

Deutsches Institut für
Wirtschaftsforschung

 **DIW** BERLIN

Data Documentation

39

Martin Kroh • Martin Spieß

**Documentation of Sample Sizes and
Panel Attrition in the
German Socio Economic Panel (SOEP)
(1984 until 2007)**

Berlin, September 2008

IMPRESSUM

© DIW Berlin, 2008

DIW Berlin
Deutsches Institut für Wirtschaftsforschung
Mohrenstr. 58
10117 Berlin
Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
www.diw.de

ISSN 1861-1532

All rights reserved.
Reproduction and distribution
in any form, also in parts,
requires the express written
permission of DIW Berlin.



Data Documentation 39

Martin Kroh*

Martin Spieß**

Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2007)

Berlin, September 2008

* DIW Berlin, Socio-Economic Panel Study. mkroh@diw.de

** DIW Berlin, Socio-Economic Panel Study and University of Hamburg mspiess@diw.de

We would like to thank Matthias Giesecke for excellent research assistance.

Contents

1 Introduction	1
2 Developments in Sample Size.....	2
2.1 Development of the Number of Successful Interviews by Cross-Section	3
2.2 Continuance and Exit: The First Wave Gross Samples and their Participatory Behavior	9
2.3 New Entrants through Birth or Move into SOEP Households and Their Participation Behavior.....	13
2.4 The Risk of Survey-Related Panel Attrition.....	17
3 Panel Attrition Due to Unsuccessful Follow-Ups	21
3.1 The Frequency of Successful Follow-Ups	21
3.2 Predicting the Probability of Successful vs. Unsuccessful Follow-Ups in the Year 2007	23
4 Panel Attrition Due to Refusals.....	25
4.1 The Frequency of Participation.....	25
4.2 Predicting the Probability of Re-Interviewing versus Refusal in the Year 2007	27
5 Summary Statistics of the Derived Longitudinal and Cross-Sectional Weights	32
6 References	36

Tables

Table 1: The Frequency of Households to be Re-Contacted and the Relative Proportion of Unsuccessful Follow-Ups by Subsample and Year.	22
Table 2: Definition of the Regressors of the Logit Model of Unsuccessful Follow-Ups.	23
Table 3: Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2007.	24
Table 4: The Frequency of Re-Contacted Households and the Relative Proportion of Refusals by Subsample and Year.	26
Table 5: Definition of the Regressors of the Logit Model of Refusal.	28
Table 6a: Estimates of Logit Models of the Probability of Re-Interviewing a Household (Relative to Refusal) in 2007.	30
Table 6b: Estimates of Logit Model for the Probability of a Drop-Out of a Household Due to Refusal in 2007.	31
Table 7a: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples A through D (Percentiles of \$HBLEIB up to Wave 24).	33
Table 7b: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples E through G (Percentiles of \$HBLEIB up to Wave 24).	34
Table 8: Summary Statistics of the Derived Cross-Sectional Weights at the Household Level (Percentiles of \$HHRF up to Wave 24).	35

Figures

Figure 1: The Number of Successful Interviews with Persons by Subsamples A through H, Waves 1 to 24.	4
Figure 2: Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 24.	5
Figure 3: Comparison of Successful Interviews with Persons and Households (Subsample C), Waves 1 to 18.	6
Figure 4: Comparison of Successful Interviews with Individuals and Households (Subsample D), Waves 1 to 13.	7
Figure 5: Comparison of successful interviews with individuals and households (subsample E), waves 1 to 10.	7
Figure 6: Comparison of Successful Interviews with Individuals and Households (Subsample F), Waves 1 to 8.	8
Figure 7: Comparison of Successful Interviews with Individuals and Households (Subsample G), Waves 1 to 6.	8
Figure 8: All First-Wave Persons (Gross Subsample A). Development up to Wave 24.	9
Figure 9: All First Wave Persons (Gross Subsample B). Development up to Wave 24.	10
Figure 10: All First Wave Persons (Gross Subsample C). Development up to Wave 18.	10
Figure 11: All First Wave Persons (Gross Subsample D). Development up to Wave 13.	11
Figure 12: All First Wave Persons (Gross Subsample E). Development up to wave 10.	11
Figure 13: All First Wave Persons (Gross Subsample F). Development up to Wave 8.	12
Figure 14: All First Wave Persons (Gross Subsample G). Development up to Wave 6.	12
Figure 15: Entrants and their Participation Behavior (Subsamples A, B).	14
Figure 16: Entrants and their Participation Behavior (Subsample C).	14
Figure 17: Entrants and their Participation Behavior (Subsample D).	15
Figure 18: Entrants and their Participation Behavior (Subsample E).	15
Figure 19: Entrants and their Participation Behavior (Subsample F).	16
Figure 20: Entrants and their Participation Behavior (Subsample G).	16

Figure 21: Successful Re-Interviewing of First-Wave Respondents by Subsamples A, B, C. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.....	18
Figure 22: Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F, G. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.....	18
Figure 23: Successful Re-Interviewing of All First-Wave Respondents by Age Categories. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.....	19
Figure 24: Successful Re-Interviewing of All First-Wave Respondents by Occupation. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.	19
Figure 25: Successful Re-Interviewing of All First-Wave Respondents by Income Quintiles. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.....	20
Figure 26: Successful Re-Interviewing of All First-Wave Respondents by Education. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.	20

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 24 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 H 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). 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 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 2007 based on the characteristics of households measured in 2006. 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 2007: XHBLEIB and XPBLEIB. Based on the inverse of the probability of observing households and persons in 2006, the staying probability in 2007, and additional post-stratification to meet benchmarks of known marginals of the underlying population in 2007, we derive the cross-sectional weightsXHHRF and XPHRF. 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 H, Waves 1 to 24
- Figure 2** Comparison for Individuals and Households in Subsamples A and B, Waves 1 to 24 (1984 – 2007).
- Figure 3** Comparison for Individuals and Households in Subsample C, Waves 1 to 18, (1990–2007).
- Figure 4** Comparison for Individuals and Households in Subsample D, Waves 1 to 13, (1995–2007).
- Figure 5** Comparison for Individuals and Households in Subsample E, Waves 1 to 10, (1998–2007).
- Figure 6** Comparison for Individuals and Households in Subsample F, Waves 1 to 8, (2000–2007).
- Figure 7** Comparison for Individuals and Households in Subsample G, Waves 1 to 6, (2002-2007).

Figure 1: The Number of Successful Interviews with Persons by Subsamples A through H, Waves 1 to 24.

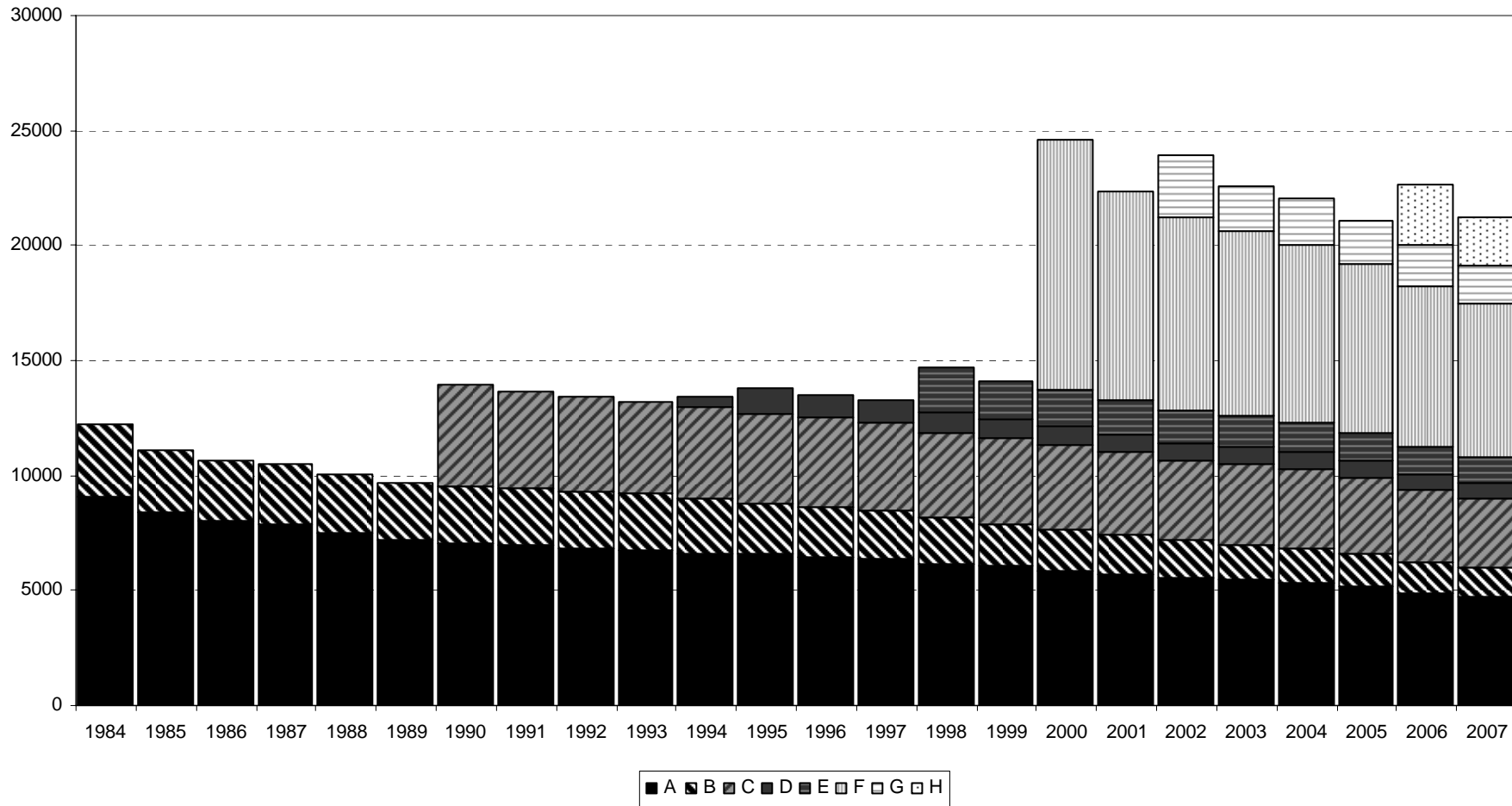
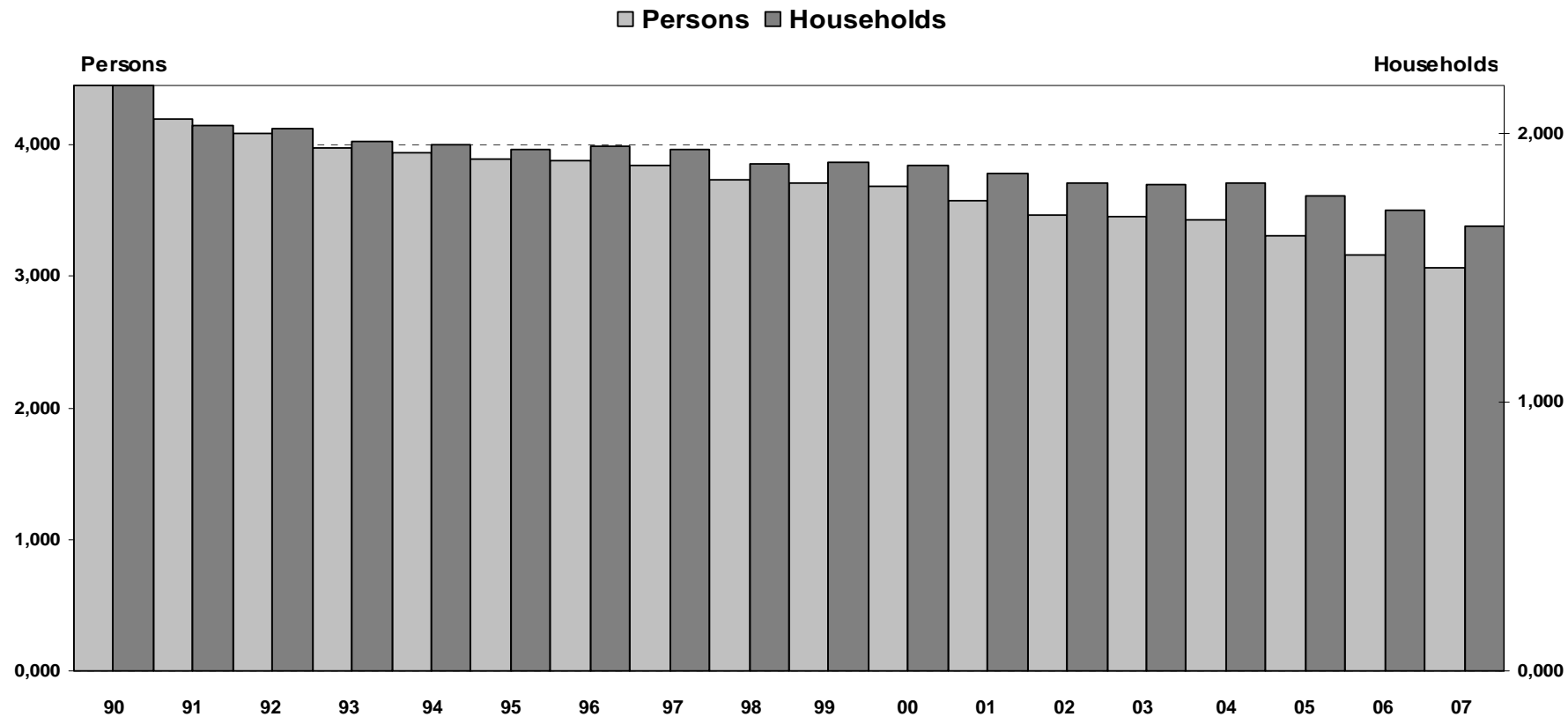


Figure 2: **Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 24.**



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
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
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

Figure 3: Comparison of Successful Interviews with Persons and Households (Subsample C), Waves 1 to 18.



Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
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
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

Data Documentation 39

2 Developments in Sample Size

Figure 4: Comparison of Successful Interviews with Individuals and Households (Subsample D), Waves 1 to 13.



Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Persons	1078	1023	972	885	838	837	789	780	789	758	734	684	658
Households	522	498	479	441	425	425	398	402	399	388	379	360	248

Figure 5: Comparison of successful interviews with individuals and households (subsample E), waves 1 to 10.

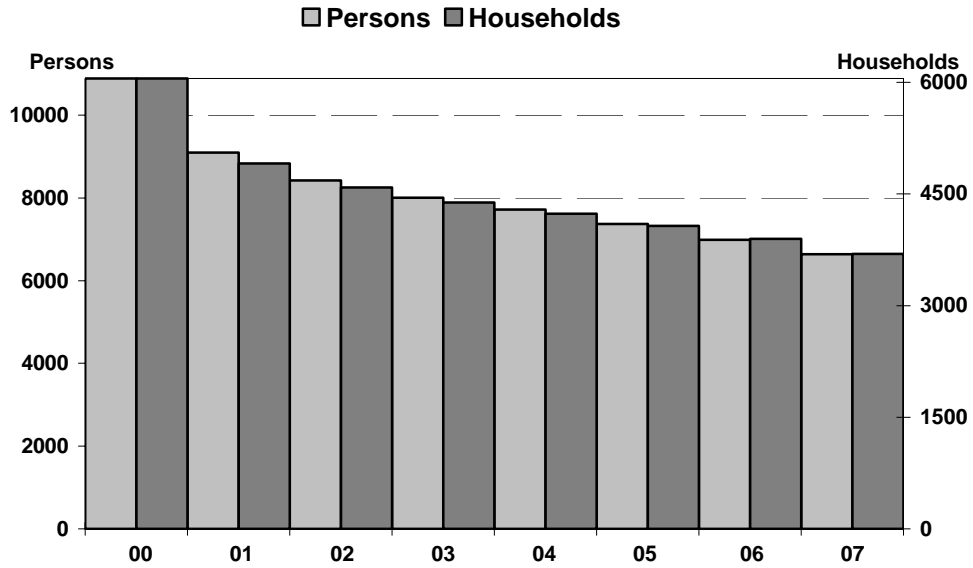


Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Persons	1910	1629	1549	1464	1373	1332	1300	1240	1198	1144
Households	1056	886	842	811	773	744	732	706	686	647

Data Documentation 39

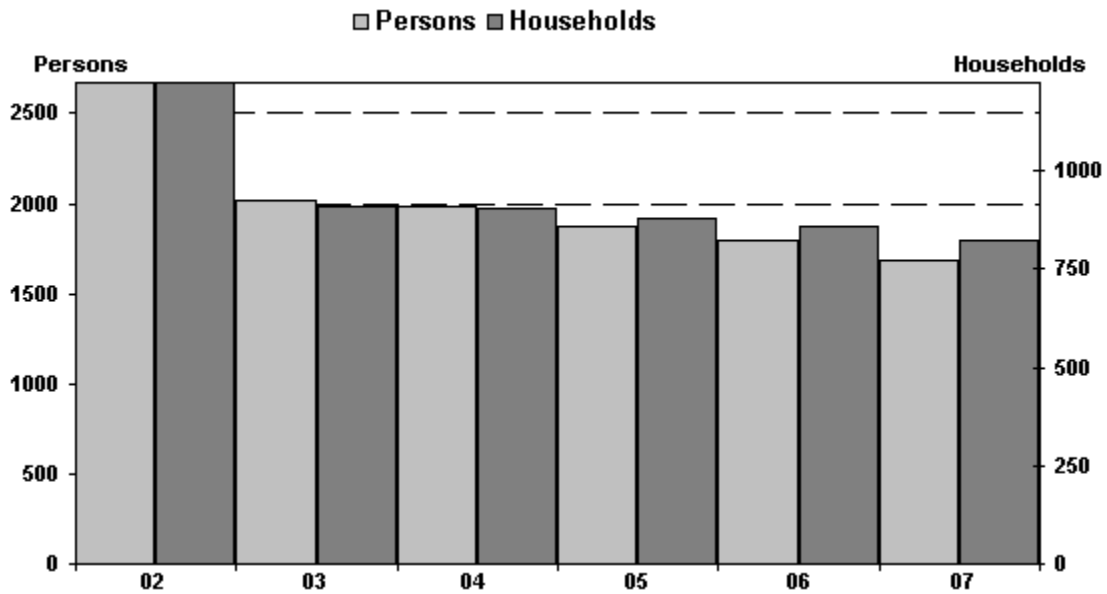
2 Developments in Sample Size

Figure 6: Comparison of Successful Interviews with Individuals and Households (Subsample F), Waves 1 to 8.



Year	2000	2001	2002	2003	2004	2005	2006	2007
Persons	10890	9098	8427	8006	7724	7371	6986	6640
Households	6052	4911	4586	4386	4234	4070	3895	3694

Figure 7: Comparison of Successful Interviews with Individuals and Households (Subsample G), Waves 1 to 6.



Year	2002	2003	2004	2005	2006	2007
Persons	2671	2013	1986	1870	1798	1682
Households	1224	911	904	879	859	824

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 8: All First Wave Persons in Subsample A. Whereabouts up to Wave 24.

Figure 9: All First Wave Persons in Subsample B. Whereabouts up to Wave 24.

Figure 10: All First Wave Persons in Subsample C. Whereabouts up to Wave 18.

Figure 11: All First Wave Persons in Subsample D. Whereabouts up to Wave 13.

Figure 12: All First Wave Persons in Subsample E. Whereabouts up to Wave 10.

Figure 13: All First Wave Persons in Subsample F. Whereabouts up to Wave 8.

Figure 14: All First Wave Persons in Subsample G. Whereabouts up to Wave 6.

Figure 8: All First-Wave Persons (Gross Subsample A). Development up to Wave 24.

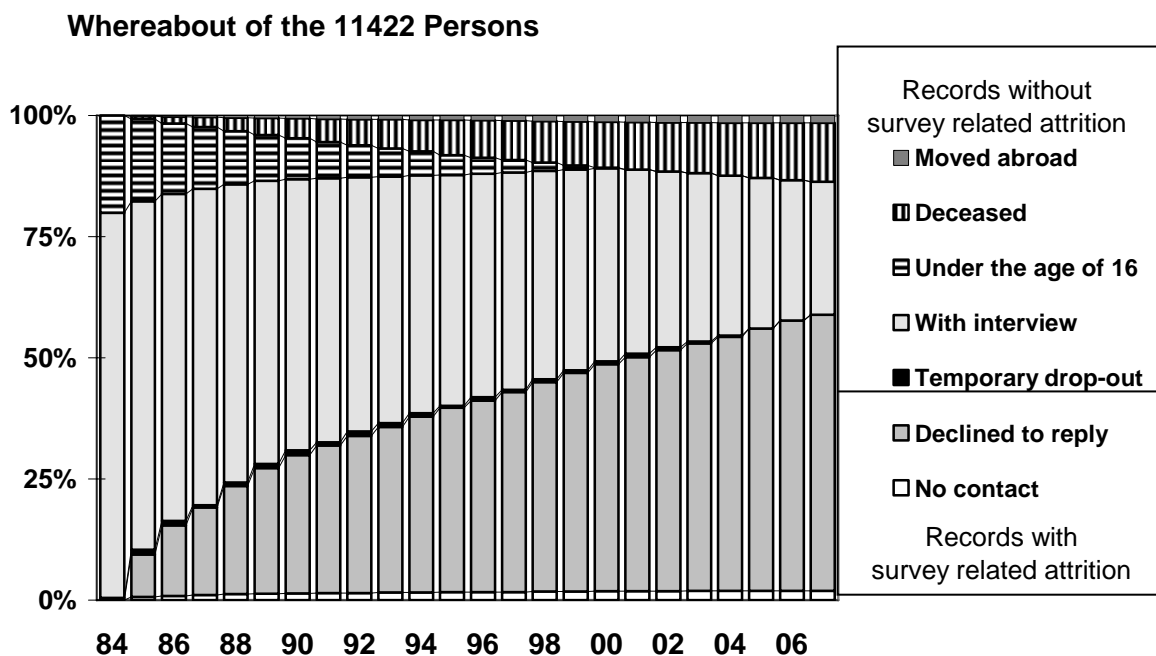


Figure 9: All First Wave Persons (Gross Subsample B). Development up to Wave 24.

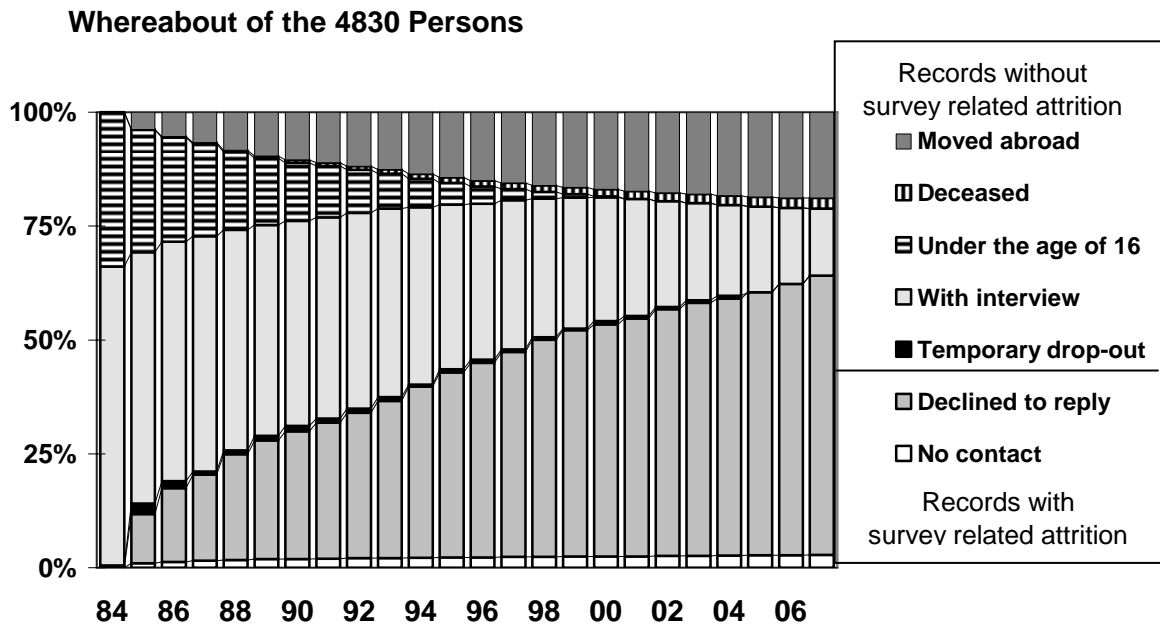


Figure 10: All First Wave Persons (Gross Subsample C). Development up to Wave 18.

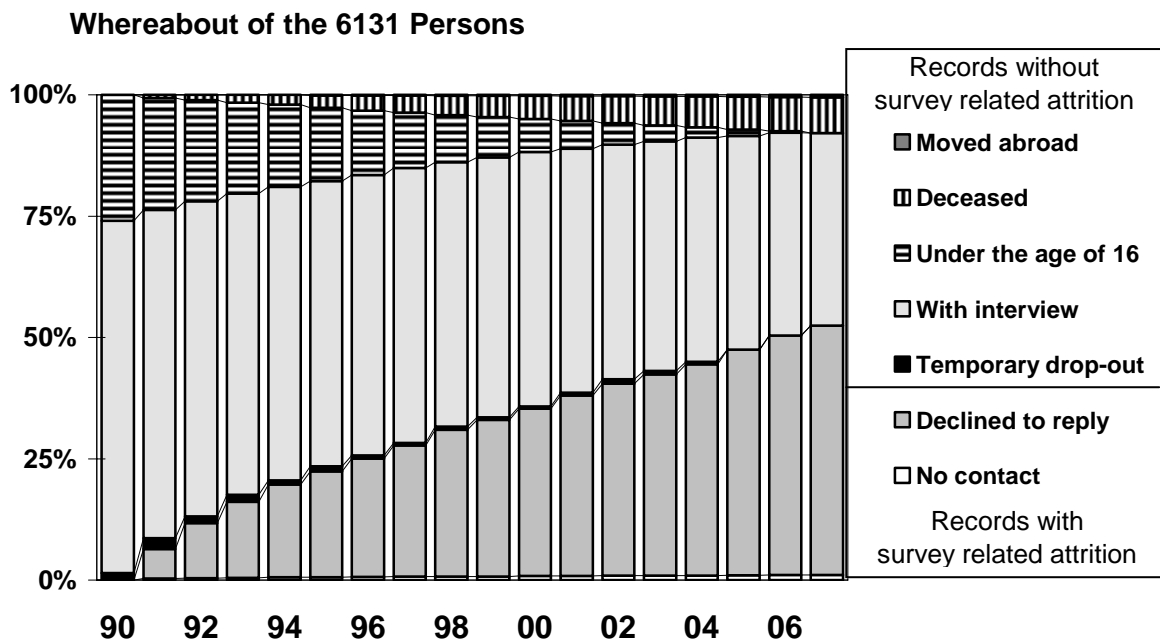


Figure 11: All First Wave Persons (Gross Subsample D). Development up to Wave 13.

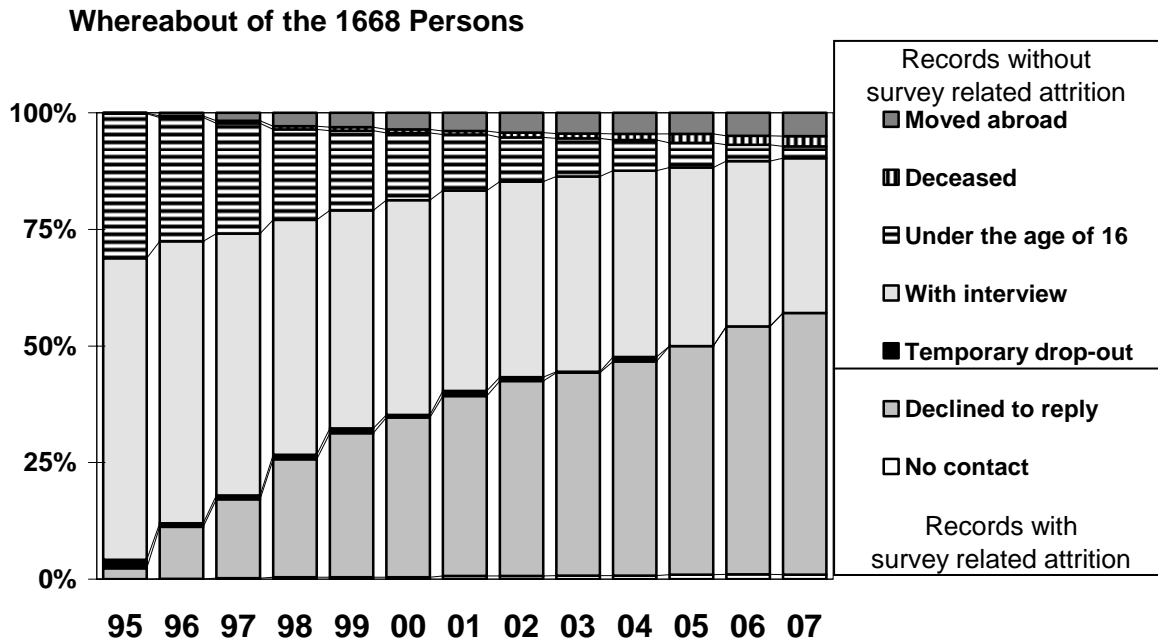


Figure 12: All First Wave Persons (Gross Subsample E). Development up to wave 10.

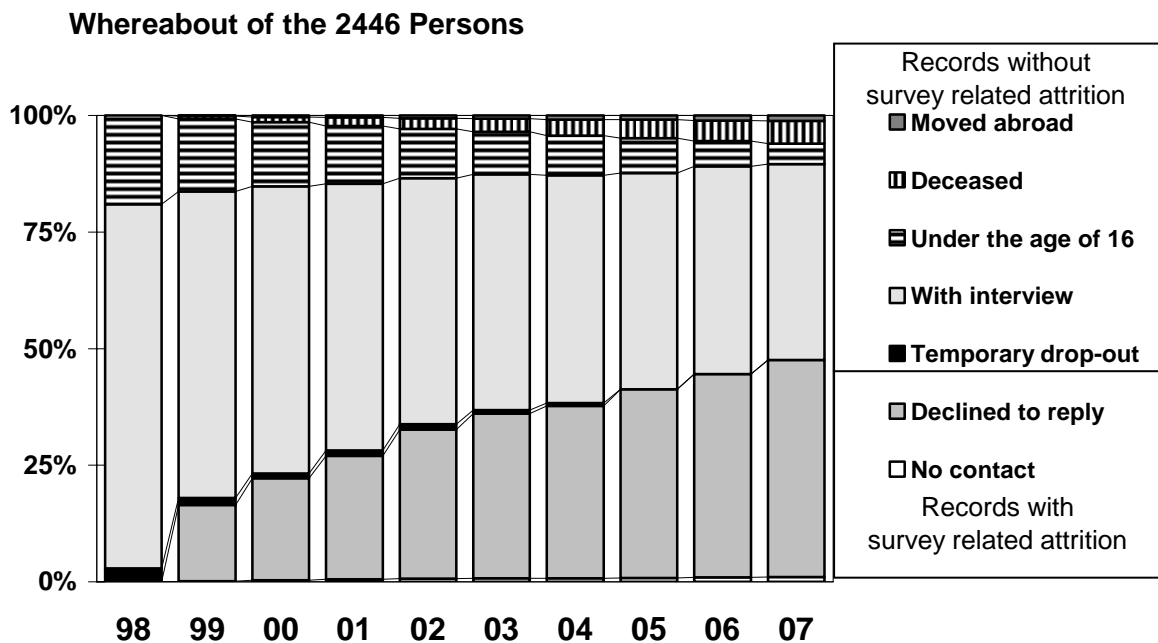


Figure 13: All First Wave Persons (Gross Subsample F). Development up to Wave 8.

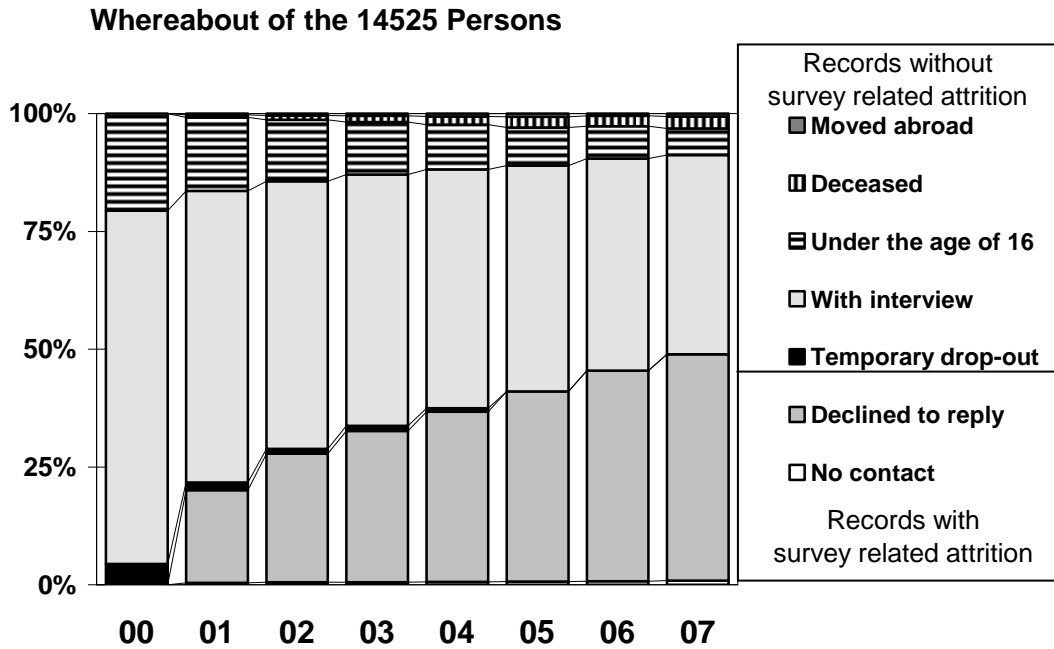
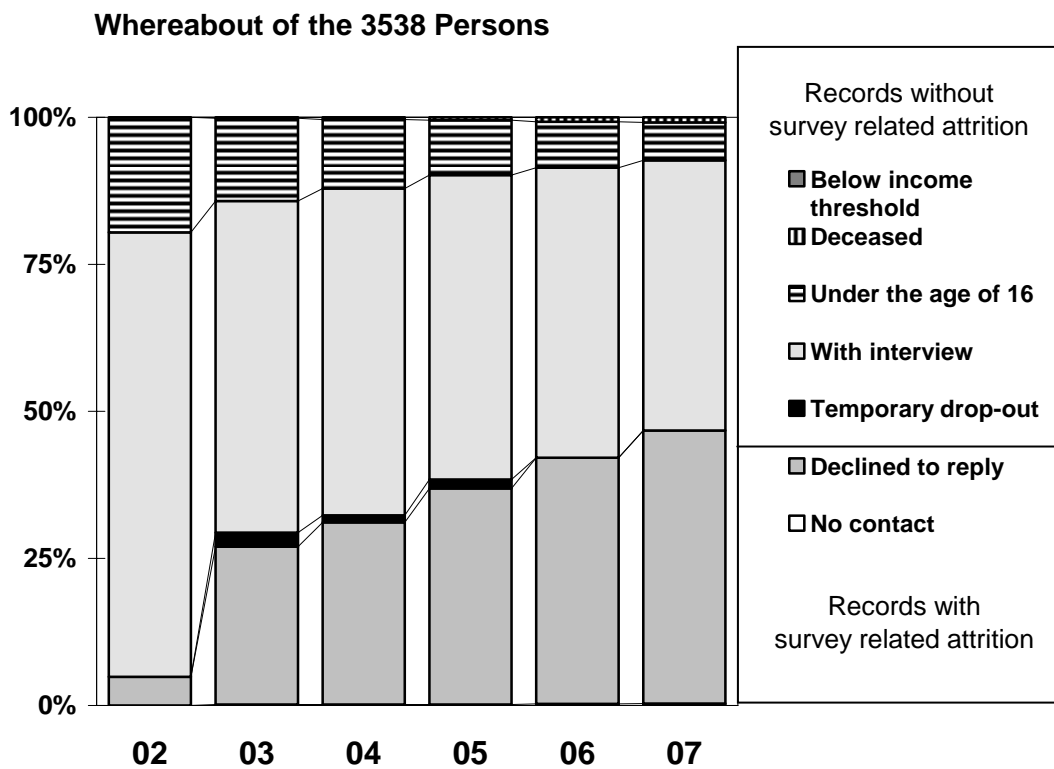


Figure 14: All First Wave Persons (Gross Subsample G). Development up to Wave 6.



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 15: Entrants Who Were Born or Moved into SOEP Households and Their Participation Behavior in Subsamples A and B

Figure 16: Entrants Who Were Born or Moved into SOEP Households and Their Participation Behavior in Subsample C

Figure 17: Entrants Who Were Born or Moved into SOEP Households and Their Participation Behavior in Subsample D

Figure 18: Entrants Who Were Born or Moved into SOEP Households and Their Participation Behavior in Subsample E

Figure 19: Entrants Who Were Born or Moved into SOEP Households and Their Participation Behavior in Subsample F

Figure 20: Entrants Who Were Born or Moved into SOEP Households and Their Participation Behavior in Subsample G

Figure 15: Entrants and their Participation Behavior (Subsamples A, B).

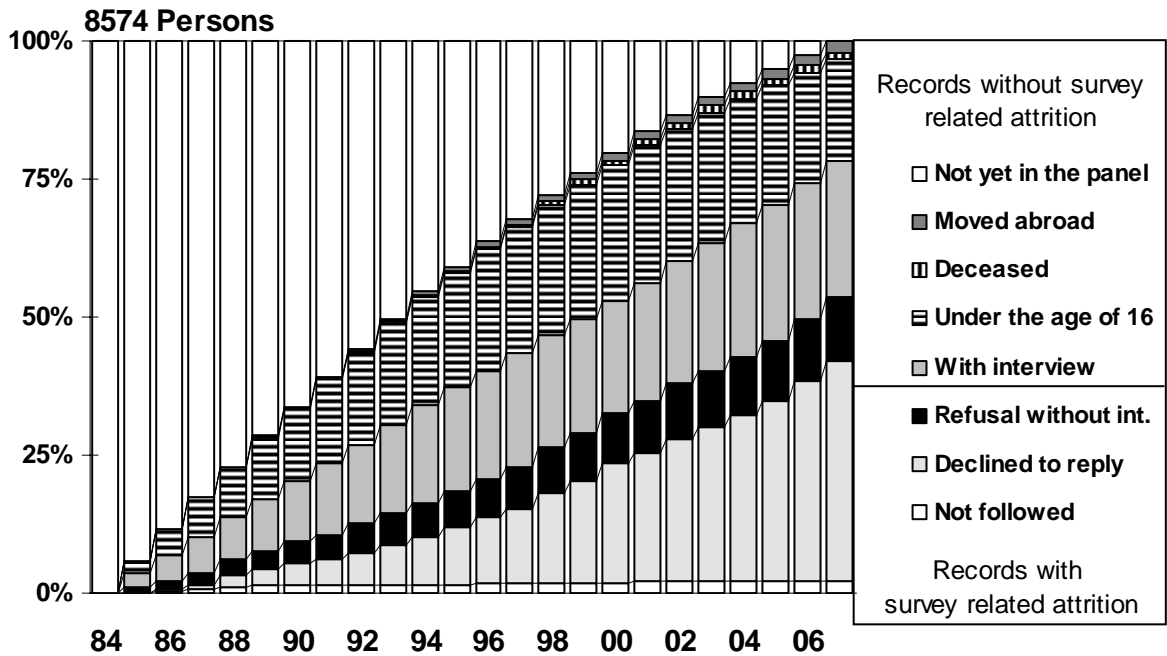


Figure 16: Entrants and their Participation Behavior (Subsample C).

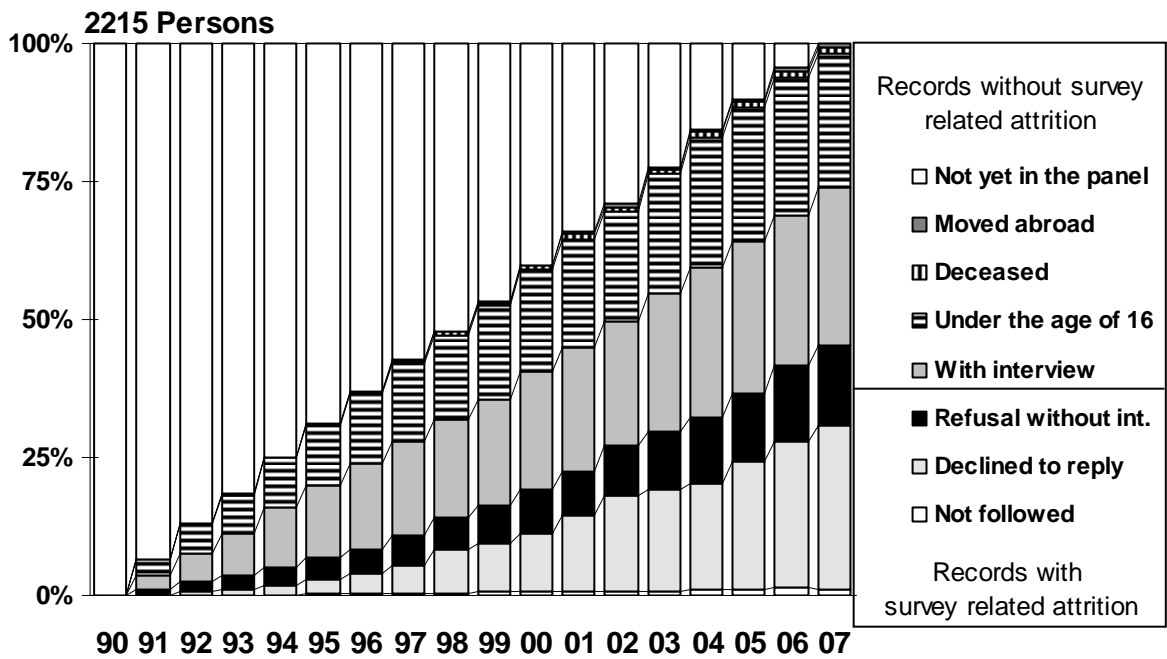


Figure 17: Entrants and their Participation Behavior (Subsample D).

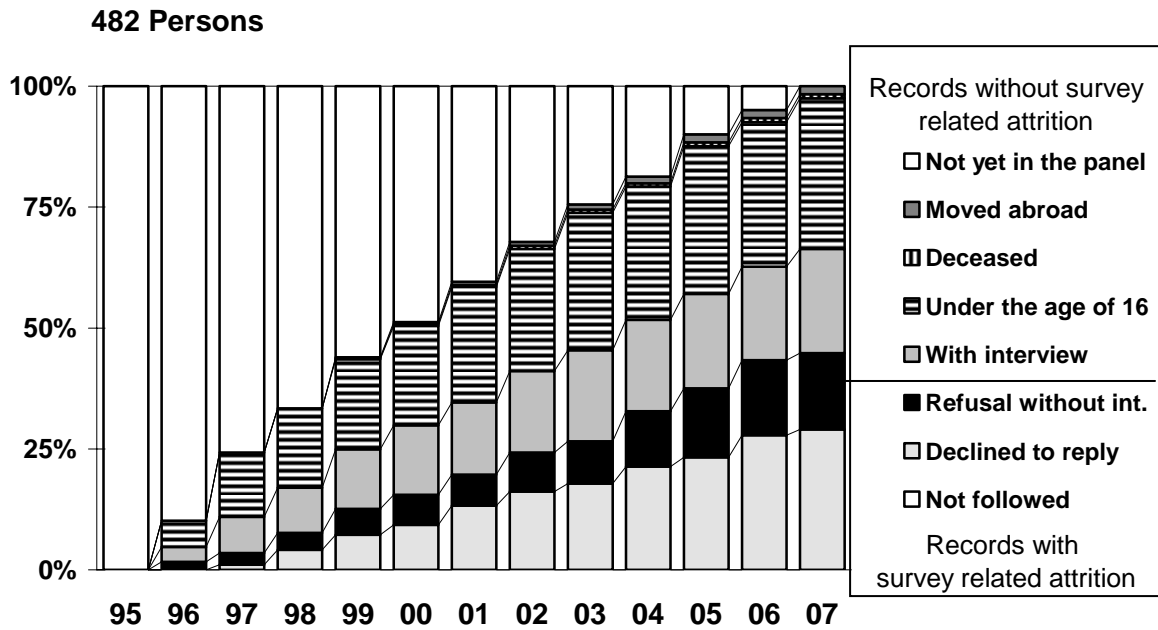


Figure 18: Entrants and their Participation Behavior (Subsample E).

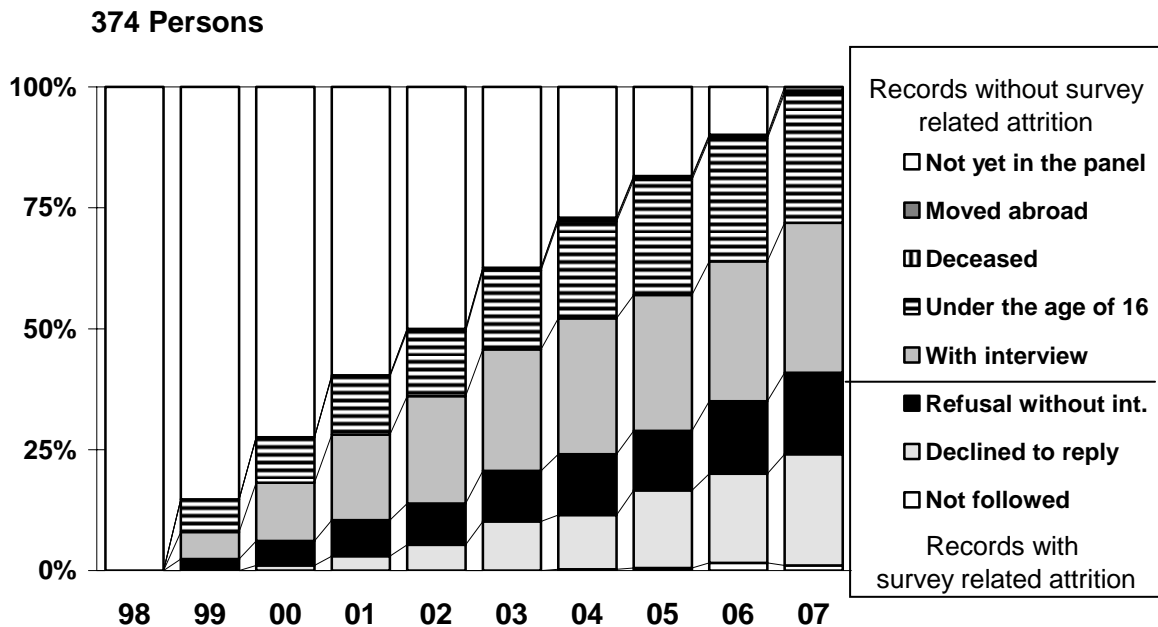


Figure 19: Entrants and their Participation Behavior (Subsample F).

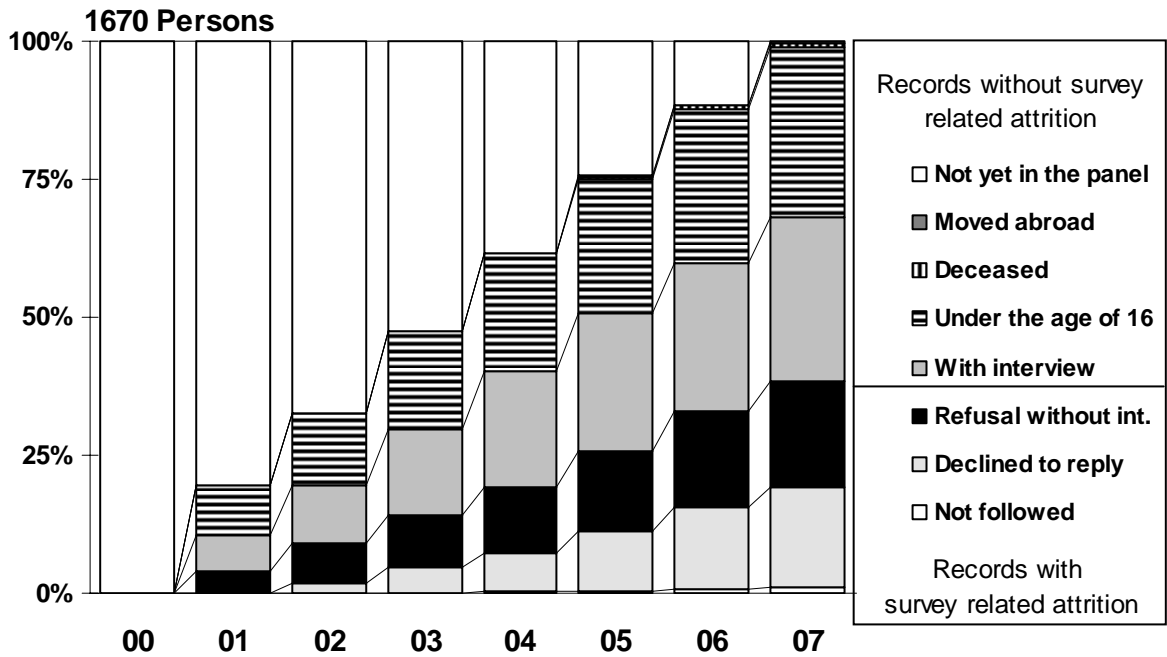
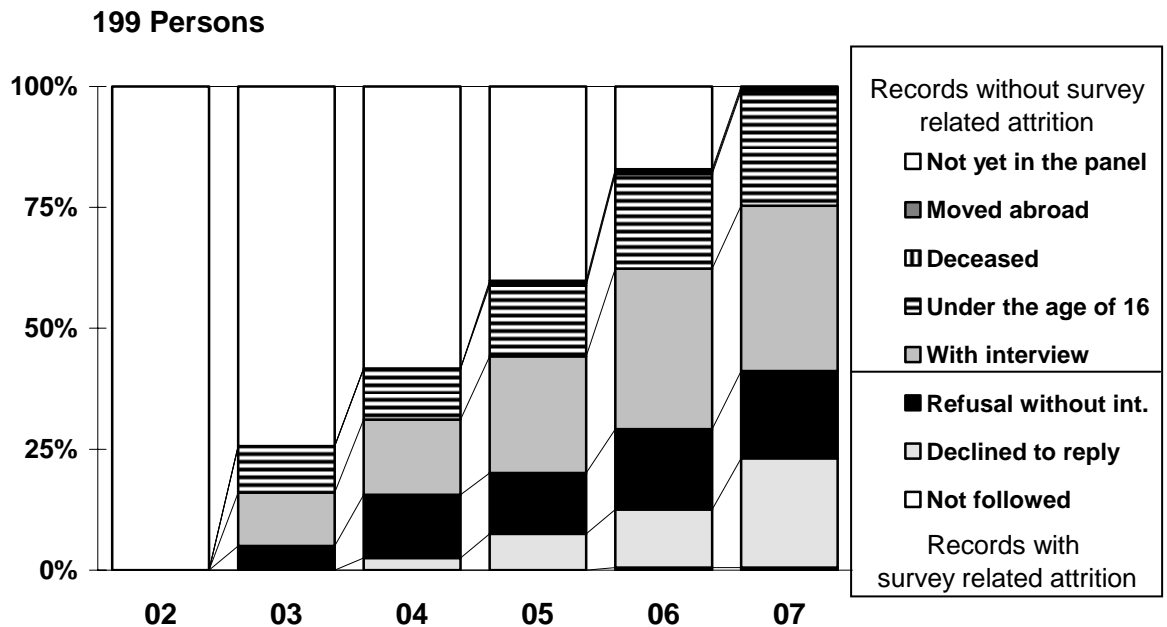


Figure 20: Entrants and their Participation Behavior (Subsample G).



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 21 and 22) and some basic socio-demographic characteristics measured in the year of sampling, such as age, occupation, income, and education (Figures 23 through 26). 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 21), 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 G (Figure 22), first-wave respondents from sample F have a somewhat lower probability of remaining in the survey than respondents from sample D.

Figure 21: Successful Re-Interviewing of First-Wave Respondents by Subsamples A, B, C.

Figure 22: Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F,

Figure 23: Successful Re-Interviewing of All First-Wave Respondents by Age Categories.

Figure 24: Successful Re-Interviewing of All First-Wave Respondents by Occupation.

Figure 25: Successful Re-Interviewing of All First-Wave Respondents by Income Quintiles.

Figure 26: Successful Re-Interviewing of All First-Wave Respondents by Education.

Figure 21: **Successful Re-Interviewing of First-Wave Respondents by Subsamples A, B, C. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

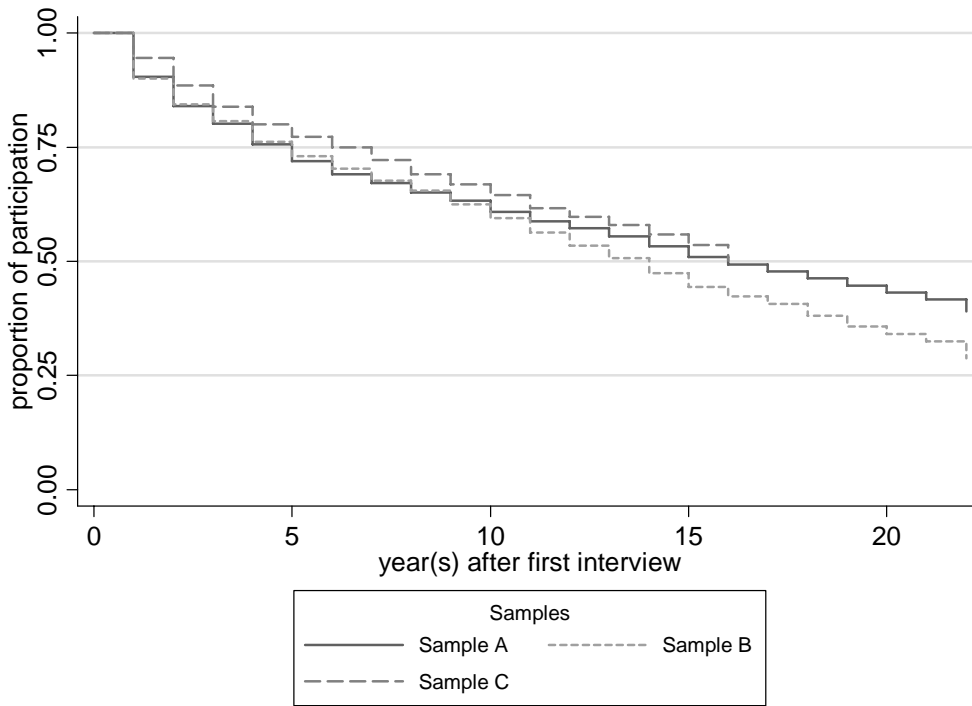
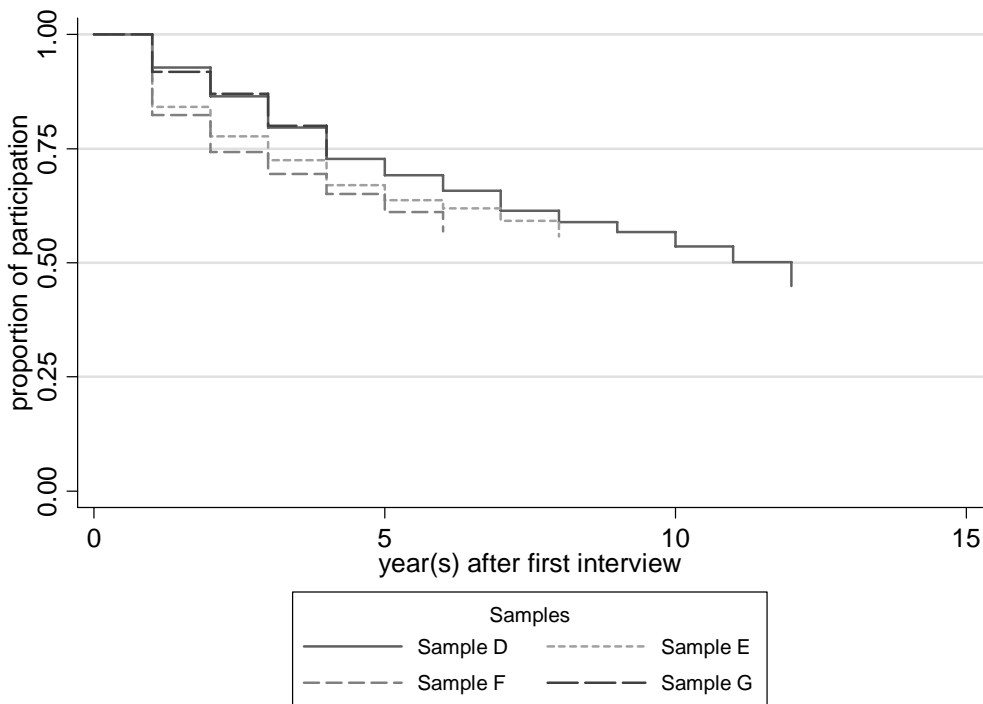


Figure 22: **Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F, G. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**



Data Documentation 39

2 Developments in Sample Size

Figure 23: Successful Re-Interviewing of All First-Wave Respondents by Age Categories. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.

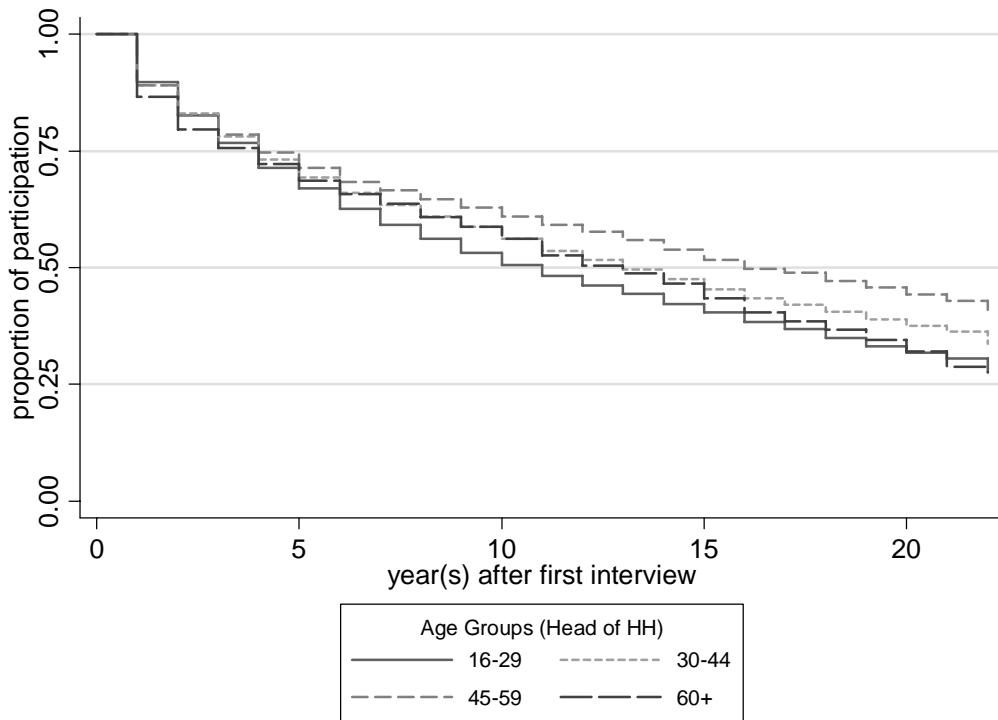


Figure 24: Successful Re-Interviewing of All First-Wave Respondents by Occupation. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.

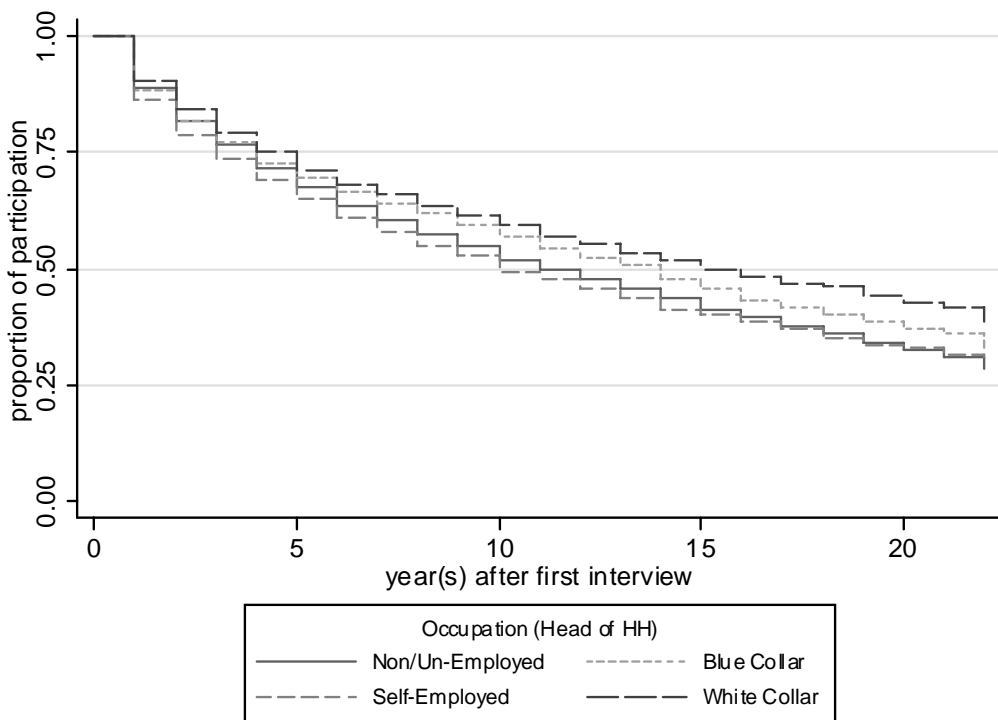


Figure 25: **Successful Re-Interviewing of All First-Wave Respondents by Income Quintiles. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

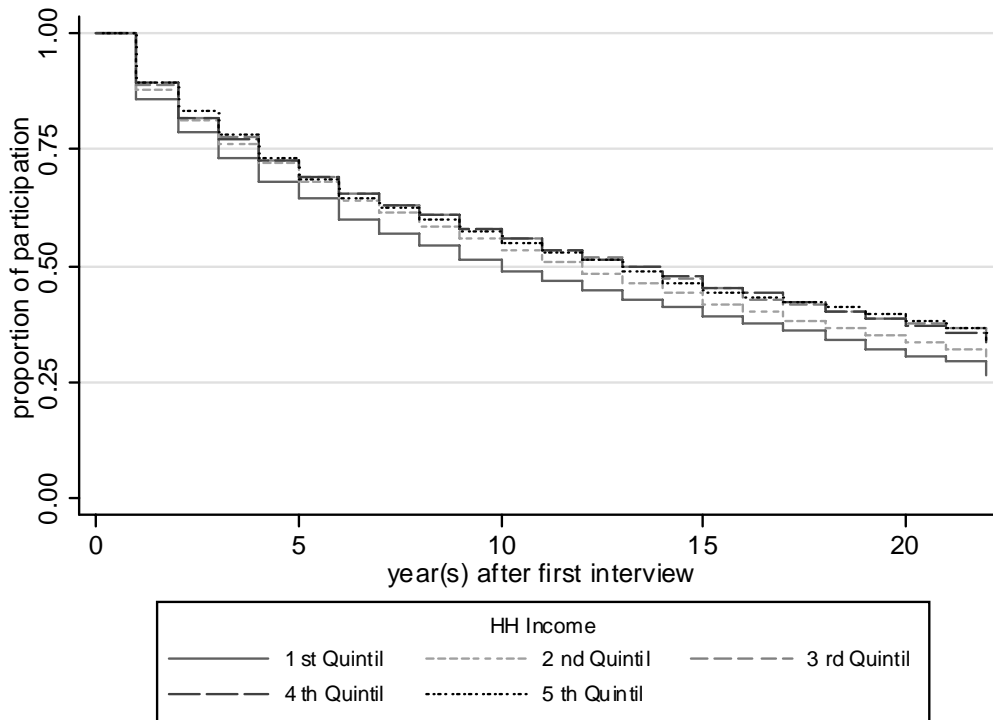
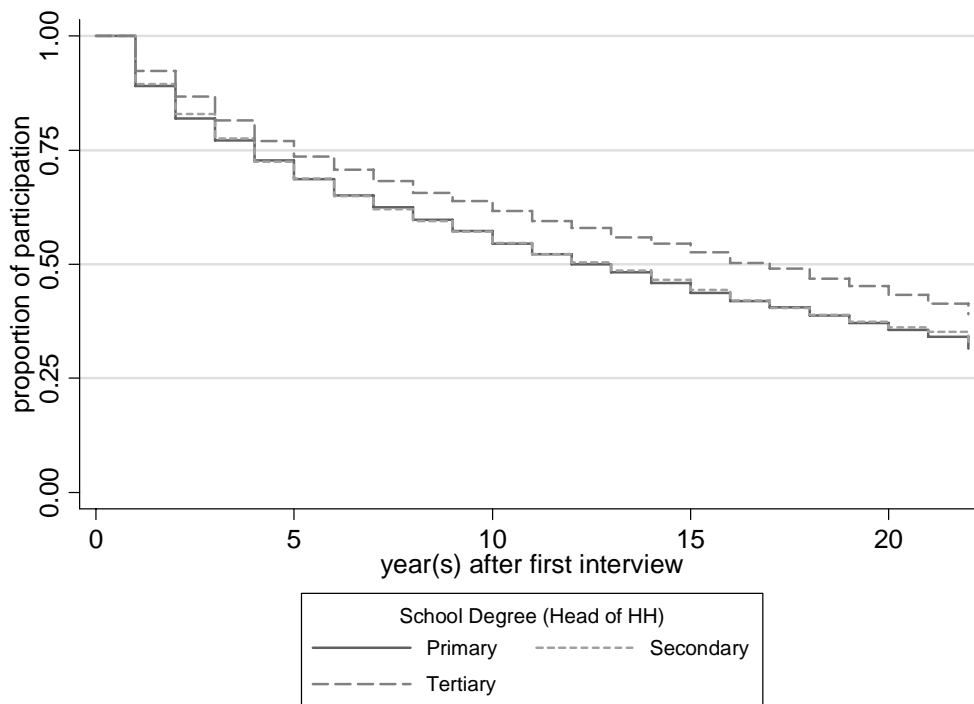


Figure 26: **Successful Re-Interviewing of All First-Wave Respondents by Education. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**



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 Sozialforschung, 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 H and waves 1985 through 2007. 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.

Data Documentation 39

3 Panel Attrition Due to Unsuccessful Follow-Ups

Table 1: The Frequency of Households to be Re-Contacted and the Relative Proportion of Successful Follow-Ups by Subsample and Year.

Year	A		B		C		D		E		F		G		H	
	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

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 2007

Based on the household and interview characteristics measured in 2006, we aim at predicting the probability of re-contacting a household relative to unsuccessful follow-up in 2007. 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 2006 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
New HH	New split off household with new address	0/1
Moved HH	Change in address of an existing household	0/1
Single HH	Single person household	0/1
Urban	Urban area (+ 100,000 inhabitants)	0/1
Mobile Area	Household situated in area with high mobility rate (Microm)	0/1
Living Apart Together (LAT)	At least one person in HH has partner outside of HH	0/1
Age Maximum 30	Oldest person in HH younger than 30	0/1
East	HH Located in East Germany = 1 / West Germany = 0	0/1

Data Documentation 39

3 Panel Attrition Due to Unsuccessful Follow-Ups

Table 3: Estimates of Logit Models of the Probability of Re-Contacting a Household (Relative to Unsuccessful Follow-Up) in 2007.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H
Intercept	-5.00 (0.41) ***	-4.86 (0.58) ***	-6.82 (0.71) ***	-4.93 (0.71) ***	-4.84 (0.72) ***	-3.04 (0.30) ***	-4.48 (0.39) ***	-2.88 (0.53) ***
New HH	-1.73 (0.36) ***				-1.71 (0.72) **	-2.08 (0.34) ***		-1.88 (0.54) ***
Moved HH		-1.60 (0.58) ***				-1.38 (0.37) ***		-1.05 (0.48) **
Single HH							-0.83 (0.39) **	
Urban						-0.56 (0.25) **		
Mobile Area	-0.81 (0.41) **							
LAT						-0.80 (0.25) ***		
Age Maximum 30								-1.61 (0.46) ***
East						-0.58 (0.25) **		
<i>Likelihood Ratio (Pr > Chisq)</i>	0.90	****	****	****	****	0.86	****	0.61

Note. *** p < 0.01; ** p < 0.05; * p < 0.10; standard errors in parentheses. **** The specified and the saturated models are the same.

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.

Data Documentation 39

4 Panel Attrition Due to Refusals

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

Year	A		B		C		D		E		F		G		H	
	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

n = Number of re-contacted households

% = Percentage of households that participated

4.2 Predicting the Probability of Re-Interviewing versus Refusal in the Year 2007

Based on the household and interview characteristics measured in 2006, we aim at predicting the probability of agreement vs. refusal to participate in the survey by the households that were re-contacted in 2007. 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).

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 2006 are not reported in the present data documentation due to space restrictions, but can be obtained from previous attrition reports.

Data Documentation 39

4 Panel Attrition Due to Refusals

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

Variable	Label	Value
First-Wave-HH	Household of the First Wave Sampling	0/1
Old HH	Household already observed in $t-1$, same address	0/1
New-HH	New split off household with new address	0/1
Freshmen	HH has been part of SOEP for a maximum of two years	0/1
HH-Move	Existing HH with new address	0/1
Face-to-Face	Face-to-face interview	0/1
CAPI	Computer Assisted Personal Interview	0/1
Change in Interviewer	Change in Interviewer between last waves	0/1
Change in Interview-Mode	HH has changed the interview-mode	0/1
Non-Regular Interview	No regular personal interview (e.g. interrupted)	0/1
Pace of Interview	Length of interview under 15 minutes	0/1
Respondent Cooperation	Low interviewer rating of respondents' cooperation	0/1
Email Disclosed	Email address known	0/1
Phone Disclosed	Telephone number known	0/1
2 Person HH	Two individuals living in HH	0/1
3 Person HH	Three individuals living in HH	0/1
Non-German HH	Head of household has non-German nationality	0/1
Gender	Gender of Head of HH (Male = 1)	0/1
Age 25-34	Head of household between 25 and 34	0/1
Age 35-64	Head of household between 35 and 64	0/1
(Age 25-34) * (Old HH)	Interaction term between respective variables	0/1
(Age 35-64) * (Old HH)	Interaction term between respective variables	0/1
Unmarried	Head of household unmarried	0/1
Married Living Apart	Married Couple living apart	0/1
Separation	Separation of couple	0/1
(Separation)*(Old HH)	Interaction term between respective variables	0/1
Hospital	Head of HH has been hospitalized at least once	0/1

Data Documentation 39

4 Panel Attrition Due to Refusals

State of Health	Head of HH's state of health is at least satisfactory	0/1
Savings	Household without savings and insurances	0/1
Tertiary Education	Head of Household with college or university degree	0/1
Vocational Education	Head of HH has vocational training	0/1
Voc. Edu. not Specified	No specification of vocational education degree of head of HH	0/1
Unemployment	At least one person in HH is unemployed	0/1
Job Worries	At least one person very concerned about own job security	0/1
Income Not Specified	No Information on HH-Income available	0/1
Welfare/Housing Subsidy	HH received welfare and/or housing subsidy	0/1
Dissatisfaction	Head of HH dissatisfied with life in general	0/1
Death	Relative/associate of HH has deceased recently	0/1
Anonymous Area (Microm)	HH located in area with high needs of anonymity	0/1
Middle-Class Area (Microm)	HH located in middle-class area ("Sinus-Milieu")	0/1
Affluent Area (Microm)	HH located in area with high socio-economic status	0/1
Newspaper Area (Microm)	HH located in area with high spreading of national newspapers	0/1
Mother/Child Questionnaire	HH with completed Mother/Child Questionnaire	0/1
Cognition Test Refusal	At least one person in HH has refused the Cognition Test	0/1
Cognition Test Participation	At least one person in HH has taken part in Cognition Test, no refusal	0/1

Data Documentation 39

4 Panel Attrition Due to Refusals

Table 6a: Estimates of Logit Models of the Probability of Re-Interviewing a Household (Relative to Refusal) in 2007.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H
Intercept	3.00 (0.28) ***	2.90 (0.35) ***	3.24 (0.30) ***	3.95 (0.80) ***	2.21 (0.48) ***	3.14 (0.17) ***	2.96 (0.33) ***	0.54 (0.16) ***
First Wave HH	0.18 (0.07) **				0.41 (0.14) ***			
New HH	-1.35 (0.29) ***					-1.15 (0.29) ***		
Freshmen			-1.36 (0.41) ***					
HH Move	-0.82 (0.25) ***					-0.76 (0.25) ***		
Face-to-Face	-1.98 (0.24) ***	-1.99 (0.33) ***	-2.11 (0.27) ***	-2.58 (0.63) ***	-1.89 (0.41) ***	-1.90 (0.16) ***	-1.79 (0.32) ***	
CAPI	-0.18 (0.08) **							
Change in Interviewer			-0.32 (0.14) **		-0.44 (0.20) **	-0.37 (0.08) ***	-0.44 (0.16) ***	
Change in Interview-Mode			-0.35 (0.15) **			-0.46 (0.11) ***		
Non-Regular Interview	-2.32 (0.24) ***	-2.21 (0.31) ***	-2.36 (0.26) ***	-2.29 (0.52) ***	-2.28 (0.39) ***	-2.34 (0.15) ***	-2.29 (0.30) ***	-1.40 (0.12) ***
Pace of Interview			0.27 (0.11) **					
Respondent Cooperation	-0.21 (0.08) **	-0.41 (0.15) ***	-0.41 (0.10) ***	-0.68 (0.30) **	-0.42 (0.17) **	-0.34 (0.07) ***		
Email Disclosed								1.20 (0.34) ***
Phone Disclosed	0.25 (0.11) **		0.36 (0.12) ***	0.84 (0.43) **	0.78 (0.27) ***			0.41 (0.13) ***
2 Person HH								-0.17 (0.07) **
3 Person HH							-0.36 (0.13) ***	
Non-German HH								-0.60 (0.22) ***
Gender				-0.99 (0.41) **				
Age 25-34	0.71 (0.28) **					0.72 (0.27) ***		
Age 35-64	0.76 (0.28) ***	0.33 (0.11) ***				0.65 (0.27) **		0.21 (0.08) ***
(Age 25-34)*(Old-HH)	-0.69 (0.29) **					-0.74 (0.29) **		
(Age 35-64)*(Old-HH)	-0.63 (0.28) **					-0.62 (0.27) **		
Unmarried			-0.19 (0.09) **				-0.35 (0.14) **	

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

Data Documentation 39

4 Panel Attrition Due to Refusals

Table 6b: Estimates of Logit Model for the Probability of Re-Interviewing a Household (Relative to Refusal) in 2007.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G	Sample H
Married Living Apart							-0.51 (0.26) **	
(Separation)*(Old-HH)			-1.22 (0.52) **					
Hospital		0.47 (0.22) **						
State of Health	0.18 (0.07) **							
Savings					-0.35 (0.16) **			
Tertiary Education						0.13 (0.06) **		
Vocational Education								-0.27 (0.09) ***
Voc. Ed. not Specified	-0.15 (0.07) **							-0.36 (0.12) ***
Unemployment						0.22 (0.09) **		
Job worries	-0.18 (0.07) ***						0.40 (0.15) **	
Income Not Specified						-0.22 (0.09) **		-0.37 (0.10) ***
Welfare/Housing Subsidy			-0.53 (0.20) ***					
Dissatisfaction							-0.46 (0.22) **	
Death								-0.33 (0.16) **
Anonymous Area	-0.16 (0.07) **							
Middle-Class Area	-0.16 (0.06) **		-0.17 (0.08) **					
Affluent Area			-0.23 (0.11) **					
Newspaper Area							0.30 (0.11) ***	
Mother/Child Quest.			0.63 (0.28) **					
Cognition Test Particip.						0.15 (0.07) **		
Cognition Test Refusal		-0.47 (0.17) ***						
Likelihood Ratio (Pr > Chisq)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

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 2007 based on characteristics measured in 2006, XHBLEIB, lends itself as a longitudinal weighting variable correcting for selective attrition between waves 2006 and 2007. Table 7 reports some sub-sample specific descriptive statistics of the longitudinal weights in each wave.

The product of the cross-sectional weight in 2006, WHHRF, and the longitudinal weight in 2007, XHBLEIB, provide the raw data for the cross-sectional weight in 2007. In a final step, reported in DIW data documentation 22 by Pischner (2007), the post-stratification of the cross-sectional weights corrects them to meet benchmarks of known marginals of the underlying population in 2007. Table 8 reports sub-sample-specific descriptive statistics of the derived cross-sectional weighting variable XHHRF and in comparison all previous cross-sectional weights AHHRF through WHHRF.

Data Documentation 39

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

Table 7a: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples A through D (Percentiles of \$HBLEIB up to Wave 24).

	bhbleib	chbleib	dhbleib	ehbleib	fhbleib	ghbleib	hhbleib	ihbleib	jhbleib	khbleib	lhbleib	mhbleib	nhbleib	ohbleib	phbleib	qhbleib	rhbleib	shbleib	thbleib	uhbleib	vhbleib	whbleib	xhbleib
<u>sample A</u>																							
p10	1.06	1.04	1.03	1.02	1.03	1.02	1.02	1.01	1.01	1.02	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.01	1.01	1.01	1.02	1.01	1.01
p50	1.1	1.07	1.03	1.04	1.04	1.02	1.02	1.02	1.01	1.02	1.01	1.03	1.02	1.03	1.02	1.02	1.02	1.02	1.03	1.01	1.02	1.04	1.03
p90	1.22	1.26	1.13	1.19	1.16	1.11	1.09	1.11	1.16	1.15	1.16	1.12	1.13	1.14	1.2	1.15	1.18	1.21	1.14	1.12	1.16	1.22	1.14
N	4141	3962	3910	3731	3647	3612	3613	3584	3603	3577	3526	3485	3458	3387	3325	3240	3168	3123	3072	3010	2937	2821	2723
<u>sample B</u>																							
p10	1.09	1.1	1.03	1.03	1.03	1.04	1.03	1.01	1.02	1.03	1.02	1.04	1.02	1.04	1.04	1.03	1.02	1.04	1.01	1.04	1.05	1.01	1.03
p50	1.1	1.1	1.03	1.04	1.04	1.04	1.03	1.03	1.03	1.05	1.05	1.04	1.04	1.07	1.04	1.03	1.02	1.04	1.03	1.04	1.05	1.05	1.07
p90	1.26	1.29	1.14	1.22	1.14	1.12	1.16	1.16	1.22	1.22	1.29	1.21	1.29	1.23	1.22	1.18	1.23	1.37	1.31	1.13	1.17	1.33	1.24
N	1181	1128	1116	1069	1043	1028	1056	1060	1064	1023	982	960	931	898	858	820	809	766	742	714	698	655	614
<u>sample C</u>																							
p10							1.03	1.06	1.03	1.02	1.03	1.01	1.02	1.02	1.01	1.01	1.02	1.01	1.01	1	1	1.01	1
p50							1.06	1.06	1.04	1.04	1.03	1.02	1.04	1.02	1.03	1.03	1.02	1.02	1.03	1.01	1.02	1.04	1.03
p90							1.18	1.22	1.17	1.12	1.11	1.15	1.12	1.2	1.1	1.13	1.16	1.21	1.14	1.12	1.15	1.24	1.16
N							2030	2020	1970	1959	1938	1951	1942	1886	1894	1879	1850	1818	1807	1813	1771	1717	1654
<u>sample D</u>																							
p10												1	1.05	1.08	1.05	1.02	1.03	1	1.01	1	1	1.03	1.01
p50												1.08	1.09	1.08	1.05	1.02	1.03	1.02	1.01	1.01	1.02	1.04	1.04
p90												1.14	1.09	1.35	1.27	1.1	1.17	1.21	1.09	1.25	1.34	1.44	1.12
N												395	336	302	296	293	273	285	290	277	273	261	248

Data Documentation 39

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

Table 7b: Summary Statistics of the Derived Longitudinal Weights at the Household Level for Subsamples E through H (Percentiles of \$HBLEIB up to Wave 24).

	bhbleib	chbleib	dhbleib	ehbleib	fhbleib	ghbleib	hhbleib	ihbleib	jhbleib	khbleib	lhbleib	mhbleib	nhbleib	ohbleib	phbleib	qhbleib	rhbleib	shbleib	thbleib	uhbleib	vhbleib	whbleib	xhbleib
<u>sample E</u>																							
p10															1	1.03	1.01	1.01	1.04	1	1.01	1	1.01
p50															1.23	1.07	1.05	1.02	1.04	1.01	1.03	1.03	1.01
p90															1.47	1.21	1.25	1.2	1.15	1.08	1.18	1.21	1.16
N															886	838	811	773	744	732	706	686	647
<u>sample F</u>																							
p10																	1.08	1.03	1.02	1.02	1.01	1.01	1.02
p50																	1.14	1.05	1.04	1.03	1.03	1.03	1.03
p90																	1.59	1.46	1.24	1.19	1.17	1.29	1.15
N																	4911	4586	4386	4235	4070	3895	3694
<u>sample G</u>																							
p10																			1.06	1.02	1.03	1	1.02
p50																			1.1	1.03	1.06	1.04	1.05
p90																			1.17	1.25	1.25	1.31	1.17
N																			911	904	879	859	824
<u>sample H</u>																							
p10																							1.04
p50																							1.16
p90																							1.46
N																							1188

Data Documentation 39

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

Table 8: Summary Statistics of the Derived Cross-Sectional Weights at the Household Level (Percentiles of \$HHRF up to Wave 24).

	ahhrf	bhhrf	chhrf	dhhrf	ehhrf	fhhrf	ghhrf	hhhrf	ihhrf	jhhrf	khhrf	lhhrf	mhhrrf	nhhrf
p5	256.92	301.98	314.52	352.65	340.69	369.13	560.49	644.36	643.87	627.38	680.89	643.6	641.49	675.41
p10	456.65	547.79	562.38	593.65	590.76	638.58	1035.65	1133.21	1149.45	1132.3	1178.01	1140.78	1128.31	1139.27
p25	1914.36	2207.28	2257.76	2281.86	2395.92	2488.34	2142.07	2204.61	2214.14	2204.54	2196.46	2170.56	2131.24	2092.16
p50	4101.62	4495.88	4611.355	4595.165	4790.225	4964.75	3745.41	3840.76	3838.29	3916.1	3939.19	3757.75	3713.38	3751.58
p75	6161.5	6970.95	7366.56	7551.34	7987.74	8258.3	6756.27	6988.9	6969.49	7083.42	7161.04	6812.035	6774.8	6850.03
p90	8555.59	9765.73	10743.81	11108.66	11987.33	12339.7	10772.53	11122.55	11251.41	11604.53	11944.66	11539.68	11856.92	12281.5
p95	10460.91	11978.65	13379.31	13838.91	14916.38	15915.27	14312.25	14935.49	15312.78	15631.78	16415.94	16348.84	17119.6	17904.04
N	5921	5322	5090	5026	4814	4690	6819	6699	6665	6637	6559	6768	6698	6617

	ohhrf	phhrf	qhhrf	rhhrf	shhrf	thhrf	uhhrf	vhhrf	whhrf	xhhrf
p5	673.22	682.64	562.28	528.14	528.75	521.24	503.21	494.59	476.17	450.62
p10	1088.45	1075.05	850.92	816.11	817.73	796.23	772.23	759.78	717.67	698.81
p25	1994.82	1941.39	1521.5	1530.61	1513.97	1466.98	1417.62	1424.14	1367.87	1305.02
p50	3825.75	3756.5	2380.28	2592.01	2586.585	2575.96	2531.555	2512.6	2470.74	2436.43
p75	6150.22	6451.12	3526.25	4044.05	4205.83	4305.96	4351.4	4445.37	3990.04	4370.14
p90	9905.59	10700.84	5280.91	6183.89	6747.815	7093.06	7490.3	7921.29	6736.14	7806.00
p95	14422.31	15628.84	7229.52	8401.11	9542.31	10295.62	11062.27	11885.03	10499.36	11996.00
N	7486	7215	13078	11783	11310	10999	10740	10416	11505	10728

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.
- Pischner, Rainer. 2007. Die Querschnittsgewichtung und die Hochrechnungsfaktoren des Sozio-oekonomischen Panels (SOEP) ab Release 2007 (Welle W): Modifikationen und Aktualisierungen. *DIW Data Documentation* 22.
- 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.
- Schupp, Jürgen and Gert Wagner. 1995. Die Zuwandererstichprobe des Sozio-oekonomischen Panels (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung* 1.
- Spiess, Martin and Ulrich Rendtel. 2000. Combining an Ongoing Panel with a New Cross-Sectional Sample. *DIW Discussion Paper* 198.
- Spiess, Martin. 2001. Derivation of Design Weights: The Case of the Socio-Economic Panel (SOEP). *DIW Research Notes* 5.
- 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.
- von Rosenblatt, B., SOEP-Gruppe, TNS Infratest Sozialforschung. 2004. *SOEP 2004 Methodenbericht zum Befragungsjahr 2004 (Welle 21) des Sozio-oekonomischen Panels*.