically represent an exit from the underlying population. The second section of this paper provides initial evidence on the risk of survey-related panel attrition in different groups of the original sample units (e.g., in different sub-samples, age, educational, and income groups).

The third section reports in more detail on the occurrence of unsuccessful follow-ups to household addresses by cross-section and sub-sample, and sub-sample-specific regression models of the probability of unsuccessful follow-ups in 2014 based on the characteristics of households measured in 2013. The fourth section does the same for the second form of survey-related attrition: refusals.

Based on the regression models of unsuccessful follow ups and refusals, we derive predicted observation probabilities. The inverse of the product of these predicted probabilities gives the longitudinal weighting variables for the year 2014: BEHBLEIB and BEPBLEIB. Based on the inverse of the probability of observing households and persons in 2013, the staying probability in 2014, and additional post-stratification to meet benchmarks of known marginals of the underlying population in 2014, we derive the cross-sectional weights BEHHRF and BEPHRF. The final section of this paper documents some summary statistics of the development of the longitudinal and the cross-sectional weights by sub-sample and wave.

¹We would like to thank Jan-Lucas Schanze and Alejandra Rodriguez Sanchez for their help and contributions.

2 Developments in Sample Size

With respect to developments in sample size, the following figures focus on (2.1) comparing the number of successful interviews by cross-section, (2.2) providing a longitudinal study of panel attrition among the original sample members, (2.3) showing entrance of new sample members by birth / moving into SOEP households and their participation behavior, and (2.4) assessing the risk of survey-related attrition of original sample respondents by social characteristics.

Note that the sample sizes of the English public use version of SOEP and the German DIW version differ by approximately 5 percent. This percentage of the original SOEP data was excluded in compliance with German data protection laws, which was accomplished technically by randomly selecting 5 percent of the first wave households and dropping these and the persons living in them from the English public-use version. Hence the difference in sample sizes is not always exactly 5 percent. The sample sizes documented below refer to the original DIW database.

2.1 Development of the Number of Successful Interviews by Cross-Section

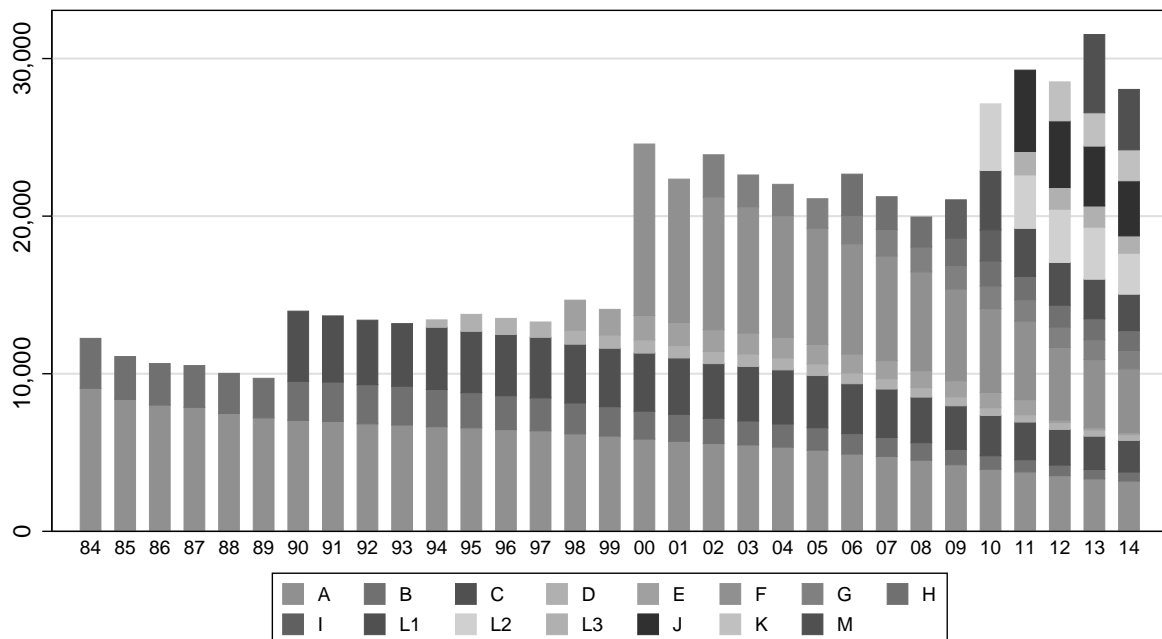


Figure 1: The Number of Successful Interviews with Persons by Subsamples A through M, Waves 1 to 31.

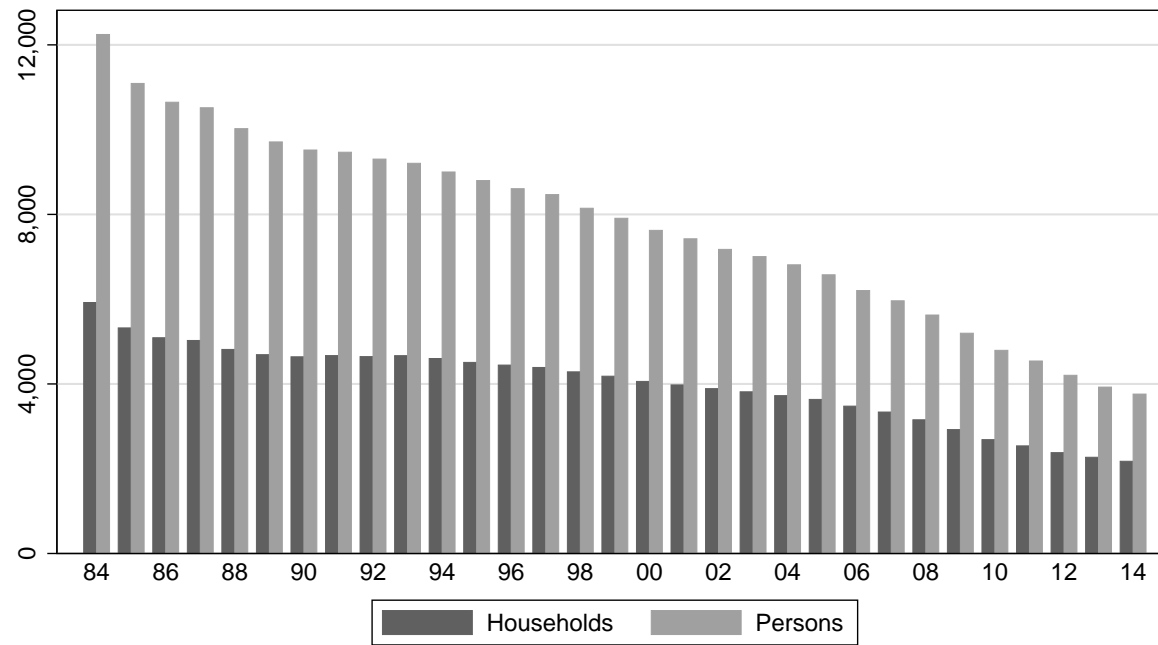


Figure 2: Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 31

Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Persons	12245	11090	10646	10516	10023	9710	9519	9467	9305	9206	9001	8798	8606	8467	8145
Households	5921	5322	5090	5026	4814	4690	4640	4669	4645	4667	4600	4508	4445	4389	4285

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	7909	7623	7424	7175	7004	6811	6575	6203	5961	5626	5197	4793	4541	4204	3926	3761
Households	4183	4060	3977	3889	3814	3724	3635	3476	3337	3154	2923	2686	2539	2379	2270	2176

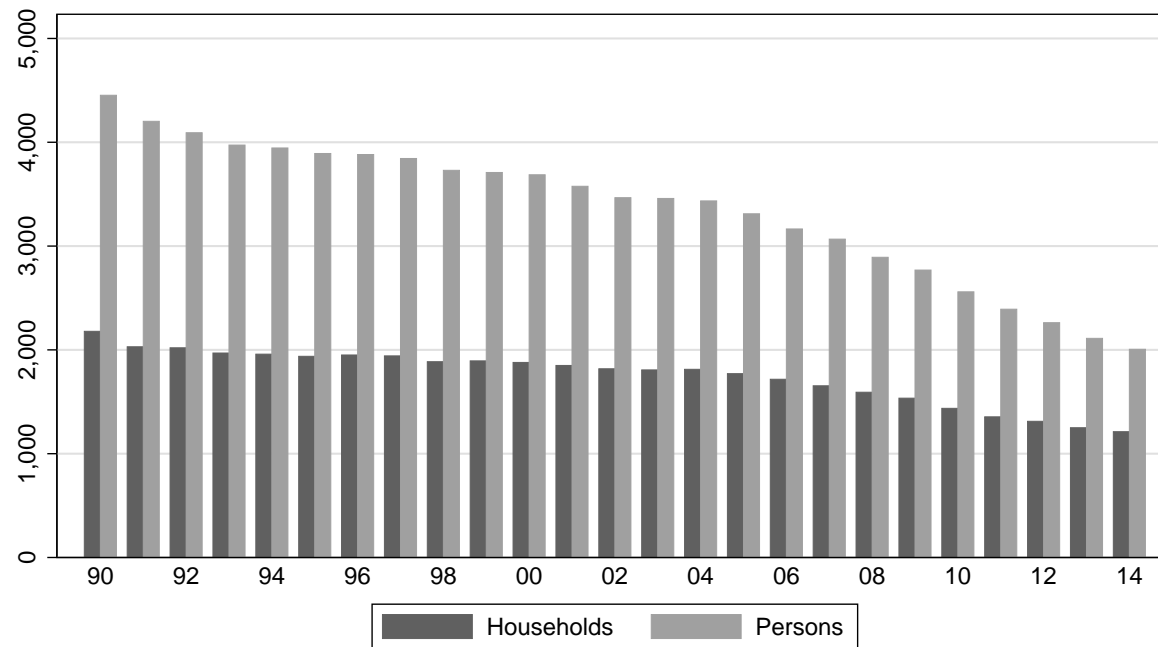


Figure 3: Comparison of Successful Interviews with Persons and Households (Subsamples C), Waves 1 to 25

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Persons	4453	4202	4092	3973	3945	3892	3882	3844	3730	3709	3687	3576
Households	2179	2030	2020	1970	1959	1938	1951	1942	1886	1894	1879	1850

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	3466	3459	3435	3311	3165	3067	2892	2769	2559	2392	2262	2111	2006
Households	1818	1807	1813	1771	1717	1654	1592	1535	1437	1355	1312	1250	1212

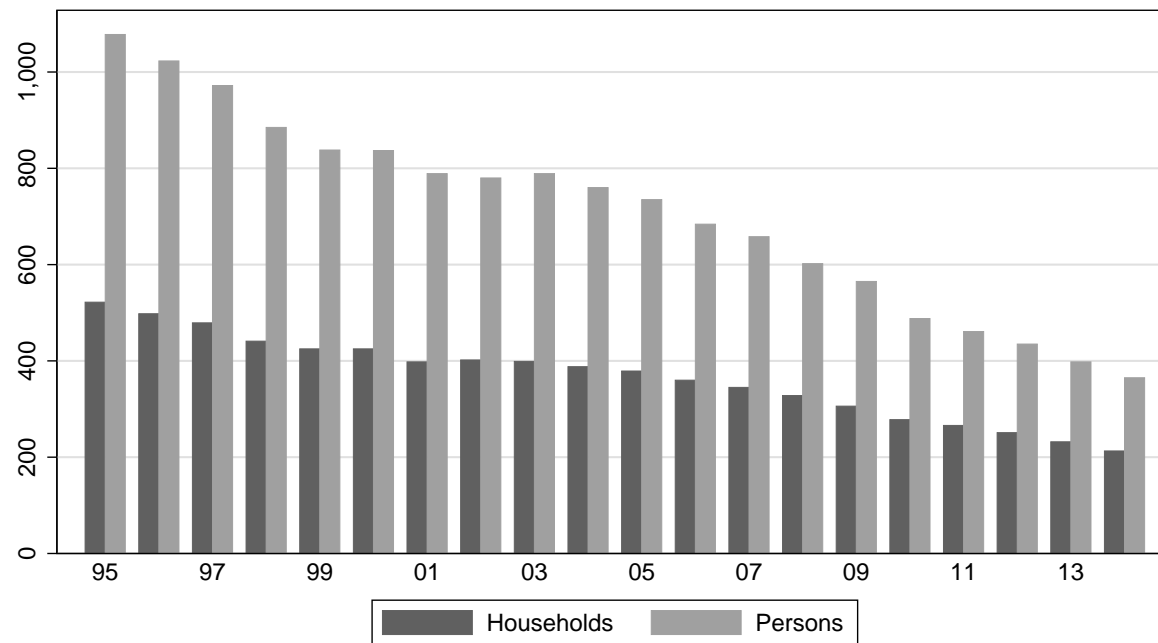


Figure 4: Comparison of Successful Interviews with Persons and Households (Subsamples D), Waves 1 to 20

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	1078	1023	972	885	838	837	789	780	789	760	735	684	658	602	565	488	461	435	398	365
Households	522	498	479	441	425	425	398	402	399	388	379	360	345	328	306	278	266	251	232	213

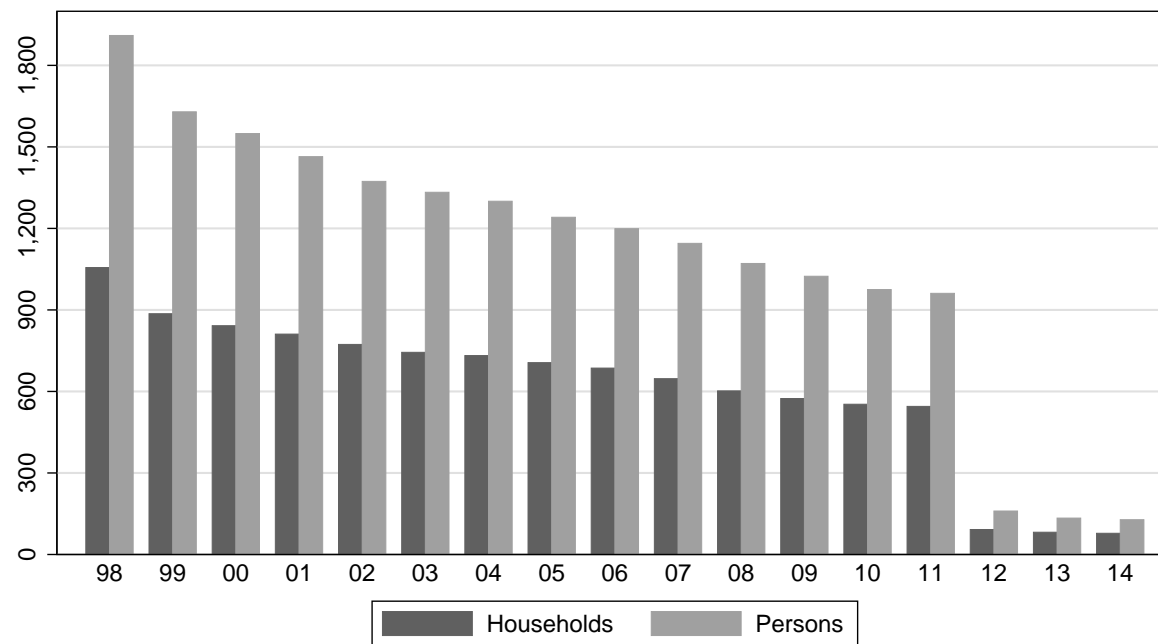


Figure 5: Comparison of Successful Interviews with Persons and Households (Subsamples E), Waves 1 to 17

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	1910	1629	1549	1464	1373	1333	1300	1241	1199	1145	1071	1024	975	961	160	134	128
Households	1056	886	842	811	773	744	732	706	686	647	602	574	553	545	92	82	78

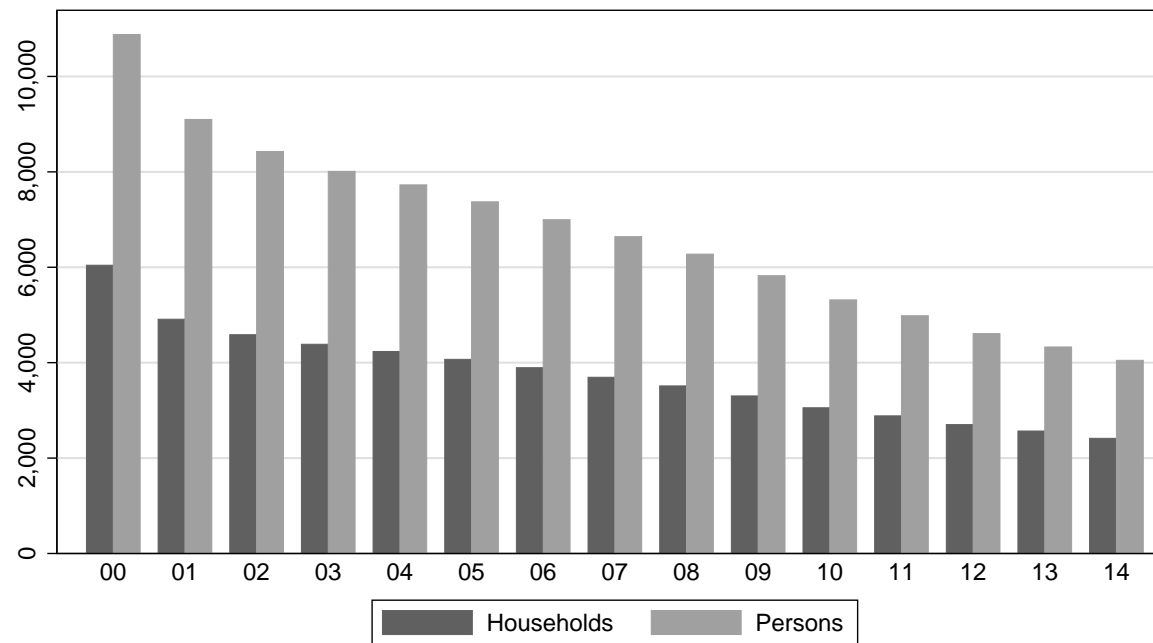


Figure 6: Comparison of Successful Interviews with Persons and Households (Subsamples F), Waves 1 to 15

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	10880	9098	8427	8010	7727	7372	6997	6642	6276	5824	5316	4984	4610	4329	4049
Households	6043	4911	4586	4386	4235	4070	3895	3694	3513	3303	3055	2885	2702	2567	2414

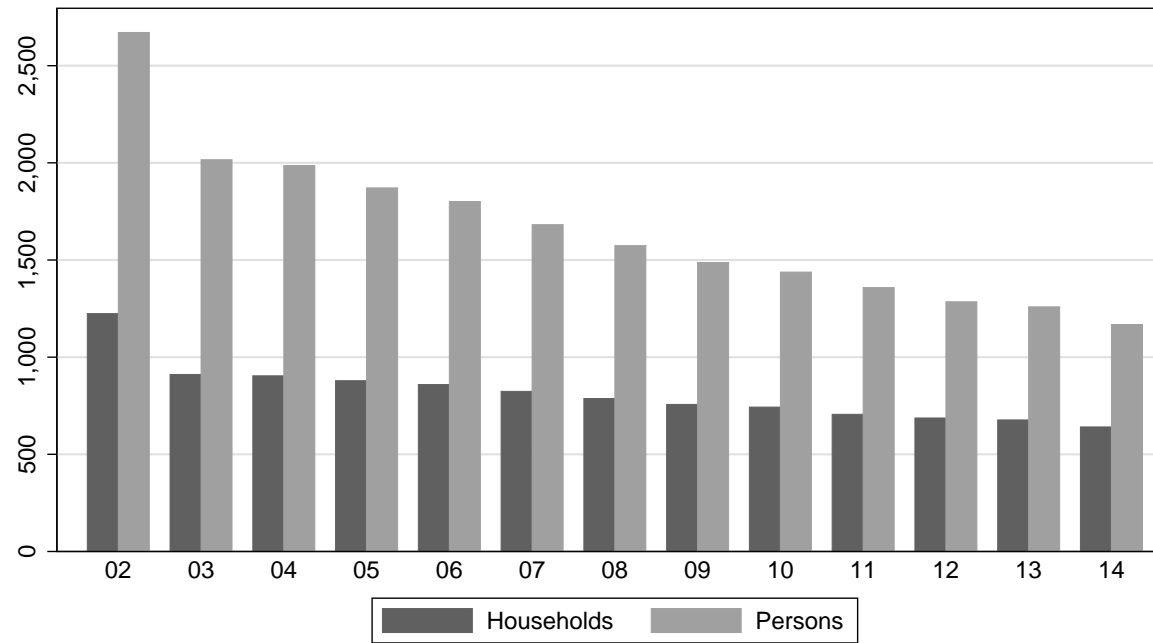


Figure 7: Comparison of Successful Interviews with Persons and Households (Subsamples G), Waves 1 to 13

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	2671	2016	1986	1871	1801	1682	1574	1487	1438	1358	1285	1259	1168
Households	1224	911	904	879	859	824	787	757	743	706	687	677	641

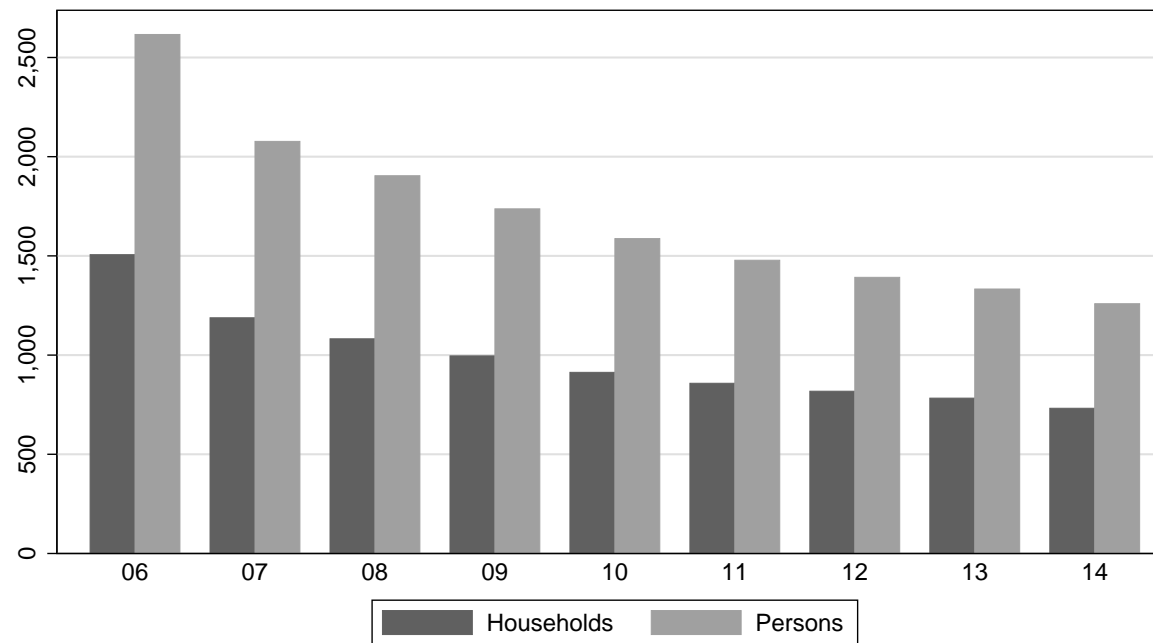


Figure 8: Comparison of Successful Interviews with Persons and Households (Subsamples H), Waves 1 to 9

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014
Persons	2616	2077	1904	1737	1587	1478	1392	1333	1259
Households	1506	1188	1082	996	913	858	818	783	732

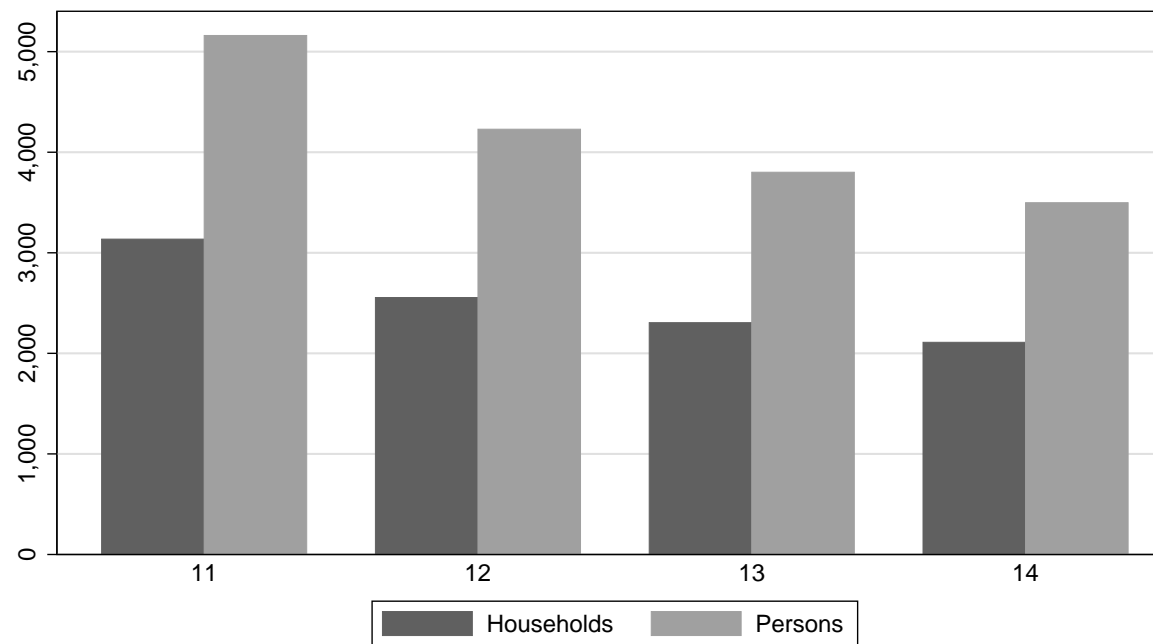


Figure 9: Comparison of Successful Interviews with Persons and Households (Subsample J), Waves 1 to 4

Year	2011	2012	2013	2014
Persons	5161	4229	3801	3498
Households	3136	2555	2305	2110

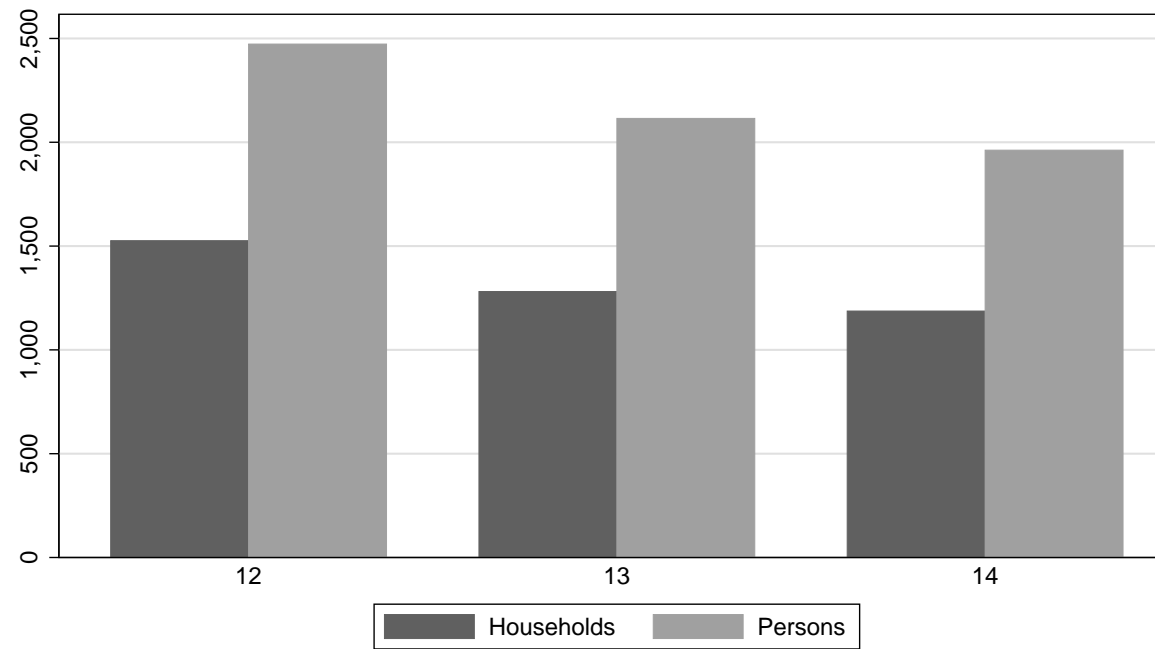


Figure 10: Comparison of Successful Interviews with Persons and Households (Subsample K), Waves 1 to 3

Year	2012	2013	2014
Persons	2473	2115	1962
Households	1256	1281	1187

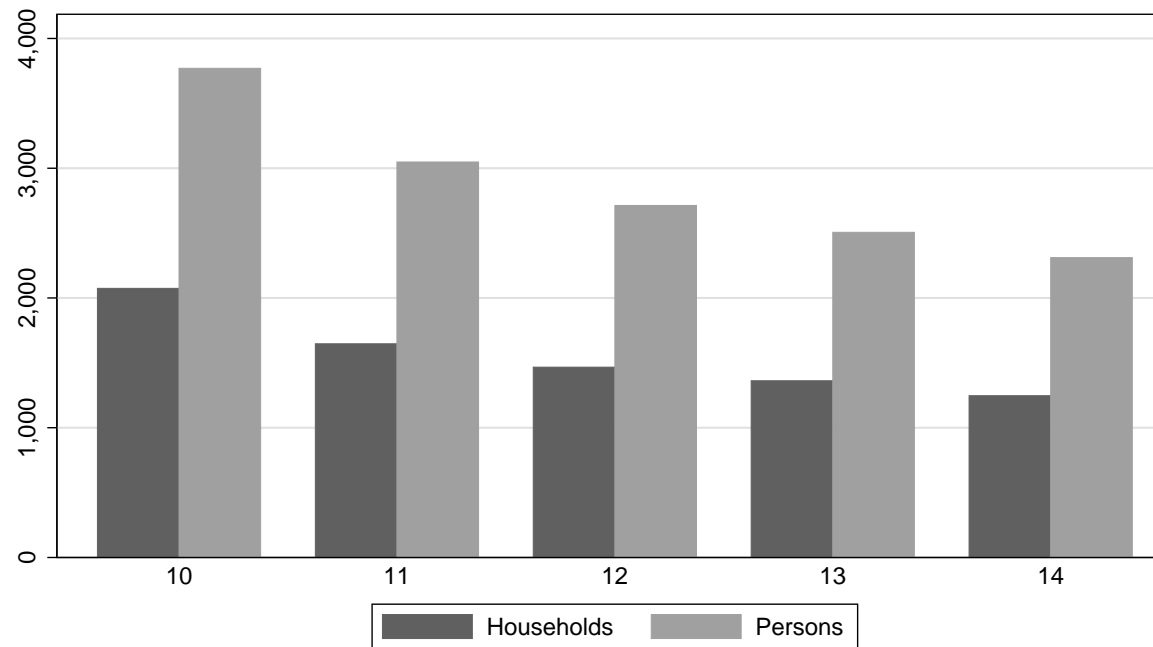


Figure 11: Comparison of Successful Interviews with Persons and Households (Subsample L1), Waves 1 to 5

Year	2010	2011	2012	2013	2014
Persons	3770	3048	2713	2506	2311
Households	2074	1647	1467	1362	1247

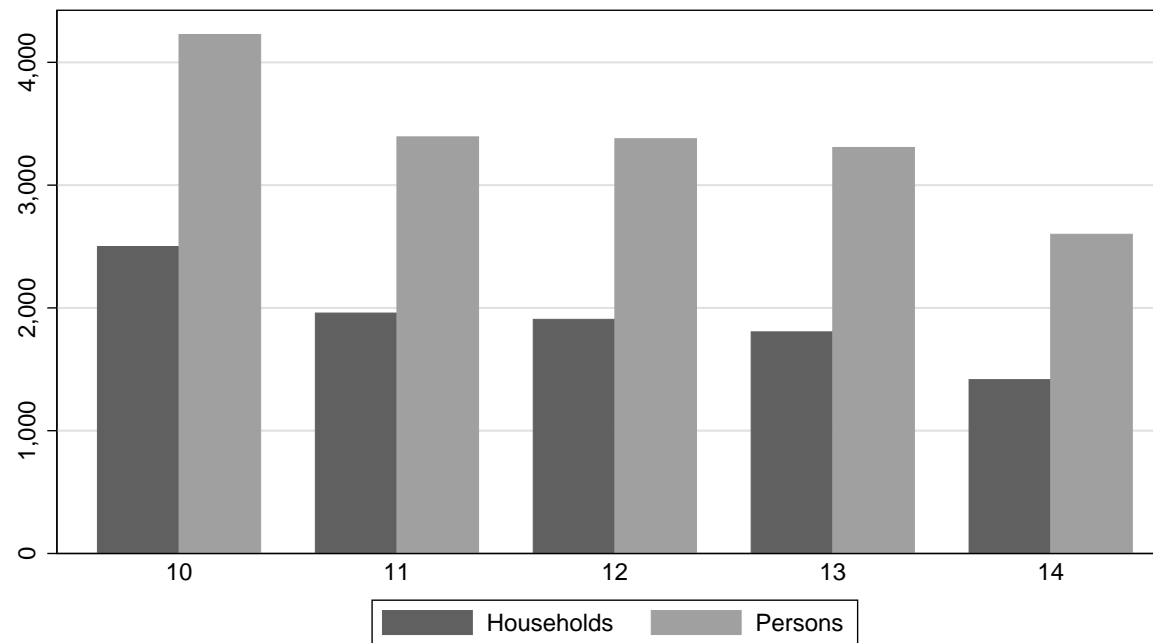


Figure 12: Comparison of Successful Interviews with Persons and Households (Subsample L2), Waves 1 to 5

Year	2010	2011	2012	2013	2014
Persons	4227	3393	3378	3307	2600
Households	2500	1958	1907	1805	1416

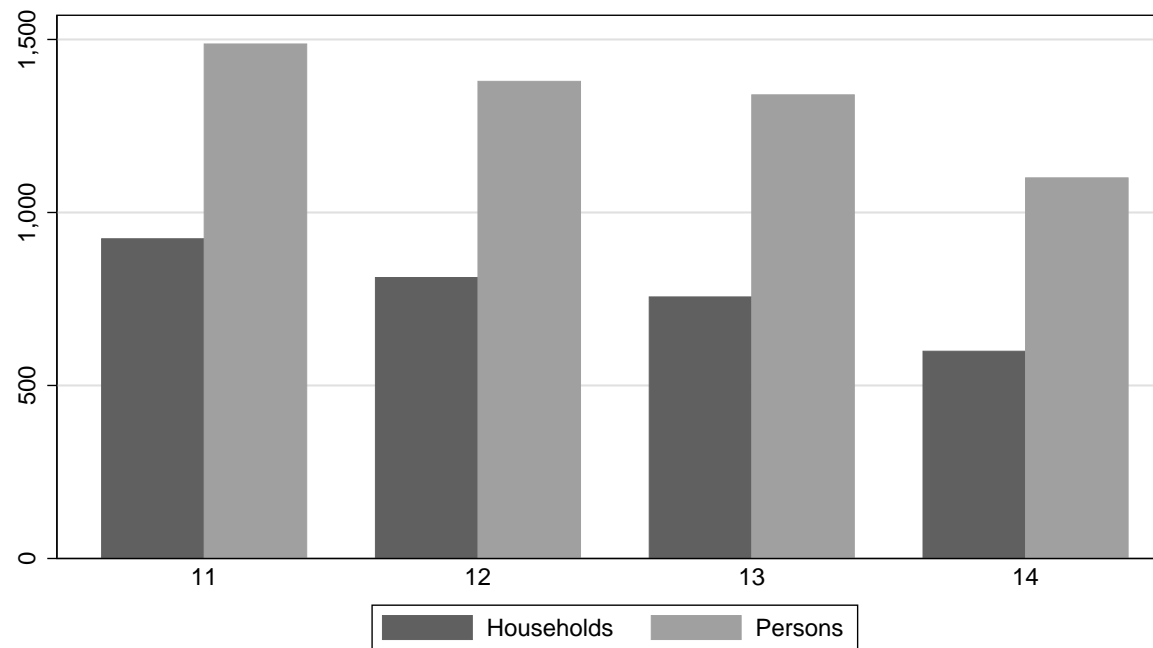


Figure 13: Comparison of Successful Interviews with Persons and Households (Subsample L3), Waves 1 to 4

Year	2011	2012	2013	2014
Persons	1487	1379	1340	1100
Households	924	812	756	599

2.2 Continuance and Exit: The First Wave Gross Samples and their Participatory Behavior

The following figures display the participation behavior of the first-wave respondents in the subsequent years distinguishing between continued participation (“With interview”), exits due to survey-unrelated attrition (“Moved abroad”, “Deceased”, “Under the age of 16”), and exits due to survey-related attrition (“Temporary drop-out”, “Drop-out”).

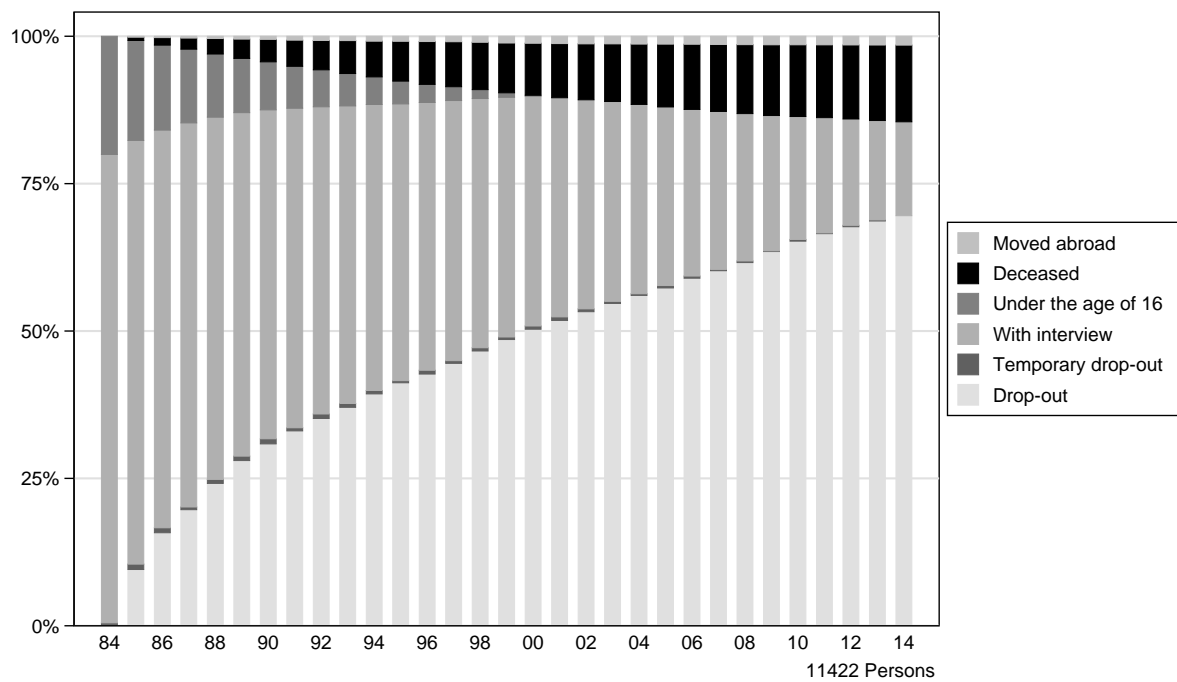


Figure 14: All First-Wave Persons (Gross Subsample A). Development up to Wave 31

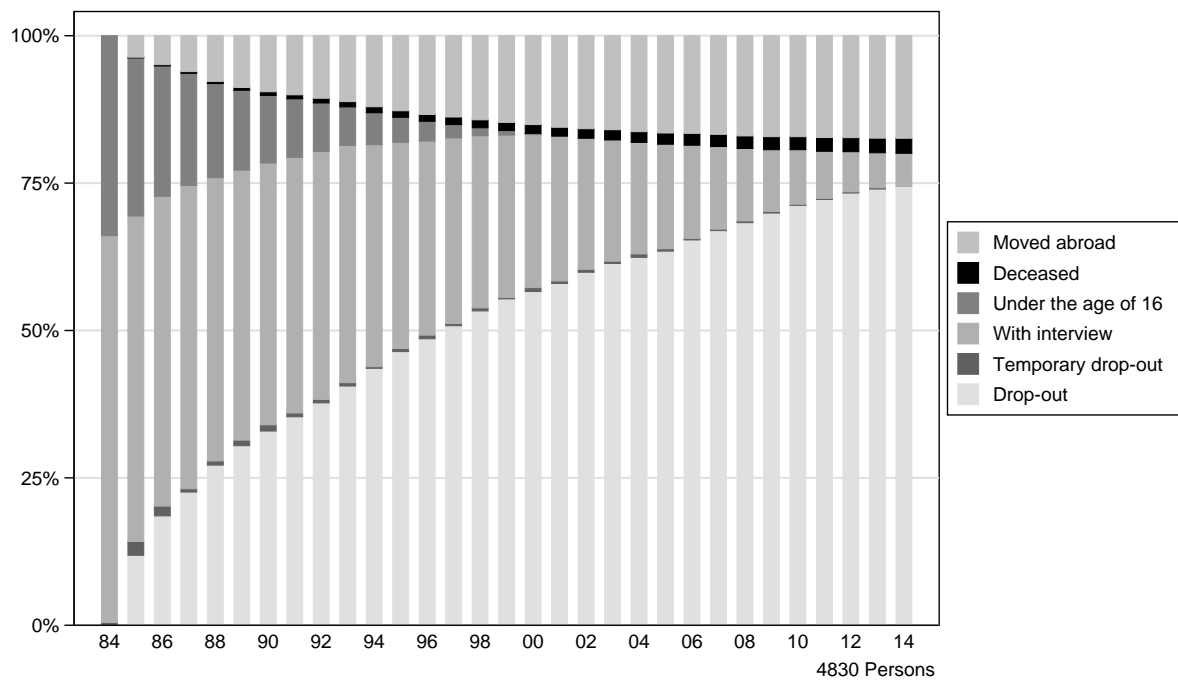


Figure 15: All First-Wave Persons (Gross Subsample B). Development up to Wave 31

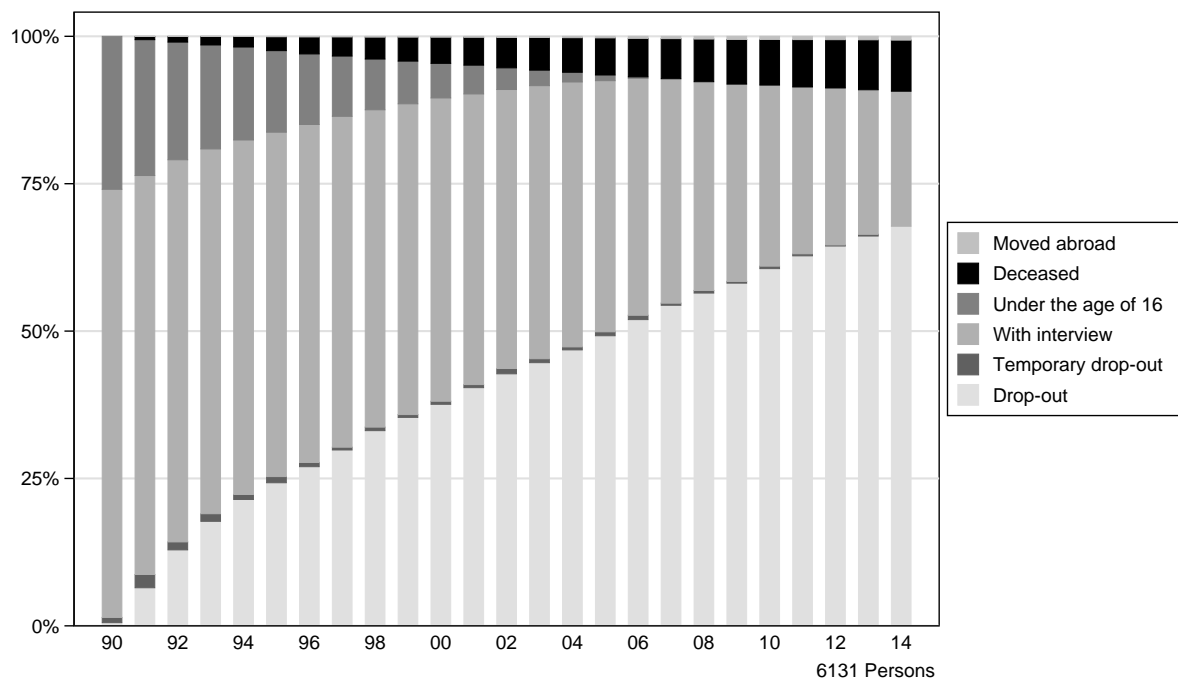


Figure 16: All First-Wave Persons (Gross Subsample C). Development up to Wave 25

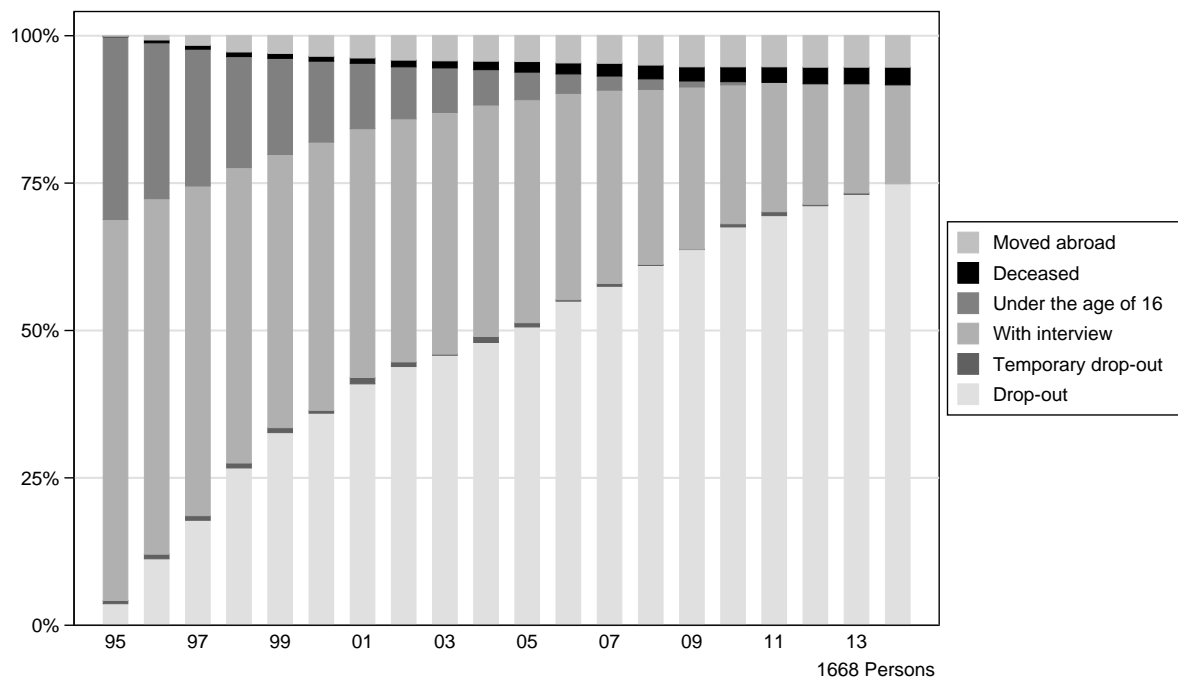


Figure 17: All First-Wave Persons (Gross Subsample D). Development up to Wave 20

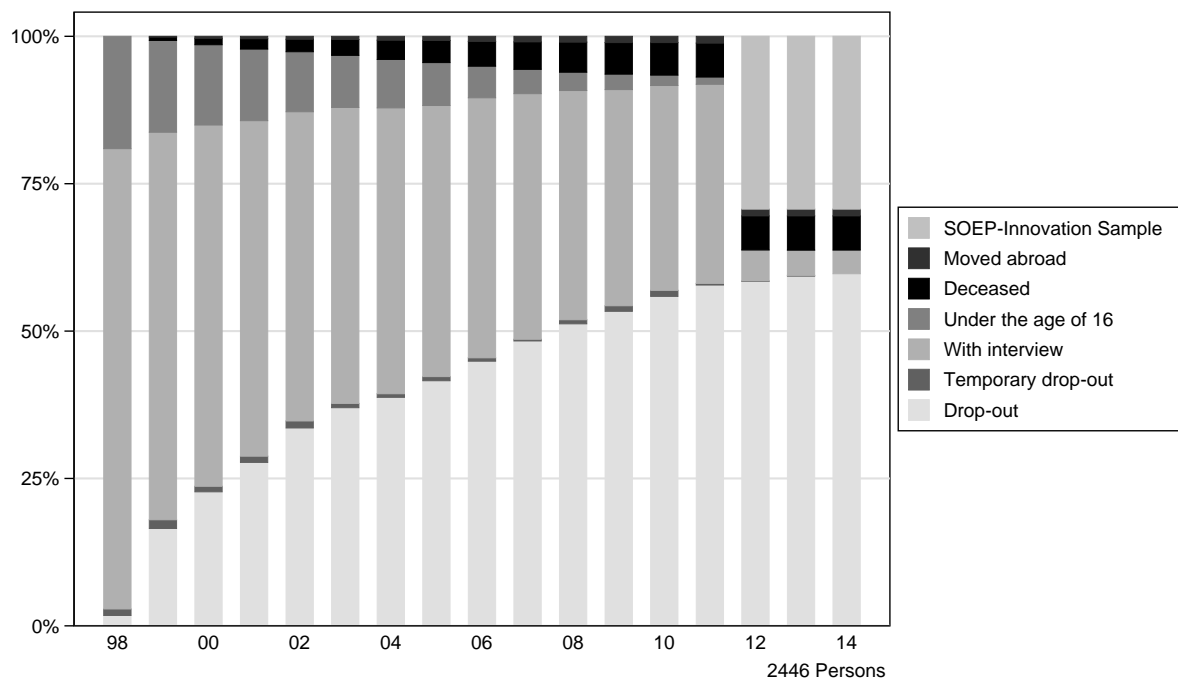


Figure 18: All First-Wave Persons (Gross Subsample E). Development up to Wave 17

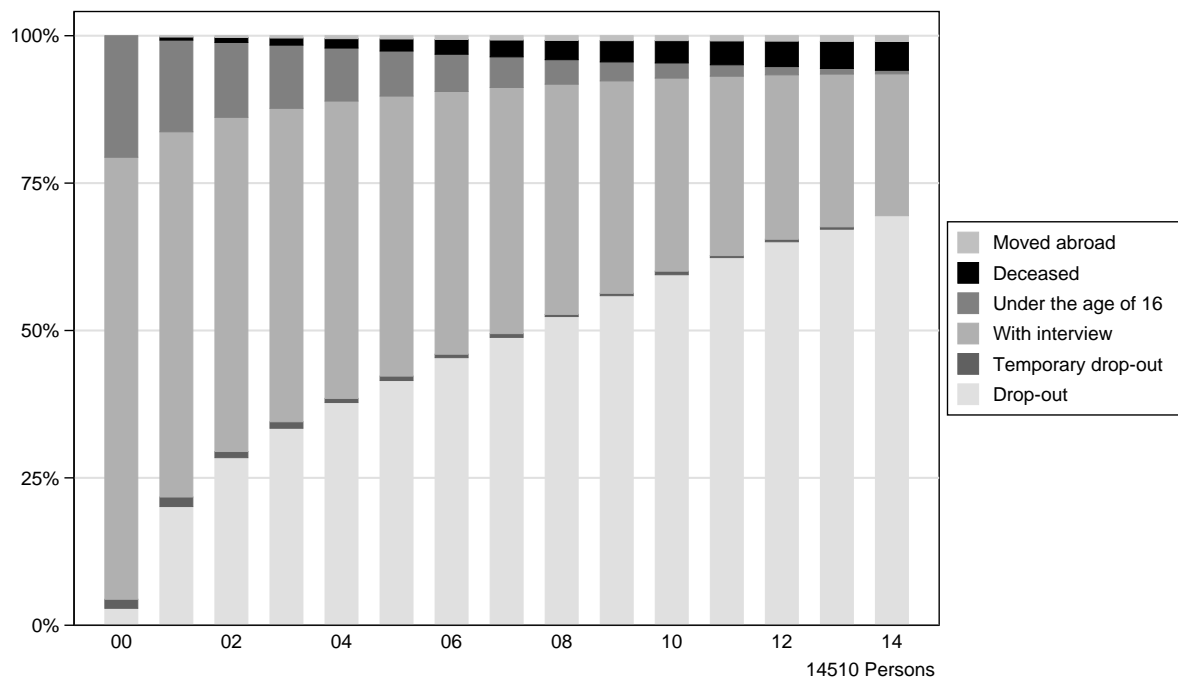


Figure 19: All First-Wave Persons (Gross Subsample F). Development up to Wave 15

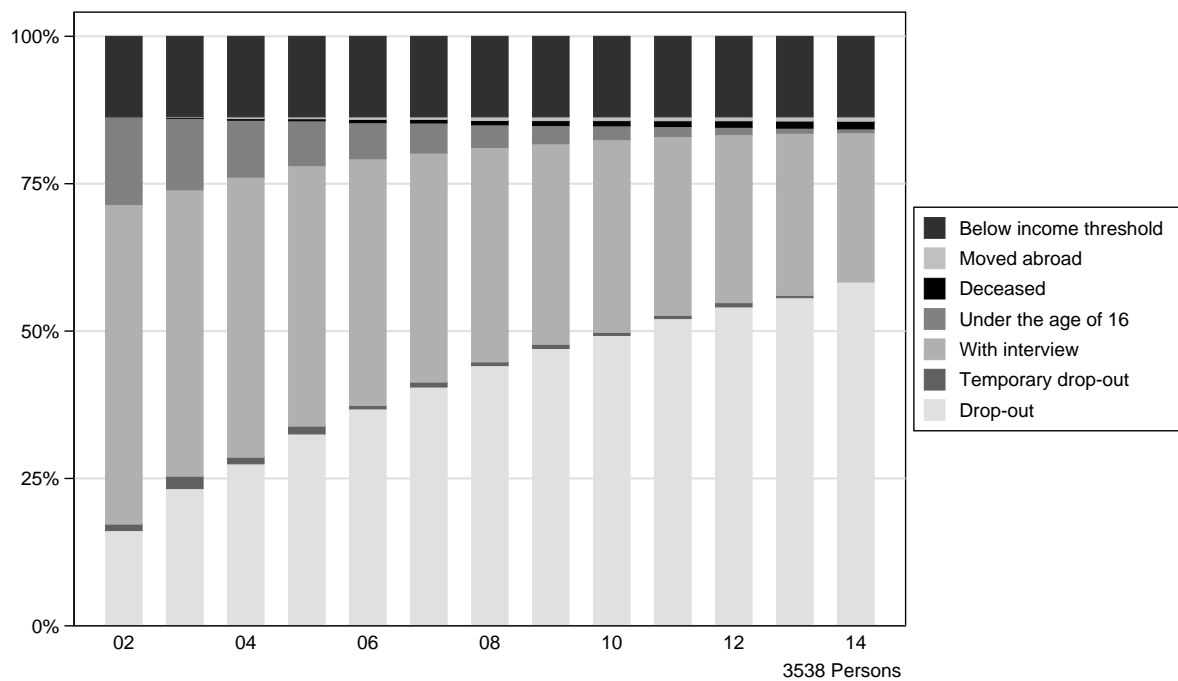


Figure 20: All First-Wave Persons (Gross Subsample G). Development up to Wave 13

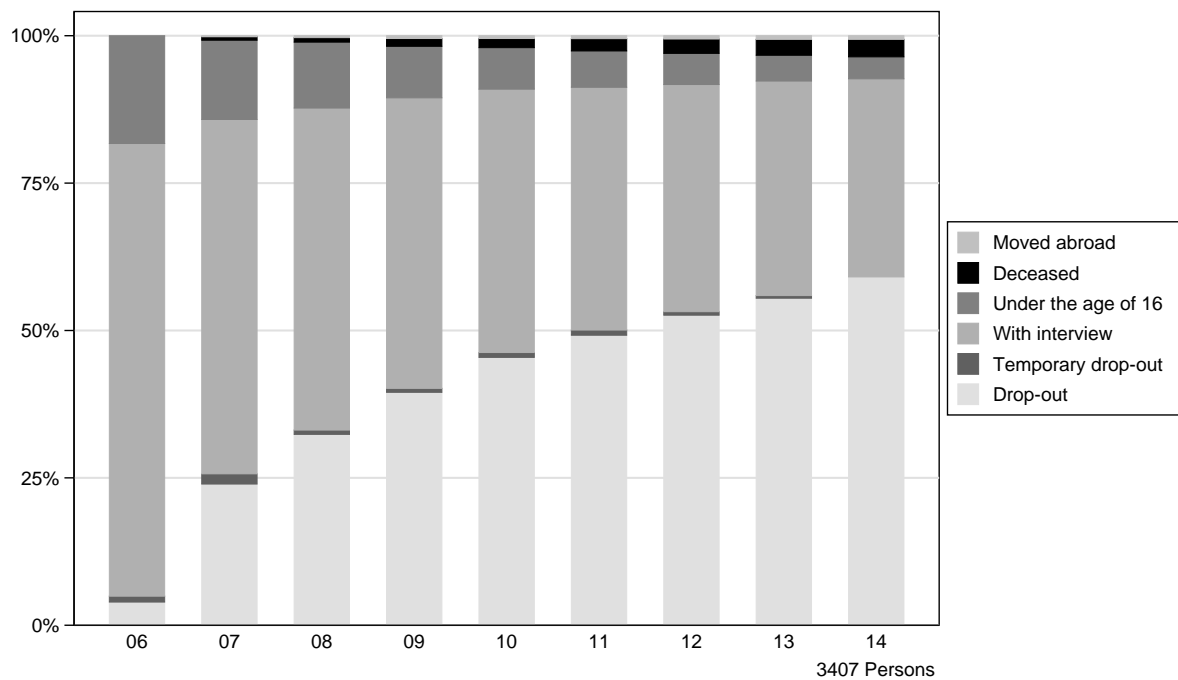


Figure 21: All First-Wave Persons (Gross Subsample H). Development up to Wave 9

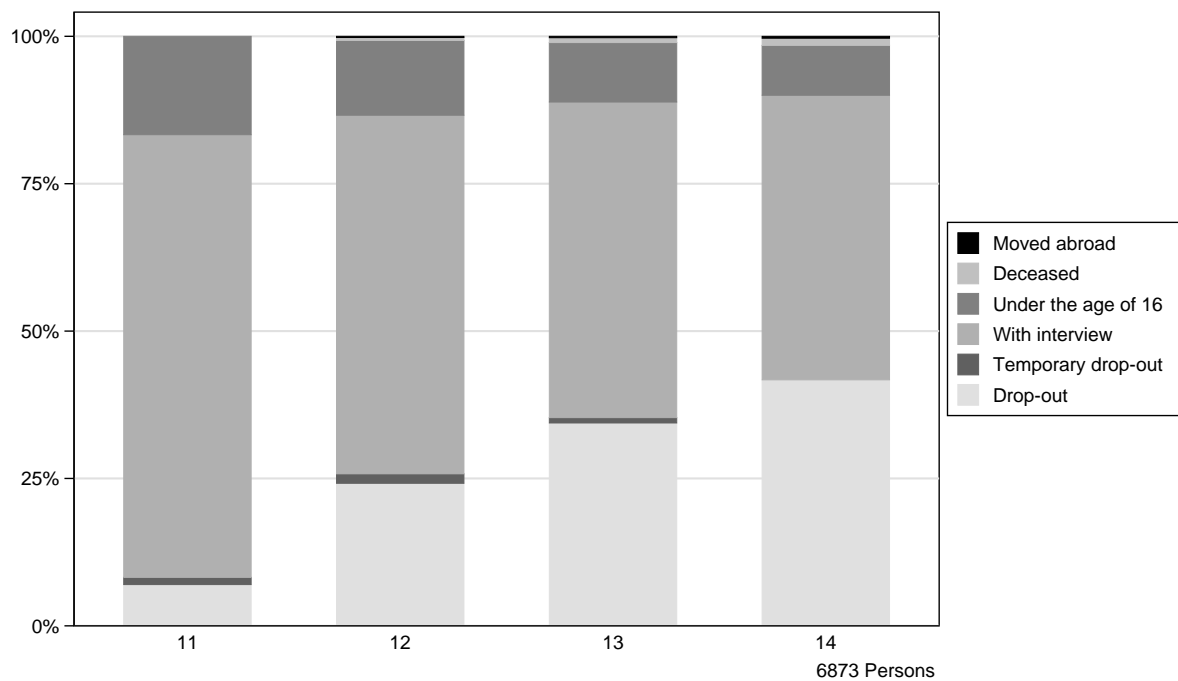


Figure 22: All First-Wave Persons (Gross Subsample J). Development up to Wave 4

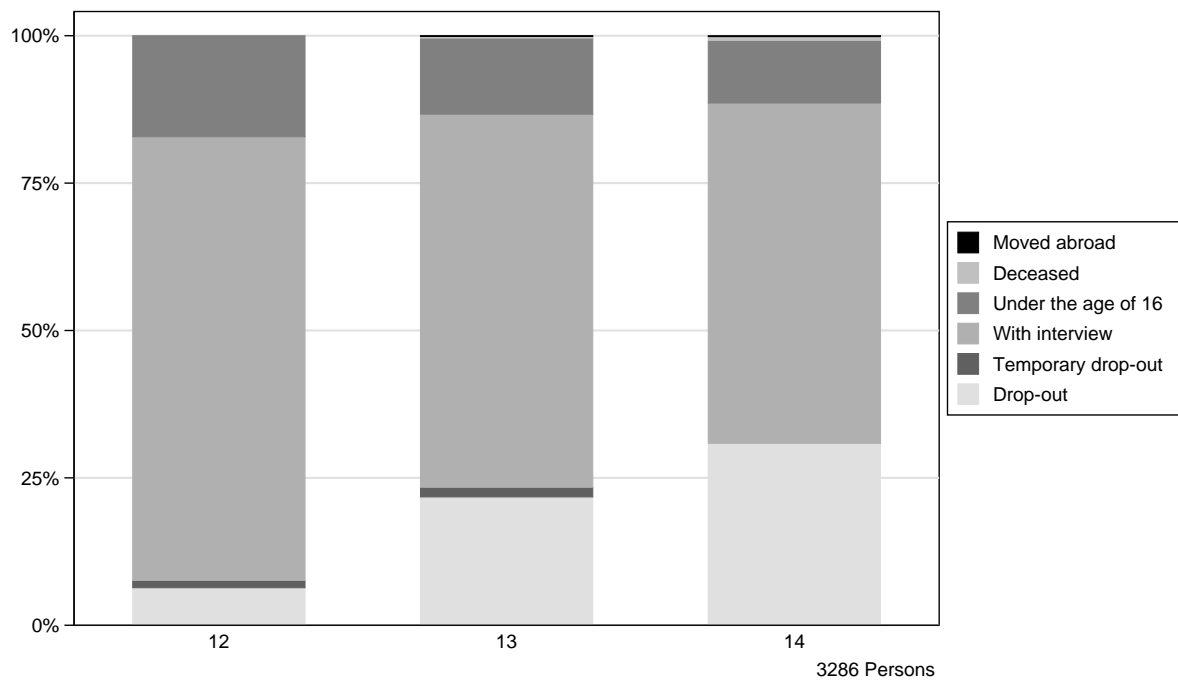


Figure 23: All First-Wave Persons (Gross Subsample K). Development up to Wave 3

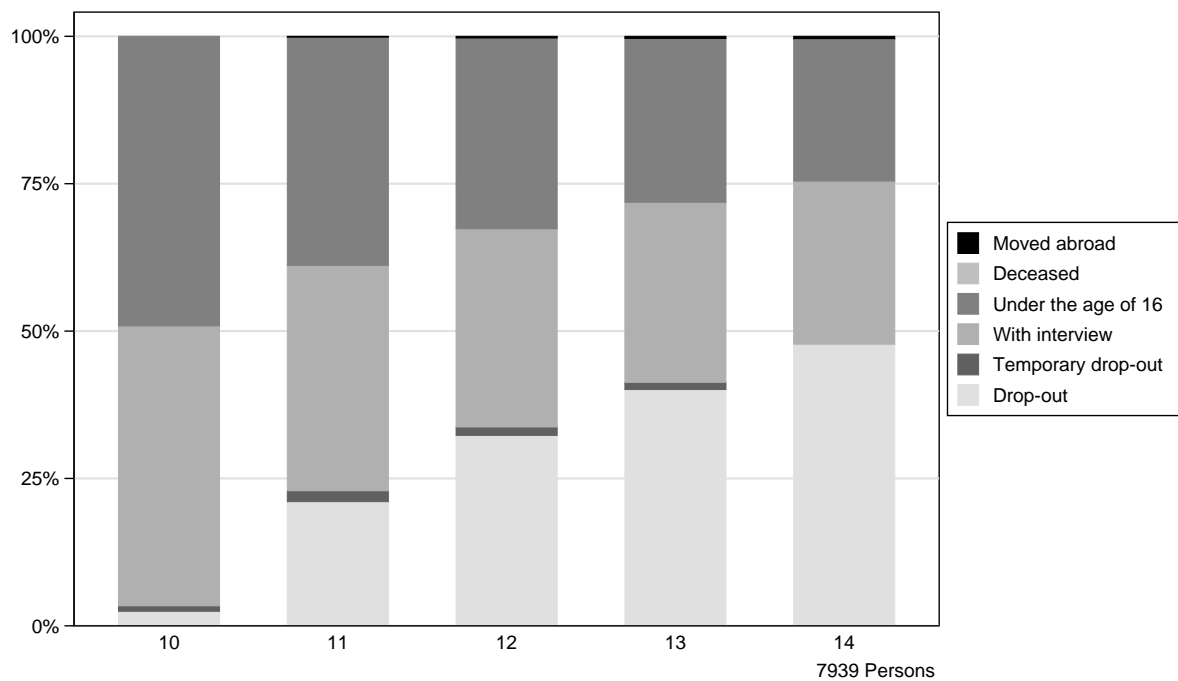


Figure 24: All First-Wave Persons (Gross Subsample L1). Development up to Wave 5

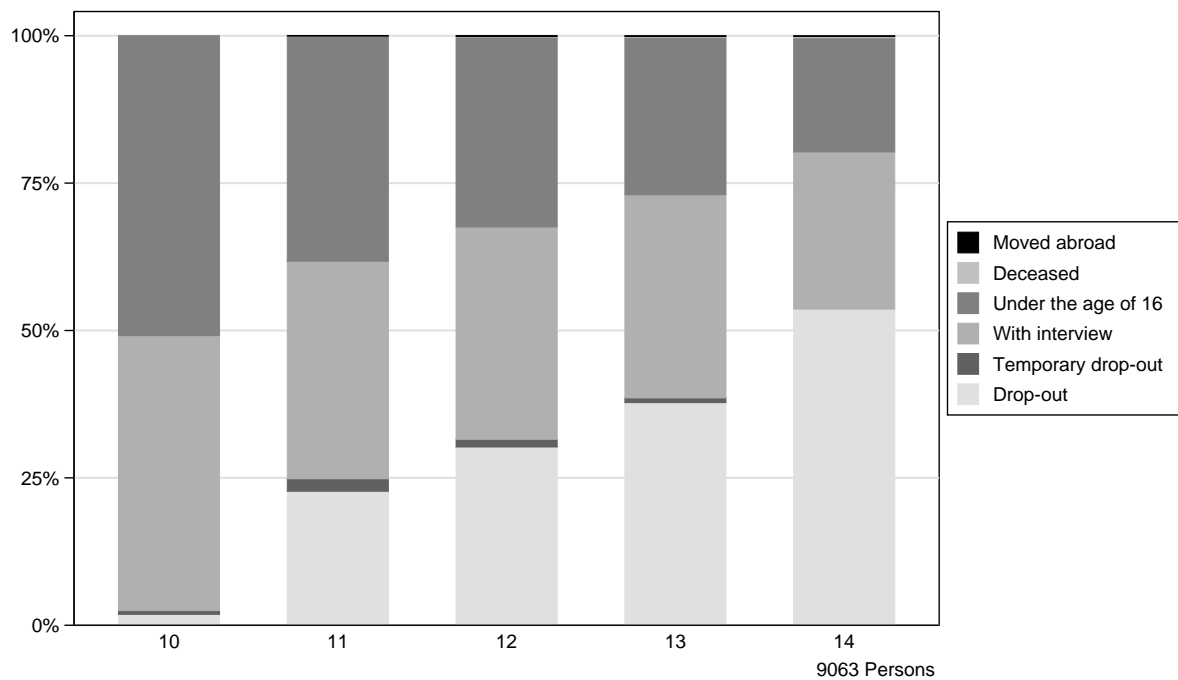


Figure 25: All First-Wave Persons (Gross Subsample L2). Development up to Wave 5

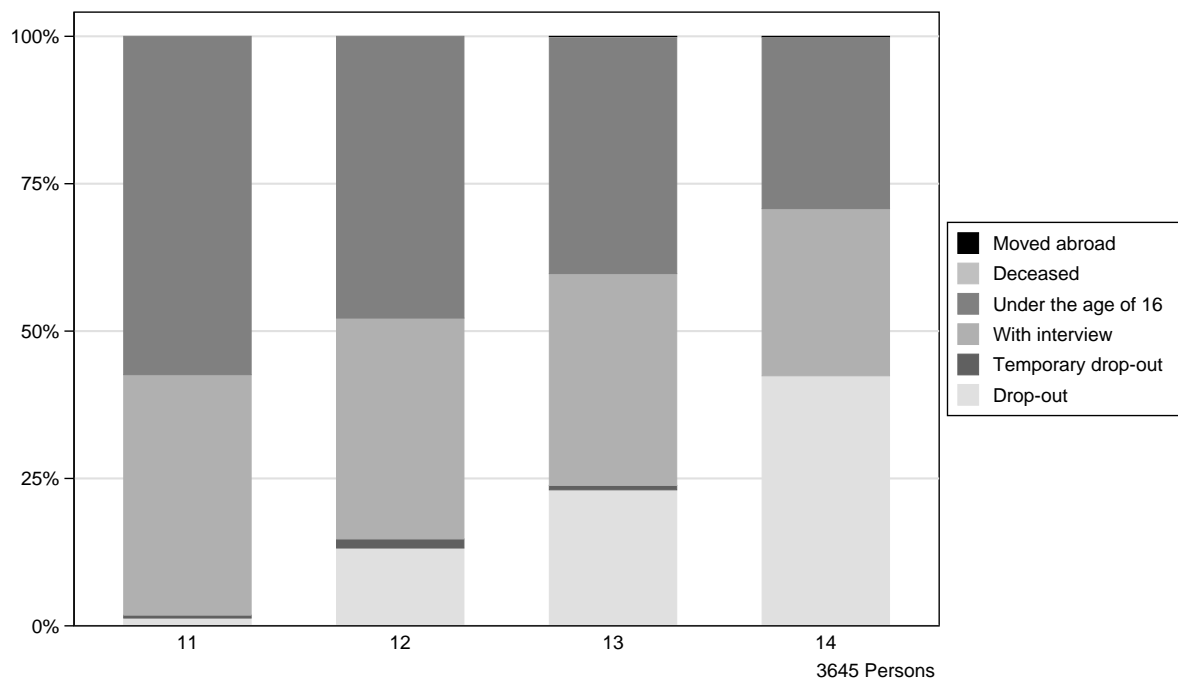


Figure 26: All First-Wave Persons (Gross Subsample L3). Development up to Wave 4

2.3 New Entrants through Birth or Move into SOEP Households and Their Participation Behavior

The following figures display the participation behavior of the non-original sample members and their entrance to the ongoing survey, distinguishing between continuation of participation, exits due to survey unrelated attrition, and exits due to survey-related attrition.

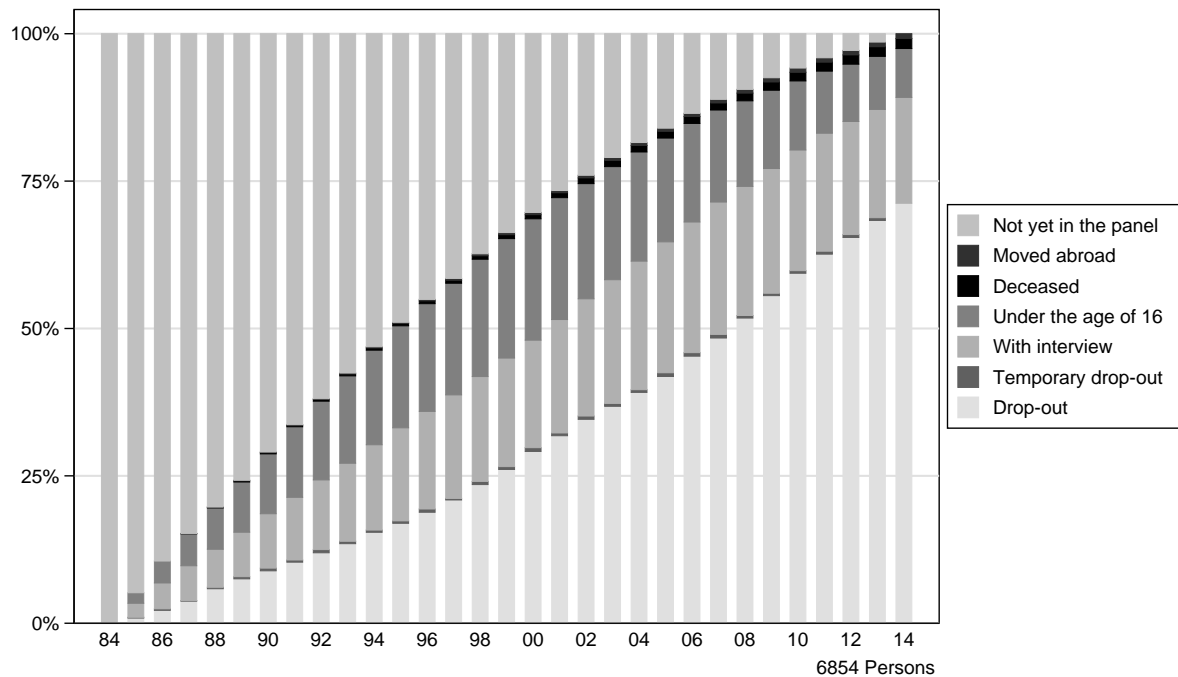


Figure 27: Entrants and their Participation Behavior (Subsample A)

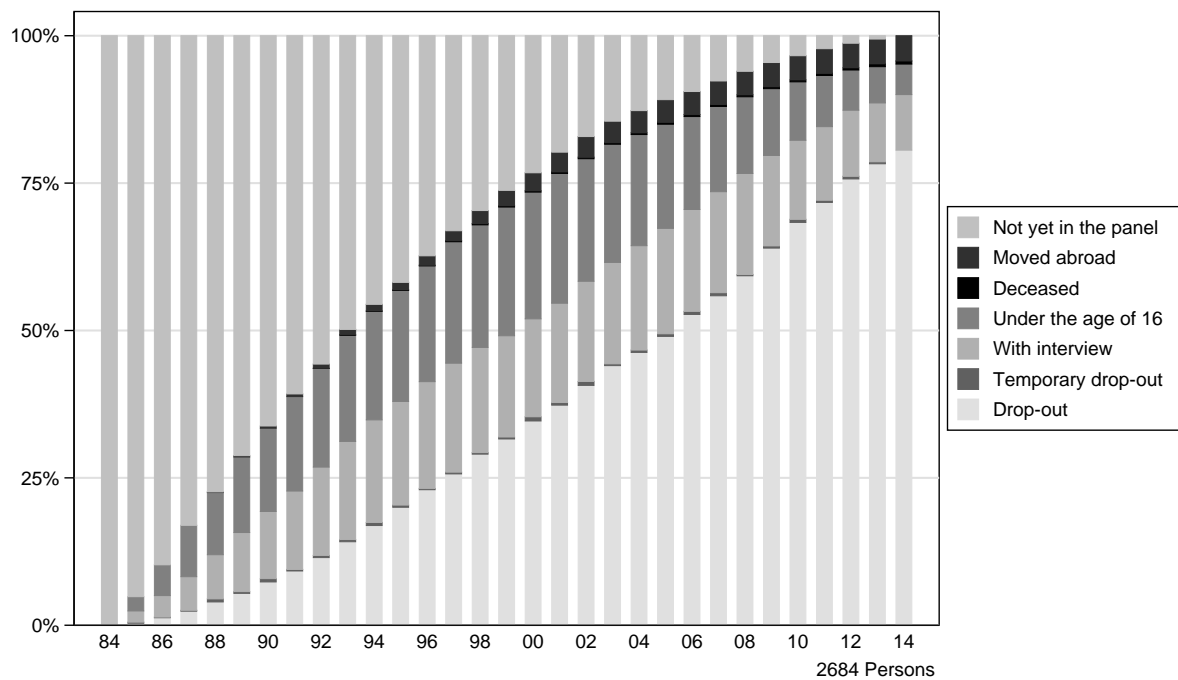


Figure 28: Entrants and their Participation Behavior (Subsample B)

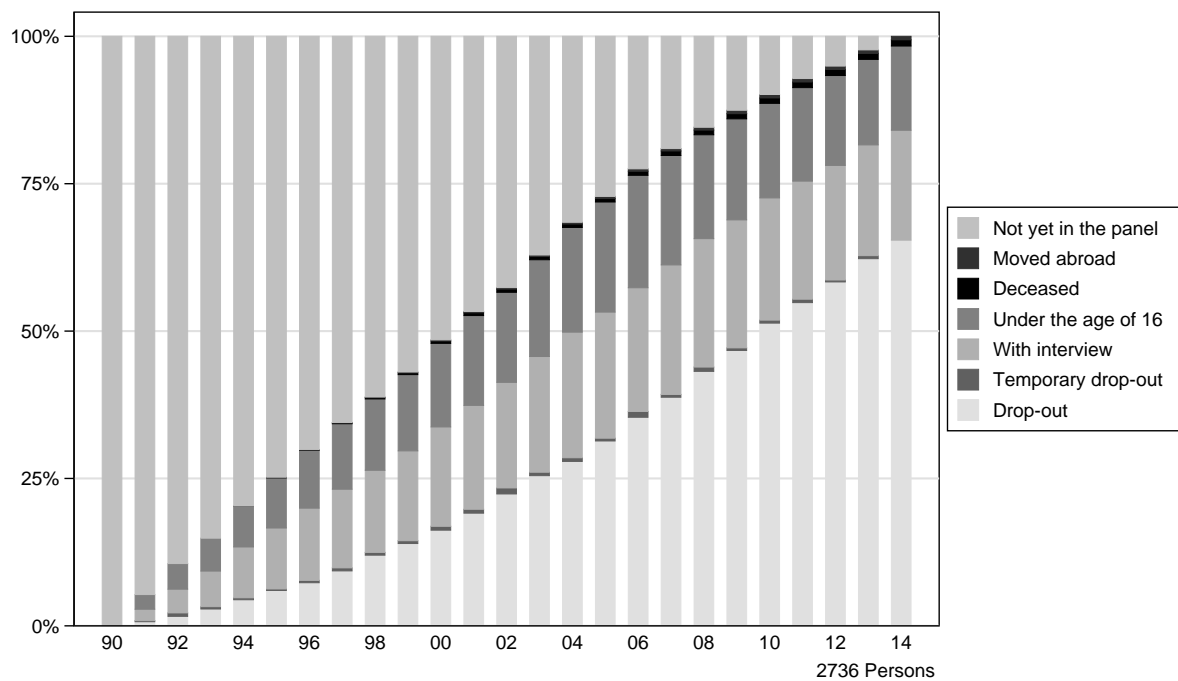


Figure 29: Entrants and their Participation Behavior (Subsample C)

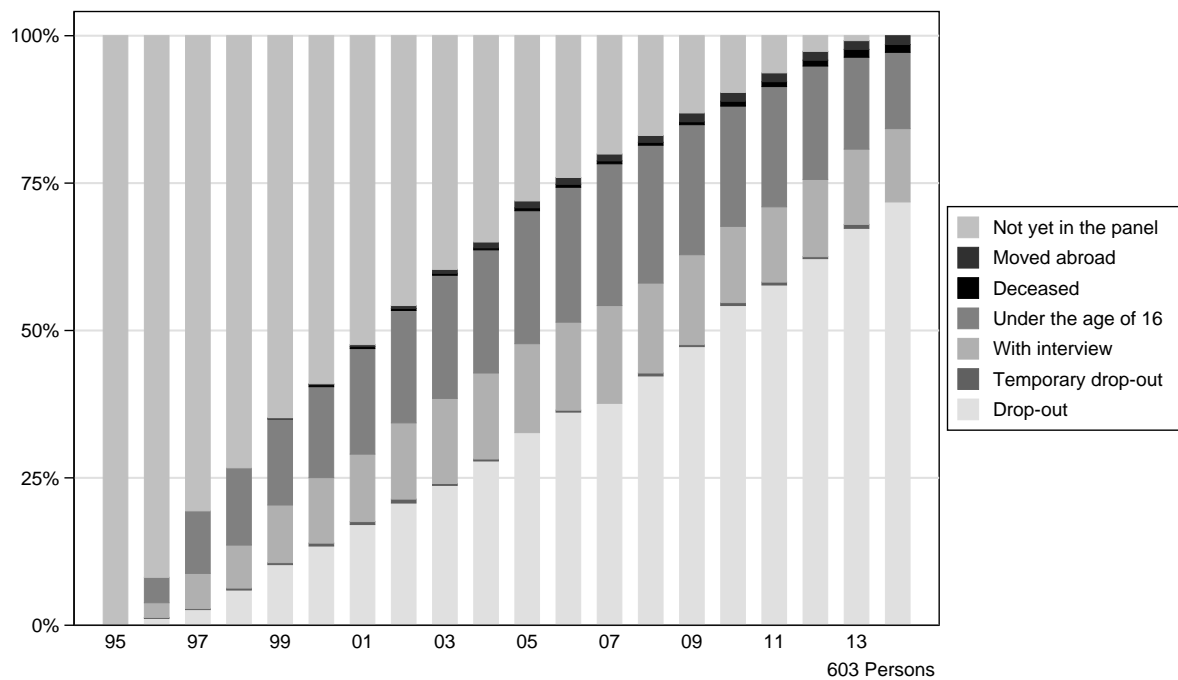


Figure 30: Entrants and their Participation Behavior (Subsample D)

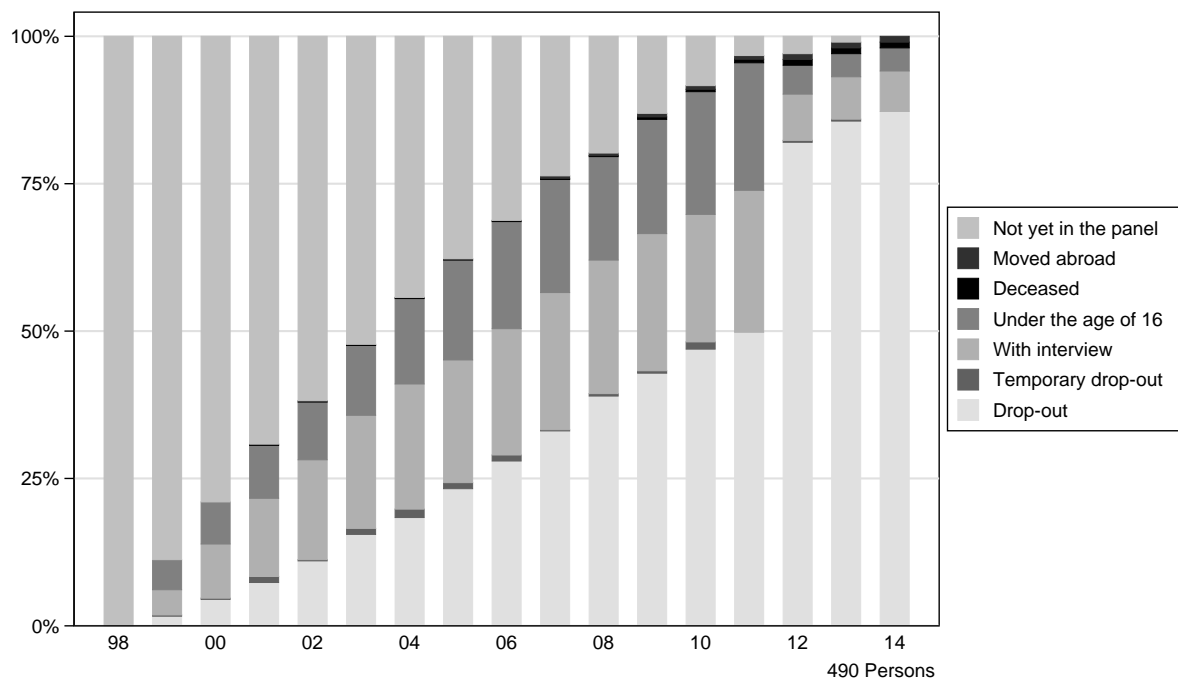


Figure 31: Entrants and their Participation Behavior (Subsample E)

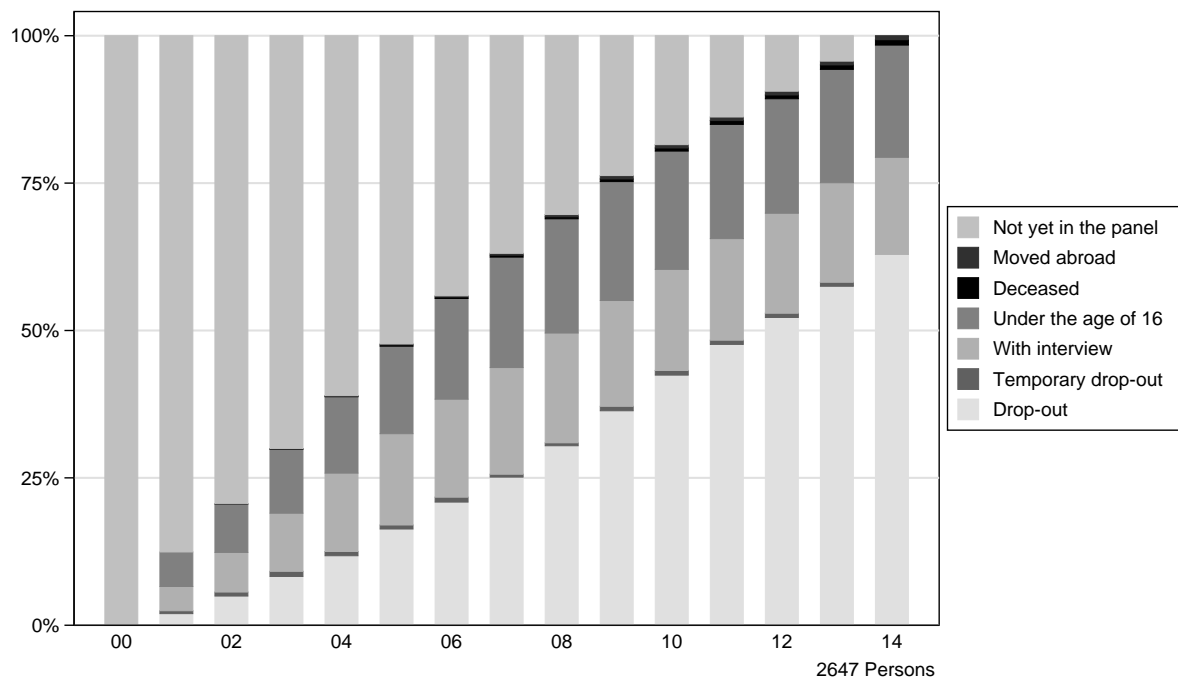


Figure 32: Entrants and their Participation Behavior (Subsample F)

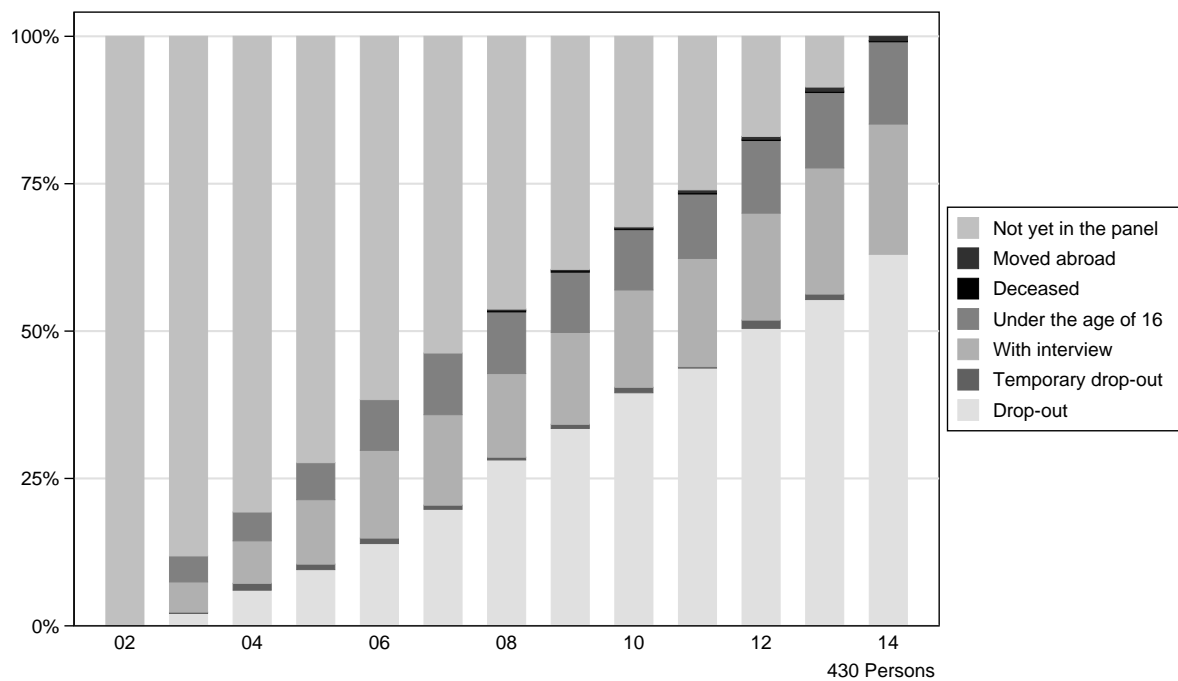


Figure 33: Entrants and their Participation Behavior (Subsample G)

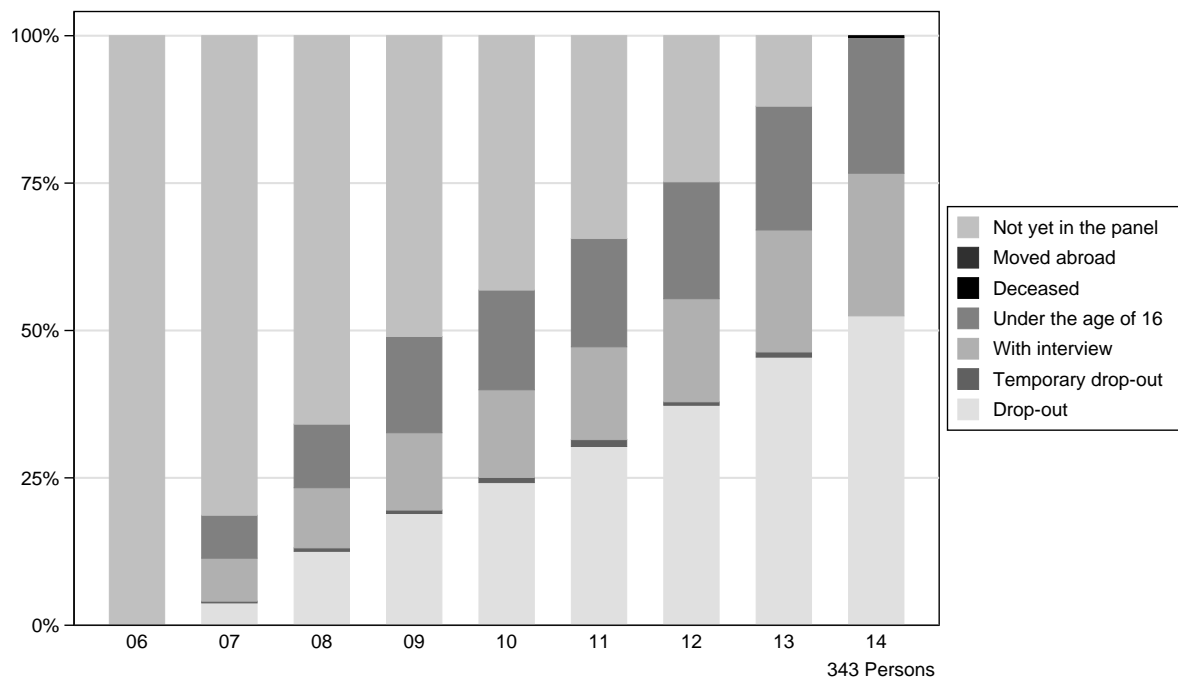


Figure 34: Entrants and their Participation Behavior (Subsample H)

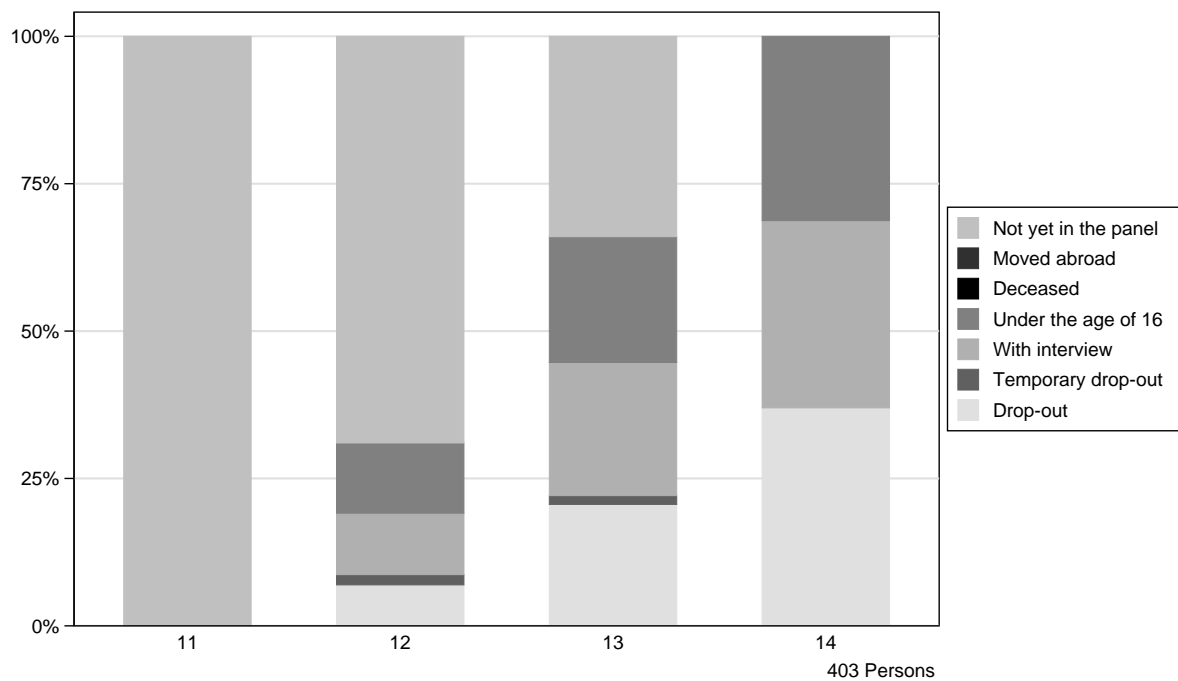


Figure 35: Entrants and their Participation Behavior (Subsample J)

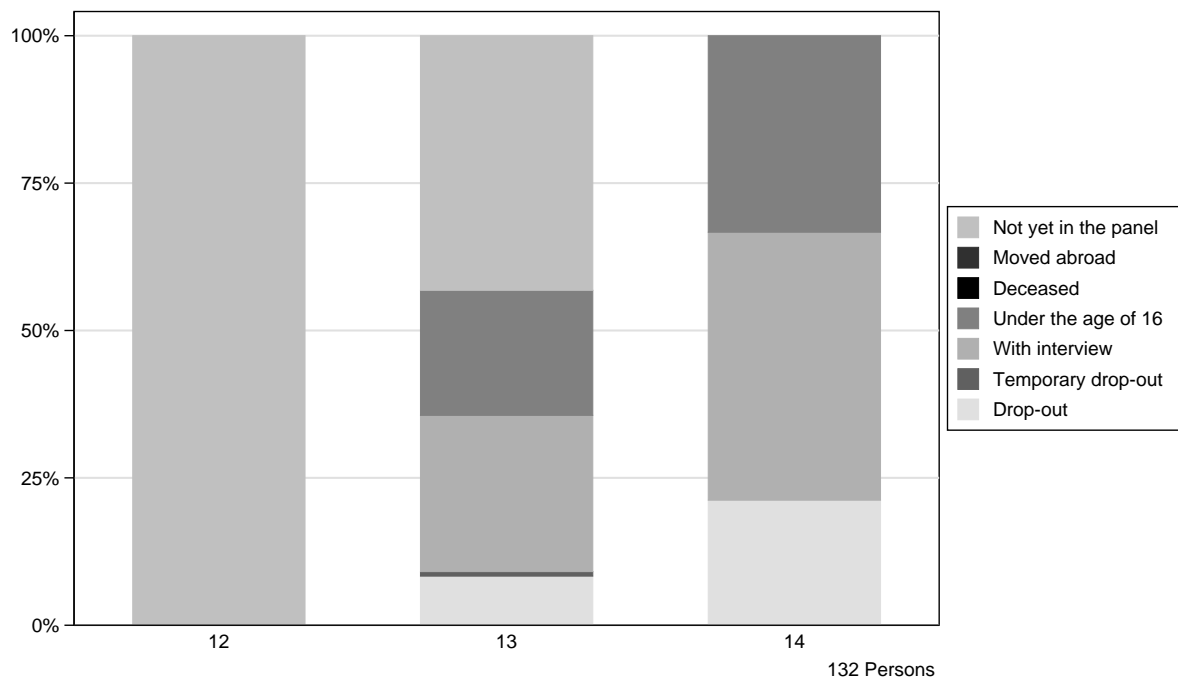


Figure 36: Entrants and their Participation Behavior (Subsample K)

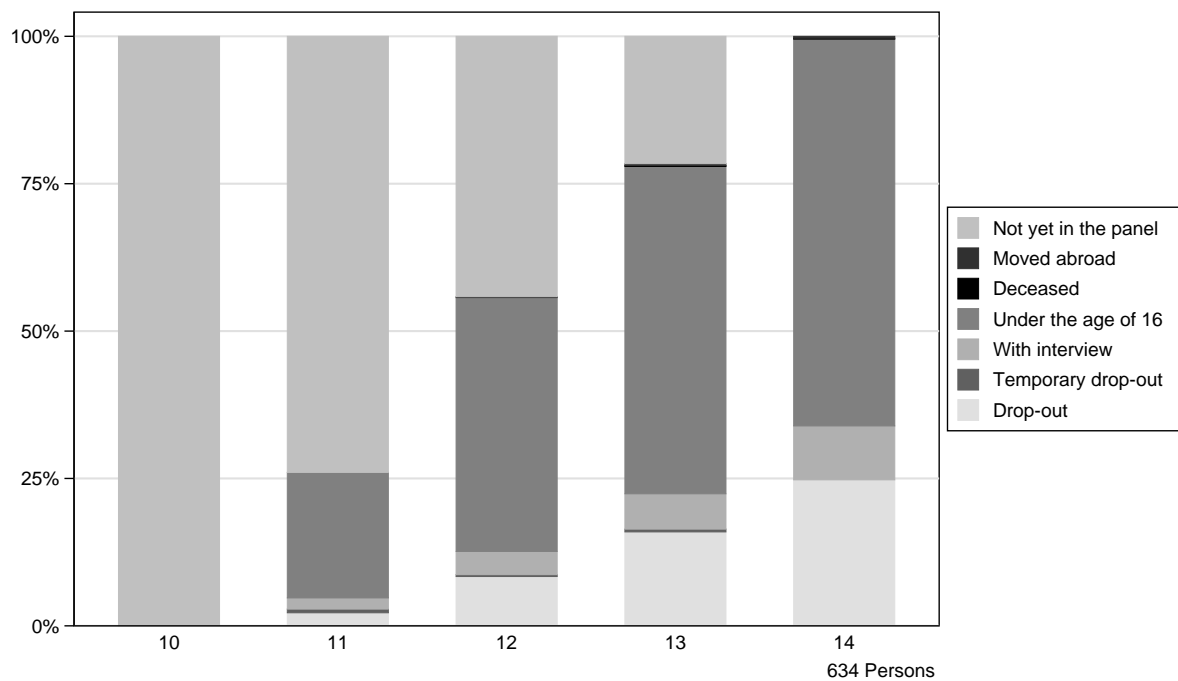


Figure 37: Entrants and their Participation Behavior (Subsample L1)

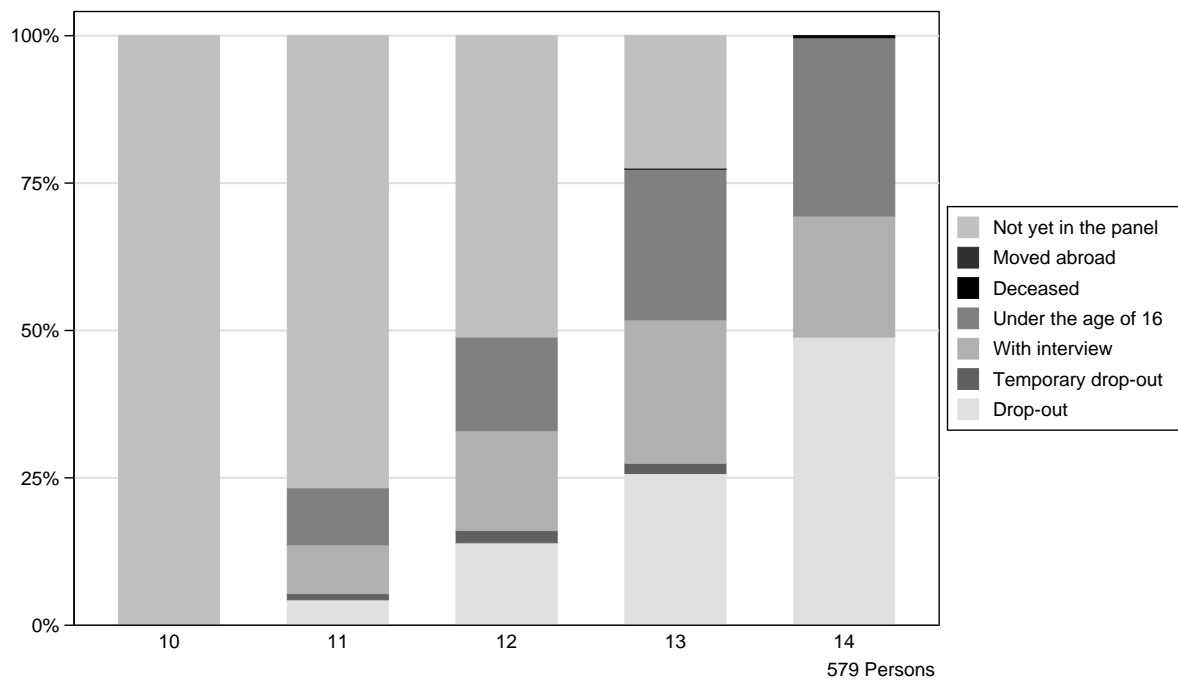


Figure 38: Entrants and their Participation Behavior (Subsample L2)

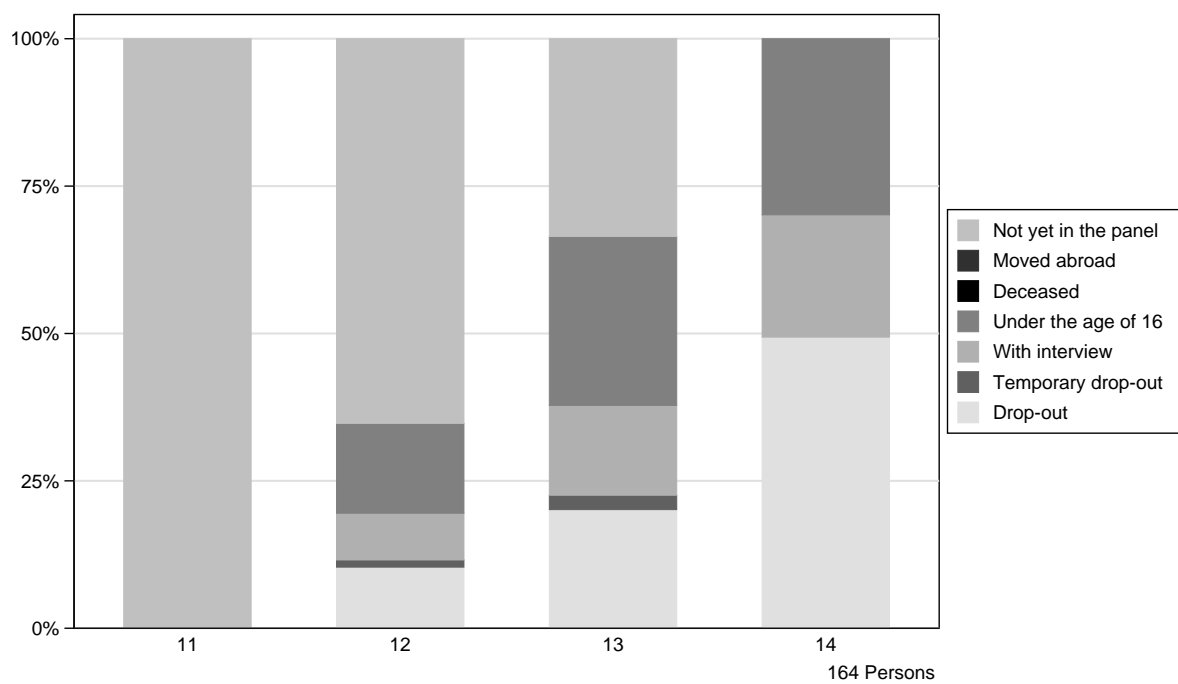


Figure 39: Entrants and their Participation Behavior (Subsample L3)

2.4 The Risk of Survey-Related Panel Attrition

The following figures display Kaplan-Meier estimates of the risk of survey related attrition (unsuccessful follow-up and refusal) of the net sample of first-wave respondents thereby ignoring survey unrelated exits (moves abroad and deaths). These figures stratify the drop-out risk in different groups of the sample defined by respondents' sample membership (Figures 40, 41, 42) and 43 and some basic socio-demographic characteristics measured in the year of sampling, such as age, occupation, income, and education (Figures 44, 45, 46 and 47). These unweighted figures show in general only moderate differences in the risk of survey related attrition between groups of the sample. Among the older samples A through C (Figure 40), for instance, first-wave respondents from sample B have a somewhat lower probability of remaining in the survey than respondents from samples A or C. In the more recent samples D through K (Figures 41 and 42), first-wave respondents from sample H have a somewhat lower probability of remaining in the survey than respondents from sample F. The latter in turn, have a lower probability of remaining in the survey than respondents from sample G.

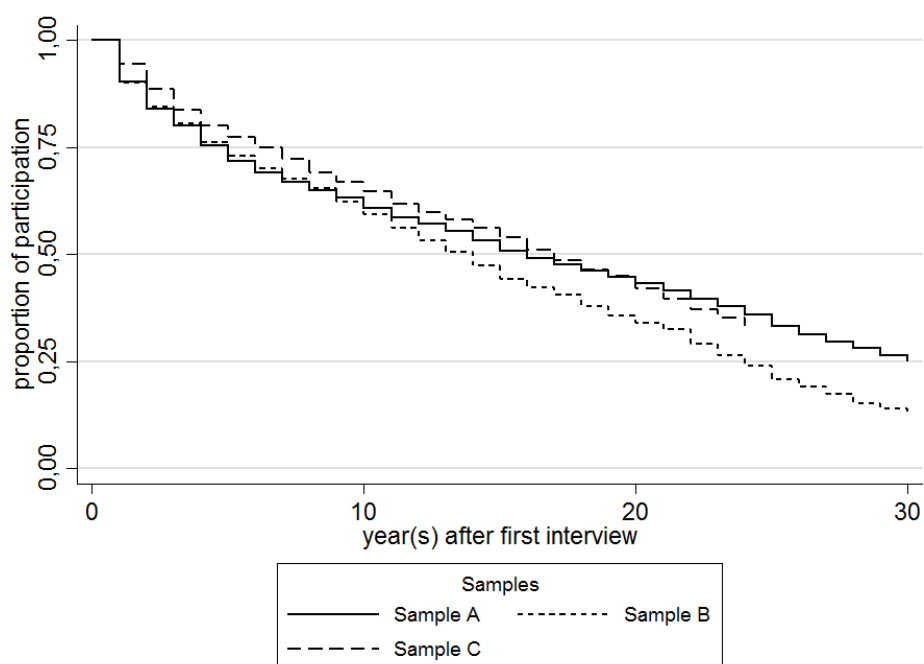


Figure 40: Successful Re-Interviewing of First-Wave Respondents by Subsamples A, B, C. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

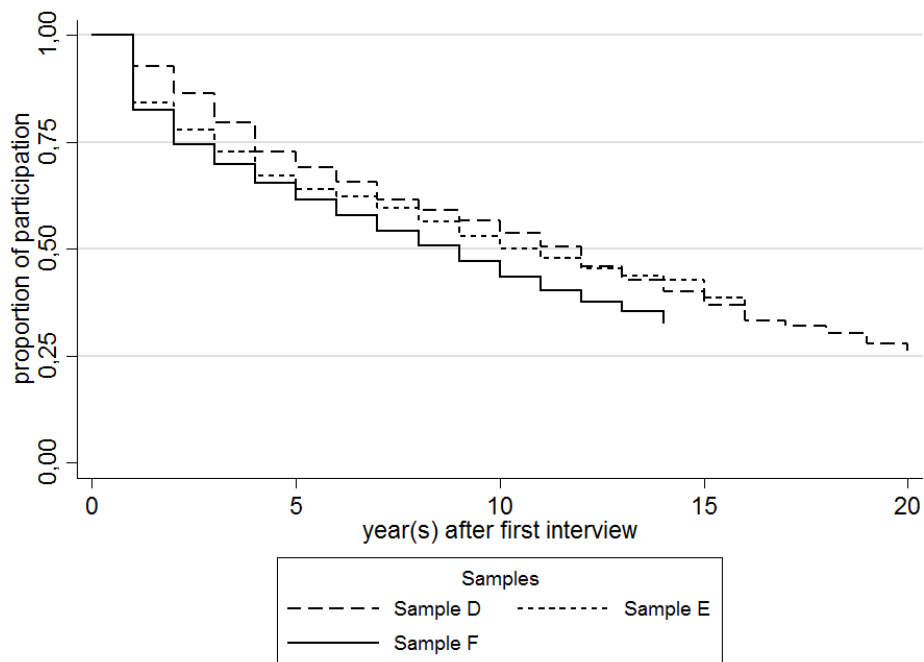


Figure 41: Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

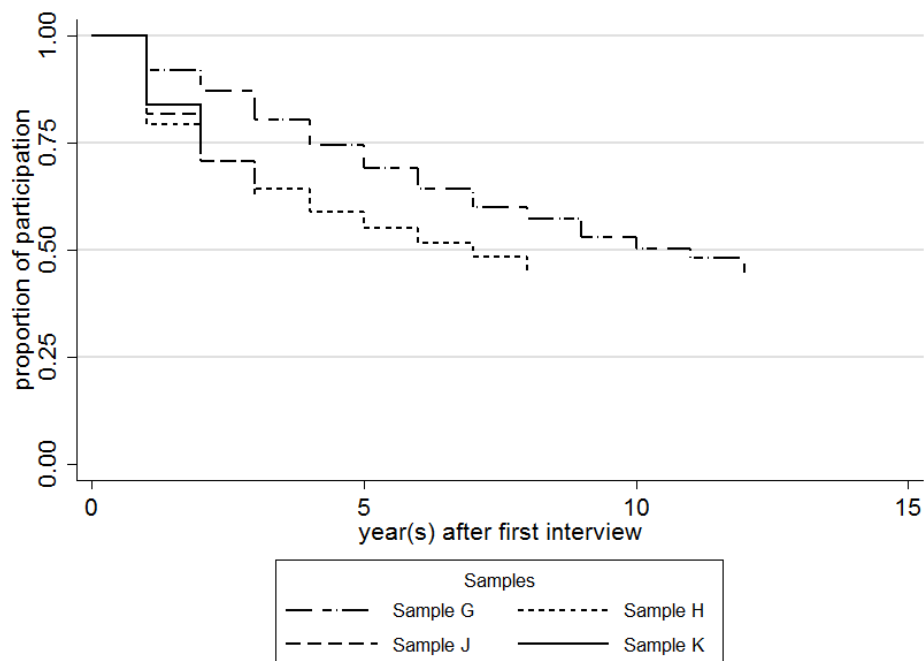


Figure 42: Successful Re-Interviewing of First-Wave Respondents by Subsamples G, H, J and K. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

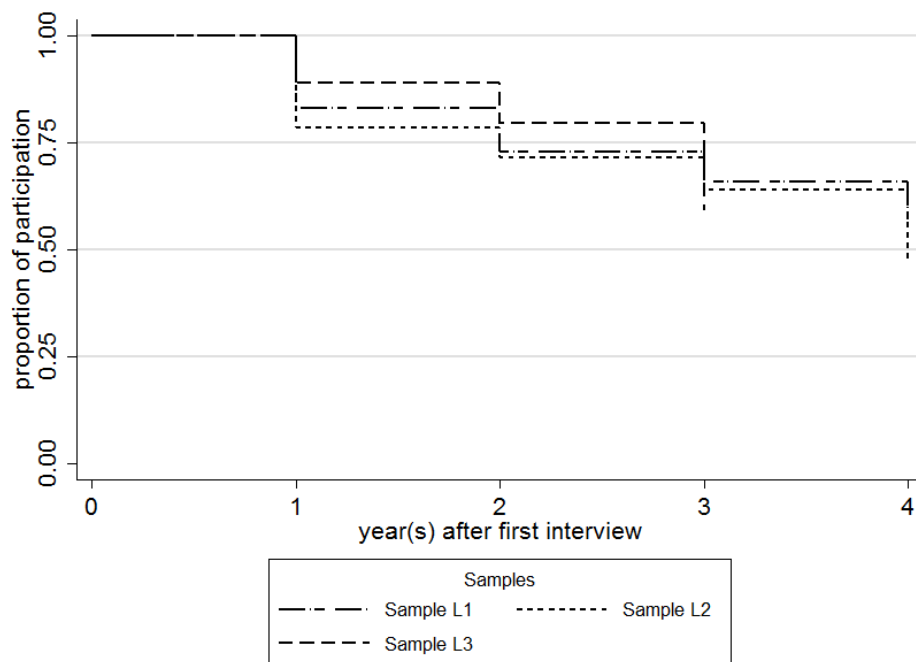


Figure 43: Successful Re-Interviewing of First-Wave Respondents by Subsamples L1, L2 and L3. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

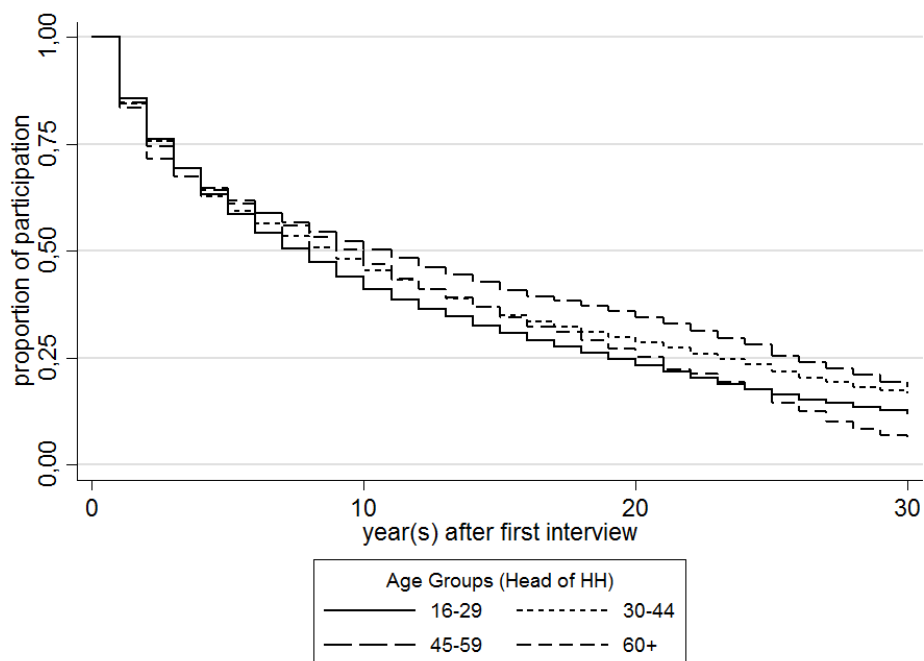


Figure 44: Successful Re-Interviewing of All First-Wave Respondents by Age Categories. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

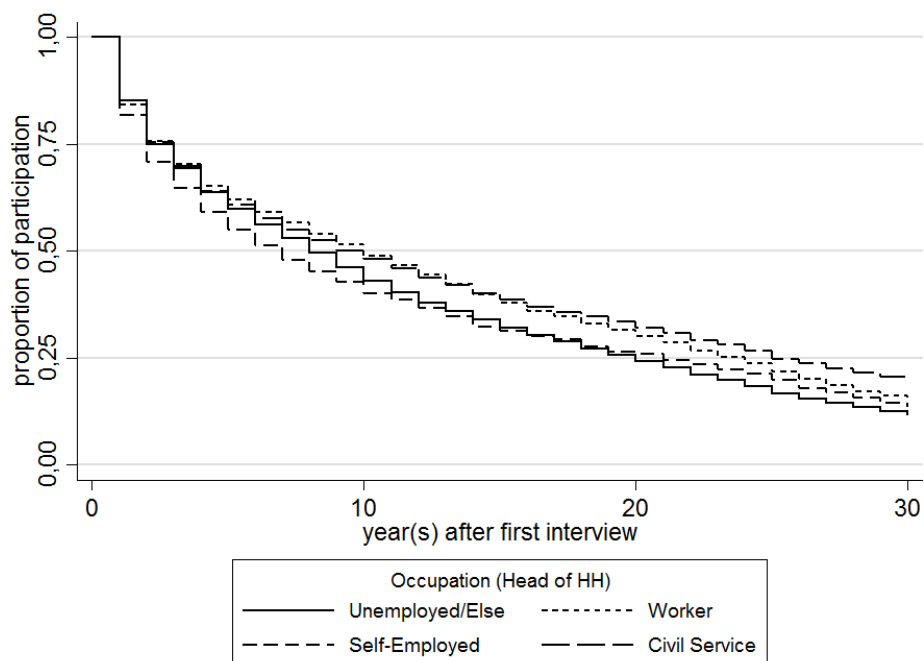


Figure 45: Successful Re-Interviewing of All First-Wave Respondents by Occupation. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

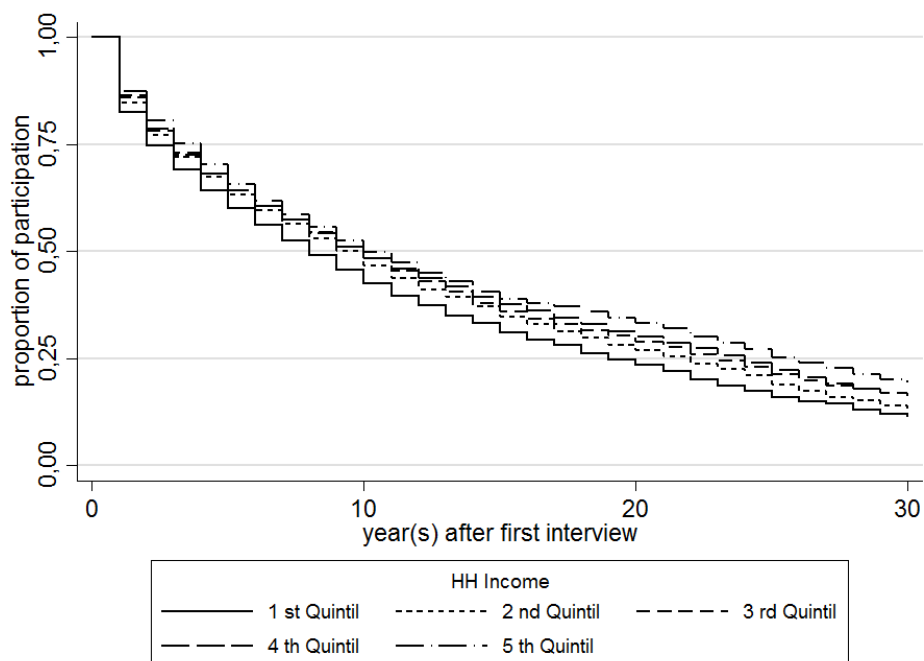


Figure 46: Successful Re-Interviewing of All First-Wave Respondents by Income Quintiles. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

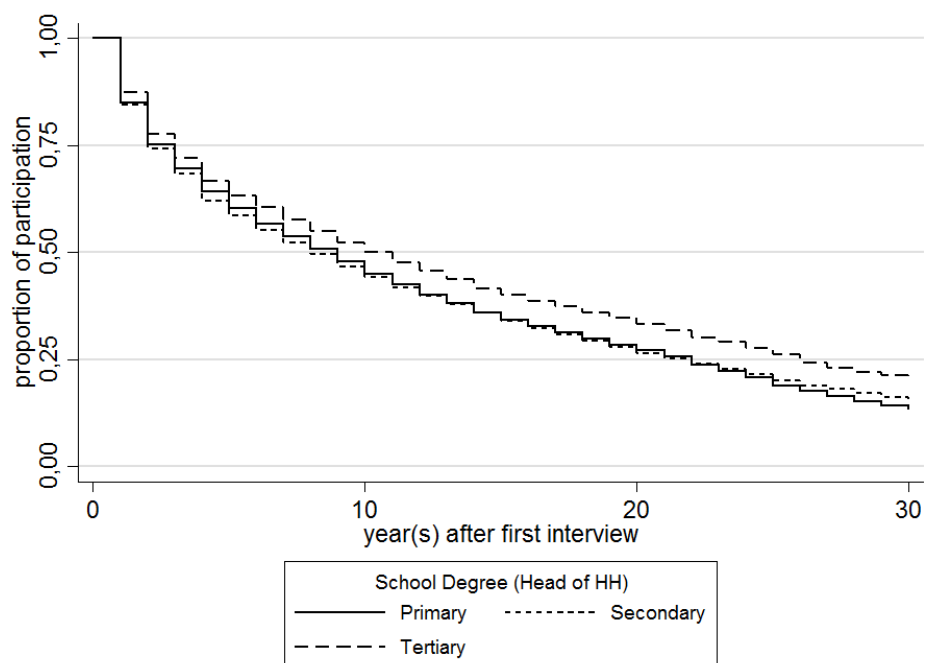


Figure 47: Successful Re-Interviewing of All First-Wave Respondents by Education. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad

3 Panel Attrition Due to Unsuccessful Follow-Ups

In each panel wave, the first step in successful re-interviewing is the identification of the place of residence of households who took part in the preceding wave. The fieldwork organization of the SOEP, TNS Infratest, identifies whether (a) a household still lives at the old address, (b) an entire household has moved or all household members have died, (c) all household members have left the sampling area, and (d) all household members have returned to an existing panel household.

3.1 The Frequency of Successful Follow-Ups

Table 3.1 the number of households of the previous waves that need to be re-contacted and the relative frequency of successful follow-ups in sub-samples A through M and waves 1985 through 2014. The re-contact rates refer to all households of the previous wave that still exist in the sampling area plus split-off households. A contact is regarded as successful if the interviewer documented a completed interview or refusal in the address protocol. Moreover, if former household members returned to an existing panel household, this is classified as a successful follow-up.

Table 3.1: The Frequency of Households to be Re-Contacted and the Percentage of Successful Follow-Ups, Subsamples A to M by Year.

Year	Sample A N	%	Sample B N	%	Sample C N	%	Sample D N	%	Sample E N	%	Sample F N	%	Sample G N	%	Sample H N	%	Sample I N	%	Sample J N	%	Sample K N	%	Sample L1 N	%	Sample L2 N	%	Sample L3 N	%	Sample M N	%
1984	4.528	100,0	1.393	100,0																										
1985	4.681	98,3	1.370	96,4																										
1986	4.486	98,9	1.325	97,0																										
1987	4.232	99,0	1.220	98,6																										
1988	4.140	99,1	1.191	99,0																										
1989	3.984	99,0	1.157	99,0																										
1990	3.902	99,1	1.124	98,8	2.179	100,0																								
1991	3.860	99,5	1.151	99,2	2.246	98,4																								
1992	3.845	99,7	1.153	99,2	2.302	99,4																								
1993	3.867	99,2	1.172	98,5	2.227	99,0																								
1994	3.849	99,2	1.150	98,9	2.134	99,4	236	100,0																						
1995	3.784	99,5	1.108	98,9	2.110	99,5	540	100,0																						
1996	3.747	99,6	1.069	99,2	2.103	99,4	544	99,6																						
1997	3.688	99,6	1.038	99,0	2.087	99,4	541	99,2																						
1998	3.667	99,4	1.019	99,3	2.079	99,3	528	98,9	1.056	100,0																				
1999	3.631	99,6	975	99,3	2.037	99,6	498	99,3	1.089	99,4																				
2000	3.549	99,6	934	99,4	2.025	99,6	467	99,8	967	99,1	6.043	100,0																		
2001	3.463	99,5	904	99,4	2.034	99,7	454	99,0	921	99,0	6.162	98,8																		
2002	3.406	99,7	877	99,0	2.005	99,5	450	99,8	873	99,4	5.447	99,4	1.224	100,0																
2003	3.330	99,6	840	99,6	1.982	99,6	434	99,5	834	99,2	4.965	99,7	1.056	99,0																
2004	3.260	99,8	803	99,6	1.962	99,6	436	99,7	797	99,7	4.736	99,6	1.010	99,7																
2005	3.220	99,8	779	99,3	1.959	99,7	429	99,2	783	99,9	4.577	99,7	1.001	99,7																
2006	3.138	99,7	770	99,5	1.941	99,4	425	98,6	775	99,0	4.401	99,2	995	99,4	1.506	100,0														
2007	3.000	99,7	725	99,4	1.834	99,9	387	99,4	727	99,7	4.157	99,5	933	99,2	1.530	99,4														
2008	2.856	99,8	676	99,1	1.767	99,5	372	99,4	680	99,7	3.962	99,3	904	99,6	1.326	99,5														
2009	2.730	99,7	620	99,2	1.695	99,9	351	99,7	636	100,0	3.760	99,6	870	99,5	1.145	99,7	1.495	100,0												
2010	2.570	99,8	548	99,3	1.627	100,0	334	99,6	605	99,8	3.538	99,5	826	99,9	1.059	99,5	1.737	97,6					2.074	100,0	2.500	100,0				
2011	2.421	99,8	495	99,0	1.541	99,8	302	99,3	589	100,0	3.318	99,6	797	99,6	992	99,5			3.136	100,0			2.082	98,4	2.271	97,9	924	100,0		
2012	2.289	99,8	440	99,7	1.466	99,9	286	100,0	116	98,9	3.076	99,9	774	99,7	928	99,9			3.201	99,1	1.526	100,0	1.865	99,5	2.254	98,3	943	98,5		
2013	2.180	99,6	392	99,1	1.417	99,7	269	99,1	98	100,0	2.880	99,7	733	99,6	877	99,5			2.869	99,4	1.564	98,8	1.752	99,1	2.177	98,6	920	99,0	2.723	100,0
2014	2.077	99,3	358	99,3	1.351	99,6	249	100,0	90	100,0	2.741	99,6	725	99,2	828	99,3			2.519	99,0	1.447	99,2	1.510	99,3	2.027	97,5	836	98,0	2.819	98,4

Note: In the case of the initial wave of a sample, table entries are the number of participating households. See also Section 2.

3.2 Predicting the Probability of Successful vs. Unsuccessful Follow-Ups in the Year 2014

Based on household and interview level characteristics measured in the 2013 previous wave, we aim at predicting the probability of re-contacting a household relative to unsuccessful follow-up in 2014. Among a very large number of regressors that we tested in preliminary analyses, we identified a small set of variables that exert a robust effect on the probability of successful follow-ups ($p < 0.05$). Table 3.2 describes the regressors and Table 3.2 reports the subsample-specific estimates of logit models for the probability of re-contacting a household relative to unsuccessful follow-up.

Note that the estimates of regression models run for the previous waves of 1985 through 2014 are due to space restrictions not reported in the present data documentation. These can be obtained from previous attrition documentations.

Table 3.2: Definition of the Regressors of the Logit Model of Unsuccessful Follow-Ups

Variable	Label	Value
Interview Characteristics		
New HH	HH is New in SOEP	0/1
Phone Unknown	Telephone Number Undisclosed	0/1
Temporary Drop-out	Temporary Drop-out in Last Wave	0/1
Change of Interviewer	Change of Interviewer	0/1
Interview in June or later	Interview in June or later	0/1
New SOEP member	New SOEP Member	0/1
Original Sample Member	Original Sample Member	0/1
Item Nonresponse on > 1 fin. quest.	Item Nonresponse on at least 1 financial question	0/1
(High) Item Nonresponse HH	High Share of Item Nonresponse in the Household	0/1
Moving in	HH Move	0/1
Separation	Move-out due to Separation in Last Wave	0/1
Region		
Northrhine Westfalia	Northrhine Westfalia	0/1
Saxony	Saxony	0/1
High share of Abitur	HH in Area with High Share of Abitur	0/1
High Share of Academics	HH in Area with High Share of Academics	0/1
High Share of Foreigners	HH in Area with High Share of Foreigners	0/1
High Share of For. f. Islam. Countr.	High Share of For. from Countries with Islam as widespread religion	0/1
High Purchasing Power	HH in Area with High Purchasing Power	0/1
Financial Situation, Real Estate		
Single HH	Single Household	0/1
House Owner	House Ownership	0/1
Single Family-House	GH living in a Single Family-House	0/1
Personality, Well-Being, and other Characteristics		
Younger than 25	Head of HH is Younger than 25	0/1
Head of HH often scared	Head of HH is Often Scared	0/1
Single	HH of HH is Single	0/1
Visited Foreigner prev. year	Head of HH Visited Foreigners in the Previous Year	0/1
Hobbies and Leisure	HH spends ≥2 Hours Per Day for Hobbies and Leisure	0/1
Unhappy	Head of HH is Often Unhappy	0/1

Table 3.3: Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2014

	Sample A	Sample B	Sample C	Sample F	Sample G	Sample H	Sample J	Sample K	Sample L1	Sample L2	Sample L3	Sample M
Intercept	2.23***	1.65***	1.98***	3.23***	1.86***	2.18***	2.45***	2.19***	2.90***	2.06***	2.50***	2.29***
<u>Interview Characteristics</u>												
New HH	-1.17***			-1.46***		-1.17***	-1.17***	-0.65**	-1.45***	-1.59***		-1.21***
Phone Unknown	-0.57***						-0.69***	-0.52**	-0.89***		-1.57***	
Temporary Drop-out				-1.04***								
Change of Interviewer										0.58***		
Interview in June and later											-0.55**	
New SOEP member					-1.41***							
Original Sample Member							-0.53***	-0.65***				
Item Nonresponse on >1 fin. quest.					-0.81***							
(High) Item Nonresponse HH						-0.60**						
Moving in	-0.75***					-0.88***	-1.01***	-0.89***		-1.05***	-1.00***	-1.34***
Separation			-1.30***	-1.02***								
<u>Region</u>												
Northrhine Westfalia				-1.17**								
Saxony				-1.21**								
High share of Abitur							-0.54***					
High Share of Academics							0.46**					
High Share of Foreigners										-0.49***		
High Share of For. f. Islam. Countr.									-0.58**			
High Purchasing Power			-0.61**									
<u>Financial Situation, Real Estate</u>												
Single HH										-0.61***		-0.49***
House Owner												-0.57***
Single Family-House										0.45**		
<u>Personality, Well-Being, and other Characteristics</u>												
Younger than 25												-0.33**
Head of HH often scared							-0.52**					
Single											-0.59**	

Note: ***p < 0.01; **p < 0.05; *p < 0.10. In Sample D, all households were re-contacted.

Table 3.3: Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2014

	Sample A	Sample B	Sample C	Sample F	Sample G	Sample H	Sample J	Sample K	Sample L1	Sample L2	Sample L3	Sample M
Visited Foreigner prev. year												0.36**
Hobbies and Leisure									-0.69**			
Unhappy									-0.68**			
<i>No. of Observations</i>	2078	361	1351	2741	725	828	2519	1448	1512	2027	836	2828
<i>Log Likelihood</i>	-44.68	-12.39	-20.73	-27.83	-20.19	-19.71	-66.05	-35.82	-25.23	-95.88	-34.53	-109.27

Note: ***p < 0.01; **p < 0.05; *p < 0.10. In Sample D, all households were re-contacted.

4 Panel Attrition Due to Refusals

In each panel wave, the second step in successful re-interviewing after having identified the location of households from the preceding wave is to obtain each household's confirmation of willingness to participate in the survey. We define successful re-interviewing relative only to survey-related panel attrition, such as refusals, and ignore survey-unrelated attrition, such as the death of a participant or her decision to move abroad, to generate the longitudinal weights.

4.1 The Frequency of Participation

Table 4.1 display the participation rates due to refusal by sub-sample and wave. The corresponding drop out rates can be then obtained following an analogous procedure. Note that in order to obtain this probability no distinction was made between the various types of refusals that can occur in a survey, such as unconditional refusals, refusals due to lack of time, or health problems, etc.

Table 4.1: The Frequency of Re-Contacted Households and the Percentage of Participation, Subsamples A to M by Year.

Year	Sample A		Sample B		Sample C		Sample D		Sample E		Sample F		Sample G		Sample H		Sample I		Sample J		Sample K		Sample L1		Sample L2		Sample L3		Sample M	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1984	4.528	100,0	1.393	100,0																										
1985	4.611	89,8	1.326	89,1																										
1986	4.442	89,2	1.290	87,4																										
1987	4.194	93,2	1.204	92,7																										
1988	4.105	91,2	1.180	90,8																										
1989	3.949	92,4	1.146	91,0																										
1990	3.871	93,3	1.111	92,5	2.179	100,0																								
1991	3.842	94,0	1.143	92,4	2.213	91,7																								
1992	3.833	93,5	1.144	92,7	2.290	88,2																								
1993	3.838	93,9	1.156	92,0	2.208	89,2																								
1994	3.821	93,6	1.139	89,8	2.122	92,3	236	100,0																						
1995	3.766	93,6	1.097	89,5	2.101	92,2	540	96,7																						
1996	3.734	93,3	1.061	90,5	2.092	93,3	542	91,9																						
1997	3.674	94,1	1.029	90,5	2.076	93,5	537	89,2																						
1998	3.645	92,9	1.013	88,6	2.066	91,3	523	84,3	1.056	100,0																				
1999	3.616	92,0	969	88,5	2.030	93,3	495	85,9	1.084	81,7																				
2000	3.535	91,7	929	88,3	2.018	93,1	466	91,2	959	87,8	6.043	100,0																		
2001	3.448	91,9	899	90,0	2.028	91,2	450	88,4	913	88,8	6.100	80,5																		
2002	3.396	92,0	869	88,1	1.996	91,1	449	89,5	868	89,1	5.420	84,6	1.224	100,0																
2003	3.318	92,6	837	88,6	1.974	91,5	432	92,4	828	89,9	4.951	88,6	1.047	87,0																
2004	3.253	92,5	800	89,2	1.955	92,7	435	89,2	795	92,1	4.719	89,7	1.007	89,8																
2005	3.214	91,4	774	90,2	1.954	90,6	426	89,0	782	90,3	4.564	89,2	998	88,1																
2006	3.130	90,1	767	85,4	1.930	89,0	420	85,7	768	89,3	4.370	89,1	990	86,8	1.506	100,0														
2007	2.992	91,0	721	85,2	1.832	90,3	385	89,6	725	89,2	4.138	89,3	926	89,0	1.523	78,0														
2008	2.850	90,7	671	84,9	1.759	90,5	370	88,6	678	88,8	3.939	89,2	901	87,3	1.321	81,9														
2009	2.723	89,0	616	81,2	1.693	90,7	350	87,4	636	90,3	3.746	88,2	866	87,4	1.142	87,2	1.495	100,0												
2010	2.565	87,5	545	80,9	1.627	88,3	333	83,5	604	91,6	3.523	86,7	825	90,1	1.054	86,6	1.708	68,8					2.074	100,0	2.500	100,0				
2011	2.417	88,9	491	79,6	1.538	88,1	300	88,7	589	92,5	3.307	87,2	794	88,9	988	86,8			3.136	100,0			2.056	80,1	2.228	87,9	924	100,0		
2012	2.285	89,0	439	78,8	1.465	89,6	286	87,8	115	80,0	3.073	87,9	772	89,0	927	88,2			3.178	80,4	1.526	100,0	1.857	79,0	2.221	85,9	931	87,2		
2013	2.172	89,7	389	82,5	1.413	88,5	267	86,9	98	83,7	2.872	89,4	730	92,7	873	89,7			2.855	80,7	1.549	82,7	1.740	78,3	2.151	83,9	912	82,9	2.723	100,0
2014	2.064	90,8	356	84,8	1.346	90,0	249	85,5	90	86,7	2.732	88,4	720	89,0	823	88,9			2.497	84,5	1.438	82,5	1.501	83,1	1.990	71,2	824	72,7	2.787	72,2

Note: In the case of the initial wave of a sample, table entries are the number of participating households. See also Section 2.

4.2 Predicting the Probability of Re-Interviewing versus Refusal in the Year 2014

Based on the household and interview characteristics measured in the year 2013, and some regional information measured in 2014, we aim at predicting the probability of agreement vs. refusal to participate in the survey for households that were re-contacted in 2014. The individual attributes refer in many cases to the head of the household in the previous wave, but for split-off households the attributes are based on the information from the person who moved out of the panel household (in the case of several persons, the first person mentioned in the address protocol). In many other cases, personal information is aggregated at the level of households, for instance, rare events, such as the presence of individuals with an acute medical condition.

As in the case of predicting successful follow-ups, we use only model specifications where all included regressors are to be considered statistically significant (that is different from zero). The definition of the regressors is given in Table 4.2. Table 4.2 reports the subsample-specific estimates of logit models for the probability of participating relative to refusing to participate. Note again that the estimates of regression models of the previous waves 1985 through 2014 are not reported in the present documentation due to space restrictions. These can as well be found in previous attrition reports.

Table 4.2: Definition of the Regressors of the Logit Model of Refusal

Variable	Label	Value
Interview Characteristics		
Original Sample Member	Head of HH is Original Sample Member	0/1
New HH	HH is New in SOEP	0/1
Partial Unit Nonresponse	Person(s) in HH did not Participate	0/1
Temporary Drop-Out	Temporary Drop-Out of HH in Previous Year	0/1
Email Known	Email Address Disclosed	0/1
Phone Unknown	Telephone Number Undisclosed	0/1
Intercom worked	Interviewer Had no Problems with Intercom	0/1
Change of Interviewer	Change of Interviewer Between Last Waves	0/1
PAP-Interview	Paper-and-Pencil-Interview, filled out by respondent	0/1
Temp. Drop Out Related HH	Temporary Drop Out of Related HH	0/1
Drop Out Related HH	Ultimate Drop Out of Related HH	0/1
Exit Related HH	Exit of Related HH (Death/Emigration)	0/1
Interviewer Related HH	Same Interviewer in Related HH	0/1
Kita-Study Participation	HH Participated in Kindergarten-Study	0/1
Kita-Study Refusal	HH Refused to Participate in Kindergarten-Study	0/1
Youth Questionnaire	Adolescents in HH Filled out the Youth Questionnaire	0/1
Negative Follow-Up	Negative Reaction to Invitation for Follow-Up Survey (Sample M)	0/1
Late Interview	Interview Done Later than May	0/1
High Item Nonresponse HH	High Item Nonresponse in HH Questionnaire	0/1
High Item Nonresponse P	High Item Nonresponse in Person Questionnaire of Head of HH	0/1
Item Nonresponse Finan. Q.	Item Nonresponse in two or more financial questions	0/1
Many Missings Finan. Q.	No. of Item Nonresponse Above the Median of MV in Financial Questions	0/1
Demographic Characteristics		
Female Gender	Head of HH is Female	0/1

Table 4.2: Definition of the Regressors of the Logit Model of Refusal

Variable	Label	Value
Age 25-34	Head of HH between 25 and 34 Years	0/1
Age 35-44	Head of HH between 35 and 44 Years	0/1
Age 55-64	Head of HH between 55 and 64 Years	0/1
Age 65-74	Head of HH between 65 and 74 Years	0/1
Single HH	One Person Living in HH	0/1
In Relationship	Head of HH Currently in a Relationship	0/1
Child Under 12	At least one Child (younger than 12 Years) in HH	0/1
Family Household	4 or More Persons Live in HH	0/1
Moving In	Current Moving In HH	0/1
Not Born in GER	Respondent Born in a Foreign Country	0/1
Foreigner in HH	At least one Foreigner Lives in HH	0/1
Foreign Language	Foreign Native Language in HH	0/1
Health Situation		
Apoplectic Stroke	At Least one Person in HH Suffers an Apoplectic Stroke	0/1
H. Blood Pressure	At Least one Person in HH Suffers from High Blood Pressure	0/1
Joint Diseases	At Least one Person in HH Suffers from Joint Diseases	0/1
Healthwise Constraints	At Least one Person in HH Is Limited in Daily Life due to Health	0/1
Disabled	At Least one Person in HH Is Severely Disabled	0/1
Sick Note	At Least one Person in HH Was Certified Sick for more than 6 Weeks	0/1
Financial Situation, Real Estate and Insurance		
High Income	High HH-Income within the 4th Quartile	0/1
Receiving ALG-II	HH Receives Unemployment Benefit (Hartz IV)	0/1
Receiving Housing Benefits	HH Receives Housing Benefit	0/1
House Owner	Head of HH Is Owner of Dwelling	0/1
No Investments	HH Did not Invest in Previous Year	0/1

Table 4.2: Definition of the Regressors of the Logit Model of Refusal

Variable	Label	Value
Work and Education		
Unskilled Occupation	Head of HH Works as Untrained Staff	0/1
Blue-Collar Worker	Head of HH Is a Blue-Collar Worker	0/1
White-Collar Worker	Head of HH Is a White-Collar Worker	0/1
Civil Servant	Head of HH Is a Civil Servant	0/1
Self-Employed	Head of HH Is Self-Employed	0/1
Other Employment	Head of HH Employed, Neither Management Nor Untrained	0/1
Open-Ended Contract	Head of HH Has Working Contract with Indefinite Duration	0/1
Evening Shift	At Least one Person in HH Works in the Evening Hours	0/1
Night Work	At Least one Person in HH Works in the Night-Time	0/1
Full Employment	All HH-Members Are Employed	0/1
Job Change	Head of HH Did Change Job Previous Year	0/1
Paternal Education	Father of Head of HH has Completed Apprenticeship or University Degree	0/1
Personality Traits, Well-Being and Other Characteristics		
Unsatisfied with Dwelling	Head of HH Is a Little or Very Dissatisfied with Dwelling	0/1
Dwelling too Small	Head of HH Thinks Apartment Is Too Small	0/1
Strong Political Interest	Head of HH Has High or Very High Political Interest	0/1
Part. in Local Initiatives	At Least one Person in HH Participates in Citizens' Initiative	0/1
Hobbies and Leisure	Head of HH Spends Much Time With Hobbies/Leisure	0/1
Artistically Inclined	Head of HH Values Artistic Experiences and Has Lively Imagination	0/1
Scared	Head of HH Often Scared	0/1
Unhappy	Head of HH Often Not Happy	0/1
No Friends	Head of HH Has No Friends	0/1
Visited by Foreigner	Head of HH Was Not Visited by Foreigner(s) in Previous Year	0/1

Table 4.2: Definition of the Regressors of the Logit Model of Refusal

Variable	Label	Value
No Holiday Trip	HH Does Not Have a Holiday Trip (at least One Week) each Year	0/1
Building, Area, and Region		
Single Family House	HH Lives in a One or Two Family House	0/1
House in Bad Condition	Dwelling House in Need of Renovation	0/1
No Internet in HH	HH Is not on the Internet	0/1
Depopulation	HH Located in Area of High Depopulation	0/1
High Fluctuation	HH Located in Area with Much Fluctuation/Anonymity	0/1
Urban Area	HH Located in Cities with More than 100,000 Inhabitants	0/1
Highrise Area	HH Located in Area with Many Multistorey Buildings	0/1
Large Apartments	HH Located in Area with Large Apartments	0/1
High Status	HH Located in Area with High Status (Microm)	0/1
H. Average Age	HH Located in Area with High Average Age	0/1
Many Abitur-Graduates	HH Located in Area with Many Abitur-Graduates	0/1
High Academics	HH Located in Area of High Academics Rate	0/1
Low Voter Turnout	HH Located in Area with Low Turnout during Federal Election 2013	0/1
High Voter Turnout	HH Located in Area with High Turnout during Federal Election 2013	0/1
H. Grüne Share of Vote	HH Located in Area of High Share of Voting “Grüne”	0/1
Low Purchasing Power	HH in Area with Low of Purchasing-Power (Microm)	0/1
High Purchasing Power	HH in Area with High of Purchasing-Power (Microm)	0/1
H. Share Foreigners	HH in Area with High Share of Foreigners (Microm)	0/1
H. Share Islamic C.	HH in Area with High Share of People from Islamic Countries	0/1
High Share Eastern EU	HH in Area with High Share of Eastern EU or Late Repatriates	0/1
Baden-Wuerttemb.	HH Located in Baden-Wuerttemberg	0/1
Lower Saxony/Bremen	HH Located in Lower Saxony or Bremen	0/1
Saxony	HH Located in Saxony	0/1

Table 4.3: Estimates of Logit Models for the Probability of Re-
Interviewing a Household (Relative to Refusal) in 2014

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J	Sample K	Sample L1	Sample L2	Sample L3	Sample M
Intercept	1.54***	0.52*	0.93***	1.26***	1.23***	1.41***	1.02***	1.60***	1.21***	0.64***	0.62***	0.17*	-0.03	0.64***
<u>Interview Characteristics</u>														
Original Sample Memb.	-0.30***			-0.52**							-0.60***			-0.58***
New HH	-0.61***		-1.17***			-1.06***			-0.55***			-2.21***		
Part. Unit Nonresponse						-0.30***			-0.23***	-0.35***		-0.34***		-0.44***
Temporary Drop-Out	-1.19***		-1.34***			-1.65***	-1.08***	-1.67***	-1.96***	-1.76***	-0.74***	-1.59***	-0.81**	
Email Known				0.68***						0.20**		0.30***	0.45***	0.63***
Phone Unknown	-0.57***		-0.29**				-0.53**	-0.77***	-0.79***	-0.81***	-0.36**			-0.85***
Intercom worked														-0.10**
Change of Interviewer	-0.71***	-1.41***	-1.12***	-1.54***		-0.91***	-0.52***	-0.45***	-0.51***	-0.24**		0.18***	0.19**	-0.16**
PAP-Interview			-0.26***											
Temp. Drop Out Related HH	0.23***													
Drop Out Related HH						-0.48***								
Exit of Related HH				-0.89**		-0.50***								
Interviewer Related HH													0.42**	
Kita-Study Participation											0.30***	0.27**		
Kita-Study Refusal	-1.39***					-1.47***			-0.93***				-0.43**	
Youth Questionnaire												-0.20**		
Negative Follow-up														-0.95***
Late Interview	0.37**		-0.66***			-0.26**			-0.21**	-0.47***	-0.30***			
High Item Nonresp. HH						-0.16**	-0.38**							-0.20***
High Item Nonresp. P		-0.80***					-0.37**	-0.44***						
Item Nonresponse Finan. Q.	-0.42***													
Many Missings Finan. Q.						-0.25**						0.32**		
<u>Demographic Characteristics</u>														
Female Gender		-0.46**												
Age 25-34		-1.04***			-1.50***									
Age 35-44				-0.72***								-0.14**		
Age 55-64	0.24**						0.48***							
Age 65-74								0.34**	0.24***					
Single HH	-0.42***					-0.25***		-0.42***	-0.31***	-0.25***			-1.04***	-0.40***

Note: ***p < 0.01; **p < 0.05; *p < 0.10.

Table 4.3: Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2014

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J	Sample K	Sample L1	Sample L2	Sample L3	Sample M
In Relationship	-0.28**								-0.23**		-0.46***			
Child Under 12									0.19**					
Family Household											0.39***	0.28***		
Moving In	-0.41***													
Not Born in GER													-0.41***	
Foreigner in HH.						-0.44***				-0.39***				
Foreign Language		0.87***												
Health Situation														
Apoplectic Stroke	-0.52***													
H. Blood Pressure						-0.14**								
Joint Diseases									0.20***					
Healthwise Constraints		0.55***					-0.34***					0.15**		
Disabled								0.36**					-0.39**	
Sick Note		-1.00***								-0.34**				
Financial Situation, Real Estate and Insurance														
High Income		-0.80***												
Receiving ALG-II									0.29**					
Receiving Housing benefit						-0.50**								
House Owner	-0.17**	0.57**												
No Investments			-0.35***						-0.25***			-0.24***		
Work and Education														
Unskilled Occupation											0.28***			
Blue-Collar Worker			-0.28**											
White-Collar Worker		0.67***												
Civil Servant											-0.49***		0.53***	
Self-Employed												-0.34***		-0.23**
Other Employment														-0.17**
Open-Ended Contract										0.25**				
Evening Shift									-0.15**					
Night Work		-0.59**				-0.21**								
Full Employment							0.35***				0.43***			
Job Change			0.27**										0.26**	

Note: ***p < 0.01; **p < 0.05; *p < 0.10.

Table 4.3: Estimates of Logit Models for the Probability of Re-
Interviewing a Household (Relative to Refusal) in 2014

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J	Sample K	Sample L1	Sample L2	Sample L3	Sample M
Paternal Education														0.17***
Personality Traits, Well-Being and Other Characteristics														
Unsatisfied with Dwelling		-0.77***							-0.31***					
Dwelling too Small						-0.21**					-0.23***			
Strong Political Interest	0.17**													
Part. in Local Initiatives						-0.16**								
Hobbies and Leisure									-0.15**					
Artistically inclined		0.61***	0.26***											
Scared											0.38**			
Unhappy								0.13**						
No Friends				-0.80**				-0.55***		-0.27**				
Visited by Foreigner						0.16***								
No Holiday Trip			0.23**							-0.17**				
Building, Area, and Region														
Single Family House												-0.23***		
House in Bad Condition							0.31**							0.17***
No Internet in HH								0.29**						
Depopulation												0.20***		
High Fluctuation			0.28**								0.27***			
Urban Area												-0.21***		-0.14**
Highrise Area									0.19**					
Large Apartments											0.30***			
High Status												0.24***		
H. Average Age	0.22**										0.33***			
Many Abitur-Grad.			0.25***					-0.39***		0.33***				
High Academics														-0.17***
Low Voter turnout								-0.36**			0.32***			
High Voter turnout						-0.13**		-0.52***	-0.15**					
H. Grüne Share of Vote										0.28***				
Low Purchasing Power							-0.41**							
High Purchasing Power														-0.21***
H.Share Foreigners					-1.03**									

Note: ***p < 0.01; **p < 0.05; *p < 0.10.

Table 4.3: Estimates of Logit Models for the Probability of Re-
Interviewing a Household (Relative to Refusal) in 2014

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J	Sample K	Sample L1	Sample L2	Sample L3	Sample M
H. Share Islamic C.										0.24**				
H. Share Eastern EU						0.35***		-0.51***						-0.11**
Baden-Wuerttemb.											0.36***	-0.23***		
Lower Saxony/Bremen			-0.53**				-0.68***							
Saxony											-0.31**			
<i>No. of Observations</i>	2065	359	1346	249	90	2732	720	823	2497	1439	1503	1990	824	2793
<i>Log Likelihood</i>	-480.51	-98.94	-313.43	-67.74	-24.81	-722.37	-203.11	-213.62	-798.9	-476.83	-558.09	-997.13	-421.65	-1432.64

Note: ***p < 0.01; **p < 0.05; *p < 0.10.

5 Summary Statistics of the Derived Longitudinal and Cross-Sectional Weights

Based on the regression models of successful vs. unsuccessful recontacts and agreements vs. refusals to participate, we derive two sets of predicted probabilities, the product of which is the household's "staying probability". The inverse of the probability of staying in the SOEP in 2014 based on characteristics measured in 2013, variable `BEHBLEIB`, lends itself as a longitudinal weighting variable which itself corrects for selective attrition between waves 2013 and 2014. Tables 5.1, Table 5.2, Table 5.3 and Table 5.4 report some sub-sample specific summary statistics of the longitudinal weights in each wave.

The product of the cross-sectional weight in 2013, variable `BDHHRF`, and the longitudinal weight in 2014, variable `BEHBLEIB`, provide the raw data for the cross-sectional weight in 2014. In a final step, the post-stratification of the cross-sectional weights corrects them to meet benchmarks of known marginal distribution characteristics of the underlying population as of the year 2013. At the household level, these variables are the states (Bundesländer), size of the community, household size, and house ownership. At the person level, SOEP weights are also adjusted to the marginal distributions of age, gender, and nationality (Non/German). With the integration of the latest migrant sample M in 2013, we also consider additional information on the country of origin of respondents and year of migration. With the integration of Samples L1, L2, and L2 in 2014, we also employ more detailed information of the micro-census on household typologies. Tables 5.5 and 5.6 report sub-sample specific summary statistics of the derived cross-sectional weighting variable `BEHHRF` and in comparison all previous cross-sectional weights `AHHRF` through `BDHHRF`.

Table 5.1: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples A through D (Percentiles of \$HBLEIB up to Wave BE).

Year	Sample A				Sample B				Sample C				Sample D			
	p10	p50	p90	N	p10	p50	p90	N	p10	p50	p90	N	p10	p50	p90	N
1985	1.06	1.10	1.22	4141	1.09	1.10	1.26	1181								
1986	1.04	1.07	1.26	3962	1.10	1.10	1.29	1128								
1987	1.03	1.03	1.13	3910	1.03	1.03	1.14	1116								
1988	1.02	1.04	1.20	3743	1.03	1.04	1.22	1071								
1989	1.03	1.04	1.16	3647	1.03	1.04	1.14	1043								
1990	1.02	1.02	1.11	3612	1.04	1.04	1.12	1028								
1991	1.02	1.02	1.09	3613	1.03	1.03	1.16	1056	1.03	1.06	1.18	2030				
1992	1.01	1.02	1.11	3585	1.01	1.03	1.16	1060	1.06	1.06	1.22	2020				
1993	1.01	1.01	1.16	3603	1.02	1.03	1.22	1064	1.03	1.04	1.17	1970				
1994	1.02	1.02	1.15	3577	1.03	1.05	1.22	1023	1.02	1.04	1.12	1959				
1995	1.01	1.01	1.16	3526	1.02	1.05	1.29	982	1.03	1.03	1.11	1938				
1996	1.01	1.03	1.12	3485	1.04	1.04	1.21	960	1.01	1.02	1.15	1951	1.00	1.08	1.16	396
1997	1.01	1.02	1.13	3458	1.02	1.04	1.29	931	1.02	1.04	1.12	1942	1.05	1.09	1.09	340
1998	1.02	1.03	1.14	3387	1.04	1.07	1.23	898	1.02	1.02	1.20	1886	1.08	1.08	1.35	308
1999	1.02	1.02	1.20	3325	1.04	1.04	1.22	858	1.01	1.03	1.10	1894	1.05	1.05	1.27	300
2000	1.02	1.02	1.15	3240	1.03	1.03	1.18	820	1.01	1.03	1.13	1879	1.02	1.02	1.10	302
2001	1.02	1.02	1.18	3168	1.02	1.02	1.23	809	1.02	1.02	1.16	1850	1.03	1.03	1.18	286
2002	1.01	1.02	1.21	3123	1.04	1.04	1.37	766	1.01	1.02	1.21	1818	1.00	1.02	1.21	289
2003	1.01	1.03	1.14	3072	1.01	1.03	1.31	742	1.01	1.03	1.14	1807	1.01	1.01	1.09	290
2004	1.01	1.01	1.12	3010	1.04	1.04	1.13	714	1.00	1.01	1.12	1813	1.00	1.01	1.25	277
2005	1.02	1.02	1.16	2937	1.05	1.05	1.17	698	1.00	1.02	1.15	1771	1.00	1.02	1.34	273
2006	1.01	1.04	1.22	2821	1.01	1.05	1.33	655	1.01	1.04	1.24	1717	1.03	1.04	1.44	261
2007	1.01	1.03	1.14	2723	1.03	1.07	1.24	614	1.00	1.03	1.16	1654	1.01	1.04	1.12	248
2008	1.02	1.05	1.13	2584	1.01	1.07	1.25	570	1.01	1.03	1.18	1592	1.02	1.07	1.22	231
2009	1.02	1.05	1.25	2423	1.01	1.05	1.60	500	1.00	1.03	1.21	1535	1.00	1.02	1.16	217
2010	1.01	1.06	1.38	2245	1.01	1.10	1.47	441	1.01	1.04	1.32	1437	1.00	1.01	1.43	278
2011	1.00	1.04	1.27	2148	1.01	1.07	1.55	391	1.01	1.05	1.24	1355	1.01	1.02	1.28	266
2012	1.02	1.08	1.27	2033	1.01	1.13	1.65	346	1.00	1.05	1.29	1312	1.00	1.04	1.45	251
2013	1.01	1.06	1.25	1949	1.01	1.09	1.58	321	1.01	1.07	1.28	1250	1.01	1.06	1.39	232
2014	1.01	1.04	1.25	1874	1.01	1.04	1.48	302	1.01	1.04	1.22	1215	1.00	1.03	1.31	213

Table 5.2: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples E through G (Percentiles of \$HBLEIB up to Wave BE).

Year	Sample E				Sample F				Sample G			
	p10	p50	p90	N	p10	p50	p90	N	p10	p50	p90	N
1999	1.00	1.23	1.47	886								
2000	1.03	1.07	1.21	838								
2001	1.01	1.05	1.25	811	1.08	1.14	1.59	4911				
2002	1.01	1.02	1.20	773	1.03	1.05	1.46	4586				
2003	1.04	1.04	1.15	744	1.02	1.04	1.24	4386	1.06	1.10	1.17	911
2004	1.00	1.01	1.08	732	1.02	1.03	1.19	4235	1.02	1.03	1.25	904
2005	1.01	1.03	1.18	706	1.01	1.03	1.17	4070	1.03	1.06	1.25	879
2006	1.00	1.03	1.21	686	1.01	1.03	1.29	3895	1.00	1.04	1.31	859
2007	1.01	1.01	1.16	647	1.02	1.03	1.15	3694	1.02	1.05	1.17	824
2008	1.00	1.01	1.19	602	1.01	1.03	1.14	3513	1.01	1.03	1.18	787
2009	1.00	1.04	1.17	574	1.02	1.04	1.34	3303	1.02	1.04	1.36	757
2010	1.01	1.04	1.25	553	1.01	1.05	1.40	3055	1.00	1.01	1.23	743
2011	1.00	1.00	1.17	545	1.01	1.05	1.34	2885	1.00	1.03	1.35	706
2012	1.05	1.24	1.66	92	1.02	1.08	1.30	2702	1.02	1.07	1.24	687
2013	1.07	1.20	1.32	82	1.01	1.06	1.21	2567	1.02	1.05	1.15	677
2014	1.03	1.03	1.42	78	1.02	1.05	1.23	2414	1.01	1.07	1.32	641

Table 5.3: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples H, J, K and M (Percentiles of \$HBLEIB up to Wave BE).

Year	Sample H				Sample J				Sample K				Sample M			
	p10	p50	p90	N	p10	p50	p90	N	p10	p50	p90	N	p10	p50	p90	N
2007	1.04	1.16	1.46	1188												
2008	1.01	1.03	1.18	1082												
2009	1.01	1.03	1.22	996												
2010	1.01	1.04	1.37	913												
2011	1.00	1.05	1.31	858												
2012	1.00	1.03	1.36	818	1.05	1.19	1.52	2555								
2013	1.00	1.05	1.27	783	1.03	1.13	1.36	2305	1.04	1.15	1.47	1281				
2014	1.01	1.05	1.27	732	1.03	1.09	1.30	2110	1.02	1.09	1.34	1187	1.08	1.28	1.81	2012

Table 5.4: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples L1, L2 and L3 (Percentiles of \$HBLEIB up to Wave BE).

Year	Sample L1				Sample L2				Sample L3			
	p10	p50	p90	N	p10	p50	p90	N	p10	p50	p90	N
2011	1.10	1.20	1.46	1,647	1.03	1.12	1.37	1,958				
2012	1.04	1.16	1.58	1,467	1.03	1.11	1.35	1,907	1.01	1.10	1.37	806
2013	1.03	1.11	1.59	1,362	1.03	1.09	1.37	1,805	1.02	1.11	1.47	750
2014	1.03	1.12	1.45	1,247	1.09	1.25	1.69	1,416	1.10	1.24	1.80	599

Table 5.5: Summary Statistics of the Derived Cross-Sectional Weights at the Household Level (Percentiles of \$HHRF up to Wave 31).

Year	p5	p10	p25	p50	p75	p90	p95	N
1984	431	597	3805	4725	5647	7130	8248	5921
1985	482	682	3905	5081	6430	8474	10033	5322
1986	538	753	3597	5303	6838	9280	11118	5090
1987	546	790	3533	5381	7044	9575	11459	5026
1988	534	804	3555	5638	7542	10350	12536	4814
1989	549	820	3603	5841	7881	10812	13275	4690
1990	696	1073	2217	4594	7044	9878	12395	6819
1991	684	1044	2331	4690	7153	10288	12873	6699
1992	670	1027	2339	4651	7136	10531	13656	6665
1993	691	1056	2403	4671	7255	10752	13977	6637
1994	711	1099	2396	4664	7279	11223	14705	6559
1995	700	1114	2384	4364	6981	11083	14834	6768
1996	740	1162	2387	4356	7017	11379	15311	6699
1997	749	1205	2402	4317	7058	11851	15877	6621
1998	983	1353	2331	3977	6218	9884	13113	7492
1999	969	1326	2311	3982	6495	10883	14351	7220
2000	804	1102	1761	2524	3569	5088	6491	13082
2001	752	1031	1756	2750	4143	6092	7842	11796
2002	507	658	1221	2553	4190	6523	8240	12320
2003	504	676	1234	2561	4329	6827	9080	11909
2004	492	670	1215	2538	4422	7262	9834	11644
2005	491	678	1237	2545	4521	7582	10863	11294
2006	457	650	1268	2392	4139	6884	9733	12361
2007	456	653	1251	2469	4461	7568	10690	11552
2008	458	656	1278	2553	4752	8226	11600	10921
2009	473	668	1301	2633	5037	9062	12431	10270
2010	222	363	664	1424	3654	7389	11050	13888
2011	214	325	610	1506	3094	5607	7807	16703
2012	217	327	633	1637	3169	5714	7585	16397
2013	177	267	516	1301	2945	5356	7444	17992
2014	202	310	617	1521	3348	6225	8561	15946

Table 5.6: Summary Statistics of the Derived Cross-Sectional Weights at the Person Level (Percentiles of \$PHRF up to Wave 31).

Year	p5	p10	p25	p50	p75	p90	p95	N
1984	386	538	1159	4365	5229	6066	6888	16173
1985	461	633	1430	4624	5713	6858	8111	14508
1986	489	678	1538	4691	6026	7585	8984	13804
1987	510	712	1601	4703	6225	7913	9360	13563
1988	489	680	1619	4878	6559	8527	10141	12872
1989	529	753	1744	5012	6873	8972	10706	12443
1990	686	1023	1908	3443	6165	8302	10185	18254
1991	735	1072	1902	3716	6204	8473	10640	17844
1992	783	1131	1982	3739	6295	8730	11148	17429
1993	843	1232	2064	3824	6363	9003	11417	17072
1994	871	1271	2091	3824	6400	9270	12127	16715
1995	769	1138	2003	3586	6079	9117	12328	17345
1996	801	1182	2017	3639	6133	9409	12813	16944
1997	844	1214	2054	3648	6210	9712	13412	16583
1998	911	1259	2036	3494	5590	8526	11459	18249
1999	904	1238	2010	3487	5809	9318	12547	17501
2000	723	968	1550	2309	3230	4576	5845	30784
2001	690	930	1524	2445	3648	5434	6930	27956
2002	431	592	1049	2191	3720	5836	7653	29101
2003	432	604	1070	2193	3814	6168	8296	27867
2004	429	601	1068	2187	3915	6512	8879	26918
2005	434	620	1100	2238	4026	6897	9623	25638
2006	406	587	1108	2181	3696	6305	8826	27442
2007	411	589	1114	2236	3936	6979	10127	25505
2008	421	603	1143	2312	4137	7634	11188	23792
2009	435	618	1179	2392	4366	8363	12617	22096
2010	175	277	532	1027	2524	5437	8570	35945
2011	164	249	454	986	2395	4413	6623	42031
2012	165	249	469	1099	2544	4500	6707	40351
2013	143	214	405	894	2246	4285	6368	44633
2014	158	244	477	1077	2596	4916	7286	38930

References

- Haisken-DeNew, J. and J. Frick (2001). *Desktop Companion to the Socio Economic Panel (SOEP)*.
- Kroh, M., R. Siegers, and S. Kühne (2015). *Gewichtung und Integration von Auffrischungsstichproben am Beispiel des Sozio-oekonomischen Panels (SOEP)*. In: Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen. Ed. by J. Schupp and C. Wolf. Wiesbaden: Springer. pp.409–444.
- Rendtel, U. (1995). *Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität*. Vol. 8. Campus Verlag.
- Schonlau, M., N. Watson, and M. Kroh (2010). “Household survey panels: how much do following rules affect sample size?” In:
- Schonlau, M., M. Kroh, N. Watson, et al. (2013). “The implementation of cross-sectional weights in household panel surveys”. In: *Statistics Surveys* 7, pp. 37–57.
- Spiess, M., M. Kroh, R. Pischner, and G. G. Wagner (2008). *On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP) – Tracing, Weighting, and Frequencies*. SOEPpapers 98, DIW/SOEP, Berlin 2008.