

Spieß, Martin; Kroh, Martin

Research Report

Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2006)

DIW Data Documentation, No. 27

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Spieß, Martin; Kroh, Martin (2008) : Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2006), DIW Data Documentation, No. 27, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/129236>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Data Documentation

27

Martin Spieß • Martin Kroh

**Documentation of Sample Sizes and Panel
Attrition in the German Socio Economic
Panel (SOEP) (1984 until 2006)**

Berlin, January 2008

IMPRESSUM

© DIW Berlin, 2008

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

ISSN 1614-6912

Alle Rechte vorbehalten.
Abdruck oder vergleichbare
Verwendung von Arbeiten
des DIW Berlin ist auch in
Auszügen nur mit vorheriger
schriftlicher Genehmigung
gestattet.

Data Documentation 27

Martin Spieß*

Martin Kroh**

Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2006)

Berlin, January 2008

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

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

We would like to thank Lars Gondolatsch 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 Unsuccessful Follow-Ups	21
3.2 Predicting the Probability of Successful vs. Unsuccessful Follow-Ups in the Year 2006	23
4 Panel Attrition Due to Refusals.....	25
4.1 The Frequency of Refusals.....	25
4.2 Predicting the Probability of Re-Interviewing versus Refusal in the Year 2006	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 2006.	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 2006.	30
Table 6b: Estimates of Logit Model for the Probability of a Drop-Out of a Household Due to Refusal in 2006.	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 23).	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 23).	34
Table 8: Summary Statistics of the Derived Cross-Sectional Weights at the Household Level (Percentiles of \$HHRF up to Wave 23).	35

Figures

Figure 1: The Number of Successful Interviews with Persons by Subsamples A through H, Waves 1 to 23.	4
Figure 2: Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 23.	5
Figure 3: Comparison of Successful Interviews with Persons and Households (Subsample C), Waves 1 to 17.	6
Figure 4: Comparison of Successful Interviews with Individuals and Households (Subsample D), Waves 1 to 12.	7
Figure 5: Comparison of successful interviews with individuals and households (subsample E), waves 1 to 9.	7
Figure 6: Comparison of Successful Interviews with Individuals and Households (Subsample F), Waves 1 to 7.	8
Figure 7: Comparison of Successful Interviews with Individuals and Households (Subsample G), Waves 1 to 5.	8
Figure 8: All First-Wave Persons (Gross Subsample A). Development up to Wave 23.	9
Figure 9: All First Wave Persons (Gross Subsample B). Development up to Wave 23.	10
Figure 10: All First Wave Persons (Gross Subsample C). Development up to Wave 17.	10
Figure 11: All First Wave Persons (Gross Subsample D). Development up to Wave 12.	11
Figure 12: All First Wave Persons (Gross Subsample E). Development up to wave 9.	11
Figure 13: All First Wave Persons (Gross Subsample F). Development up to Wave 7.	12
Figure 14: All First Wave Persons (Gross Subsample G). Development up to Wave 5.	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 23 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”. The SOEP-team currently prepares an additional DIW data documentation that outlines the rules for inclusion of new sample units and their treatment within the weighting framework. 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 2006 based on the characteristics of households measured in 2005. 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 2006: WHBLEIB and WPBLEIB. Based on the inverse of the probability of observing households and persons in 2005, the staying probability in 2006, and additional post-stratification to meet benchmarks of known marginals of the underlying population in 2006, we derive the cross-sectional weights WHHRF and WPHRF. 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 23
- Figure 2** Comparison for Individuals and Households in Subsamples A and B,
 Waves 1 to 23 (1984 – 2006).
- Figure 3** Comparison for Individuals and Households in Subsample C,
 Waves 1 to 17, (1990–2006).
- Figure 4** Comparison for Individuals and Households in Subsample D,
 Waves 1 to 12, (1995–2006).
- Figure 5** Comparison for Individuals and Households in Subsample E,
 Waves 1 to 9, (1998–2006).
- Figure 6** Comparison for Individuals and Households in Subsample F,
 Waves 1 to 7, (2000–2006).
- Figure 7** Comparison for Individuals and Households in Subsample G,
 Waves 1 to 5, (2002-2006).

Figure 1: The Number of Successful Interviews with Persons by Subsamples A through H, Waves 1 to 23.

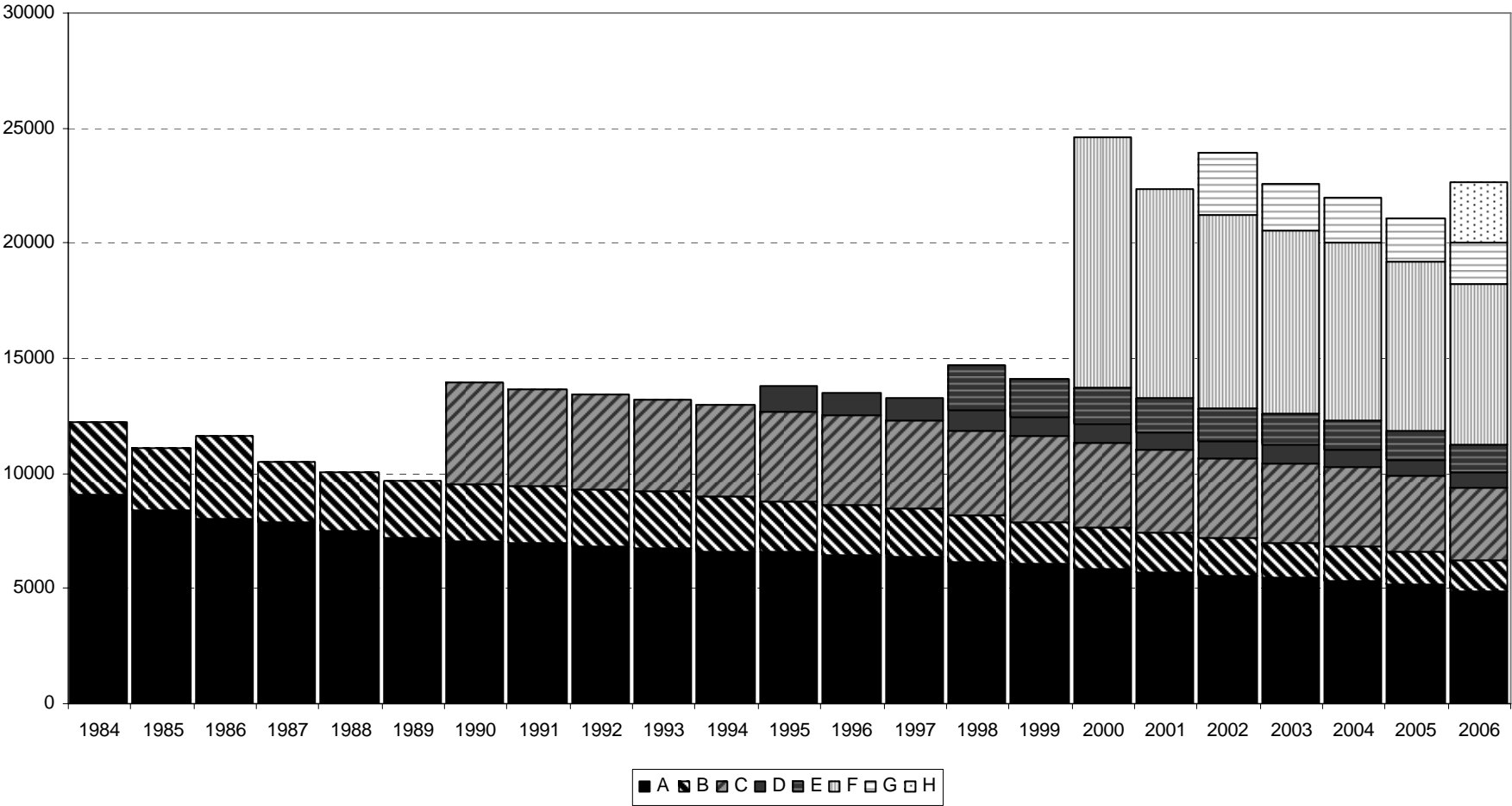
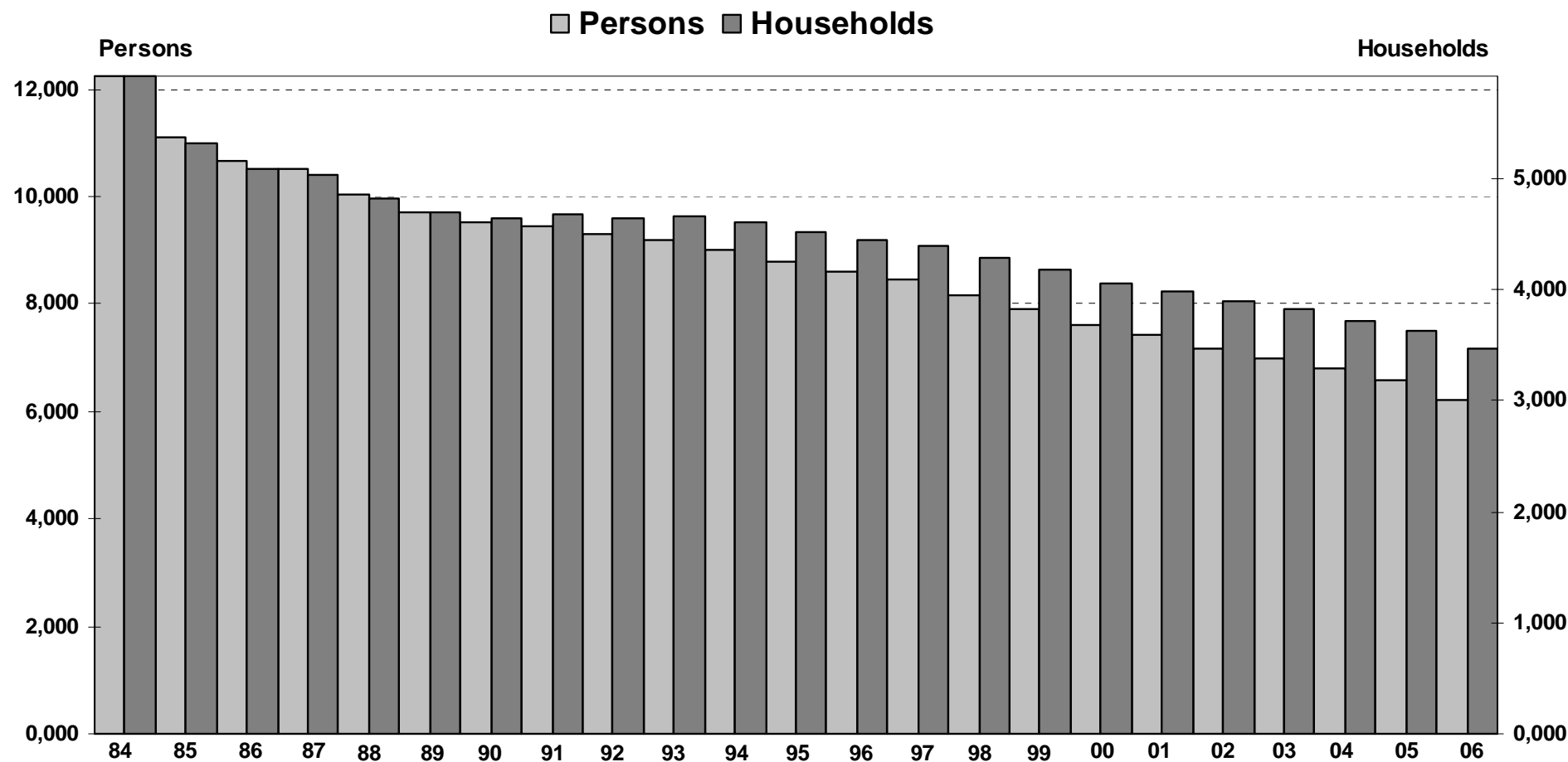


Figure 2: Comparison of Successful Interviews with Persons and Households (Subsamples A and B), Waves 1 to 23.



Year	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
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
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

Figure 3: Comparison of Successful Interviews with Persons and Households (Subsample C), Waves 1 to 17.

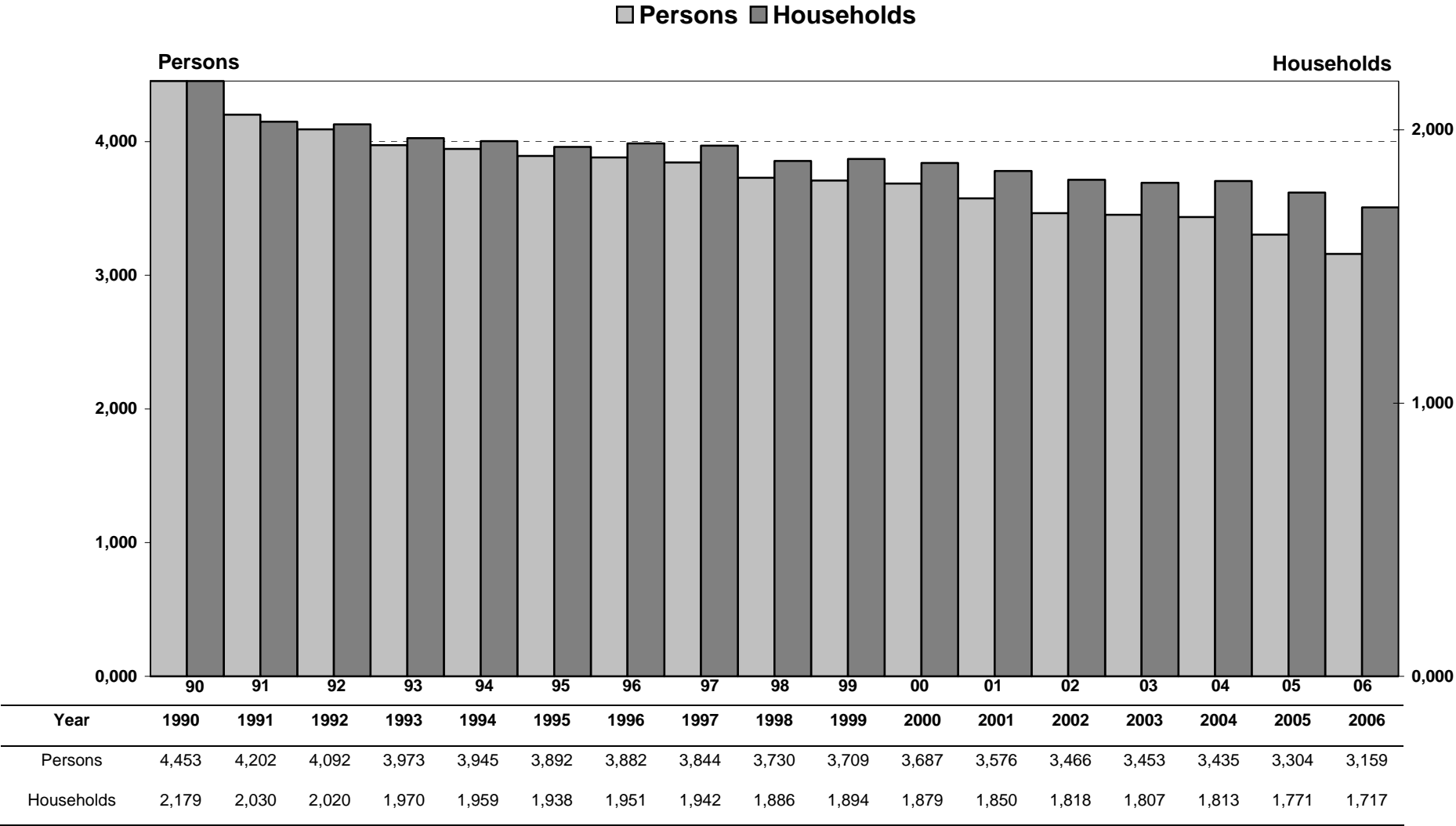


Figure 4: Comparison of Successful Interviews with Individuals and Households (Subsample D), Waves 1 to 12.

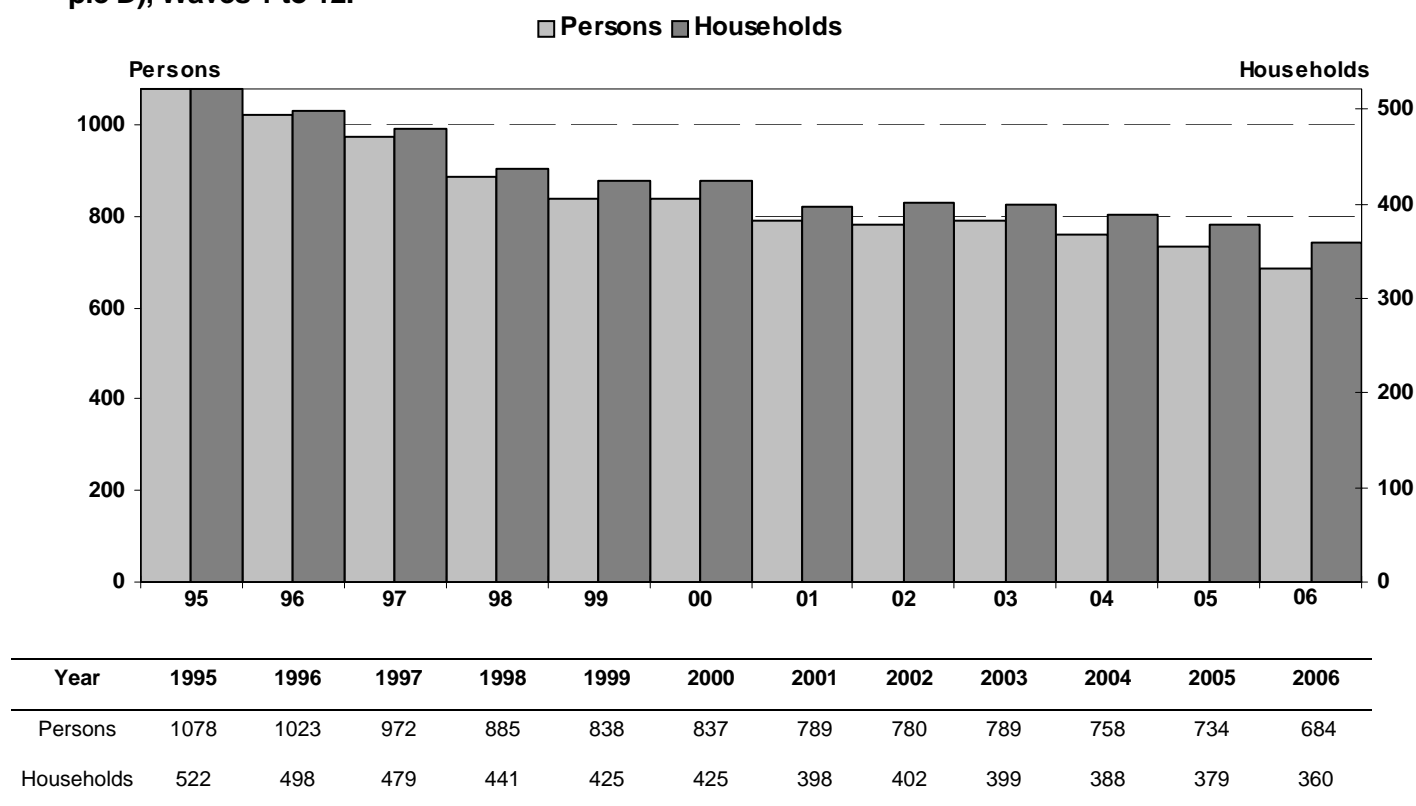
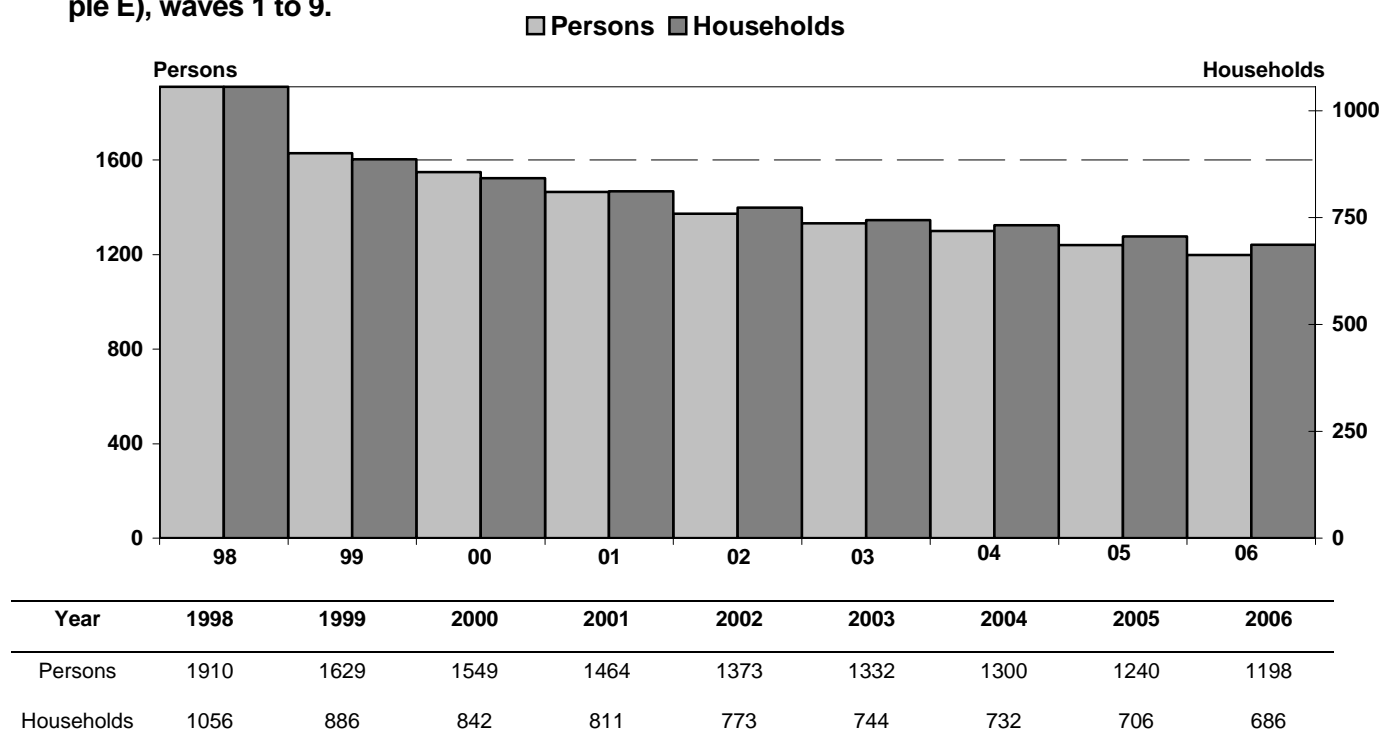


Figure 5: Comparison of successful interviews with individuals and households (subsample E), waves 1 to 9.



Data Documentation 27

2 Developments in Sample Size

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

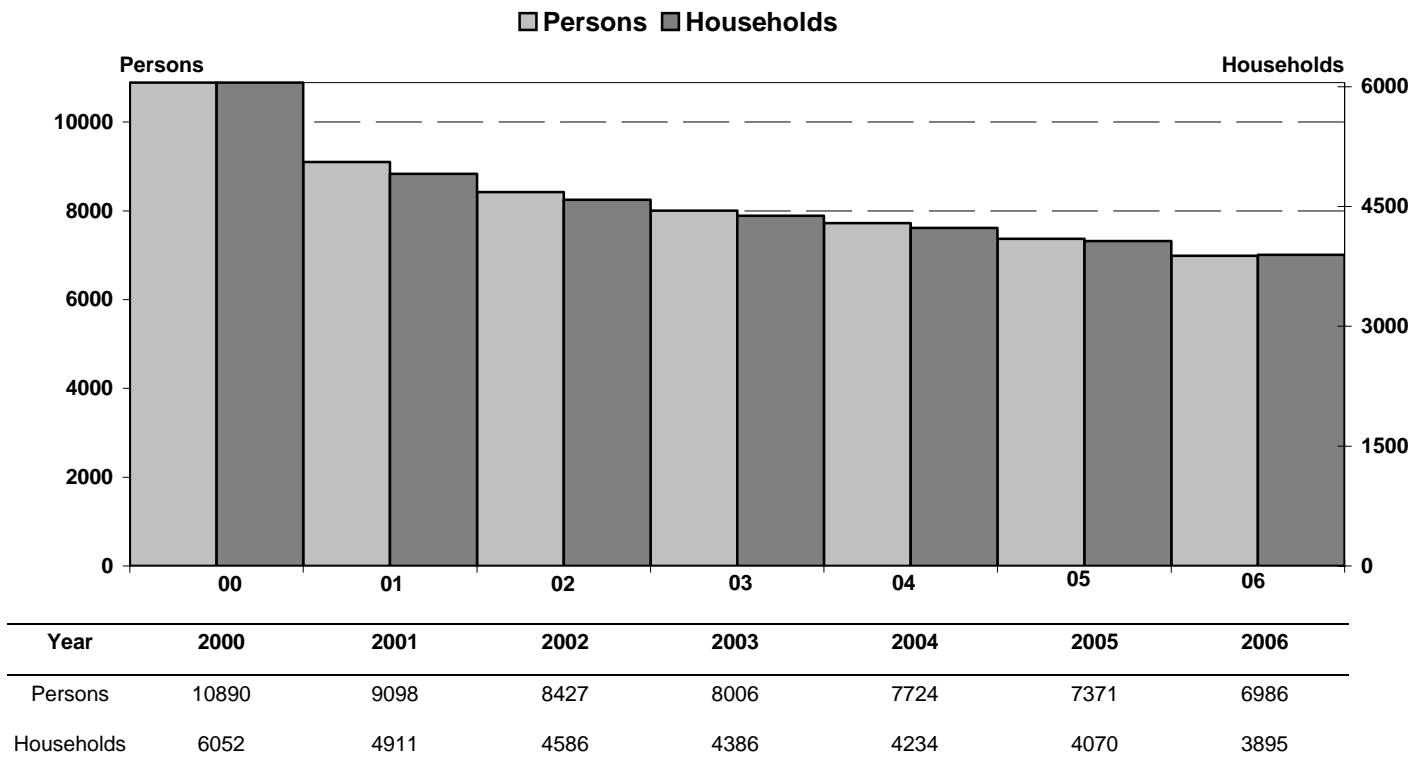
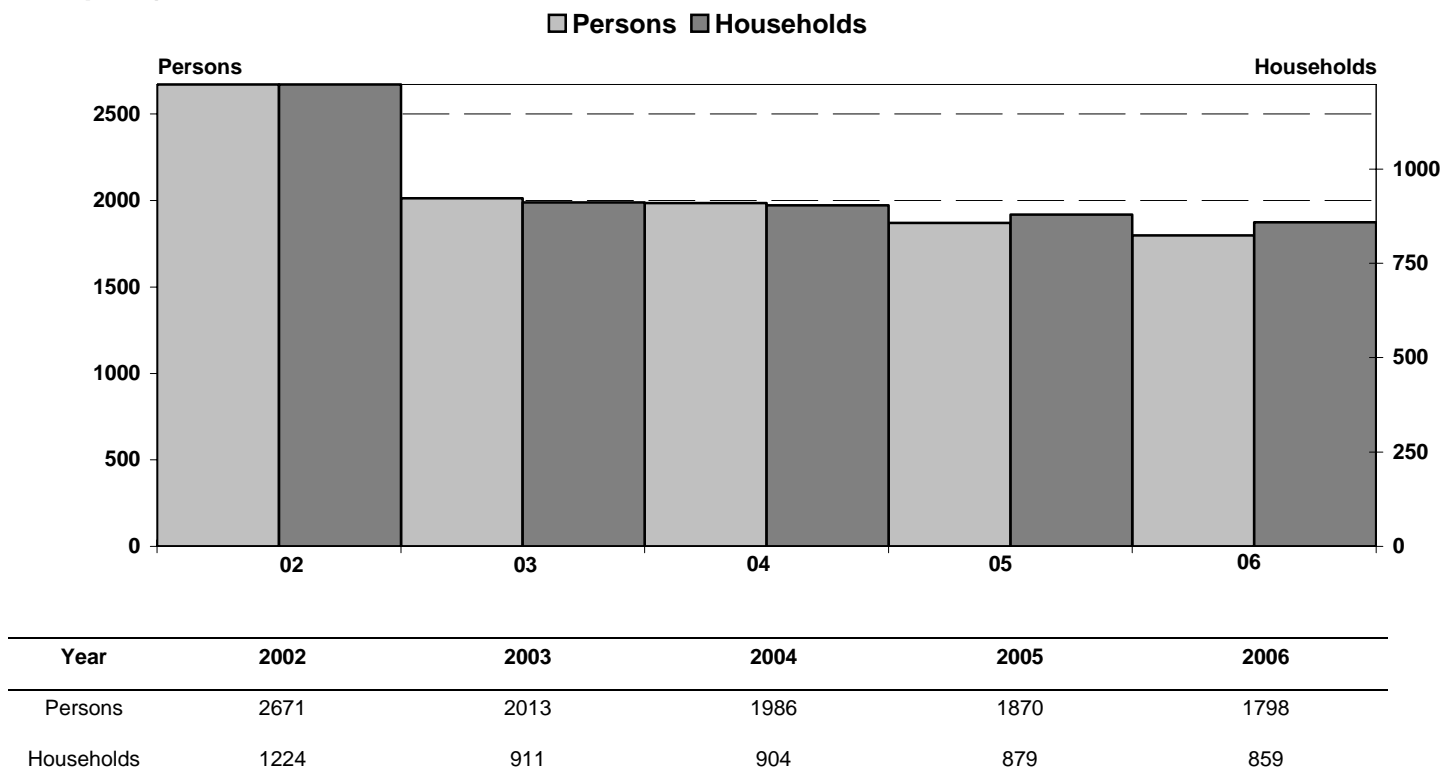


Figure 7: Comparison of Successful Interviews with Individuals and Households (Subsample G), Waves 1 to 5.



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 23.

Figure 9: All First Wave Persons in Subsample B. Whereabouts up to Wave 23.

Figure 10: All First Wave Persons in Subsample C. Whereabouts up to Wave 17.

Figure 11: All First Wave Persons in Subsample D. Whereabouts up to Wave 12.

Figure 12: All First Wave Persons in Subsample E. Whereabouts up to Wave 9.

Figure 13: All First Wave Persons in Subsample F. Whereabouts up to Wave 7.

Figure 14: All First Wave Persons in Subsample G. Whereabouts up to Wave 5.

Figure 8: All First-Wave Persons (Gross Subsample A). Development up to Wave 23.

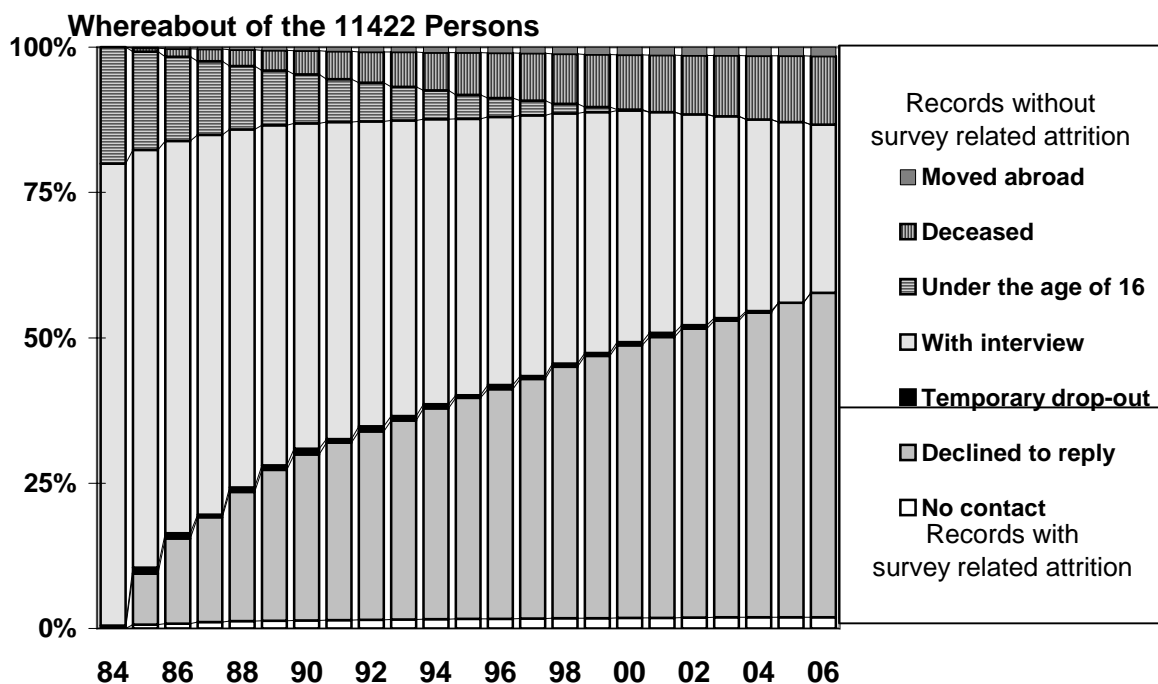


Figure 9: All First Wave Persons (Gross Subsample B). Development up to Wave 23.

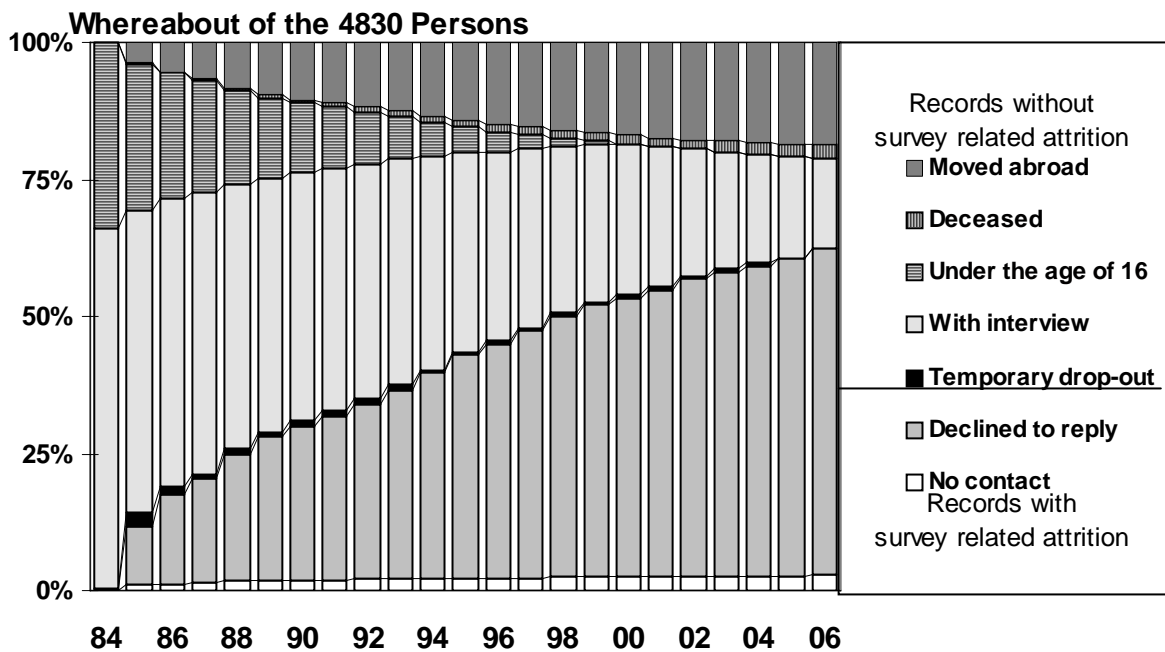


Figure 10: All First Wave Persons (Gross Subsample C). Development up to Wave 17.

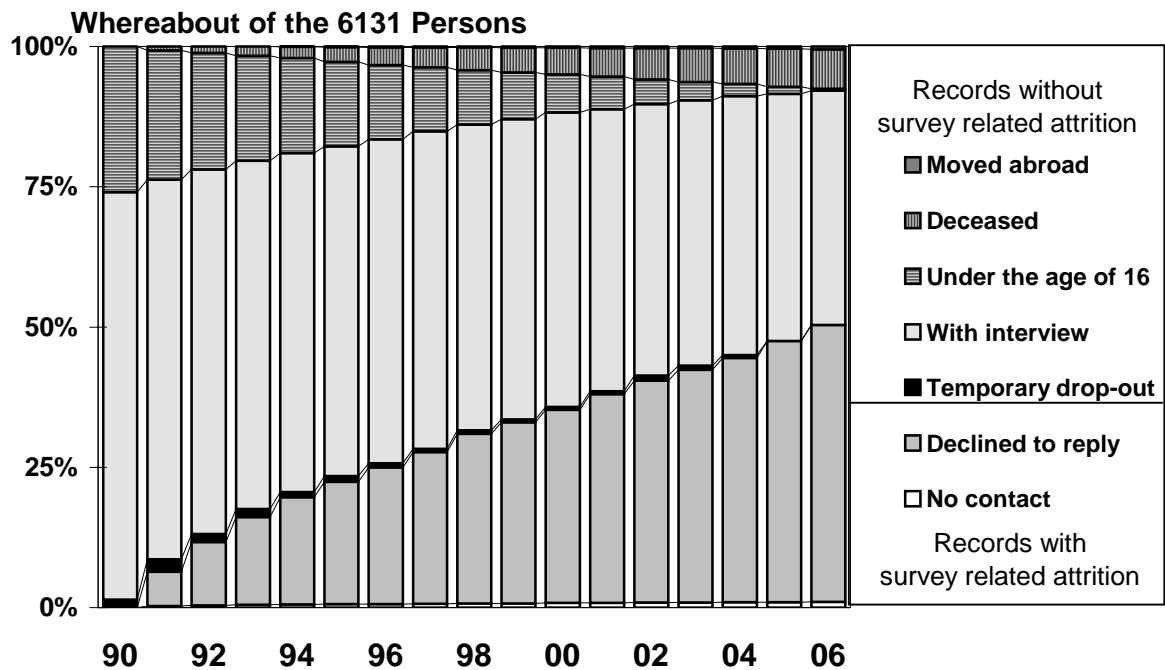


Figure 11: All First Wave Persons (Gross Subsample D). Development up to Wave 12.

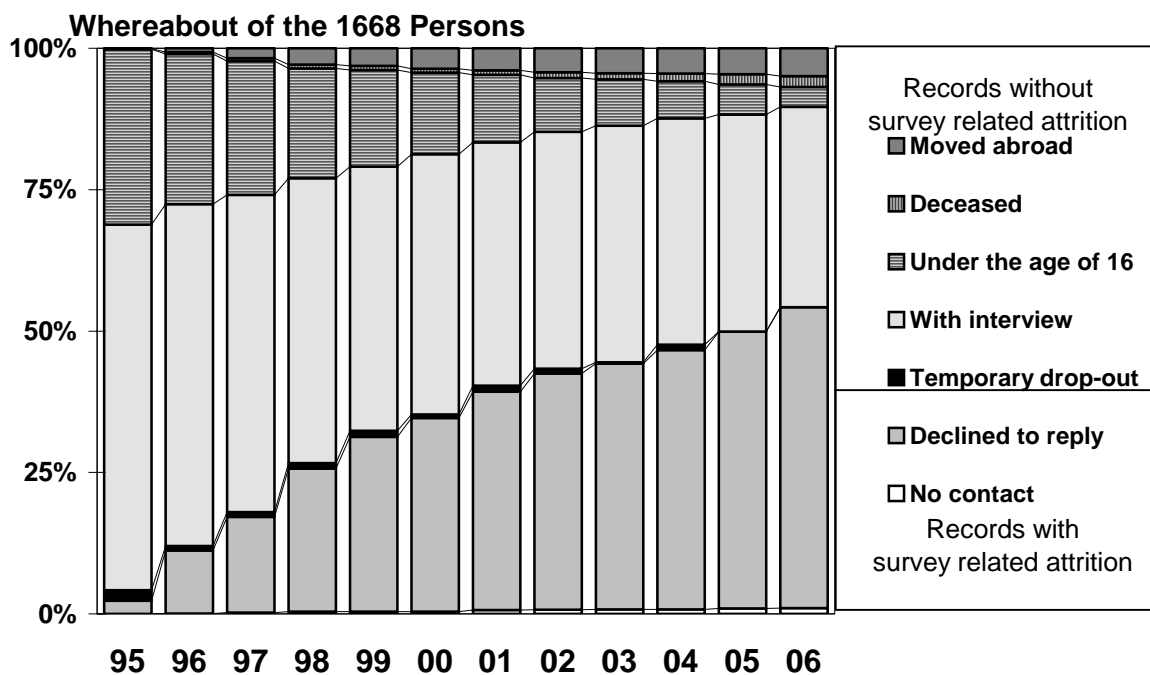


Figure 12: All First Wave Persons (Gross Subsample E). Development up to wave 9.

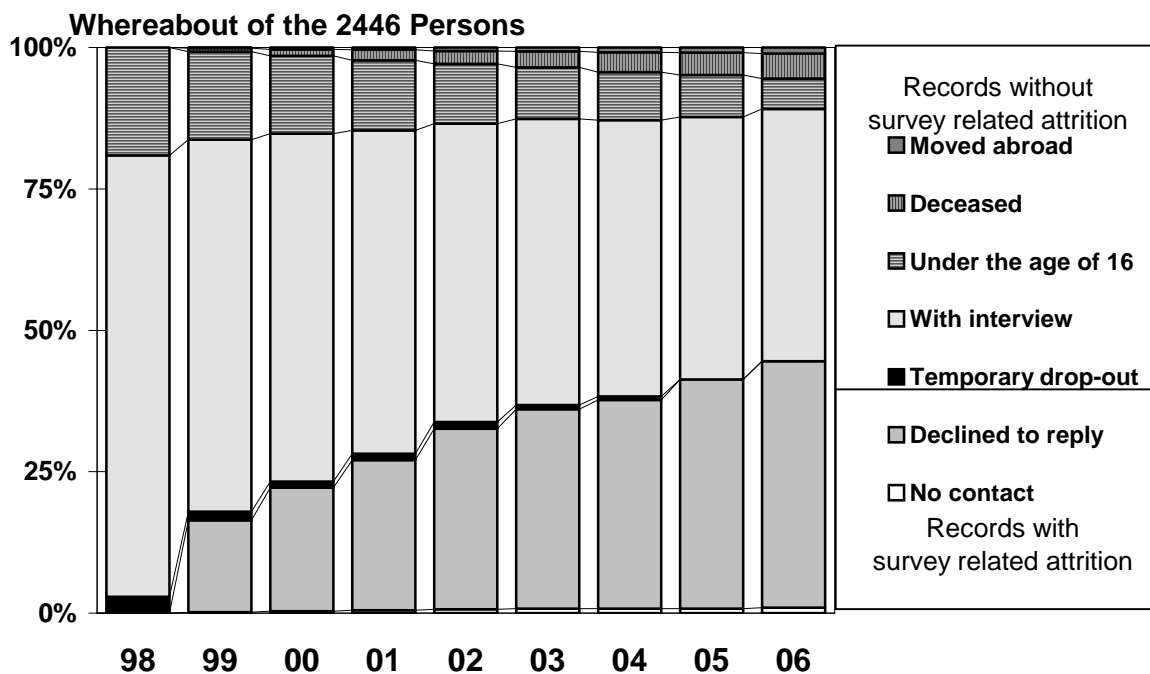


Figure 13: All First Wave Persons (Gross Subsample F). Development up to Wave 7.

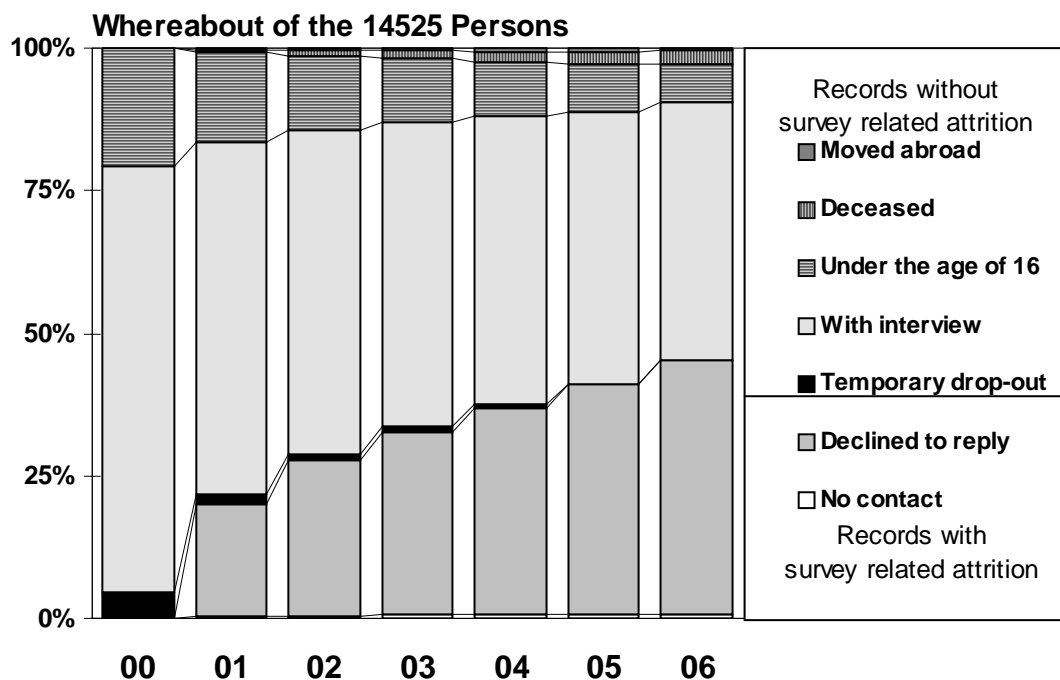
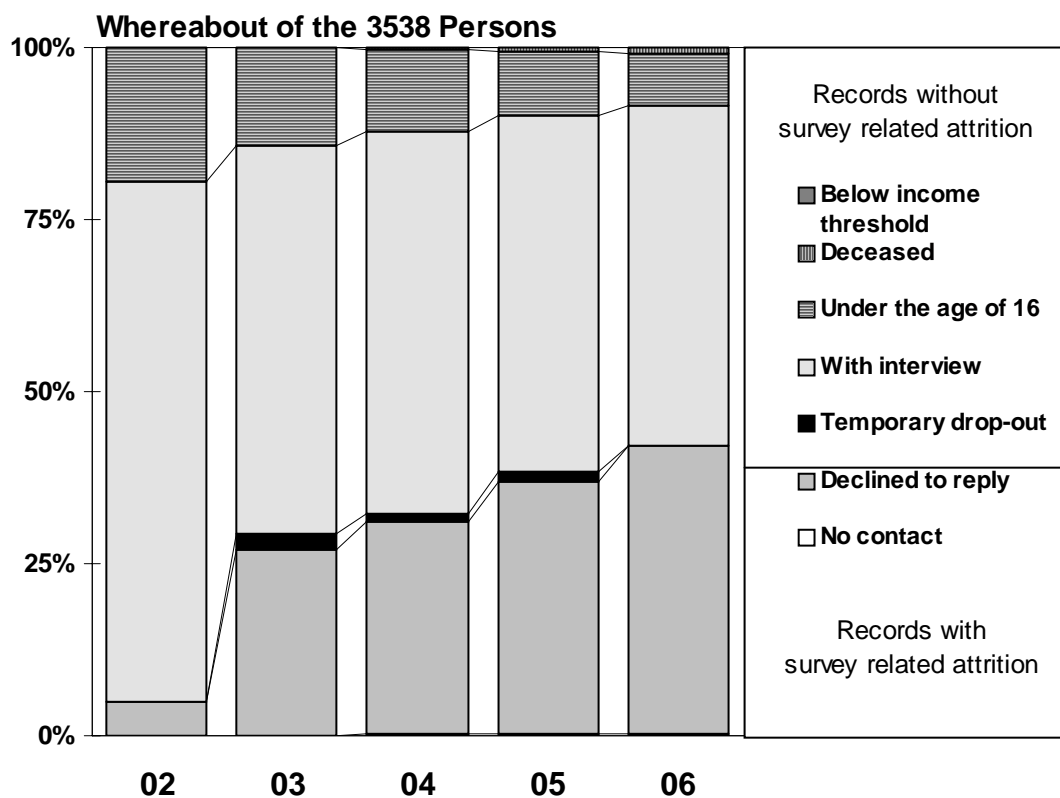


Figure 14: All First Wave Persons (Gross Subsample G). Development up to Wave 5.



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).

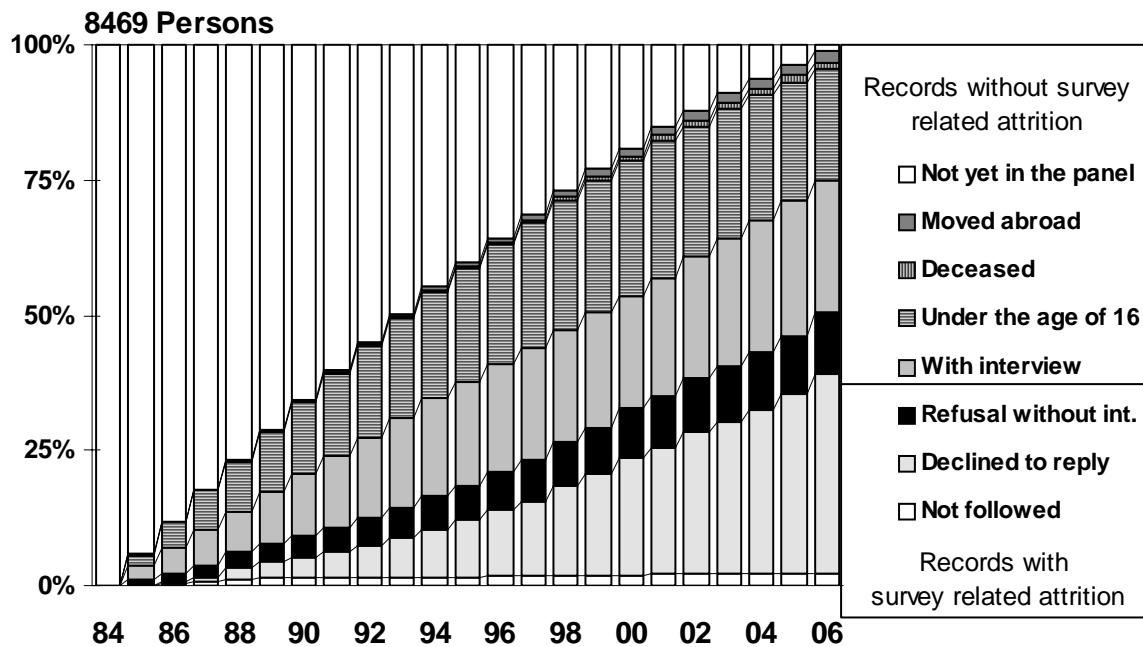


Figure 16: Entrants and their Participation Behavior (Subsample C).

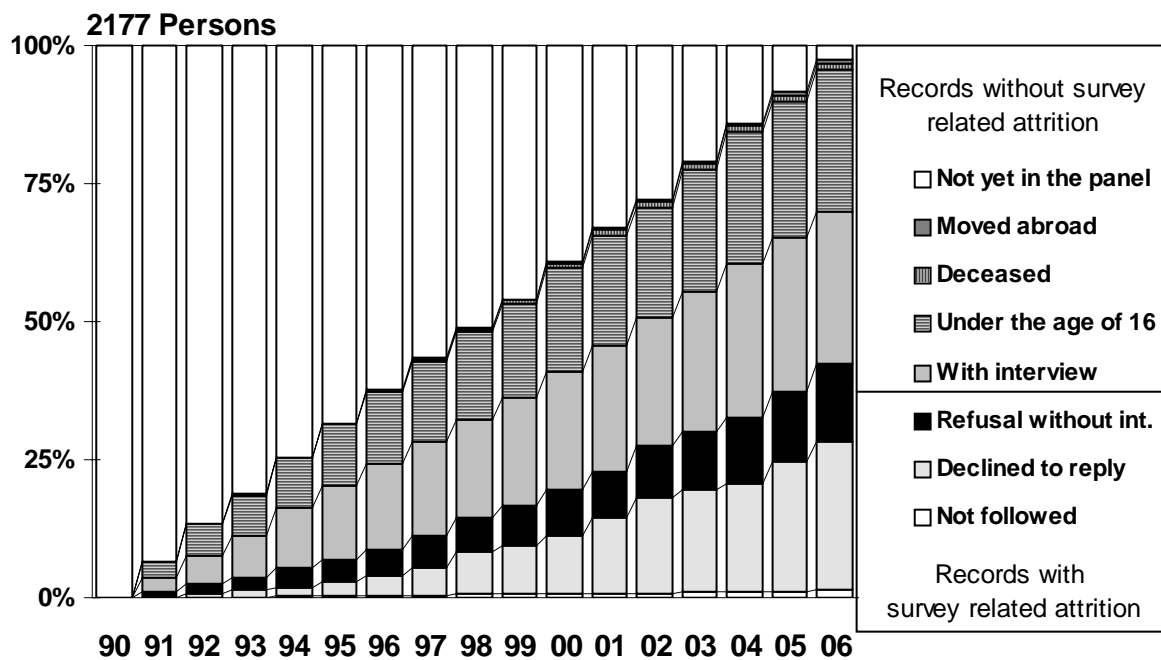


Figure 17: Entrants and their Participation Behavior (Subsample D).

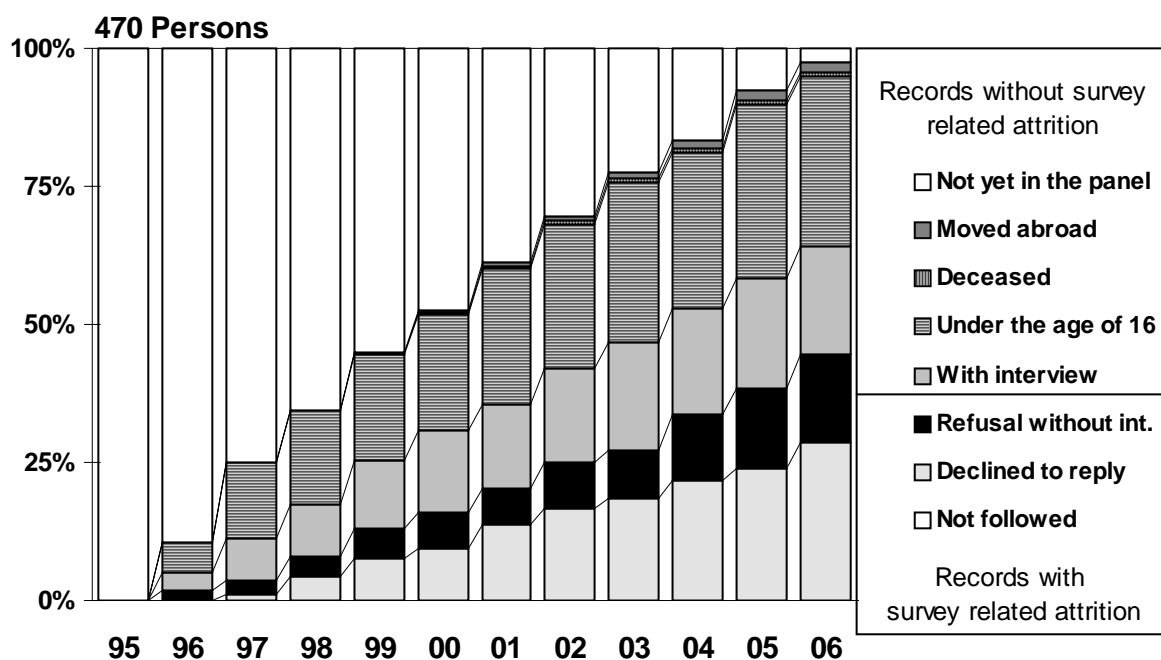


Figure 18: Entrants and their Participation Behavior (Subsample E).

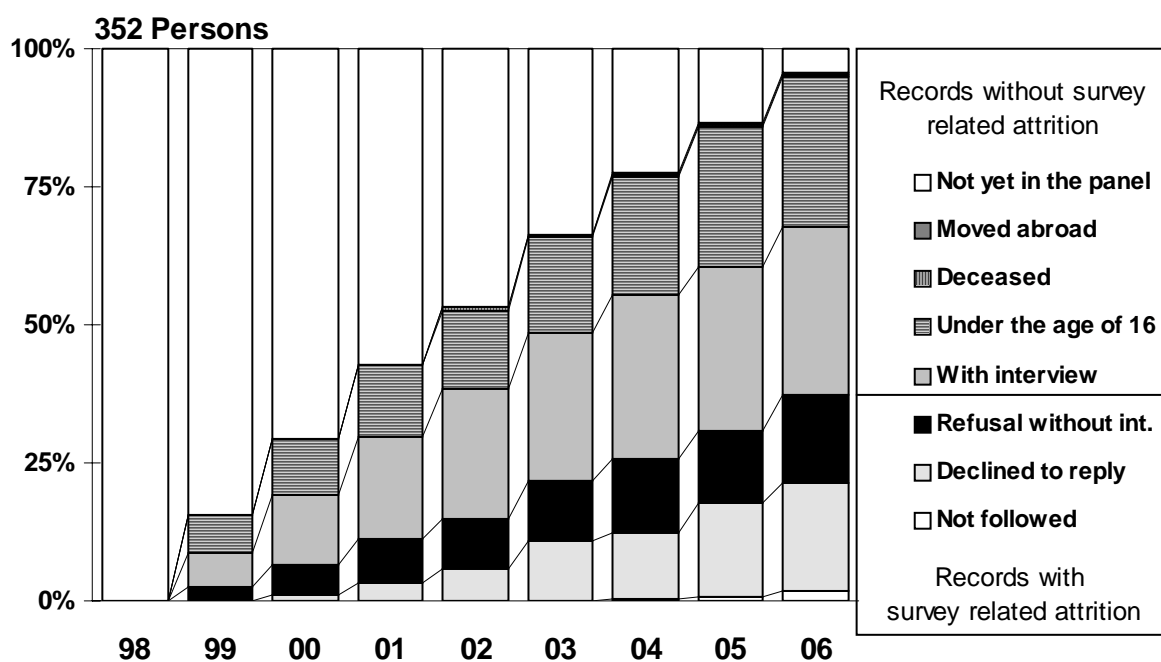


Figure 19: Entrants and their Participation Behavior (Subsample F).

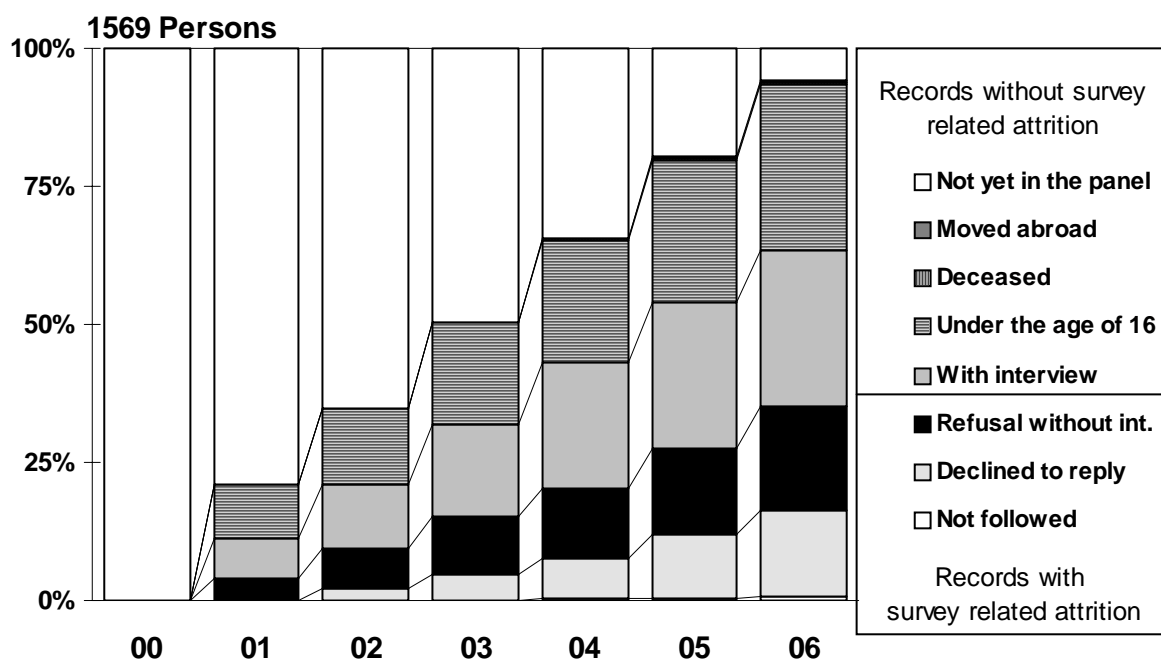
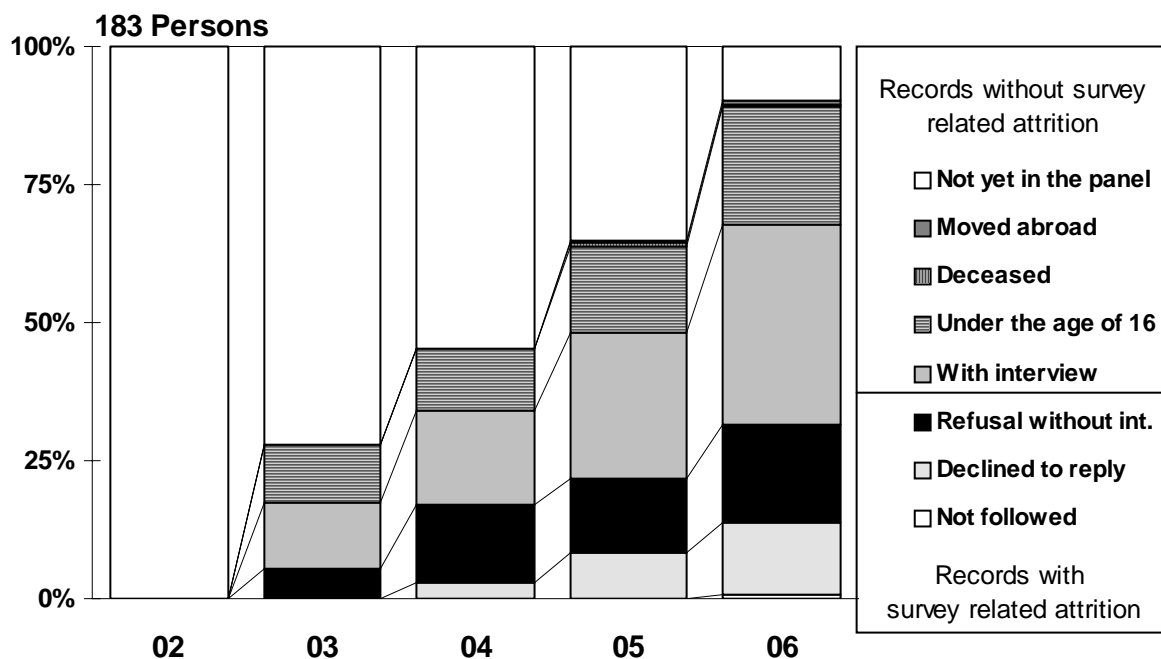


Figure 20: Entrants and their Participation Behavior (Subsample G).



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.**

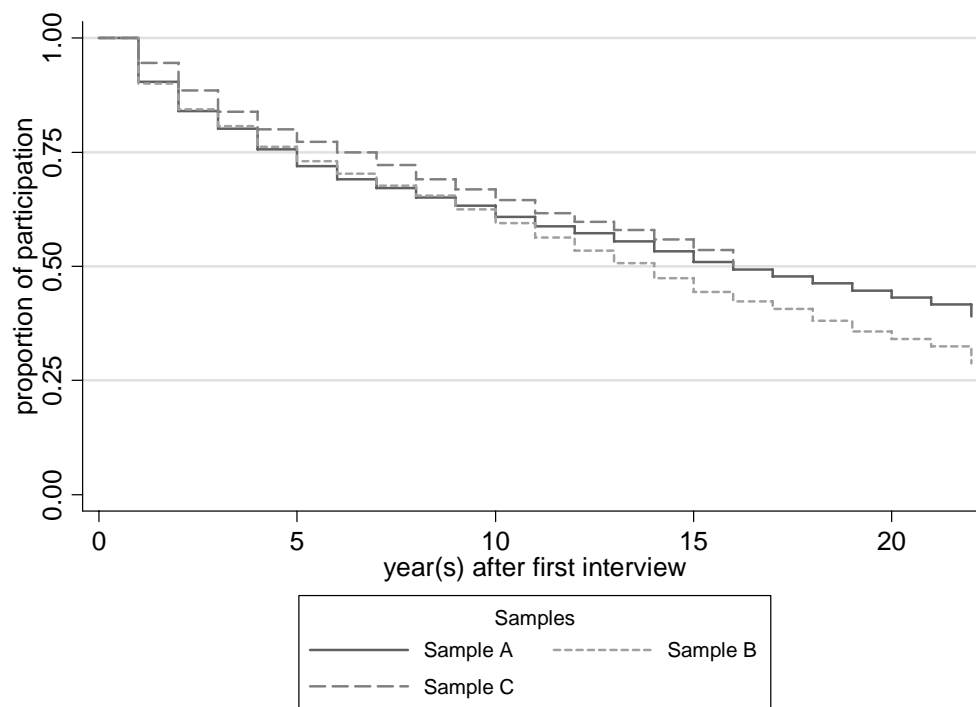


Figure 22: **Successful Re-Interviewing of First-Wave Respondents by Subsamples D, E, F, G. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

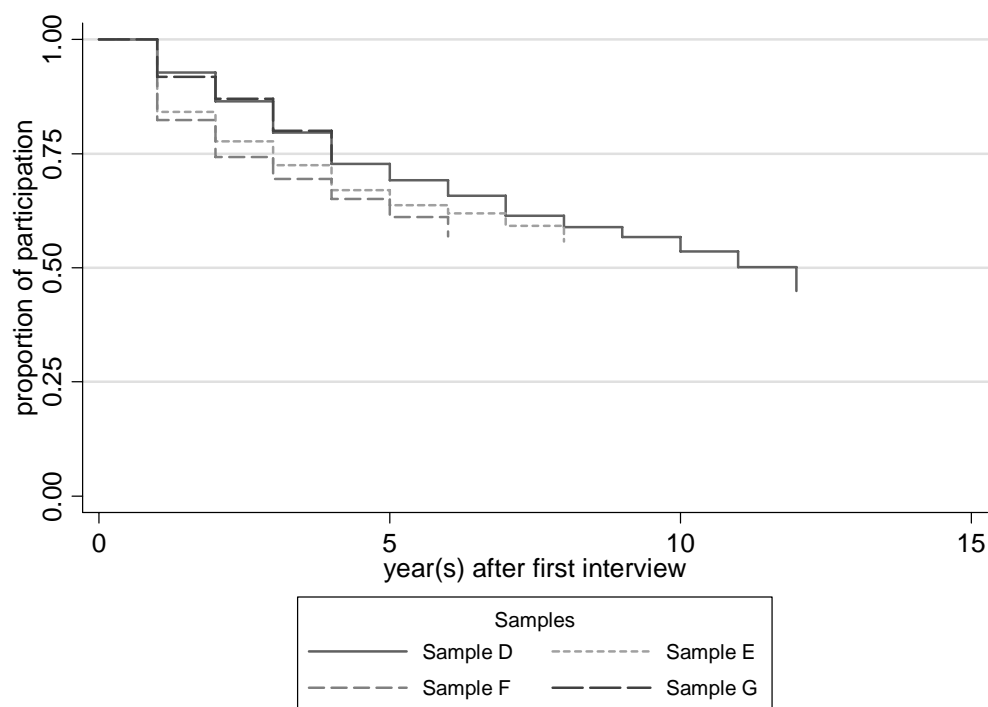


Figure 23: **Successful Re-Interviewing of All First-Wave Respondents by Age Categories. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

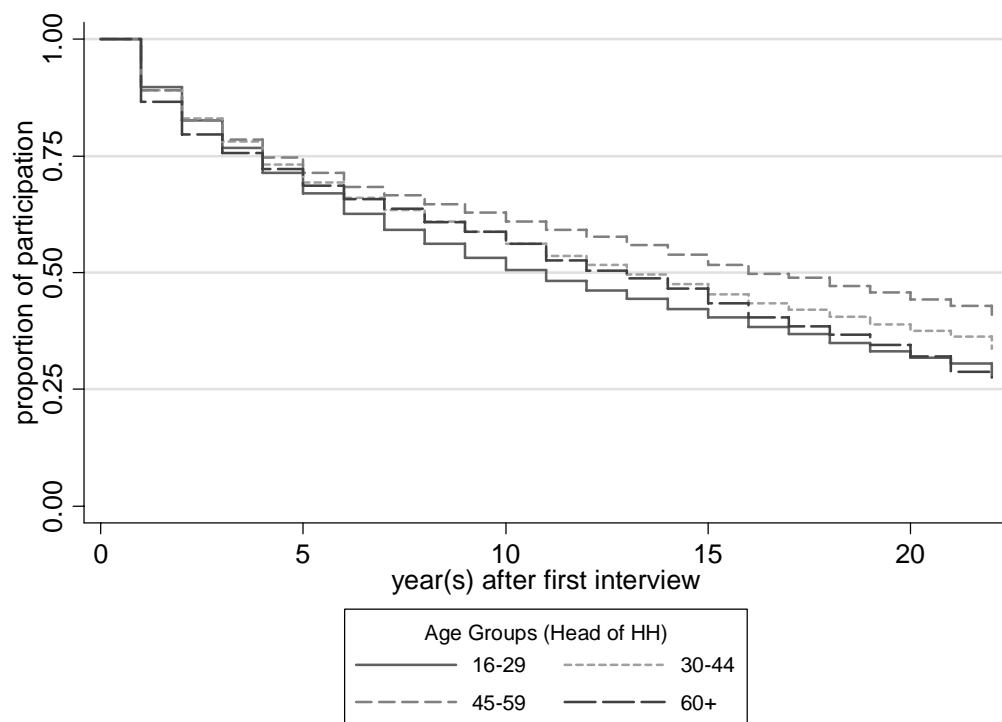


Figure 24: **Successful Re-Interviewing of All First-Wave Respondents by Occupation. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

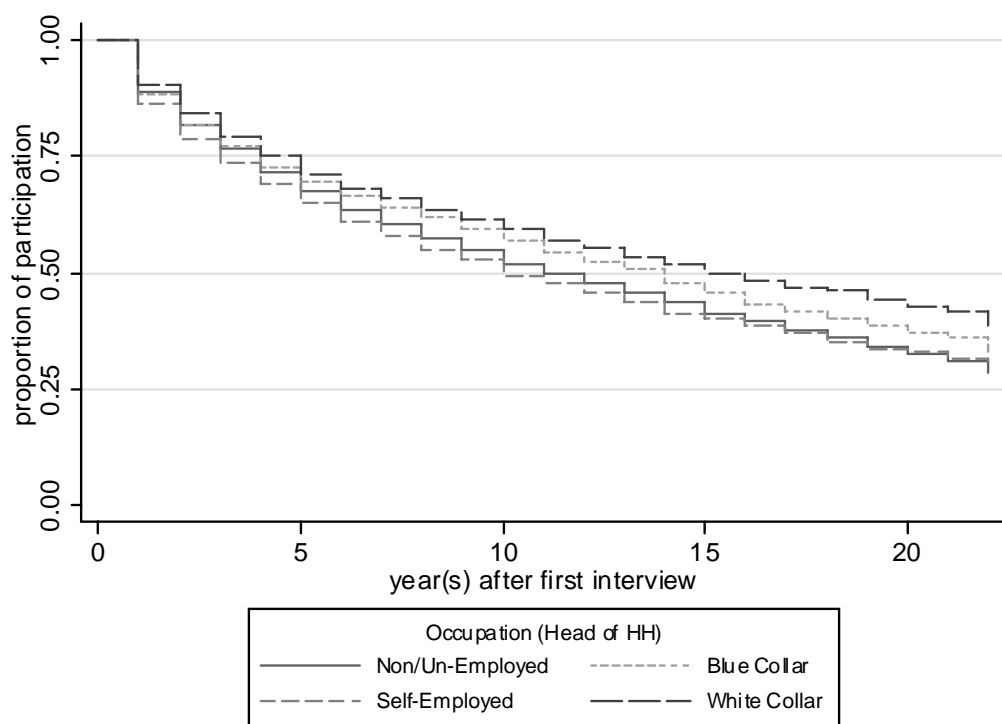


Figure 25: **Successful Re-Interviewing of All First-Wave Respondents by Income Quintiles. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**

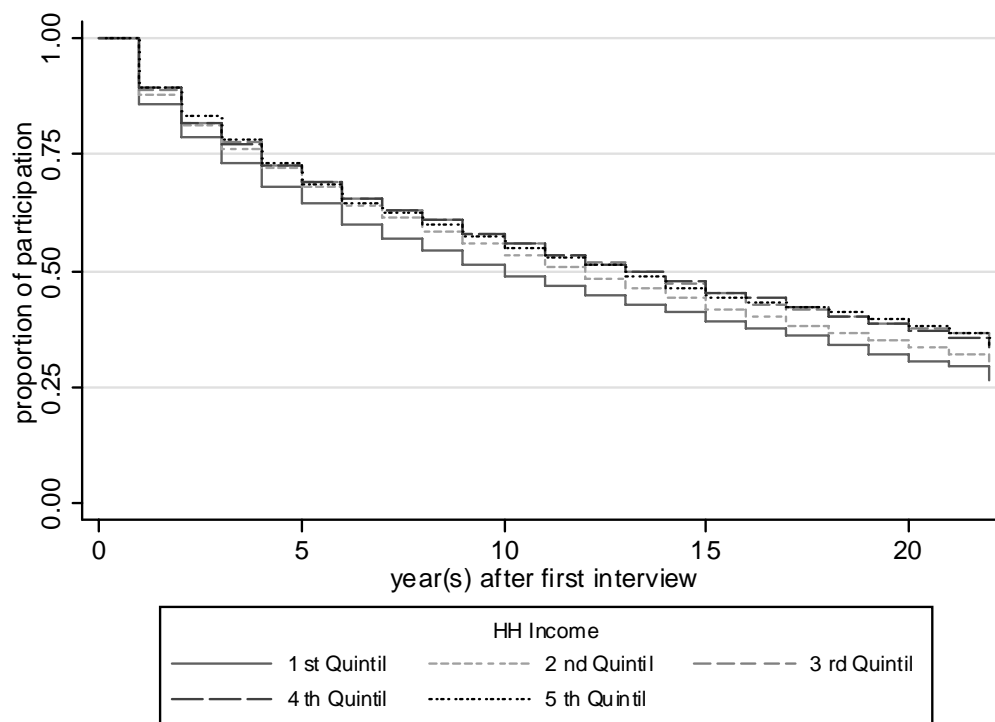
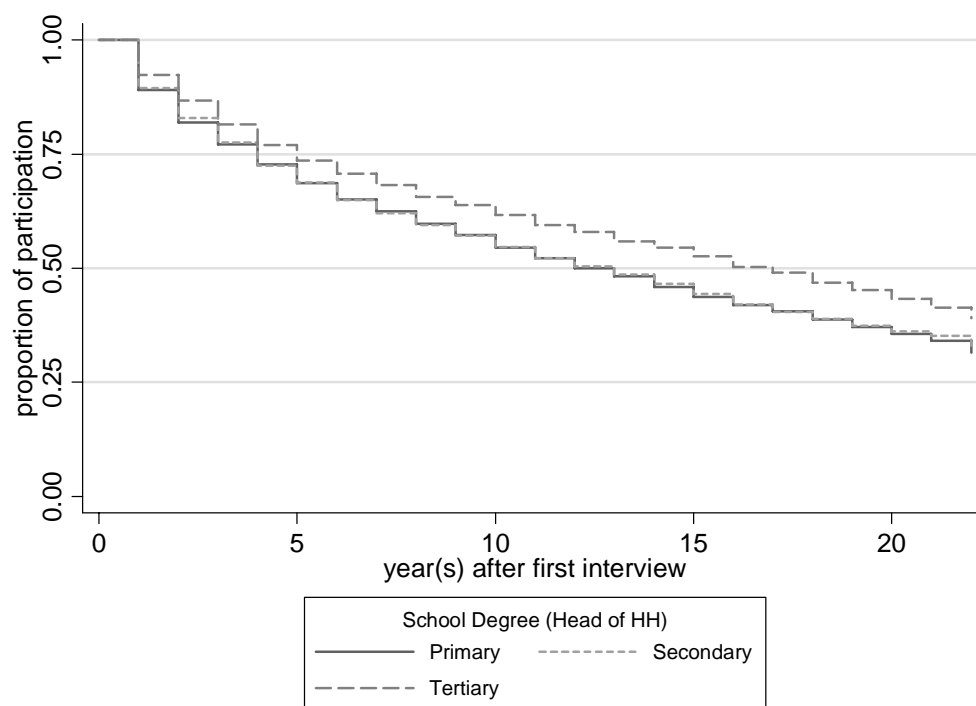


Figure 26: **Successful Re-Interviewing of All First-Wave Respondents by Education. Kaplan-Meier Estimates of Survey-Related Attrition Ignoring Deaths and Moves Abroad.**



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 Unsuccessful Follow-Ups

Table 1 displays the number of households of the previous waves that need to be re-contacted and the relative frequency of unsuccessful follow-ups in subsamples A through G and waves 1985 through 2006. The drop-out 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 27

3 Panel Attrition Due to Unsuccessful Follow-Ups

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

	A/B		C		D		E		F		G	
	n	%	n	%	n	%	n	%	n	%	n	%
1985	6051	1.9										
1986	5814	1.4										
1987	5465	1.0										
1988	5342	0.9										
1989	5156	0.9										
1990	5044	0.9										
1991	5029	0.5	2246	1.5								
1992	5006	0.4	2304	0.5								
1993	5049	0.9	2227	0.9								
1994	5008	0.8	2136	0.6								
1995	4900	0.6	2113	0.4								
1996	4817	0.4	2104	0.5	544	0.4						
1997	4733	0.5	2091	0.5	542	0.7						
1998	4695	0.6	2081	0.6	498	0.6						
1999	4616	0.5	2041	0.3	529	0.9	1100	0.5				
2000	4495	0.4	2028	0.4	467	0.2	968	0.8				
2001	4371	0.5	2036	0.3	454	0.9	922	0.87	6172	1.0		
2002	4290	0.4	2010	0.5	450	0.2	875	0.57	5451	0.5		
2003	4170	0.4	1982	0.4	434	0.5	834	0.72	4965	0.3	1056	0.9
2004	4063	0.3	1962	0.4	436	0.2	797	0.25	4736	0.4	1010	0.3
2005	3999	0.3	1959	0.3	429	0.7	783	0.1	4577	0.3	1001	0.3
2006	3909	0.3	1941	0.6	425	1.2	775	0.9	4401	0.7	995	0.5

n = Number of households to be recontacted

% = Percentage of households without contact

3.2 Predicting the Probability of Successful vs. Unsuccessful Follow-Ups in the Year 2006

Based on the household and interview characteristics measured in 2005, we aim at predicting the probability of re-contacting a household relative to unsuccessful follow-up in 2006. 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 2005 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
(Moved HH)*(SingleHH)	Interaction term between respective variables	0/1
Single HH	Single person household	0/1
3+ Person HH	Household with more than three individuals	0/1
Non-Germ. Nationality	At least one HH-member with Non-Germ. nationality	0/1
Large Building	Neighborhood with large buildings	0/1
Rural	Rural neighborhood	0/1
Urban	Urban area (+ 100,000 inhabitants)	0/1

Data Documentation 27

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 2006.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G
Intercept	-3.59 (0.44) ***	-4.83 (0.62) ***	-3.14 (0.34) ***	-2.50 (0.63) ***	-2.65 (0.46) ***	-0.77 (0.36) **	-3.85 (0.48) ***
New HH	-1.98 (0.44) ***	-1.76 (0.62) ***	-2.25 (0.55) ***	-1.73 (0.66) ***	-2.03 (0.59) ***	-2.83 (0.35) ***	-1.72 (0.48) ***
Moved HH			-1.94 (0.55) ***	-1.33 (0.64) **	-2.03 (0.59) ***	-1.99 (0.25) ***	
Single HH				-1.57 (0.59) ***		-1.23 (0.31) ***	
(Moved HH)*(Single HH)	-1.75 (0.50) ***						
3+ Person HH						-0.87 (0.30) ***	
Non-German Nationality						-1.14 (0.29) ***	
Large Building						-0.51 (0.25) **	-1.12 (0.49) **
Urban			-0.71 (0.33) **				
Rural					-0.91 (0.44) **		

<i>Likelihood Ratio (Pr > Chisq)</i>	****	****	0.43	0.22	0.65	0.53	0.89
---	------	------	------	------	------	------	------

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 Refusals

Table 4 displays the drop-out rates due to refusal by sub-sample and wave. 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 27

4 Panel Attrition Due to Refusals

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

	A		B		C		D		E		F		G	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
1985	4,611	10.19	1,326	10.94										
1986	4,442	10.81	1,290	12.56										
1987	4,194	6.77	1,204	7.31										
1988	4,105	8.82	1,180	9.24										
1989	3,949	7.65	1,146	8.99										
1990	3,871	6.69	1,111	7.47										
1991	3,842	5.96	1,143	7.61	2,213	8.27								
1992	3,833	6.47	1,144	7.34	2,290	11.79								
1993	3,838	6.12	1,156	7.96	2,208	10.78								
1994	3,821	6.39	1,139	10.18	2,122	7.68								
1995	3,766	6.37	1,097	10.48	2,101	7.76	634	17.67						
1996	3,734	6.67	1,061	9.52	2,092	6.74	542	8.12						
1997	3,674	5.88	1,029	9.52	2,076	6.45	537	10.80						
1998	3,645	7.08	1,013	11.35	2,066	8.71	523	15.68						
1999	3,616	8.05	969	11.46	2,030	6.70	495	14.14	1,084	18.27				
2000	3,535	8.35	929	11.73	2,018	6.89	466	8.80	959	12.20				
2001	3,448	8.12	899	10.01	2,028	8.78	450	11.56	913	11.17	6,109	19.61		
2002	3,396	8.04	869	11.85	1,996	8.92	449	10.47	868	10.94	5,420	15.39		
2003	3,318	7.41	837	11.35	1,974	8.46	432	7.64	828	10.14	4,951	11.41	1,047	12.99
2004	3,253	7.47	800	10.75	1,955	7.26	435	10.80	795	7.92	4,719	10.26	1,007	10.23
2005	3,214	8.62	774	9.82	1,954	9.37	426	11.03	782	9.72	4,564	10.82	998	11.92
2006	3,130	9.87	767	14.60	1,930	11.04	420	14.29	768	10.68	4,370	10.87	990	13.23
n = Number of recontacted households														
% = Percentage of households that refuse to participate														

4.2 Predicting the Probability of Re-Interviewing versus Refusal in the Year 2006

Based on the household and interview characteristics measured in 2005, we aim at predicting the probability of agreement vs. refusal to participate in the survey by the households that were re-contacted in 2006. The individual attributes refer in most 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 2005 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
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
Face-to-Face	Face-to-face interview in $t - 1$	0/1
CAP1	Random CAP1-Sample (vs. PAPI) in Sample E	0/1
Experiment	Participated in behavioral experiment (sample F only)	0/1
Change in Interviewer	Change in Interviewer between last and current wave	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
SOEP-Experience	Number of successful interviews	1/22
Respondent Cooper.	Low interviewer rating of respondents' cooperation	0/1
Email Disclosed	Email address known	0/1
Phone Disclosed	Telephone number known	0/1
Gender	Female Gender of head of household	0/1
2 Person HH	Two individuals living in HH	0/1
4+ Person HH	More than 3 individuals living in household	0/1
Non-German HH	Head of household has non-German nationality	0/1
Age 35-64	Head of household was between 35 and 64 in $t - 1$	0/1
Age 25-34	Head of household was between 25 and 34 in $t - 1$	0/1
(Age 25-34) * (Old HH)	Interaction term between respective variables	0/1
Unmarried	Head of household unmarried	0/1
Separation	Separation of couple	0/1

Data Documentation 274 Panel Attrition Due to Refusals

(Separation)*(Old HH)	Interaction term between respective variables	0/1
Rural	Rural neighborhood	0/1
Care	Household member in need of care	0/1
Savings	Household without savings and insurances	0/1
Tertiary Education	Head of Household with college or university degree	0/1
No Vocational Educ.	No vocational education degree of head of hh	0/1
Unemployed	Head of household registered unemployed in t – 1	0/1
Irregular Employment	Military service, maternity leave of head of hh	0/1
Job Worries	Very concerned about own job security	0/1
Extraversion	Compound scale: extraversion of head of hh (big-5)	1/20
Neuroticism	Compound scale: neuroticism of head of hh (big-5)	1/20
Reciprocity	Compound scale: positive reciprocity of head of hh	1/20
Dissatisfaction	Dissatisfied with life in general (head of hh)	0/1

Data Documentation 27

4 Panel Attrition Due to Refusals

Table 6a: Estimates of Logit Models of the Probability of Re-Interviewing a Household (Relative to Refusal) in 2006.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G
Intercept	1.06 (0.07) ***	0.34 (0.24)	1.63 (0.20) ***	3.87 (0.76) ***	-0.42 (0.28)	0.98 (0.11) ***	-0.17 (0.36)
First Wave HH	0.25 (0.06) ***						
New HH	-0.70 (0.16) ***		-0.90 (0.15) ***			-0.70 (0.13) ***	
Face-to-Face				-2.62 (0.75) ***			
CAPI					0.30 (0.13) **		0.28 (0.13) **
Experiment						0.14 (0.06) **	
Change in Interviewer	-0.75 (0.08) ***	-0.62 (0.18) ***	-0.89 (0.13) ***		-0.62 (0.20) ***	-0.86 (0.07) ***	-0.64 (0.15) ***
Non-Regular Interview	-0.28 (0.07) ***	-0.59 (0.14) ***	-0.55 (0.09) ***	-2.73 (0.71) ***		-0.91 (0.07) ***	-0.80 (0.18) ***
Pace of Interview	0.14 (0.62) **		-0.22 (0.09) **			-0.12 (0.05) **	
SOEP Experience		0.03 (0.01) ***					
Low Cooperation	-0.31 (0.08) ***	-0.49 (0.13) ***	-0.46 (0.09) ***	-1.09 (0.24) ***	-0.57 (0.16) ***	-0.27 (0.07) ***	-0.42 (0.14) ***
Email Disclosed	0.15 (0.07) **						
Phone Disclosed					1.37 (0.22) ***	0.27 (0.10) ***	0.89 (0.32) ***
Gender							0.32 (0.13) **
2 Person HH							0.29 (0.11) **
4+ Person HH						-0.15 (0.07) **	
Non-German HH					-0.95 (0.30) ***		
Age 25-34	0.26 (0.09) ***						
Age 35-64						0.22 (0.06) ***	
(Age 25-34)*(Old-HH)		0.43 (0.18) **					

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

Data Documentation 27

4 Panel Attrition Due to Refusals

Table 6b: Estimates of Logit Model for the Probability of a Drop-Out of a Household Due to Refusal in 2006.

	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F	Sample G
Unmarried	-0.19 (0.08) **					0.15 (0.07) **	
Separation	-0.55 (0.18) ***	-0.59 (0.27) **			-0.77 (0.24) ***		
(Separation)*(Old-HH)			-1.54 (0.43) ***				
Rural							-0.27 (0.10) ***
Care	-0.41 (0.15) ***						
Savings							-0.51 (0.26) **
Tertiary Education			0.17 (0.08) **				
No Vocational Education	-0.15 (0.07) **						
Unemployed			-0.43 (0.19) **				
Irregular Employment	-0.14 (0.59) **	0.40 (0.14) ***	-0.40 (0.19) **				
Job worries						-0.12 (0.06) **	
Extraversion	0.02 (0.01) **		-0.02 (0.01) **				
Neuroticism							-0.04 (0.01) ***
Reciprocity			0.03 (0.01) **				
Dissatisfaction				-0.72 (0.29) **			
Likelihood Ratio ($Pr > \chi^2$)	< 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 2006 based on characteristics measured in 2005, WHBLEIB, lends itself as a longitudinal weighting variable correcting for selective attrition between waves 2005 and 2006. Table 7 reports some sub-sample specific descriptive statistics of the longitudinal weights in each wave.

The product of the cross-sectional weight in 2005, VHHRF, and the longitudinal weight in 2006, WHBLEIB, provide the raw data for the cross-sectional weight in 2006. 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 2006. Table 8 reports sub-sample-specific descriptive statistics of the derived cross-sectional weighting variable WHHRF and in comparison all previous cross-sectional weights AHHRF through VHHRF.

Data Documentation 27

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 23).**

	bhbleib	chbleib	dhbleib	ehbleib	fhbleib	ghbleib	hhbleib	ihbleib	jhbleib	khbleib	lhbleib	mhbleib	nhbleib	ohbleib	phbleib	qhbleib	rhbleib	shbleib	thbleib	uhbleib	vhbleib	whbleib
<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
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
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
N	4141	3962	3910	3731	3647	3612	3613	3584	3603	3577	3526	3485	3458	3387	3325	3240	3168	3123	3072	3010	2937	2821
<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
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
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
N	1181	1128	1116	1069	1043	1028	1056	1060	1064	1023	982	960	931	898	858	820	809	766	742	714	698	655
<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
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
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
N							2030	2020	1970	1959	1938	1951	1942	1886	1894	1879	1850	1818	1807	1813	1771	1717
<u>sample D</u>																						
p10												1	1.05	1.08	1.05	1.02	1.03	1	1.01	1	1	1.03
p50												1.08	1.09	1.08	1.05	1.02	1.03	1.02	1.01	1.01	1.02	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
N												395	336	302	296	293	273	285	290	277	273	261

Data Documentation 27

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 G (Percentiles of \$HBLEIB up to Wave 23).

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

Data Documentation 27

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 23).

	ahhrf	bhhrf	chhrf	dhhrf	ehhrf	fhhrrf	ghhrf	hhhrf	ihhrf	jhhrf	khhrf	lhhrf	mhhrf	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
p5	673.22	682.64	562.28	528.14	528.75	521.24	503.21	494.59	476.17
p10	1088.45	1075.05	850.92	816.11	817.73	796.23	772.23	759.78	717.67
p25	1994.82	1941.39	1521.5	1530.61	1513.97	1466.98	1417.62	1424.14	1367.87
p50	3825.75	3756.5	2380.28	2592.01	2586.585	2575.96	2531.555	2512.6	2470.74
p75	6150.22	6451.12	3526.25	4044.05	4205.83	4305.96	4351.4	4445.37	3990.04
p90	9905.59	10700.84	5280.91	6183.89	6747.815	7093.06	7490.3	7921.29	6736.14
p95	14422.31	15628.84	7229.52	8401.11	9542.31	10295.62	11062.27	11885.03	10499.36
N	7486	7215	13078	11783	11310	10999	10740	10416	11505

6 References

- Arbeitsgemeinschaft ADM-Stichproben und Bureau Wendt* 1994: Das ADM-Stichproben - System (Stand 1993). In: *Gabler, S.; Hoffmeyer-Zlotnik, J.; Krebs, D.* (Hg): Gewichtung in der Umfragepraxis. Westdeutscher Verlag, Opladen, S. 188-203.
- SOEP Group* 2001: The Socio-Economic Panel (SOEP) after more than 15 Years. In: Vierteljahrshefte zur Wirtschaftsforschung, 2001/1.
- Haiken-DeNew, John and Joachim Frick* (eds.) 2001: Desktop Companion to the Socio Economic Panel (SOEP), Version 5, Berlin.
- 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, S. 225-244.
- Rendtel, Ulrich* 1995: Panelausfälle und Panelrepräsentativität, Campus Verlag, Frankfurt/M.- New York.
- Rendtel, Ulrich, Markus Pannenberg and Stefan Daschke* 1997: Die Gewichtung der Zuwanderer-Stichprobe des Sozio-ökonomischen Panels (SOEP), in: Vierteljahrshefte zur Wirtschaftsforschung, Heft 2, S. 271-286.
- Schupp, Jürgen and Gert Wagner* 1995: Die Zuwandererstichprobe des Sozio-oekonomischen Panels (SOEP). Vierteljahrshefte zur Wirtschaftsforschung, Heft 1.
- Spiess, Martin and Ulrich Rendtel* 2000: Combining an ongoing panel with a new cross-sectional sample, DIW Discussion Paper No. 198.
- Spiess, Martin* 2001: Derivation of design weights: The case of the Socio-Economic Panel (SOEP), DIW Research Notes No. 5.
- von Rosenblatt, B., SOEP-Gruppe, TNS Infratest Sozialforschung* 2004: SOEP 2004 Methodenbericht zum Befragungsjahr 2004 (Welle 21) des Sozio-oekonomischen Panels.