SOEP The German Socio-Economic Panel Study

177

SOEP Survey Papers

Series D - Variable Description and Coding

SOEP — The German Socio-Economic Panel Study at DIW Berlin 2014

Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2012)

Martin Kroh

DIW SOEP

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 Documentations (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:
Prof. Dr. Gert G. Wagner, DIW Berlin and Technische Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Kroh, Martin. 2014. Documentation of Sample Sizes and Panel Attrition in the German Socio-Economic Panel (SOEP) (1984 until 2012). SOEP Survey Papers 177: Series D. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soeppapers@diw.de

MARTIN KROH

DOCUMENTATION OF SAMPLE SIZES AND PANEL ATTRITION IN THE GERMAN SOCIO-ECONOMIC PANEL (SOEP) (1984 UNTIL 2012)

Berlin, 2013

Reprint 2014

Contents

Tables

Figures

1 Introduction

This data documentation is meant to provide SOEP users with a general overview of the longitudinal development of the survey over the past 29 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 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. (2011). 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 and Schonlau et al. 2013 and for a general overview, Haisken-DeNew & 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 2012 based on the characteristics of households measured in 2011. 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 2012: BCHBLEIB and BCPBLEIB. Based on the inverse of the probability of observing households and persons in 2011, the staying probability in 2012, and additional post-stratification to meet benchmarks of known marginals of the underlying population in 2012, we derive the cross-sectional weights BCHHRF and BCPHRF. 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.

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 in 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. Five percent 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 original wave 1 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

The following figures display the number of successful interviews considering different aspects:

Figure 1: **The Number of Successful Interviews with Persons by Subsamples A through K, Waves 1 to 29.**

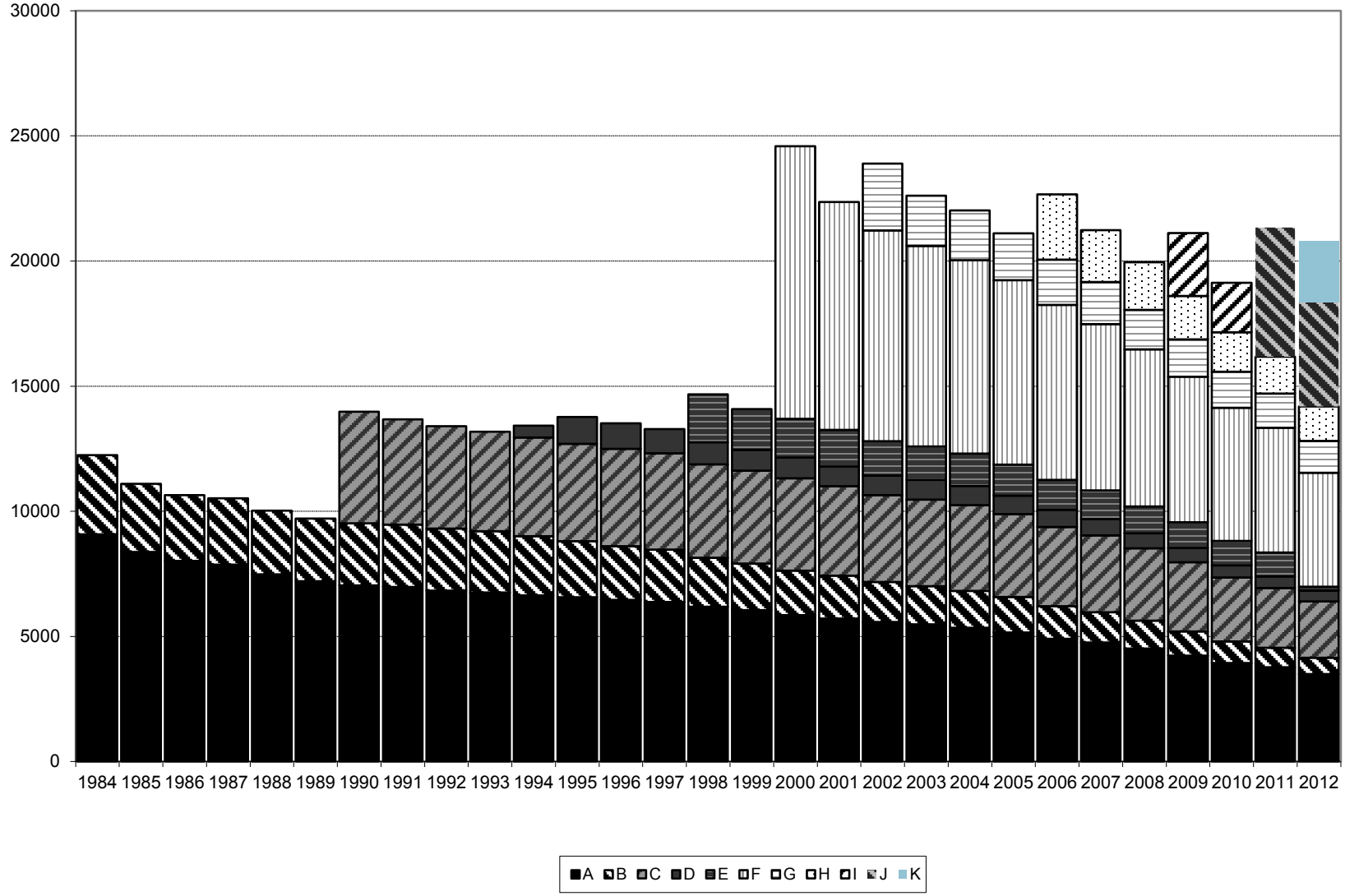
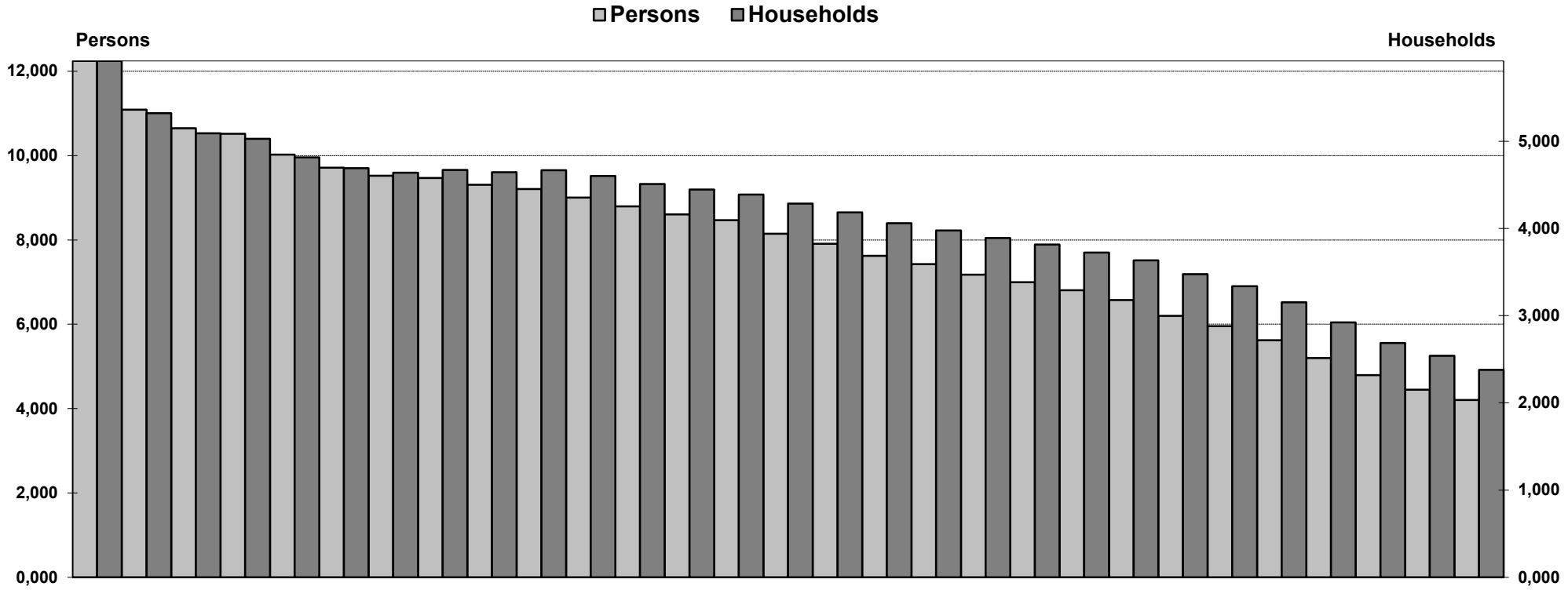
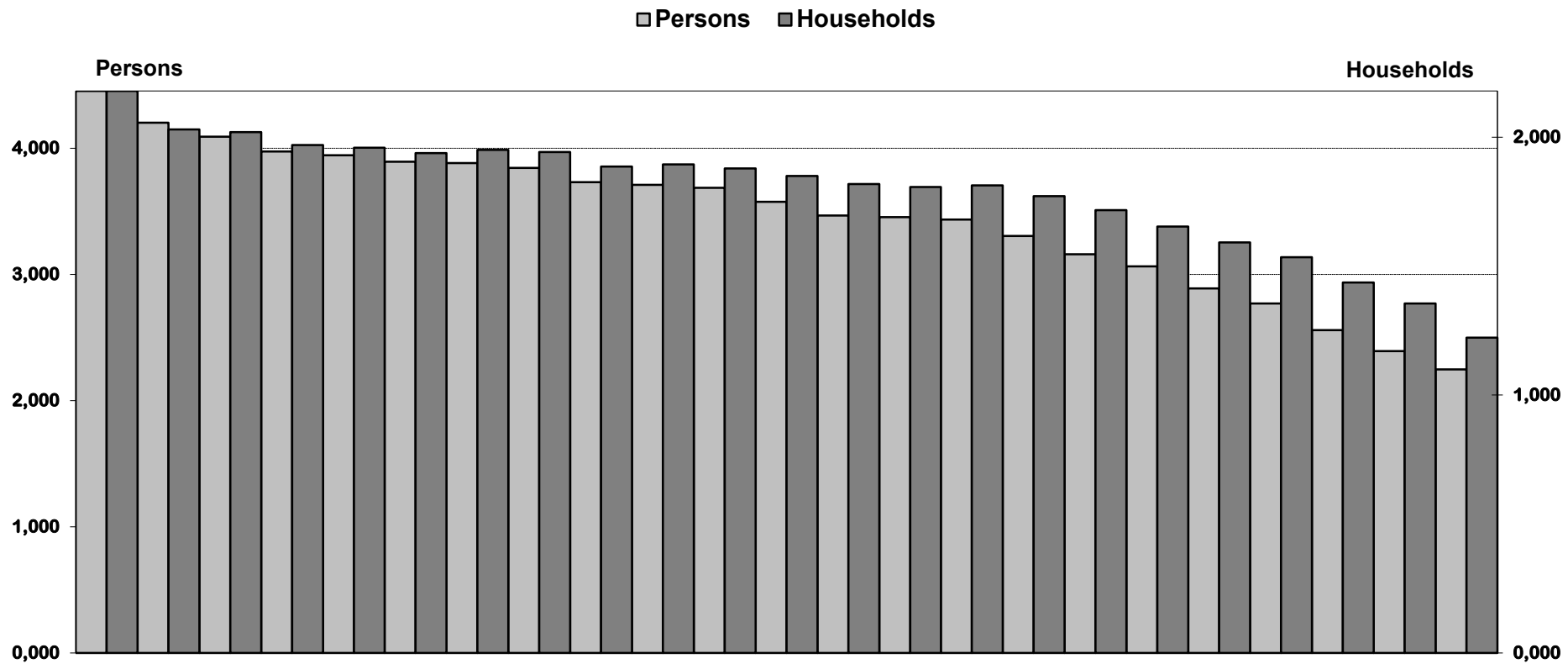


Figure 2**: Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 29.**



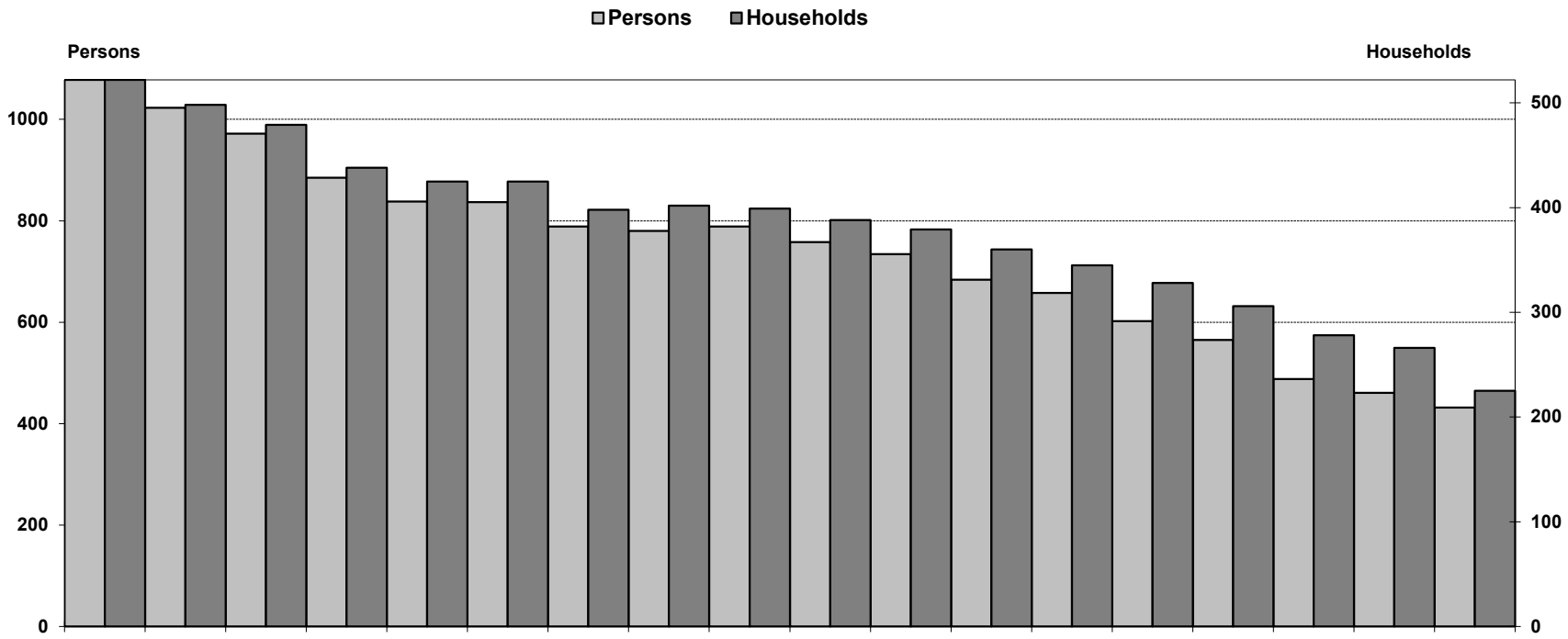
Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Persons	12,245	11,090	10,646	10,516	10,023	9,710	9,519	9,467	9,305	9,206	9,001	8,798	8,606	8,467	8,145	7,909	7,623	7,424	7,175	6,999	6,809	6,572	6,198	5,957	5,619	5,197	4.739	4,451	4,204
Households	5,921	5,322	5,090	5,026	4,814	4,690	4,640	4,669	4,645	4,667	4,600	4,508	4,445	4,389	4,285	4,183	4,060	3,977	3,889	3,814	3,724	3,635	3,476	3,337	3,154	2,923	2.686	2,539	2,379

Figure 3:**Comparison of Successful Interviews with Persons and Households (Subsample C), Waves 1 to 23.**



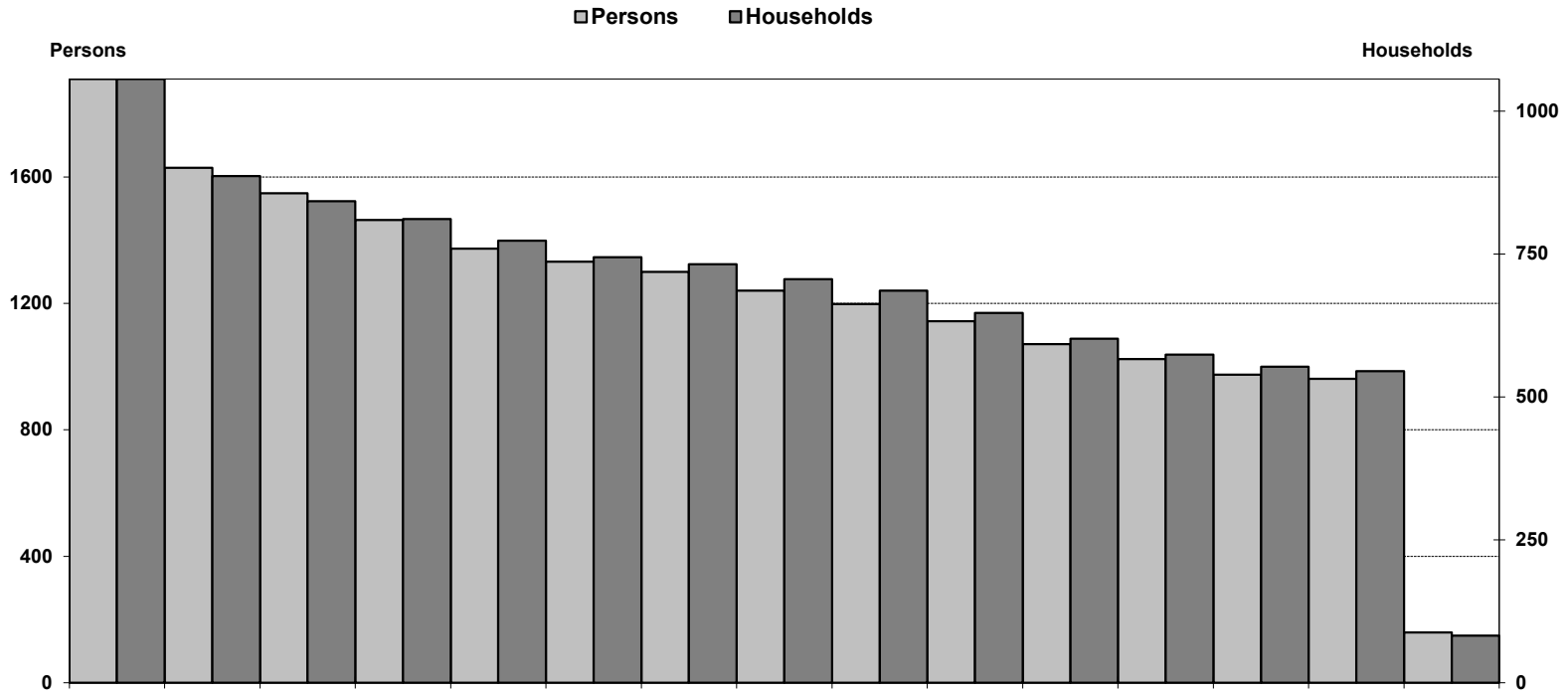
Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Persons	4,453	4,202	4,092	3,973	3,945	3,892	3,882	3,844	3,730	3,709	3,687	3,576	3,466	3,453	3,435	3,304	3,159	3,063	2,889	2,769	2,559	2,392	2,247
Households	2,179	2,030	2,020	1,970	1,959	1,938	1,951	1,942	1,886	1,894	1,879	1,850	1,818	1,807	1,813	1,771	1,717	1,654	1,592	1,535	1,437	1,355	1,223

Figure 4: **Comparison of Successful Interviews with Individuals and Households (Subsample D), Waves 1 to 18.**



Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Persons	1,078	1,023	972	885	838	837	789	780	789	758	734	684	658	602	565	488	461	432
Households	522	498	479	441	425	425	398	402	399	388	379	360	345	328	306	278	266	225

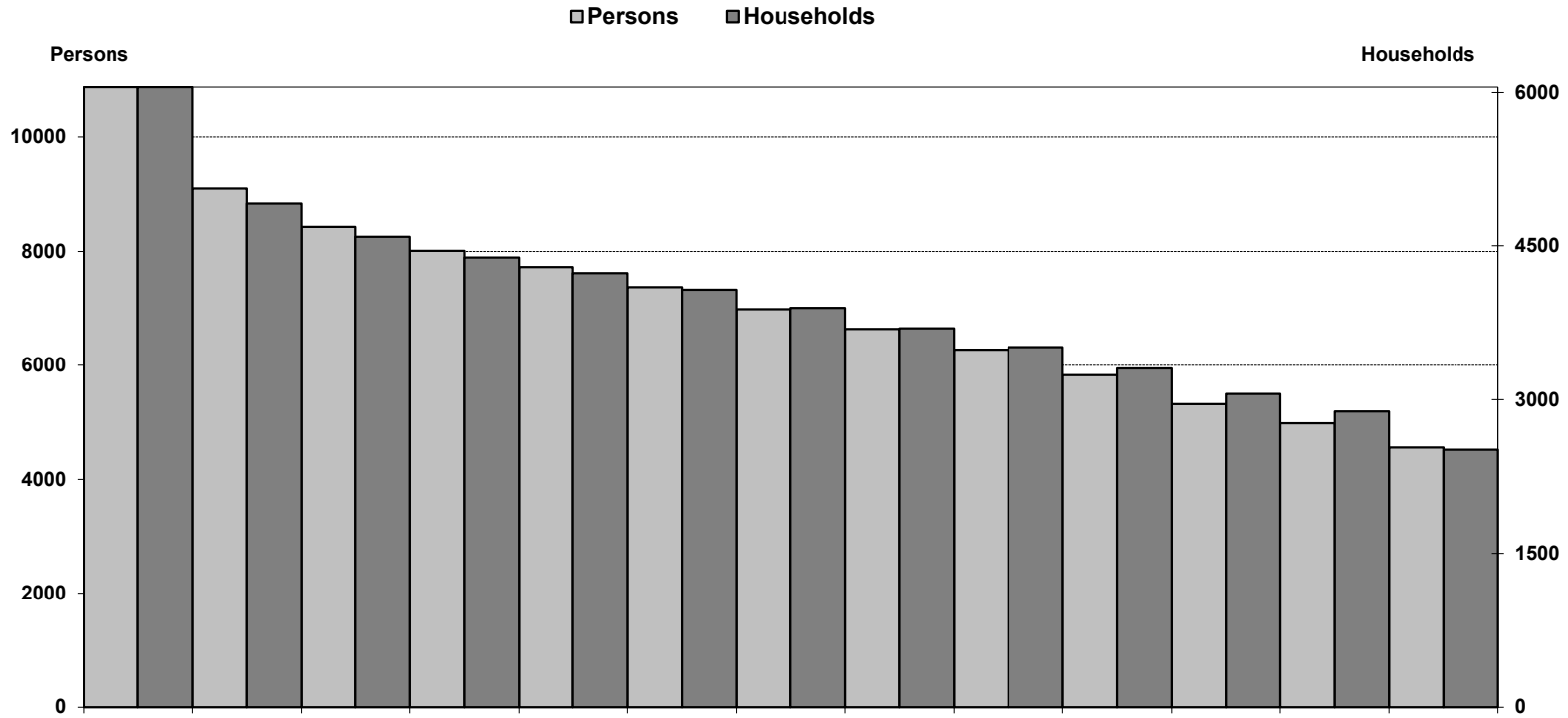
Figure 5: **Comparison of successful interviews with individuals and households (Subsample E), Waves 1 to 15.**



Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Persons	1,910	1,629	1,549	1,464	1,373	1,332	1,300	1,240	1,198	1,144	1,071	1,024	975	961	159
Households	1056	886	842	811	773	744	732	706	686	647	602	574	553	545	82

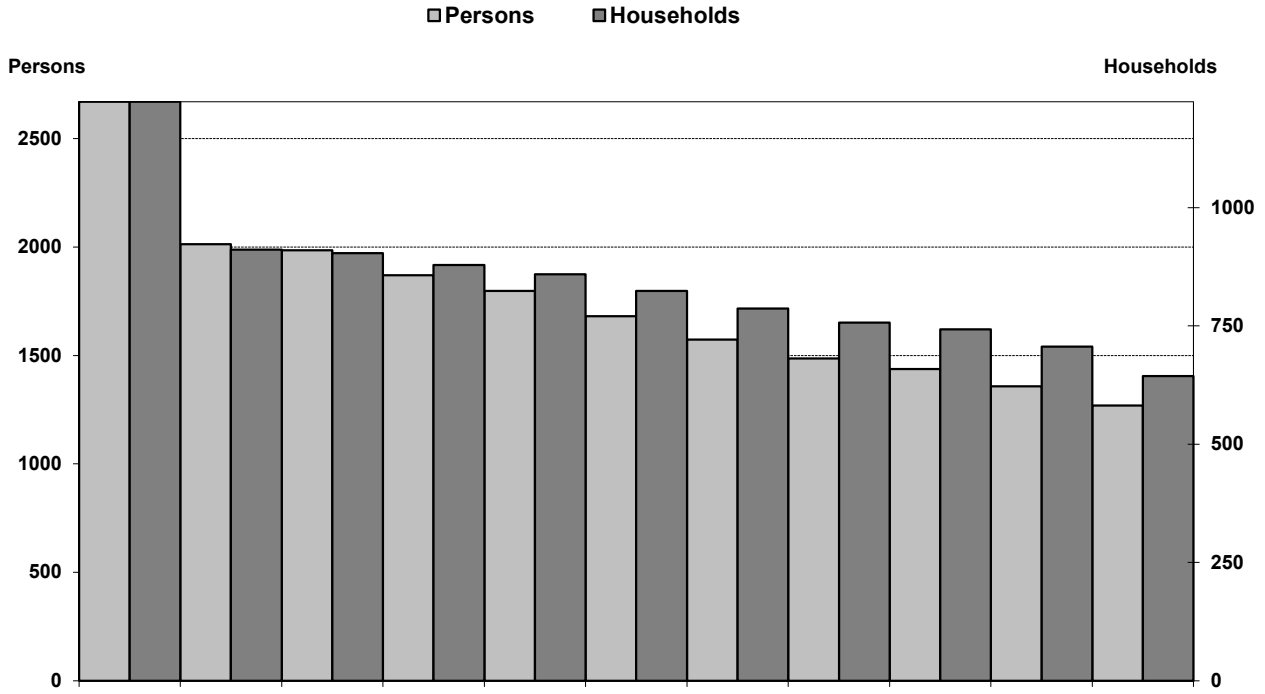
Note. In Wave BC (2012), Sample E was split between SOEP-Innovation Sample (SOEP-IS) and SOEP-Core (N=82 Households).

Figure 6: **Comparison of Successful Interviews with Individuals and Households (Subsample F), Waves 1 to 13.**



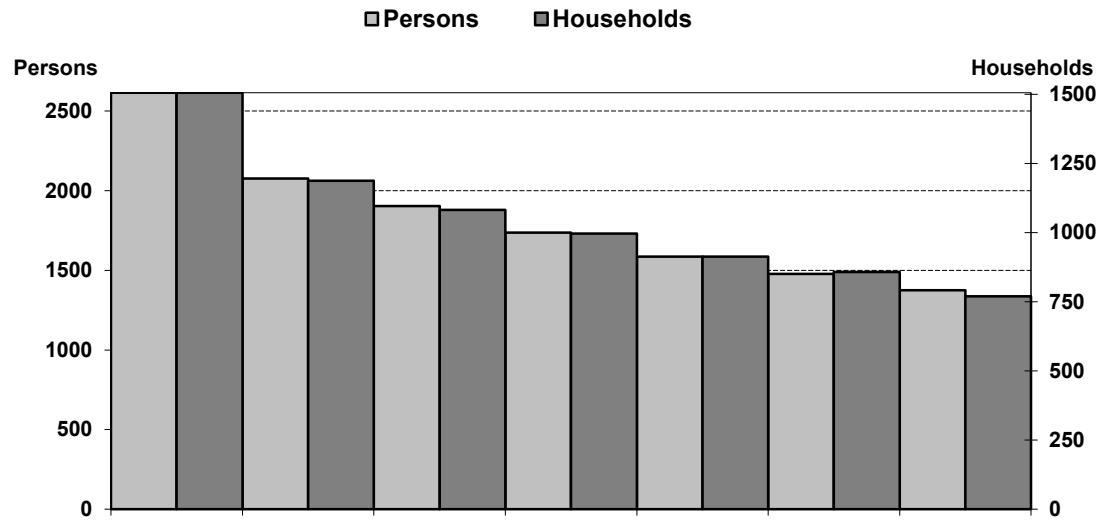
Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Persons	1,0890	9,098	8,427	8,006	7,724	7,371	6,986	6640	6,274	5,824	5,316	4,984	4,556
Households	6,052	4,911	4,586	4,386	4,234	4,070	3,895	3,694	3,513	3,303	3,055	2,885	2,511

Figure 7: **Comparison of Successful Interviews with Individuals and Households (Subsample G), Waves 1 to 11.**



Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Persons	2,671	2,013	1,986	1,870	1,798	1,682	1,574	1,487	1,438	1,358	1,270
Households	1,224	911	904	879	859	824	787	757	743	706	644

Figure 8: **Comparison of Successful Interviews with Individuals and Households (Subsample H), Waves 1 to 7.**



Year	2006	2007	2008	2009	2010	2011	2012
Persons	2,616	2,077	1,904	1,737	1,587	1,478	1,376
Households	1,506	1,188	1,082	996	913	858	770

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, exits due to survey-unrelated attrition, and exits due to survey-related attrition.

Figure 9: **All First-Wave Persons (Gross Subsample A). Development up to Wave 29.**

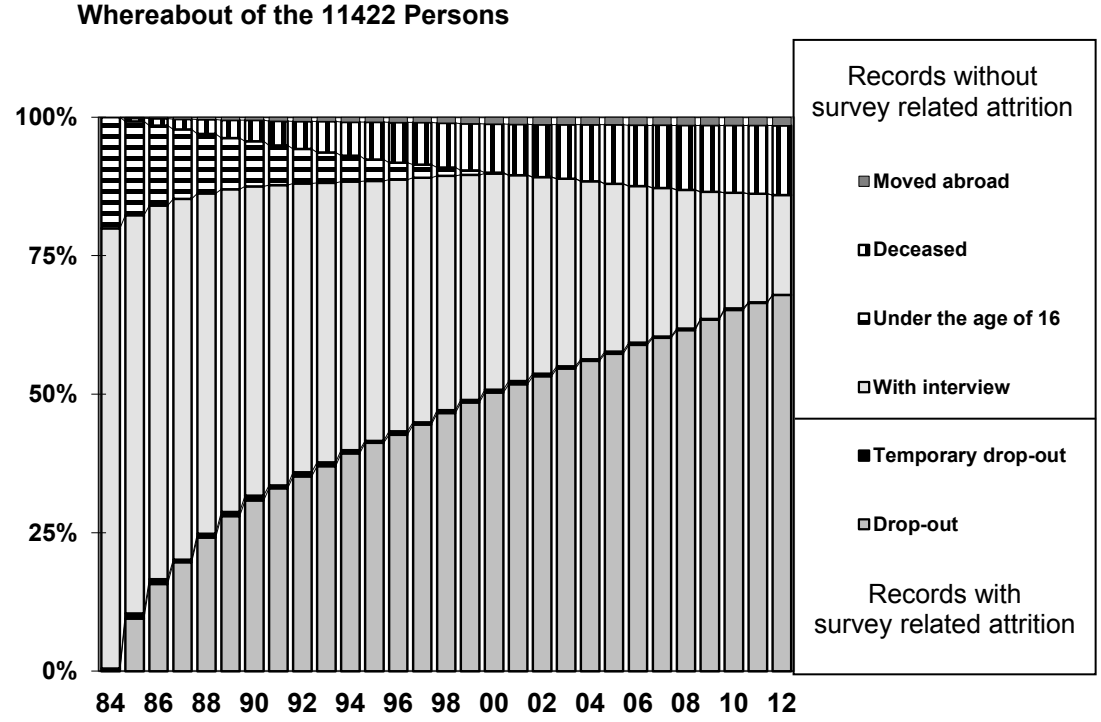


Figure 10: **All First Wave Persons (Gross Subsample B). Development up to Wave 29.**

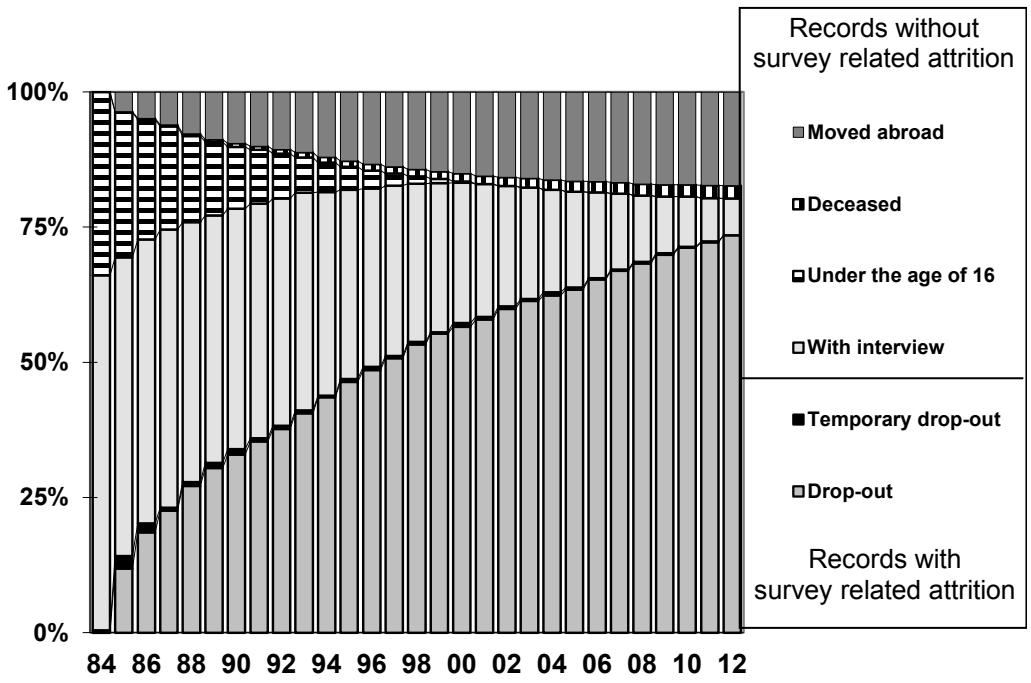


Figure 11: **All First Wave Persons (Gross Subsample C). Development up to Wave 23.**

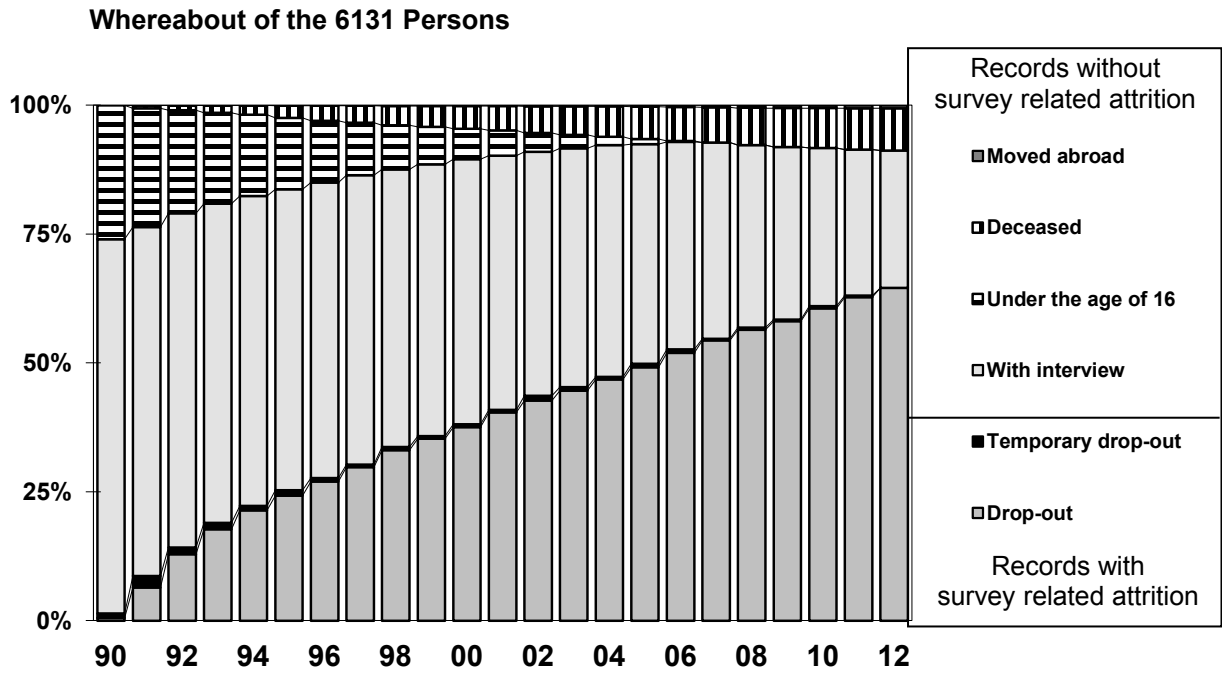


Figure 12: **All First Wave Persons (Gross Subsample D). Development up to Wave 18.**

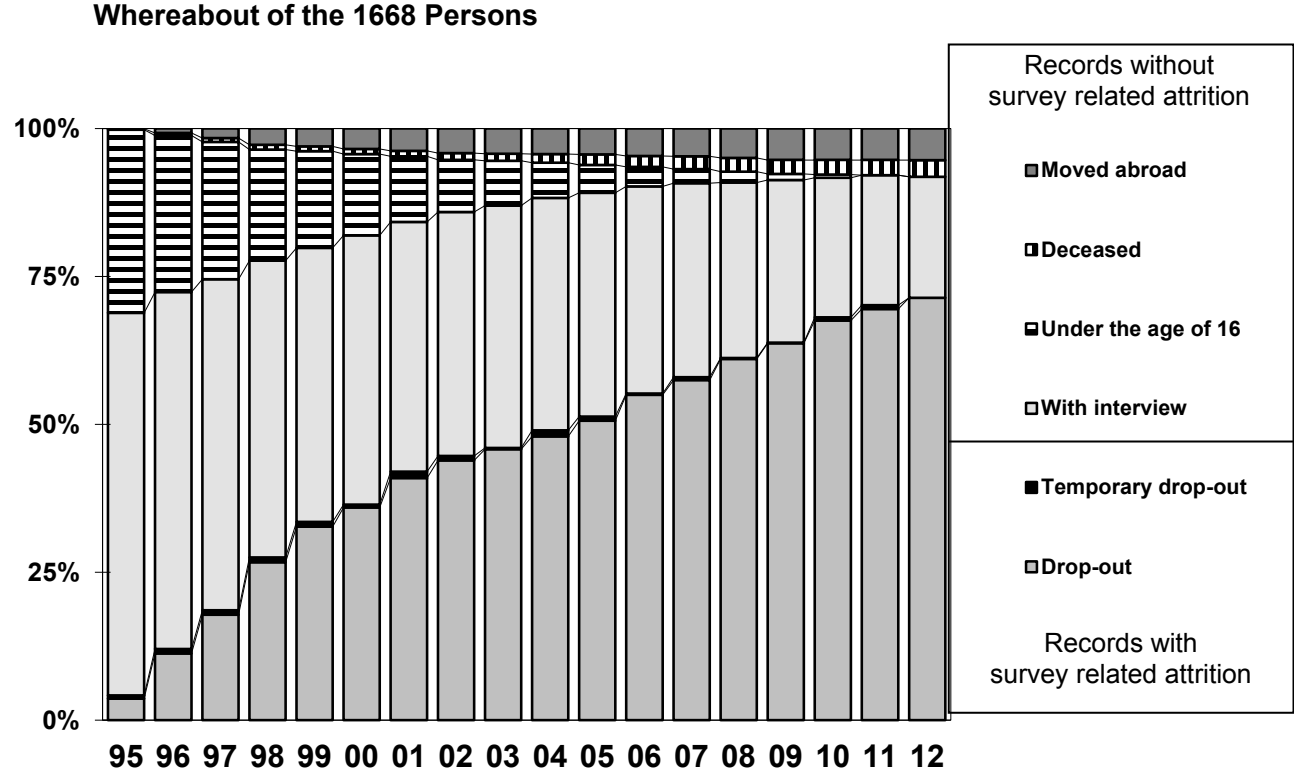


Figure 13: **All First Wave Persons (Gross Subsample E). Development up to Wave 15.**

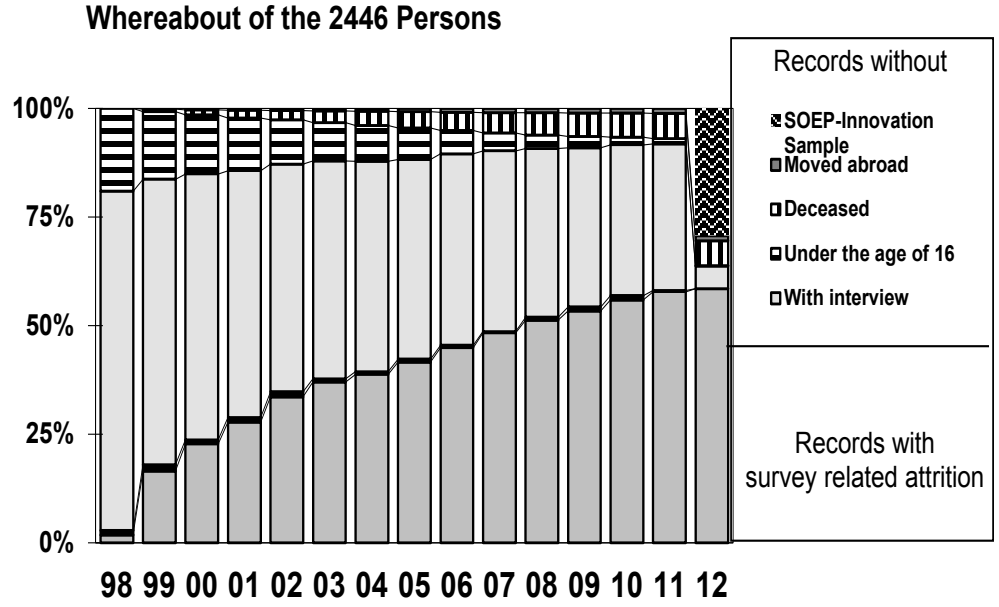


Figure 14: **All First Wave Persons (Gross Subsample F). Development up to Wave 13.**

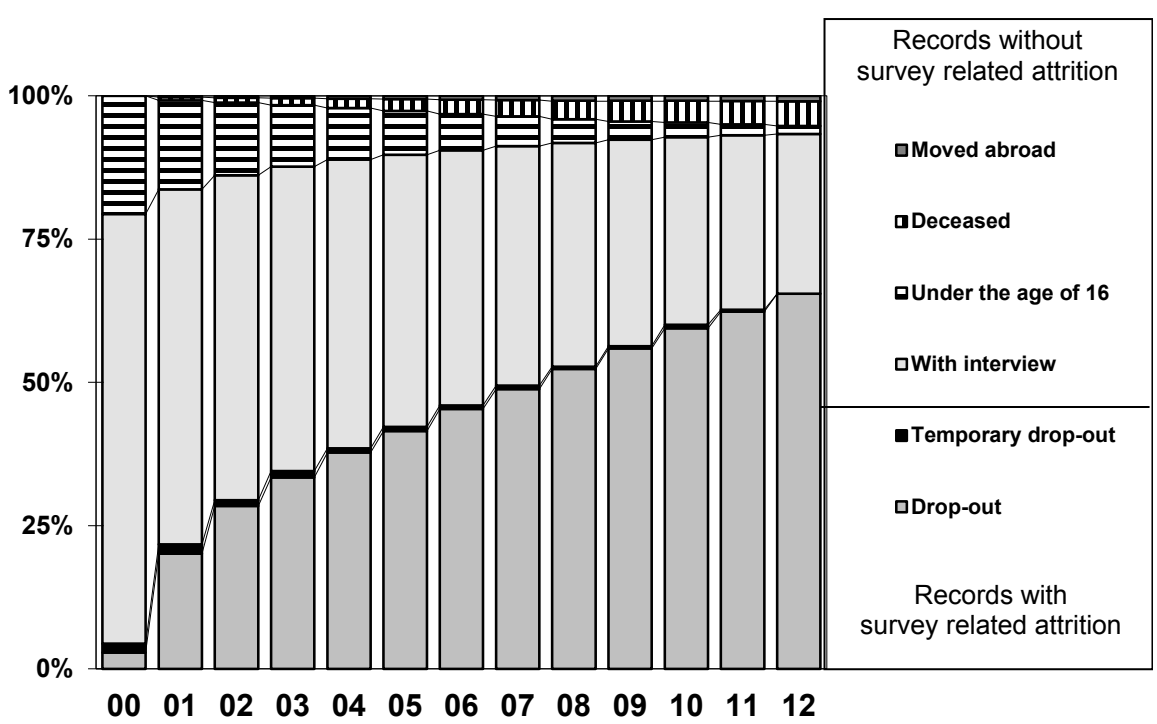


Figure 15: **All First Wave Persons (Gross Subsample G). Development up to Wave 11.**

Whereabout of the 3538 Persons

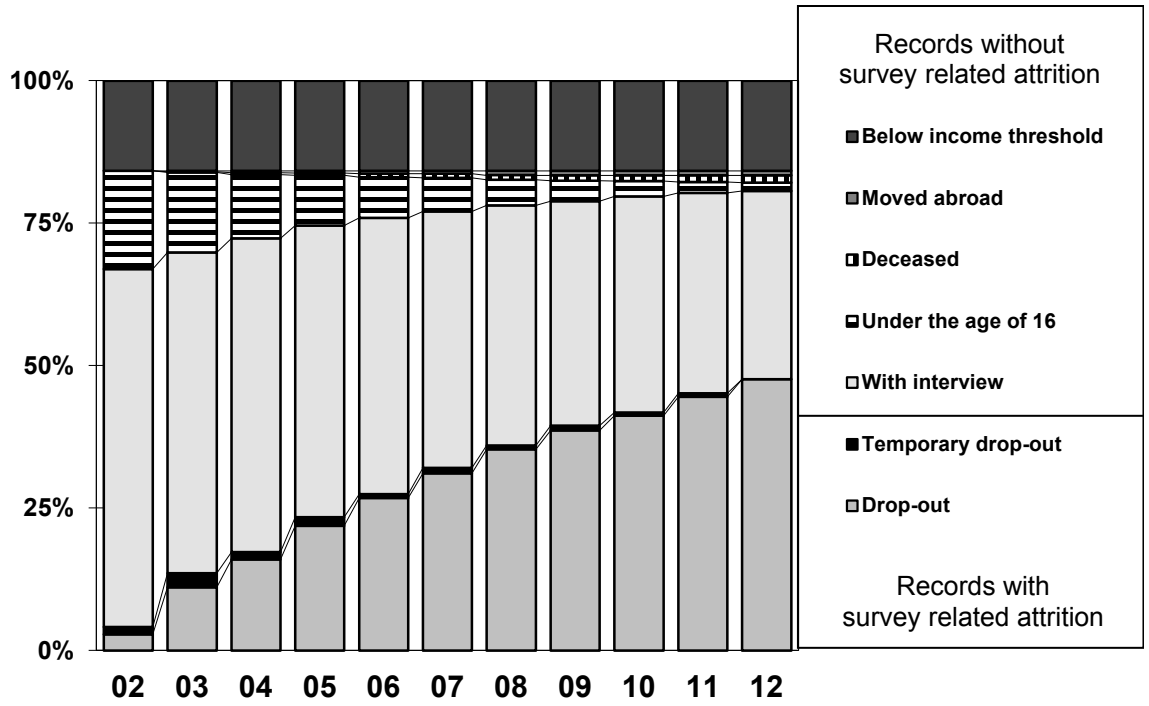
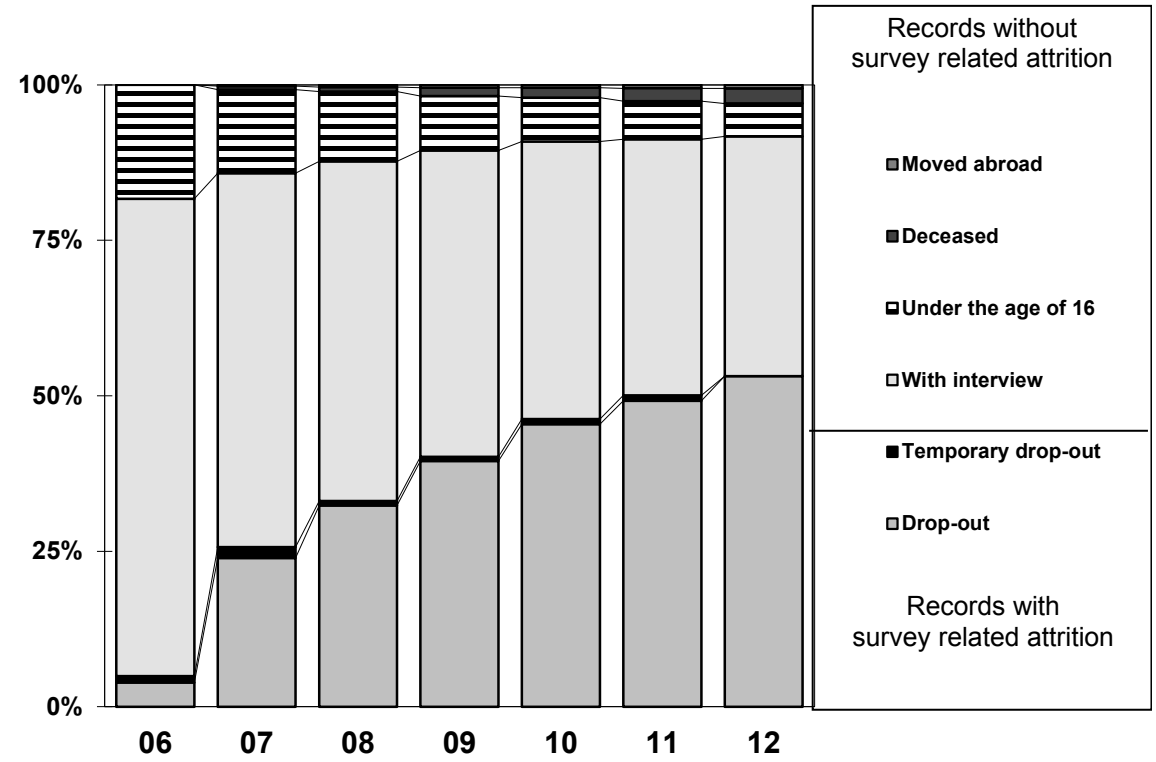


Figure 16: **All First Wave Persons (Gross Subsample H). Development up to Wave 7.**

Whereabout of the 3407 Persons



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 17: **Entrants and their Participation Behavior (Subsample A).**

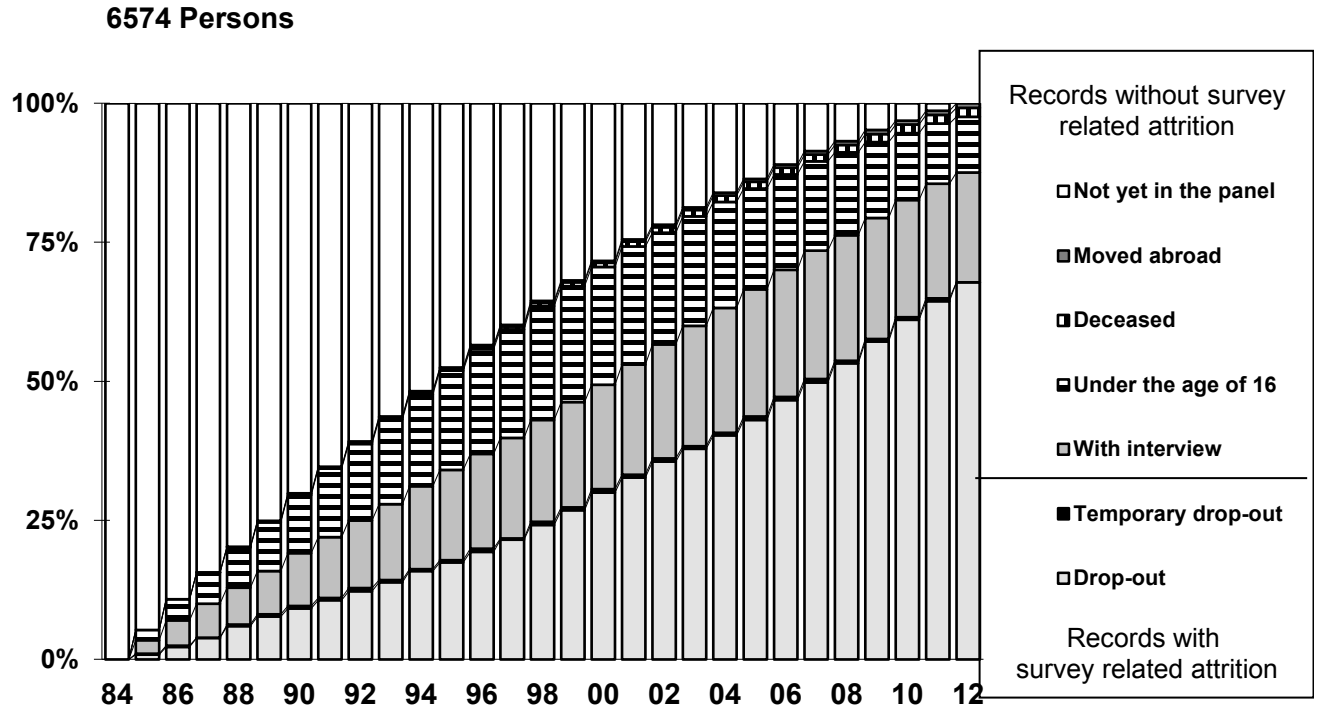


Figure 18: **Entrants and their Participation Behavior (Subsample B).**

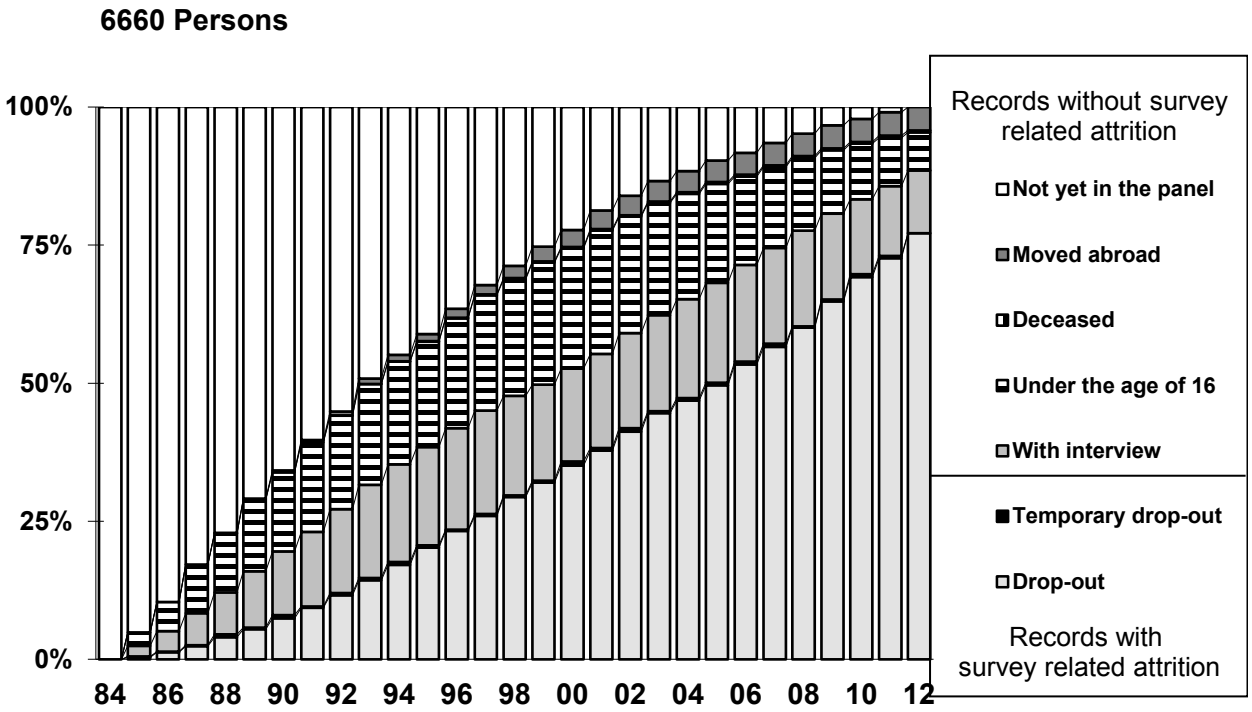


Figure 19: **Entrants and their Participation Behavior (Subsample C).**

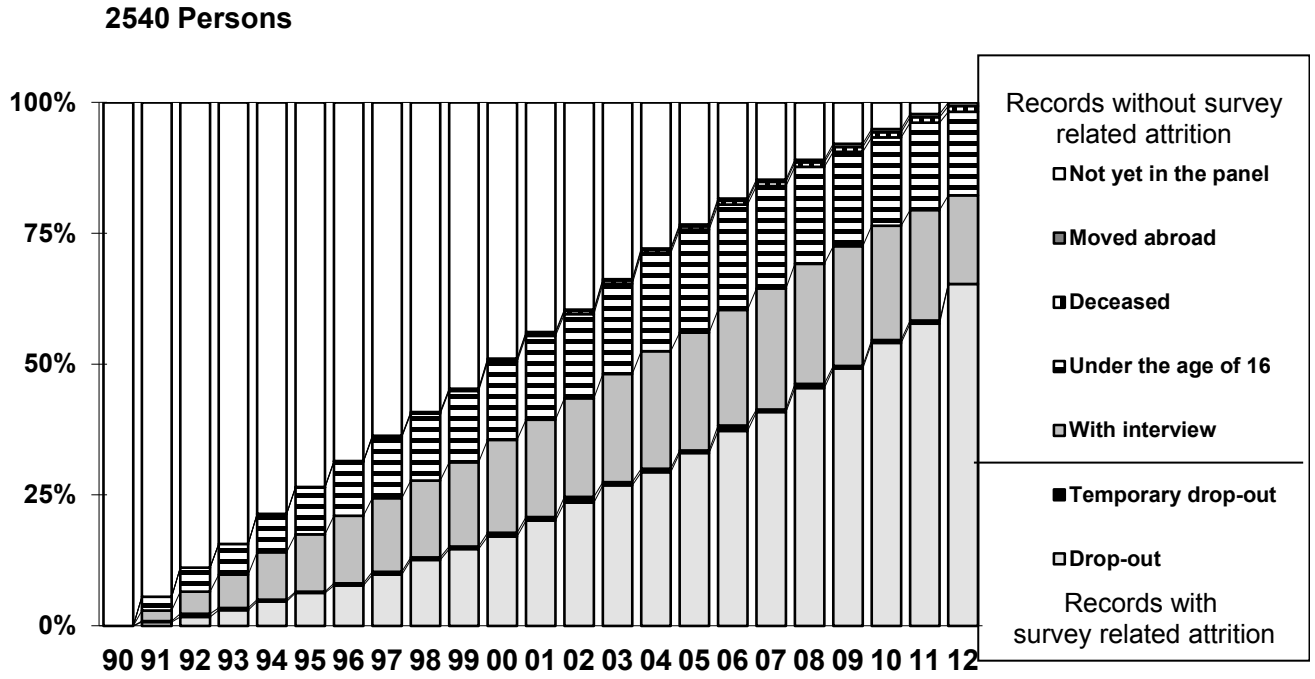


Figure 20: **Entrants and their Participation Behavior (Subsample D).**

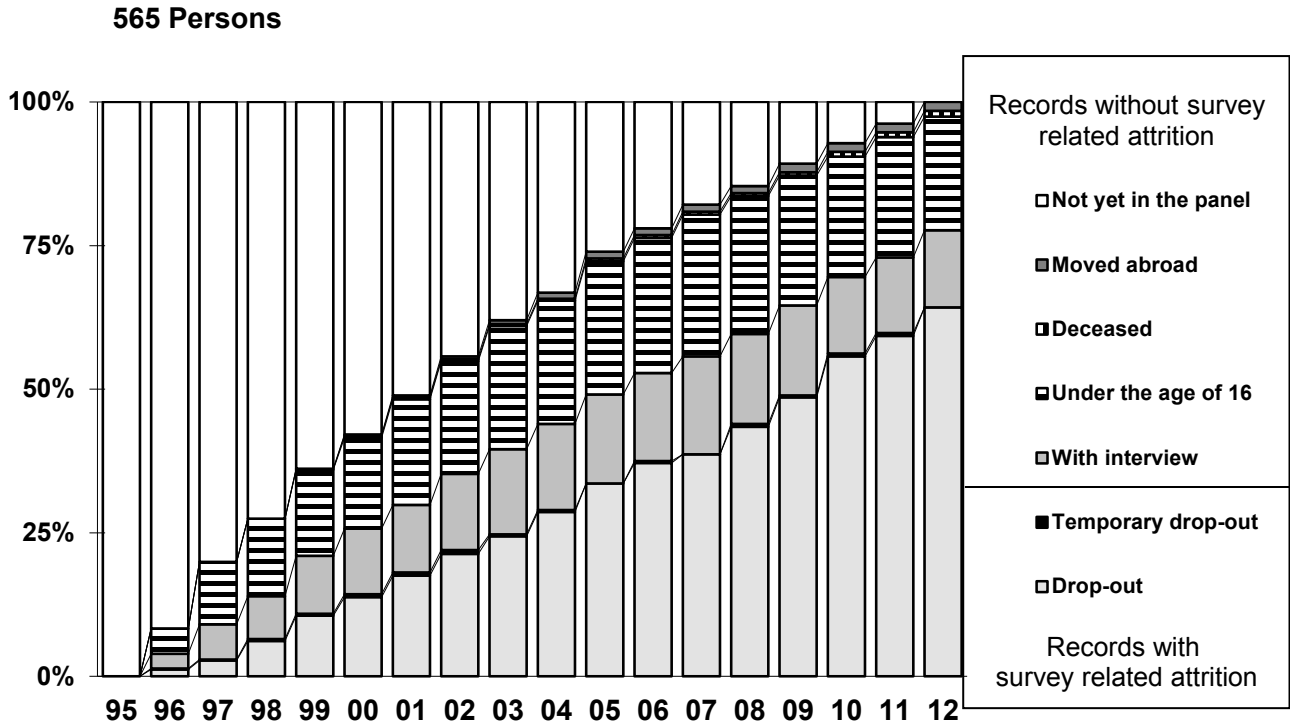


Figure 21: **Entrants and their Participation Behavior (Subsample E).**

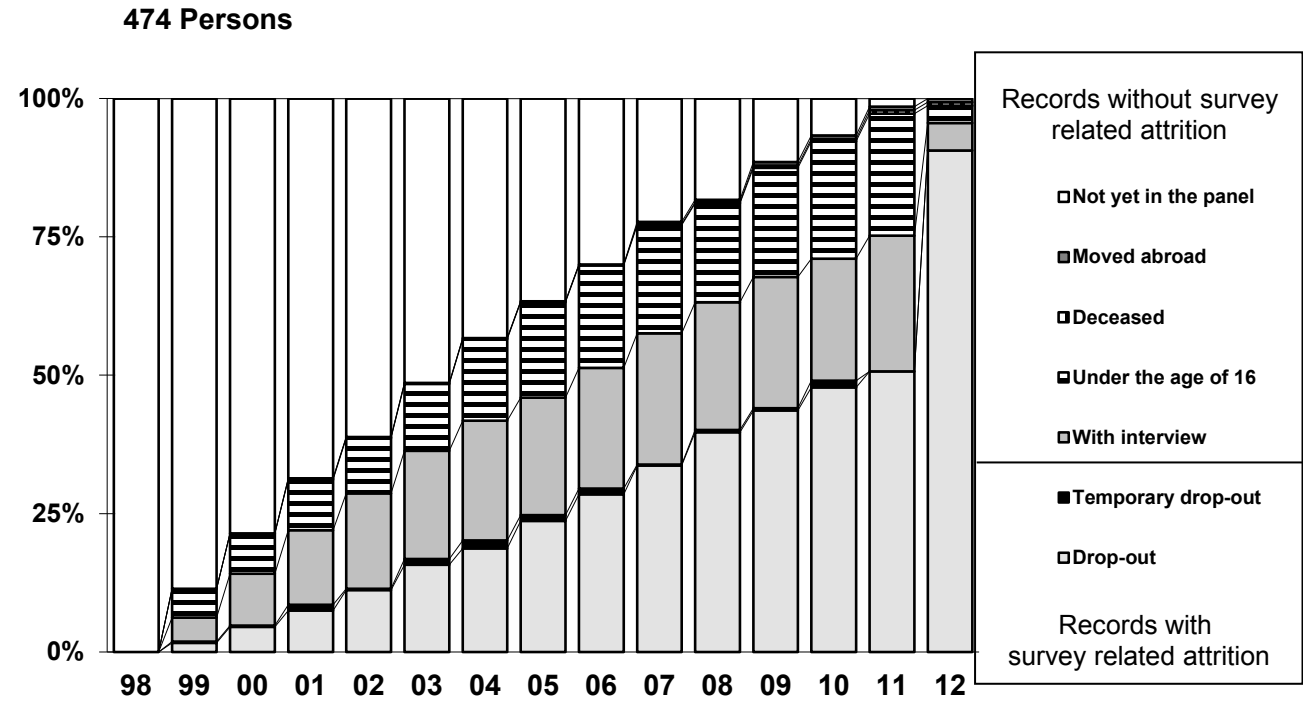


Figure 22: **Entrants and their Participation Behavior (Subsample F).**

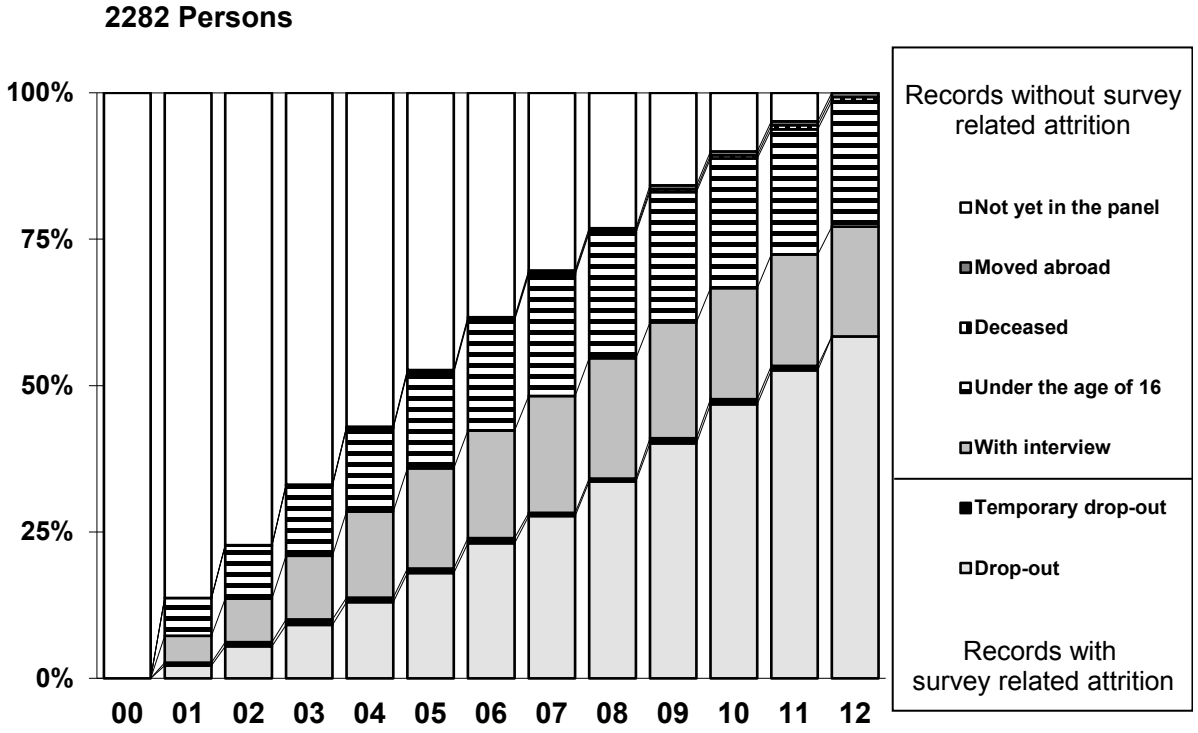


Figure 23: **Entrants and their Participation Behavior (Subsample G).**

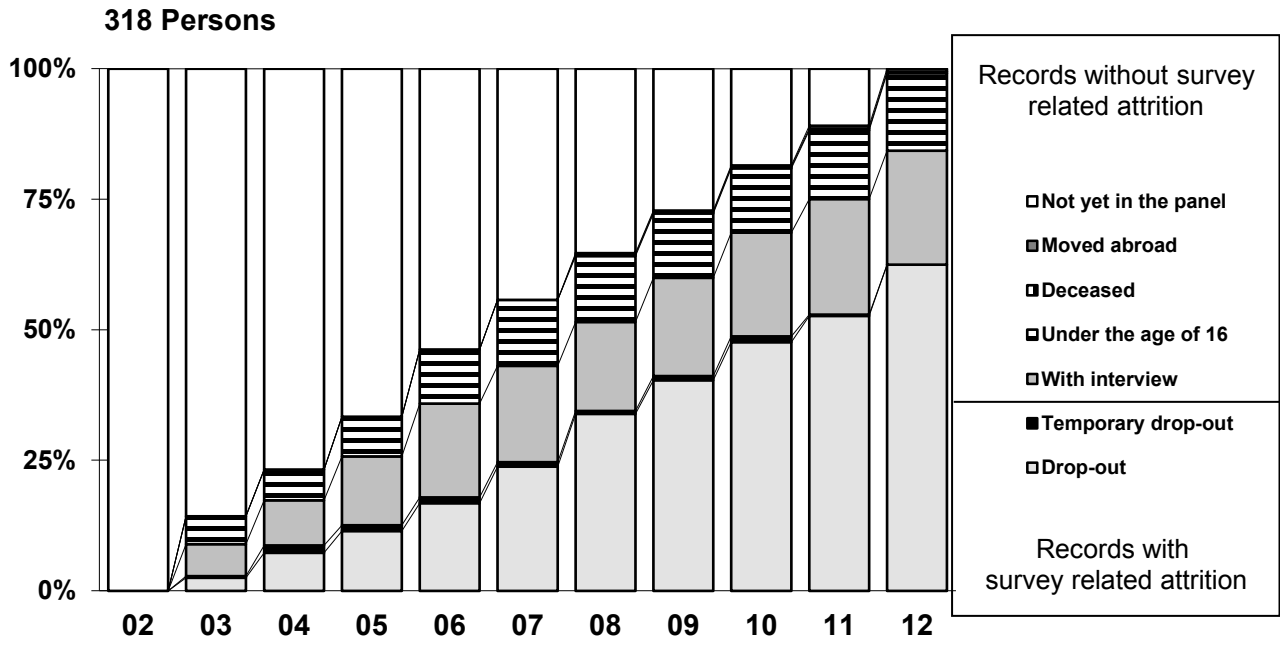
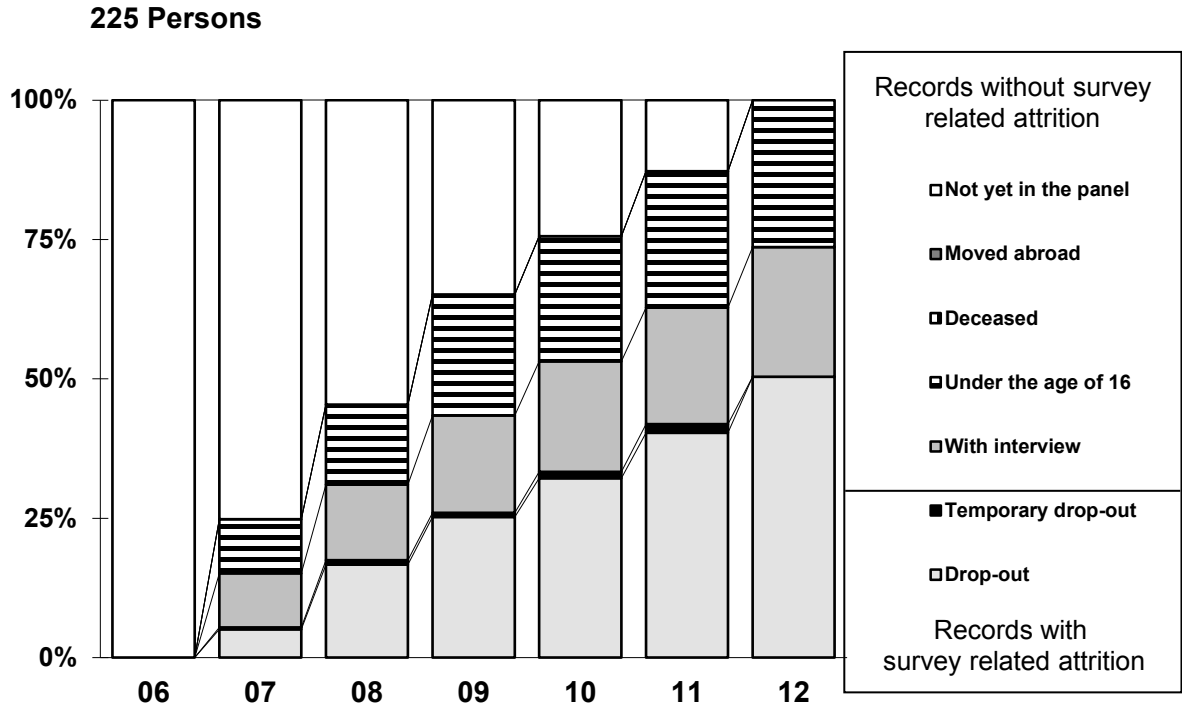


Figure 24: **Entrants and their Participation Behavior (Subsample H).**



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 26, 27, and 28) and some basic socio-demographic characteristics measured in the year of sampling, such as age, occupation, income, and education (Figures 29 through 32). 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 26), for instance, first-wave respondents from sample B have a somewhat lower probability of remaining in the survey than respondents from sample A and C. In the more recent samples D through J (Figures 27 and 28), 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 25: **Successful Re-Interviewing of First-Wave Respondents by Subsamples A, B, C. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

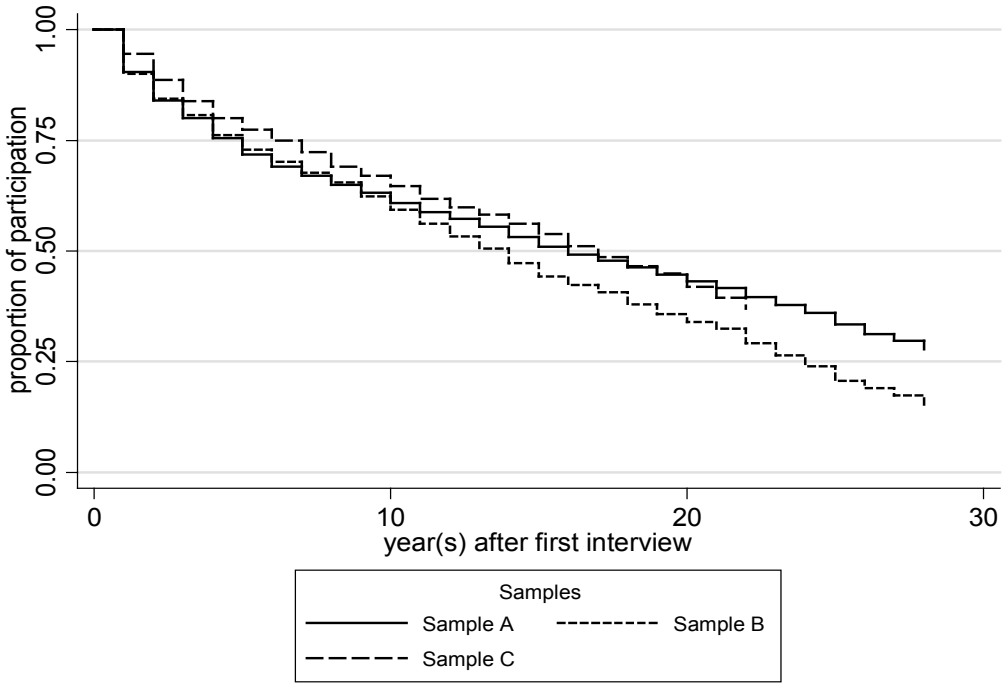


Figure 26: **Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

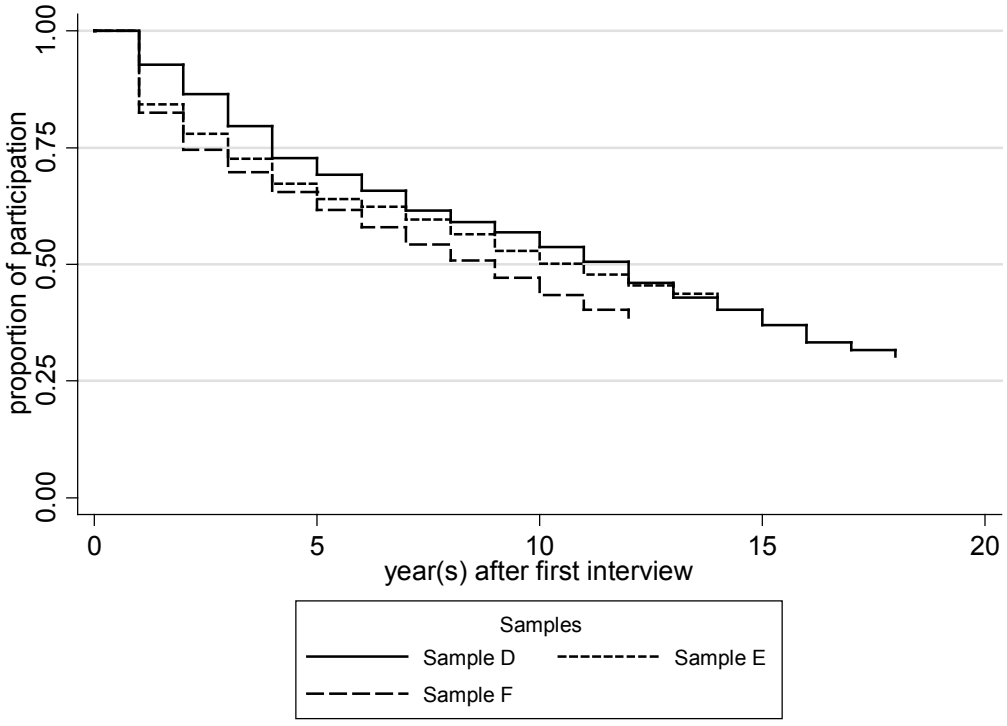


Figure 27: **Successful Re-Interviewing of First-Wave Respondents by Subsamples G, H, J. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

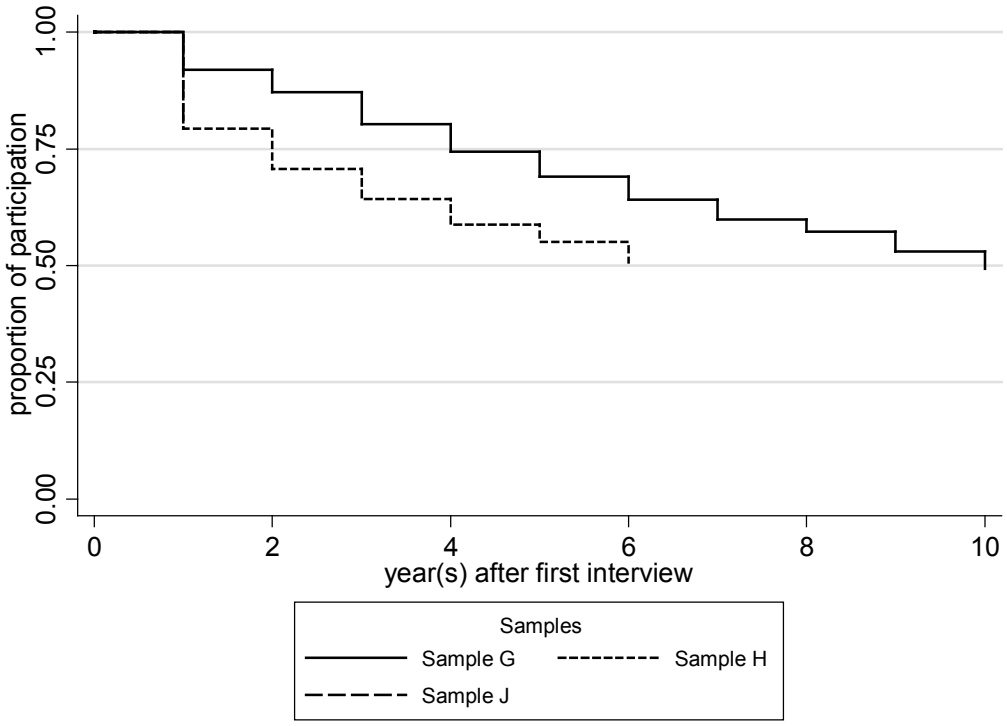


Figure 28: **Successful Re-Interviewing of All First-Wave Respondents by Age Categories. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

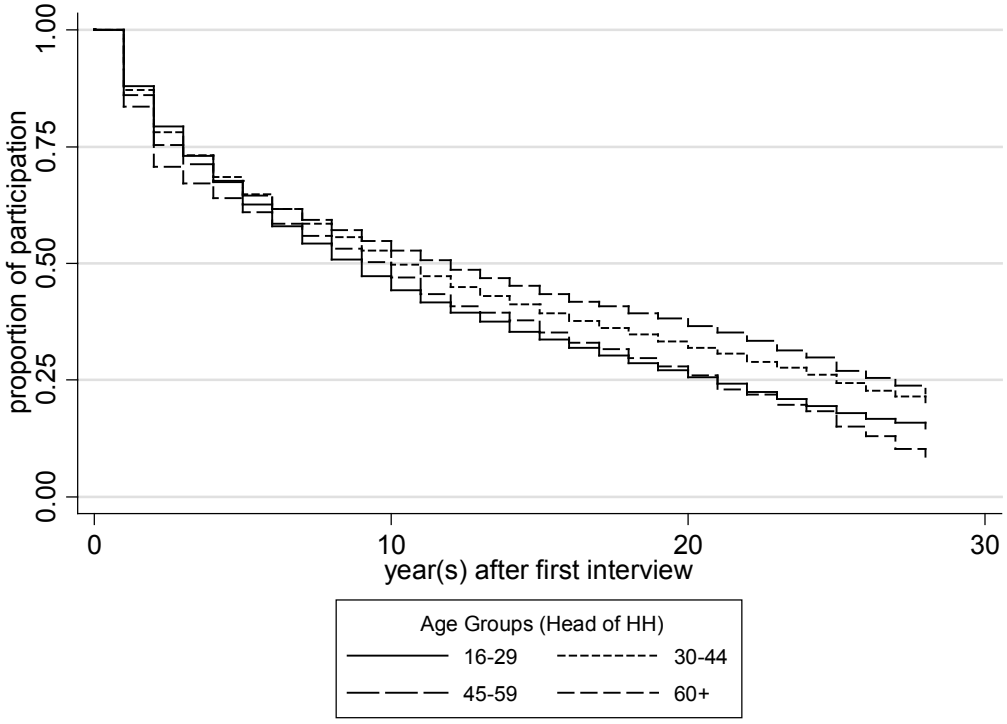


Figure 29: **Successful Re-Interviewing of All First-Wave Respondents by Occupation. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

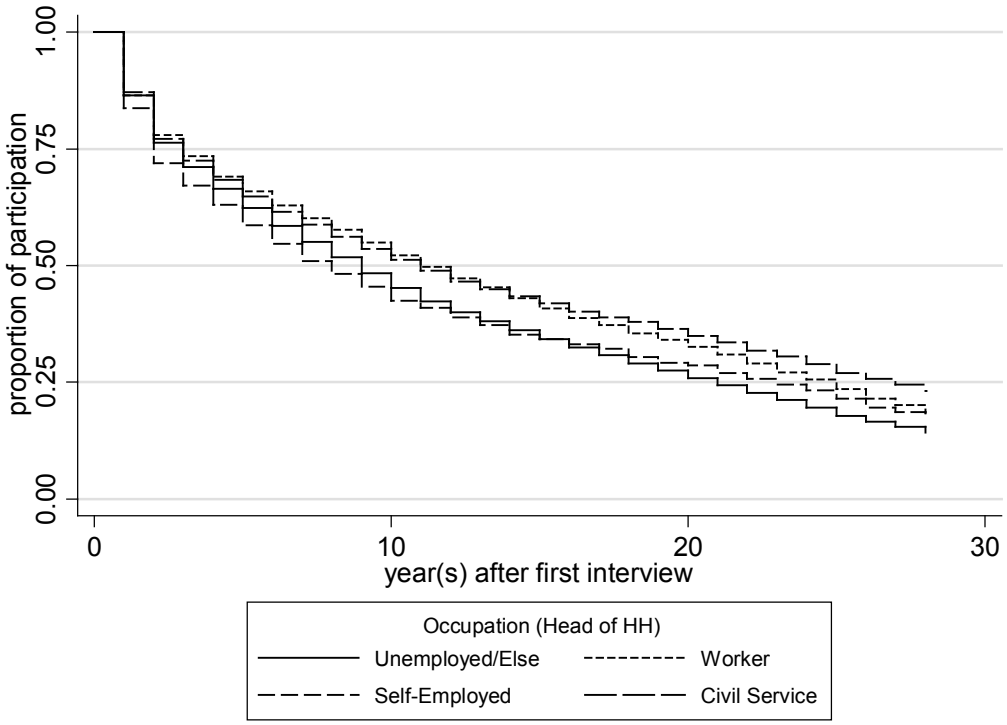


Figure 30: **Successful Re-Interviewing of All First-Wave Respondents by Income Quintiles. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

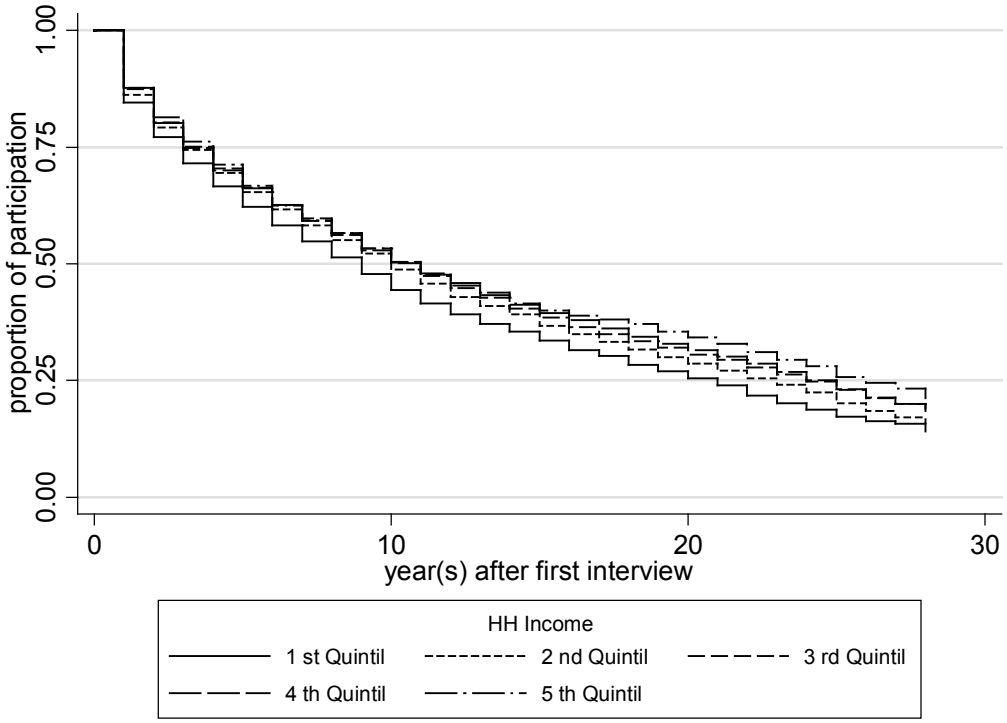
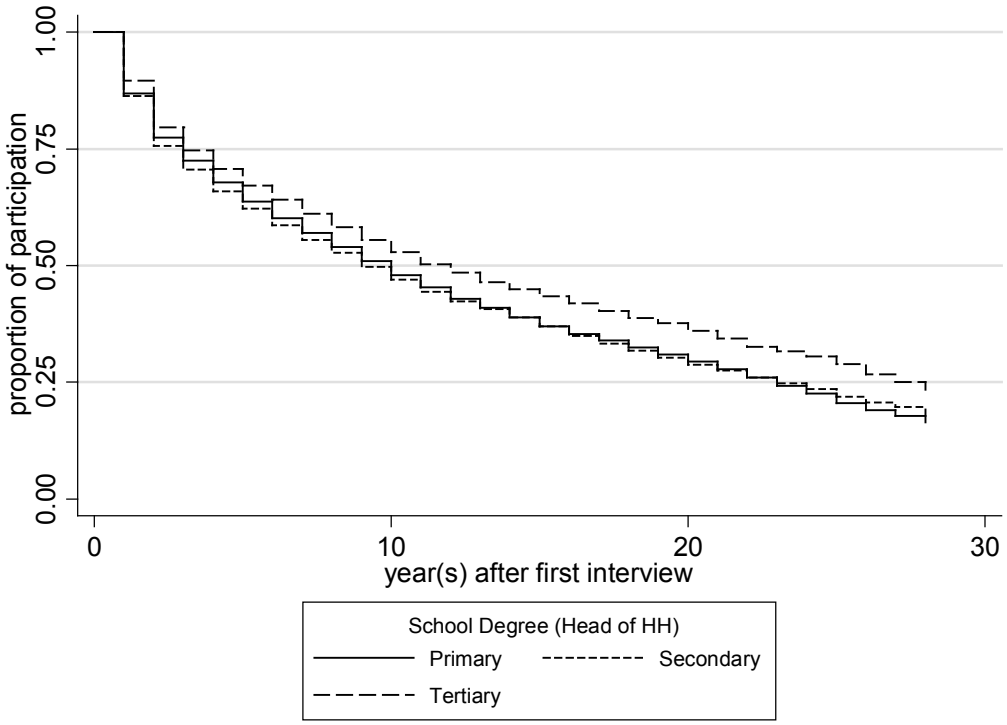


Figure 31: **Successful Re-Interviewing of All First-Wave Respondents by Education. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**



3 Panel Attrition Due to Unsuccessful Follow-Ups

In each panel wave, the first step in successful re-interviewing is the relocation of the households of 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 1 displays the number of households of the previous waves that need to be re-contacted and the relative frequency of successful follow-ups in subsamples A through J and waves 1985 through 2012. 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 1: **The Frequency of Households to be Re-Contacted and the Relative Proportion of Successful Follow-Ups by Subsample and Year.**

	Sample A		Sample B		Sample C		Sample D		Sample E		Sample F		Sample G		Sample H		Sample J	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1985	4681	98.5	1370	96.9														
1986	4486	99.0	1325	97.4														
1987	4232	99.1	1220	98.7														
1988	4140	99.2	1191	99.1														
1989	3984	99.1	1157	99.1														
1990	3902	99.2	1124	98.9														
1991	3860	99.5	1151	99.3	2246	98.5												
1992	3845	99.7	1153	99.2	2304	99.5												
1993	3867	99.3	1172	98.7	2227	99.1												
1994	3849	99.3	1150	99.1	2136	99.4												
1995	3784	99.5	1108	99.0	2113	99.6												
1996	3747	99.7	1069	99.3	2104	99.5	544	99.6										
1997	3688	99.6	1038	99.1	2091	99.5	542	99.3										
1998	3667	99.4	1019	99.4	2081	99.4	498	99.4										
1999	3631	99.6	975	99.4	2041	99.7	529	99.1	1100	99.5								
2000	3549	99.6	934	99.5	2028	99.6	467	99.8	968	99.2								
2001	3463	99.6	904	99.5	2036	99.7	454	99.1	922	99.1	6172	99.0						
2002	3406	99.7	877	99.1	2010	99.5	450	99.8	875	99.4	5451	99.5						
2003	3330	99.6	840	99.6	1982	99.6	434	99.5	834	99.3	4965	99.7	1056	99.1				
2004	3260	99.8	803	99.6	1962	99.6	436	99.8	797	99.7	4736	99.6	1010	99.7				
2005	3220	99.8	779	99.4	1959	99.7	429	99.3	783	99.1	4577	99.7	1001	99.7				
2006	3138	99.7	770	99.6	1941	99.4	425	98.8	775	99.1	4401	99.3	995	99.5				
2007	3000	99.7	725	99.5	1834	99.9	387	99.5	727	99.7	4157	99.5	933	99.3	1530	99.5		
2008	2856	99.7	676	99.1	1767	99.5	372	99.4	680	99.7	3962	99.4	904	99.6	1326	99.6		
2009	2730	99.7	620	99.2	1695	99.8	351	99.7	636	100	3760	99.6	870	99.5	1145	99.7		
2010	2570	99.8	548	99.3	1627	100	334	99.6	605	99.8	3538	99.5	826	99.9	1059	99.5		
2011	2421	99.8	495	99.0	1541	99.8	302	99.3	589	100	3318	99.6	797	99.6	992	99.5		
2012	2289	99.8	440	99.7	1466	0.99	286	100	116	98.9	3076	99.9	774	99.7	928	99.9	1526	99.1

n = Number of households to be recontacted

% = Percentage of households with successful recontact

3.2 Predicting the Probability of Successful vs. Unsuccessful Follow-Ups in the Year 2012

Based on the household and interview characteristics measured in 2011, we aim at predicting the probability of re-contacting a household relative to unsuccessful follow-up in 2012. Among a very large number of regressors that we tested in preliminary analyses, we identified a smaller number of variables that exert a robust effect on the probability of successful follow-ups ($p < 0.05$). Table 2 describes the regressors and Table 3 reports the subsample-specific estimates of logit models of the probability of re-contacting a household relative to unsuccessful follow-up.

Note that the estimates of regression models of the previous waves 1985 through 2011 are due to space restrictions not reported in the present data documentation, but can be obtained from previous attrition documentations.

Table 2: **Definition of the Regressors of the Logit Model of Unsuccessful Follow-Ups.**

Variable	Label	Value
Interview Characteristics		
Email Disclosed	Email Address Known	0/1
Move HH	HH Move	0/1
Interviewer Related HH	Same Interviewer in Related HH	0/1
Incentives	Incentives-Experiment in Old-Samples	0/1
Demographics and Health		
Age below 25	Head of HH Younger than 25 Years	0/1
Single HH	One Person Living in HH	0/1
Living Apart Together	Partner Living Apart HH	0/1
Female Head of HH	Head of HH Male/Female	0/1
Thuringia	HH located in Thuringia	0/1
Religion	Affiliated with Any Religion	0/1

Table 2: **Definition of the Regressors of the Logit Model of Unsuccessful Follow-Ups.**

Variable	**Label**	**Value**
Financial Situation, Real Estate and Insurance		
Change Job	Changed Job in Previous Year	0/1
Income 100	HH in Fourth Quartile of Income Distribution	0/1
Investment	No Investment in Previous Year	0/1
Personality Traits and Well-Being		
Job Security	Concerned about Job Security	0/1

Table 3: **Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2012.**

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J
Intercept	2.47	1.81	1.99		1.56	2.07	**1.78**	1.92	3.16
Interview Characteristics									
Move HH	-0.60								-1.30
Interv. Related HH									-0.93
Email Disclosed									0.59
Incentives	-0.86								
Demographics and Health									
Age below 25						-0.54			
Single HH									-0.99
Female head of HH									0.66
Livi. Apart Togeth.									0.48
Religion									-0.52
Region									
Thuringia						-0.52			
Financial Situation, Real Estate and Insurance									
Change Job	-0.72								
Income 100									-0.57
Investment									-0.39
Personality, Traits and Well-Being									
Job Security									-0.55
Likel. Ratio (Pr > 0.45)		****	****	****		****	****	****	

*Note. *** p < 0.01; ** p < 0.05; * p < 0.10; standard errors in parentheses.*

4 Panel Attrition Due to Refusals

In each panel wave, the second step in successful re-interviewing after relocating 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 deaths, and moves abroad, to generate the longitudinal weights.

4.1 The Frequency of Participation

Table 4 displays the participation rates due to refusal by sub-sample and wave. In reverse one can derive the corresponding drop-out rates. Note that we did not distinguish between various types of refusals such as unconditional refusals, refusals due to lack of time, or health problems, etc.

Table 4: **The Frequency of Re-Contacted Households and the Relative Proportion of Participation by Subsample and Year.**

	Sample A		Sample B		Sample C		Sample D		Sample E		Sample F		Sample G		Sample H		Sample J	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1985	4611	89.8	1326	89.1														
1986	4442	89.2	1290	87.4														
1987	4194	93.2	1204	92.7														
1988	4105	91.1	1180	90.8														
1989	3949	92.4	1146	91.0														
1990	3871	93.3	1111	92.5														
1991	3842	94.0	1143	92.4	2213	91.7												
1992	3833	93.5	1144	92.7	2290	88.2												
1993	3838	93.9	1156	92.0	2208	89.2												
1994	3821	93.6	1139	89.8	2122	92.3												
1995	3766	93.6	1097	89.5	2101	92.2	634	82.3										
1996	3734	93.3	1061	90.5	2092	93.3	542	91.9										
1997	3674	94.1	1029	90.5	2076	93.6	537	89.2										
1998	3645	92.9	1013	88.6	2066	91.3	523	84.3										
1999	3616	92.0	969	88.5	2030	93.3	495	85.9	1084	81.7								
2000	3535	91.7	929	88.3	2018	93.1	466	91.2	959	87.8								
2001	3448	91.9	899	90.0	2028	91.2	450	88.4	913	88.8	6109	80.4						
2002	3396	92.0	869	88.1	1996	91.1	449	89.5	868	89.1	5420	84.6						
2003	3318	92.6	837	88.6	1974	91.5	432	92.4	828	89.9	4951	88.6	1047	87.0				
2004	3253	92.5	800	89.25	1955	92.7	435	89.2	795	92.1	4719	89.7	1007	89.8				
2005	3214	91.4	774	90.2	1954	90.6	426	89.0	782	90.3	4564	89.2	998	88.1				
2006	3130	90.1	767	85.4	1930	89.0	420	85.7	768	89.3	4370	89.1	990	86.8				
2007	2992	91.0	721	85.2	1832	90.3	385	89.6	725	89.2	4138	89.3	926	89.0	1523	78.0		
2008	2850	90.7	671	84.9	1759	90.5	370	88.6	678	88.8	3939	89.2	901	87.3	1321	81.9		
2009	2723	89.0	616	81.2	1693	90.7	350	87.4	636	90.3	3746	88.2	866	87.4	1142	87.2		
2010	2565	87.5	545	80.9	1627	88.3	333	83.5	604	91.6	3523	86.7	825	90.1	1054	86.6		
2011	2417	88.9	491	79.6	1538	88.1	300	88.7	589	92.5	3307	87.2	794	88.9	988	86.8		
2012	2285	89.0	439	78.8	1456	89.6	286	87.8	115	80.0	3073	87.9	772	89.0	927	88.2	3178	80.4

n = Number of re-contacted households

% = Percentage of households that participated

4.2 Predicting the Probability of Re-Interviewing versus Refusal in the Year 2012

Based on the household and interview characteristics measured in 2011, and some regional information measured in 2012, we aim at predicting the probability of agreement vs. refusal to participate in the survey by the households that were re-contacted in 2012. The individual attributes refer in many cases to the head of the household in the previous wave, but for split-off households the attributes refer to 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 are aggregated within the households, for instance, rare events, such as acute medical conditions.

As in the case of predicting successful follow-ups, we use only model specifications where all included regressors are significantly different from zero. The definition of the regressors is given in Table 5. Table 6 reports the subsample-specific estimates of logit models of the probability of participating relative to refusal. Note that the estimates of regression models of the previous waves 1985 through 2011 are not reported in the present data documentation due to space restrictions, but can be obtained from previous attrition reports.

Table 5: **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 Disclosed	Email Address Known	0/1
Phone Disclosed	Telephone Number Known	0/1
Change in Interviewer	Change in Interviewer Between Last Waves	0/1

Table 5: **Definition of the Regressors of the Logit Model of Refusal.**

Variable	Label	Value
SAQ	Self-Administered Questionnaire	0/1
Change in Interview Mode	Change in Interview Mode Between Last Waves	0/1
Temp. Related HH	Related HH Temporary Drop-Out	0/1
Refusal Related HH	Related HH Refusal	0/1
Interviewer Related HH	Same Interviewer in Related HH	0/1
Add. Questionnaire Bio	Additional Questionnaire Answered (Bio)	0/1
Add. Questionnaire Dj	Additional Questionnaire Answered (Dj)	0/1
Add. Questionnaire Muki	Additional Questionnaire Answered (Muki)	0/1
Short Interview	Interview Duration Short	0/1
Long Interview	Interview Duration Long	0/1
Demographics and Health		
Female Head of HH	Head of HH is Female	0/1
Younger than 25	Head of HH is Younger than 25	0/1
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
Religion	Affiliated with Any Religion	0/1
Birth in HH	Baby was Born In HH	0/1
Death in HH	Someone Deceased In HH	0/1
Single HH	One Person Living in HH	0/1
Family HH	Family Living in HH	0/1
Divorced	Head of HH Divorced	0/1
Widowed	Head of HH Widowed	0/1
Moving In	Current Moving In HH	0/1
Separation	Current Separation in HH	0/1
Marriage	Marriage in HH	0/1
Partner Apart Together	Partner Living Apart Together	0/1
Partner	Married or Unmarried Couple in HH	0/1

Table 5: **Definition of the Regressors of the Logit Model of Refusal.**

Variable	Label	Value
Previous Partner	Married or Unmarried Couple Previous Year in HH	0/1
Health Situation		
Diabetes	Person Having Diabetes	0/1
Depression	Person Having Depression	0/1
Chronic Backache	Person Having Chronic Backache	0/1
Asthma	Person Having Asthma	0/1
Heart Condition	Person Having a Heart Condition	0/1
Stroke	Person had a Stroke	0/1
Care	At least one HH-Member is in need of care	0/1
Disability	Level of Disability	0/1
Care Insurance Benefits	Benefits from Care insurance in HH	0/1
Financial Situation, Real Estate and Insurance		
Subtenant	Subtenant of Dwelling	0/1
No Income	HH is in no Quartile of Income Distribution	0/1
Income 25	HH in First Quartile of Income Distribution	0/1
Income 75	HH in Third Quartile of Income Distribution	0/1
Income 100	HH in Fourth Quartile of Income Distribution	0/1
Insurances	More than 4 Insurances	0/1
Investment	No Investment in Previous Year	0/1
Homeowner	Owner of Dwelling	0/1
Work and Education		
Commuting	At Least one Member of HH is Commuting	0/1
Unemployed	Head of HH is Unemployed	0/1
White-Collar Worker	Head of HH is 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
Blue-Collar Worker	Head of HH is an Blue-Collar Worker	0/1

Table 5: **Definition of the Regressors of the Logit Model of Refusal.**

Variable	Label	Value
Other Labour Status	At least one Person in HH in Maternity, Community Service, Military Service or Irregular Job	0/1
Not Employed in HH	At least one Person in HH is not Employed	0/1
Unemployment HH	One or More Persons Unemployed in HH	0/1
University Degree	Head of HH has a University Degree	0/1
Change Job	Changed Job in Previous Year	0/1
Personality Traits, Well-Being and Other Characteristics		
Volunteer Work	At least one Person in HH is Engaged in Volunteer Work	0/1
Union	At least one Person in HH is Part of an Union	0/1
Health Happiness	At least one Person in HH has High Level of Health Happiness	0/1
Employee Organ.	At least one Person in HH is Engaged in an Employee Organization	0/1
Friends	Head of HH has a lot of Friends	0/1
No Friends	Head of HH does not have any Friends	0/1
Education Friend	Head of HH knows Education Level of Close Friend	0/1
Labour Status Friend	Head of HH knows Labor Status of Friend	0/1
Strong Political Interest	Head of HH has very High or No Political Interest	0/1
Building, Area, and Region		
Prosperous Area	HH Located in Area of High Prosperity	0/1
Family Area	HH Located in Area with Mostly Families	0/1
Single Area	HH Located in Area with Mostly Singles	0/1
High Rise Block	HH Located in Area with High Rate of High Rise Buildings	0/1
Low Academics	HH Located in Area of Low Academics Rate	0/1
High Academics	HH Located in Area of High Academics Rate	0/1
Low Turnout	HH Located in Area of Low Voter Participation	0/1
High Turnout	HH Located in Area of High Voter Participation	0/1
Depopulation	HH Located in Area of High Depopulation	0/1

Table 5: **Definition of the Regressors of the Logit Model of Refusal.**

Variable	**Label**	**Value**
Small Flat	HH Located in Area of High Share of Small Flats	0/1
Big Town	HH in Town with More Than 100.000 Inhabitants	0/1
High CDU Share of Vote	HH Located in Area of High Share of Voting “CDU”	0/1
High Grüne Share of Vote	HH Located in Area of High Share of Voting “Grüne”	0/1
High Linke Share of Vote	HH Located in Area of High Share of Voting “Linke”	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
Berlin/Brandenburg	HH Located in Berlin/Brandenburg	0/1
Thuringia	HH Located in Thuringia	0/1
Saxony-Anhalt	HH Located in Saxony-Anhalt	0/1
Baden-Wuerttemberg	HH Located in Baden-Wuerttemberg	0/1
Hessen	HH Located in Hessen	0/1
Mecklenburg-Vorpommern	HH Located in Mecklenburg-Vorpommern	0/1
West	HH Located in Western/Eastern Germany	0/1
Few Abitur-Graduates	HH Located in Area with Few Abitur-Graduates	0/1
Old Area	HH Located in Area with High Average Age	0/1
Newspaper	HH Located in Area with High Level of supra-regional Newspaper Readers	0/1

Table 6: **Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2012.**

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J
Intercept	0.09	0.77	0.56	0.98	-0.09	-0.63	0.84	-1.11	-0.43
Interview Characteristics									
Partial Unit Nonresponse							-0.43		-0.41
Original Sample Member	0.13			1.06		0.39		1.05	
New HH	-0.71					-0.74			-0.62
Refusal Related HH						-0.47			1.23
Email Disclosed									0.45
Phone Disclosed	0.57	0.63	0.78			0.77	0.60	1.21	0.80
Change in Interviewer	-0.38					-0.39			
Change Interview Mode			-0.39						
Interviewer Related HH	0.24		0.37			0.32			
Temp. Drop-Out		-1.89		-1.81			-1.65		
Add. Questionnaire Bio		1.68							
Add. Questionnaire Dj		-1.54		-2.06					
Add. Questionnaire Muki		-1.80							
SAQ						0.24			
Temp. Related HH	-0.77					-0.50			
Short Interview		-0.34	0.27						
Long Interview				-1.42				-0.58	

Table 6: **Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2012.**

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J
Demographic and Relational Characteristics									
Female Head of HH		-0.40							
Religion								0.51	
Younger than 25		-1.28							
Age 25-34								-0.63	
Age 35-44				-0.58					
Birth in HH									-0.30
Death in HH			-0.68						
Single HH			-0.43			-0.20	-0.64		-0.24
Family HH						-0.26			
Divorced									0.25
Widowed									0.24
Moving In								-1.01	
Separation			-0.69						
Marriage		-0.98						1.61	
Partner							-0.44		
Previous Partner			-0.37					-0.41	
Partn. Apart Togeth.								-0.46	

Table 6: **Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2012.**

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J
Health Situation									
Disability			0.39						
Care Ins. Benefits	-0.42								
Diabetes	0.22				-1.74				
Asthma						0.20	-0.39		
Depression			0.40						
Stroke				-2.01					
Chronic Backache					1.42	-0.18			
Heart Condition								0.39	
Care			-0.48						-0.25
Fin. Situation, Real Est., Insurance									
Subtenant		-0.91							
Homeowner			0.24						
No Income	-0.31		-0.82	-1.62					-0.25
Income 25				-0.88		0.26			
Income 75		-0.32							
Income 100									-0.21
Insurances									0.19
Investment	-0.19		0.31			-0.21		-0.50	

Table 6: **Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2012.**

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J
Work and Education									
Commuting					1.33				
White-Collar Worker			0.29			0.39			
Civil Servant			0.82						
Self-Employed						0.39			
Blue-Collar Worker				-0.83		0.44		0.44	
Other Labour Status									0.27
Unemployed	0.45								0.26
Change Job			-0.46						
Not Employed in HH	0.14					0.41			0.20
Unemployment HH							0.96		
Univ. Degree			0.26					0.32	0.21
Personality Traits, Well-Being and Other Characteristics									
Employee Orga.						0.49			
Union	0.25							0.39	
Volunteer Work									0.13
Strong Pol. Interest	0.20				1.19	0.18			
Health Happiness									-0.31
Friends		-0.82			-0.87				
No Friends								0.65	-0.25
Educ. Friend								-0.56	
Lab. Status Friend								0.55	
Building, Area and Region									
Prosperous Area				0.55					
Family Area			0.29						
Single Area			0.39						
Low Academics		-0.73							
High Academics			-0.51						

Table 6: **Estimates of Logit Models for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2012.**

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H	Sample J
High Rise Block			0.27						
Low Academics		-0.73							
High Academics			-0.51						
Low Turnout								-0.58	
High Turnout						0.19			
Small Flat		-0.58	-0.40					-0.43	
Depopulation			-0.26			-0.15			
Big Town						0.21			
Newspaper		0.57						0.46	
High Grüne Share						-0.17			
High Linke Share			0.58					0.55	
High CDU Share			0.51						
Low Purch. Power									0.16
High Purch. Power								-0.47	
Saxony-Anhalt			-0.93						
Baden-Wuerttemberg									-0.24
Berlin/Brandenburg			-0.29					0.87	
Thuringia			-0.74						
Hessen								-0.53	
Mecklenburg-Vor.			-0.84						
West						0.28			0.26
Few Abitur-Grad.							0.32		0.17
Old Area					0.75				
Likel. Ratio (Pr > Chisq)									

Note. *** p < 0.01; **p < 0.05; * p < 0.10; standard errors in parentheses.

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 this probability of staying in the SOEP in 2012 based on characteristics measured in 2011, BCHBLEIB, lends itself as a longitudinal weighting variable correcting for selective attrition between waves 2011 and 2012. Table 7 reports some sub-sample specific descriptive statistics of the longitudinal weights in each wave.

The product of the cross-sectional weight in 2011, BBHHRF, and the longitudinal weight in 2012, BCHBLEIB, provide the raw data for the cross-sectional weight in 2012. In a final step, the post-stratification of the cross-sectional weights corrects them to meet benchmarks of known marginals of the underlying population in 2012. These are at the household level 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). Table 8 reports sub-sample-specific descriptive statistics of the derived cross-sectional weighting variable BCHHRF and in comparison all previous cross-sectional weights AHHRF through BBHHRF.

Table 7: **Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples A through D (Percentiles of $HBLEIB up to Wave 29).**

	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	220
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	2147	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	1.12	1.65	346	1.00	1.05	1.27	1312	1.00	1.04	1.45	251

Table 8: **Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples E through H (Percentiles of $HBLEIB up to Wave 29).**

	sample E				sample F				sample G				sample H			
	p10	**p50**	**p90**	**N**	**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	1.04	1.16	1.46	1188
2008	1.00	1.01	1.19	602	1.01	1.03	1.14	3513	1.01	1.03	1.18	787	1.01	1.03	1.18	1082
2009	1.00	1.04	1.17	574	1.02	1.04	1.34	3303	1.02	1.04	1.36	757	1.01	1.03	1.22	996
2010	1.01	1.04	1.25	553	1.01	1.05	1.40	3055	1.00	1.01	1.24	743	1.01	1.04	1.37	913
2011	1.00	1.00	1.17	546	1.01	1.05	1.31	2886	1.00	1.03	1.33	706	1.00	1.05	1.31	857
2012	1.04	1.20	1.62	92	1.02	1.08	1.3	2702	1.03	1.07	1.23	687	1.00	1.03	1.39	818

Table 9: **Summary Statistics of the Derived Cross-Sectional Weights at the Household Level (Percentiles of $HHRF up to Wave 29).**

	p5	p10	p25	p50	p75	p90	p95	*N*
1984	431	597	3805	4725	5647	7130	8248	*5921*
1985	483	683	3891	5084	6431	8471	10031	*5322*
1986	537	752	3598	5301	6841	9281	11117	*5090*
1987	546	792	3529	5382	7048	9580	11464	*5026*
1988	531	804	3528	5632	7542	10356	12546	*4814*
1989	551	820	3591	5841	7884	10812	13276	*4690*
1990	694	1073	2217	4602	7046	9882	12398	*6819*
1991	680	1043	2333	4691	7156	10285	12871	*6699*
1992	667	1027	2340	4653	7135	10532	13655	*6665*
1993	681	1056	2402	4668	7259	10748	13970	*6637*
1994	706	1099	2404	4666	7281	11220	14682	*6559*
1995	708	1108	2292	4389	7016	11087	14902	*6768*
1996	741	1162	2305	4375	7023	11373	15400	*6699*
1997	751	1187	2326	4332	7070	11836	15928	*6621*
1998	982	1332	2307	3984	6231	9954	13051	*7492*
1999	977	1309	2282	4000	6533	10893	14253	*7220*
2000	795	1095	1767	2529	3571	5087	6444	*13082*
2001	747	1030	1758	2755	4141	6084	7827	*11796*
2002	504	658	1218	2559	4185	6503	8250	*12320*
2003	501	671	1232	2564	4322	6789	9082	*11909*
2004	489	665	1216	2535	4423	7214	9839	*11644*
2005	489	674	1239	2539	4521	7506	10860	*11294*
2006	449	646	1264	2382	4132	7095	9723	*12361*
2007	446	645	1250	2454	4455	7658	10608	*11552*
2008	447	649	1274	2541	4729	8298	11541	*10921*
2009	461	665	1308	2630	4998	9138	12384	*10270*
2010	499	709	1395	2817	5459	10161	13287	*9551*
2011	421	636	1354	2480	4347	7404	9624	*12183*
2012	392	609	1377	2571	4388	7141	9699	*12217*

Table 10: **Summary Statistics of the Derived Cross-Sectional Weights at the Person Level (Percentiles of $PHRF up to Wave 29).**

	p5	p10	p25	p50	p75	p90	p95	N
1984	386	538	1159	4365	5229	6066	6888	*16173*
1985	457	635	1429	4623	5717	6859	8109	*14508*
1986	489	679	1536	4695	6027	7583	8984	*13804*
1987	508	712	1599	4705	6227	7914	9360	*13563*
1988	489	679	1619	4878	6556	8530	10145	*12872*
1989	526	754	1747	5012	6874	8976	10709	*12443*
1990	684	1023	1908	3440	6171	8299	10184	*18254*
1991	727	1070	1902	3712	6203	8466	10643	*17844*
1992	779	1133	1979	3737	6295	8726	11138	*17429*
1993	844	1233	2064	3827	6366	9007	11413	*17072*
1994	869	1272	2091	3824	6408	9259	12126	*16715*
1995	792	1133	1958	3559	6110	9176	12329	*17345*
1996	824	1172	1962	3600	6156	9440	12928	*16944*
1997	864	1207	2024	3628	6225	9696	13427	*16583*
1998	909	1242	2024	3503	5595	8536	11490	*18249*
1999	909	1226	1994	3478	5805	9319	12622	*17501*
2000	708	963	1553	2310	3229	4568	5851	*30784*
2001	677	924	1529	2449	3648	5433	6929	*27956*
2002	431	589	1043	2196	3718	5841	7627	*29101*
2003	431	600	1062	2198	3817	6169	8256	*27867*
2004	428	598	1064	2186	3924	6502	8839	*26918*
2005	433	612	1098	2242	4030	6879	9583	*25638*
2006	397	576	1099	2188	3686	6316	8868	*27442*
2007	402	584	1111	2242	3920	7068	10138	*25505*
2008	412	594	1140	2321	4137	7664	11196	*23792*
2009	428	610	1175	2401	4366	8393	12538	*22096*
2010	459	655	1259	2582	4775	9347	13702	*20281*
2011	396	589	1233	2322	4009	6914	9475	*25337*
2012	385	574	1276	2449	4116	6860	9457	*24987*

6 References

Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt. 1994. Das ADM-Stichproben - System (Stand 1993). In: S. Gabler, S., J. Hoffmeyer-Zlotnik, J. and D. Krebs, D. (eds): *Gewichtung in der Umfragepraxis*. Westdeutscher Verlag, Opladen: 188-203.

SOEP Group. 2001. The Socio-Economic Panel (SOEP) after more than 15 Years. *Vierteljahrshefte zur Wirtschaftsforschung* 1.

Haisken-DeNew, John and Joachim Frick (eds.) 2001: *Desktop Companion to the Socio Economic Panel (SOEP)*. Berlin, DIW.

Rendtel, Ulrich. 1991. Die Schätzung von Populationswerten in Panelerhebungen. *Allgemeines Statistisches Archiv* 75: 225-244.

Rendtel, Ulrich. 1995. *Panelausfälle und Panelrepräsentativität*. Frankfurt, Campus Verlag.

Rendtel, Ulrich, Markus Pannenberg and Stefan Daschke. 1997. Die Gewichtung der Zuwanderer-Stichprobe des Sozio-ökonomischen Panels (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung* 2: 271-286.

Schonlau, Matthias, Watson, Nicole, and Martin Kroh. 2011. Household Survey Panels: How Much do Following Rules Affect Sample Size? *Survey Research Methods* 5: 53-61.

Schonlau, Matthias, Kroh, Martin and Nicole Watson. 2013. The Implementation of Cross-Sectional Weights in Household Panel Surveys. *Statistics Surveys*.

Spiess, Martin and Ulrich Rendtel. 2000. Combining an Ongoing Panel with a New Cross-Sectional Sample. *DIW Discussion Paper* 198.

Spiess, Martin, Kroh, Martin, Pischner, Rainer, and Gert G. Wagner. 2008. On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP) - Tracing, Weighting, and Frequencies. *SOEP Paper* 98.