SOEP

**The German
Socio-Economic
Panel study**

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP-Core v33.1 – PPFAD

SOEP Group

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 Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials
The SOEP Survey Papers are available at http://www.diw.de/soepsurveypapers

Editors:
Dr. Jan Goebel, DIW Berlin
Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

SOEP-Core v33.1 – PPFAD

SOEP Group

2018

Contents

1 General Information

This file contains one row for each person which ever lived in a SOEP household. Each person saved in at least one $PBRUTTO file is also in PPFAD. The primary key is PERSNR, the never changing Person ID. It indentifies each row of the file.

PPFAD is the best source, if one wants to reconstruct the survey history of a person, the household membership at a given point in time or needs some (almost) time-independent information like sample membership, sex, year and country of origin. It is recommended that these variables are used for cross-sectional and longitudinal analysis. These variables are adjusted on a wave-by-wave basis in the framework of demographic testing.

2 Primary Key, Foreign Keys and Sample Information

persnr – Never Changing Person ID

This is the primary key in this dataset.

psample – Sample Member

1	[1] A 1984 Initial Sample (West)	18421
2	[2] B 1984 Migration (until 1983, West)	7538
3	[3] C 1990 Initial Sample (East)	8947
4	[4] D 1994/5 Migration (1984-1994, West)	2282
5	[5] E 1998 Refreshment	2941
6	[6] F 2000 Refreshment	17341
7	[7] G 2002 High Income	4034
8	[8] H 2006 Refreshment	3799
9	[9] I 2009 Innovation Sample	3702
10	[10] J 2011 Refreshment	7469
11	[11] K 2012 Refreshment	3541
12	[12] L1 2010 Birth Cohort (2007-2010)	8726
13	[13] L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	9834
14	[14] L3 2011 Family Type (Single-Parent, Large Families)	3888
15	[15] M1 2013 Migration (1995-2011)	9082
16	[16] M2 2015 Migration (2009-2013)	3139
17	[17] M3 2016 Refugee (2013-2015)	4671
18	[18] M4 2016 Refugee/family (2013-2015)	6796
19	[19] M5 2017 Refugee (2013-2016)	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

The sample membership does never change, even if the person moves from West-Germany to East-Germany (c.p. $SAMPREG) or aquires German citizenship.

For more information, contact: Peter Krause (Tel. 030-89789-690)

hhnr – Original Household Number

The ID of the household the person belongs to in the survey year, when the person enters the SOEP.

$hhnr – Current Wave Household Number [generic]

Waves: all

The ID of the household in which a person lived at the time of the interview. [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)

3 Survey History

eintritt – Year First Contacted, Netto=10-99

1984		16252
1985		485
1986		515
1987		504
1988		459
1989		476
1990		6595
1991		609
1992		582
1993		574
1994		1301
1995		1454
1996		564
1997		552
1998		2986
...	(3 rows omitted)	16515
2002		4189
2003		811
2004		711
2005		679
2006		4062
2007		655
2008		556
2009		3993
2010		17736
2011		11269
2012		4191
2013		9522
2014		1114
2015		3929
2016		12311

The year a person joined the SOEP.

For more information, contact: Peter Krause (Tel. 030-89789-690)

erstbefr – Year First Surveyed, Netto=10-99

1984		12245
1985		527
1986		503
1987		464
1988		382
1989		386
1990		4815
1991		506
1992		499
1993		476
1994		936
1995		1069
1996		480
1997		487
1998		2326
...	(4 rows omitted)	16053
2003		779
2004		693
2005		655
2006		3151
2007		679
2008		487
2009		2857
2010		8619
2011		7403
2012		3391
2013		5964
2014		1189
2015		2620
2016		5438
-2		40072

The year of a person's first interview.
For more information, contact: Peter Krause (Tel. 030-89789-690)

austritt – Year Of Last Contact, Netto=10-99

1985	1384
1986	1398
1987	837
1988	1072
1989	929
1990	562
1991	743
1992	886
1993	1072
1994	976
1995	984

1996		1015
1997		887
1998		1112
1999		1366
...	(2 rows omitted)	4231
2002		3032
2003		2108
2004		1880
2005		1654
2006		2440
2007		2527
2008		2433
2009		2395
2010		6844
2011		4415
2012		4065
2013		5011
2014		5329
2015		5156
2016		57408

The year a person ultimately left SOEP (entered in YBRUTTO).
For more information, contact: Peter Krause (Tel. 030-89789-690)

letztbef – Year Of Last Survey, Netto=10-99

1984		1451
1985		962
1986		692
1987		836
1988		663
1989		525
1990		730
1991		756
1992		720
1993		757
1994		724
1995		702
1996		757
1997		901
1998		1111
...	(4 rows omitted)	7520
2003		1391
2004		1512
2005		1668
2006		2090
2007		1842
2008		1791
2009		2457
2010		4988

2011	4040
2012	3040
2013	4240
2014	3205
2015	4295
2016	29713
-2	40072

The year of a person's most recent interview.
For more information, contact: Peter Krause (Tel. 030-89789-690)

$netto – Current Wave Survey Status [generic]

10	[10] Interviewee With Succesful Interview (_P)	11957
12	[12] Individual Questionnaire And Person Biography	0
13	[13] Individual Questionnaire And Youth Biography	0
14	[14] Individual Questionnaire And Other Questionnaires	0
15	[15] Individual Questionnaire And Experiments, Test	0
16	[16] Individual Questionnaire, First Time Surveyed, Age 17	280
17	[17] Youth Biography First Time Surveyed, Age 17	0
18	[18] Individual Questionnaire And Child under age 17	8
19	[19] Individual Questionnaire Without Household Interview	0
20	[20] Children in Succesfully Interviewed Households (_Kind)	3928
21	[21] Children With Mother-Child Questionnaire_I, Age 0-1	0
22	[22] Children With Mother-Child Questionnaire_II, Age 2-3	0
23	[23] Children With Mother-Child Questionnaire_III, Age 5-6	0
24	[24] Children age 7-8, with parental questionnaire	0
25	[25] Children age 9-10, with parental questionnaire	0
…	(16 rows omitted)	79
89	[89] Repatriate - (was drop out [90])	0
90	[90] Individual Dropouts PBR_EXIT	0
91	[91] Moved abroad	0
92	[92] Moved abroad (abroad)	0
93	[93] Moved abroad (exit)	0
94	[94] Person Gap with advices	0
97	[97] advice to dead person (exit)	0
98	[98] advice to dead person (_VP)	0
99	[99] Has Died	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	109899
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all
The connection between persons and survey instrument can be retraced with the help of the variables $NETTO in PPFAD or $HNETTO in HPFAD: They show which survey modules (questionnaieres) have been completed or the reason for non-response are stored in this variable.

With the year 2006 (wave W, 23rd survey wave) the compilation of data on the survey population has changed fundamentally. Previously, an individual interview was carried out with all household members above the age of 16. As of 2006, the regular individual interviews based on the standard adult questionnaire are introduced one year later when household members reach the age of 18. Seventeen-year-olds instead receive an expanded youth questionnaire in their first year as SOEP respondents. (This applies to the old samples A-G; for the new sample H, distribution of this youth questionnaire will start next year, while this year's 17-year-olds have received the regular individual questionnaire, in line with the old system).

This means that we now have two instruments instead of one to obtain data on respondents: the individual and the youth questionnaire. To ensure a consistent differentiation over time, it will therefore be necessary either to include the youth population of the current year or to increase the age limit for all previous years. The $NETTO variable assists retrospectively in both differentiations for the entire survey period.

$NETTO has more detailed information than $NETOLD: The Codes 10-39 describe the population in realized and not realized households if these households were part of the current gross population. [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)

$netold – Current Wave Survey Status [generic]

0	[0] Person Gap PBR_EXIT	0
1	[1] Successful Interview _P	12245
2	[2] Below Survey Age _KIND	3928
3	[3] Did Not Participate _PBRUTTO	79
4	[4] Missing This Wave _PLUECKE	0
5	[5] Interviewee Without Household Interview	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	109899
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

A coarsened version of $NETTO. The code 1 refers (from 2006 on) not only to persons with a realized personal questionnaire and includes also persons with a realized youth questionnaire.

Based on $NETTO o-Variable the following populations can be identified:

- Respondents aged 17 and older: $netto 10-19 [$netold 1,5]
- in realized houdeholds: $netto 10-18 [$netold 1]
- 17-years old with Youth-Questionnaire: $netto 17 [$netold 1]
- Respondents aged 18 and older: (only Person Questionnaire): $netto 10-16 [$netold 1]
- Childs in realized households: $netto 20-29 [$netold 2]
- Persons in current address-log: $netto 10-39 [$netold 1-4]

You need the additional variable $HNETTO (HPFAD) if you want to differentiate between person in realized and non-realized households. The variable $POP includes this differenti-

ation and identifies additionally private and institutionalized households. [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)

$casemat – Case Match, Combined Panel Households [generic]

0	[0] CASE With HH Details	0
27367		2
27430		1
-1	[-1] No Answer	0
-2	[-2] Does not apply	126148
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: since j

It is possible that Individuals from different original households (HHNR) move together in one common household. Then people with identical values for $HHNR in one wave have different values for HHNR. Only for those persons moving together $CASEMAT contains the HHNR of the other household members. This information is not relevant when linking person and household data based on the current household number $HHNR. [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)

4 Basic Demographic Information

sex – Sex

1	[1] Male	63012
2	[2] Female	63096
-1	[-1] No Answer	43
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Respondent's (last) sex, plausibility longitudinally validated.
For more information, contact: Peter Krause (Tel. 030-89789-690)

sexor – Sexual Orientation

0	[0] probably heterosexual	64139
1	[1] probably bi/homosexual	612
-1	[-1] No Answer	61400
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0

| -5 | [-5] Not included in this version of the questionnaire | 0 |
| -6 | [-6] Version of questionnaire with modified filtering | 0 |

The variable SEXOR combines information on the sexual orientation of respondents from various sources in the SOEP. In 2016 (wave BG), (1) a direct question about sexual orientation was introduced (self-rep). Questions on marital status in the SOEP distinguish between same-sex civil unions and different-sex marriages. This distinction has been introduced in the household questionnaire since waves 2002 (wave S), in the person questionnaire since 2011 (wave BB), and in the biographical questionnaire since2012 (wave BC). Starting with these years respectively, we use information of (2) the head of household on marital status of all household members (civil-hh), information on the marital status (3) reported by individuals in the person questionnaire (civil-p), as well as reported (4) in the partnership biography (civil-bio). Finally, the SOEP team provides pointers to the partner of each person in the SOEP households since 1984 (see partnr documentation, page xxx). Combining information on the gender of both partners cohabiting in the SOEP household provides (5) the final source of information on the sexual orientation of adults in the SOEP (pointer).

Self-reports on sexual orientation surveyed in 2016 distinguish between the response options heterosexual, bisexual, and homosexual. It is however impossible to clearly identify bisexual respondents from data on same-sex and different-sex partnerships even in longitudinal studies like the SOEP. This is because some bisexual respondents may be observed at periods of no-cohabitation, only same-sex, and only different-sex partnerships. Without any observed change in the partner's gender, we are unable to identify respondents as bisexual. Our approach to this problem is as follows: first, we do not seek to distinguish between homo- and bisexuals in the generated SEXOR variable. That is, we code individuals with (at least) one observation of a same-sex partnership as homo/bisexual. We code individuals with information from at least two years (arbitrary threshold) on only different-sex relationships as heterosexual. Since bisexuals in stable/multiple different-sex partnerships are misclassified as heterosexuals instead of homo/bisexuals, we add the label "probably" to our generated variable to indicate that this information is potentially erroneous. In the case of no information on partnerships or only one year of information on different-sex partnerships we consider this insufficient to make any inferences on sexual orientation in these individuals on the basis of their observed partnerships.

Finally, the sexor variable integrates both the self-reported as well as the partnership-obtained information on sexual orientation.

sexorinfo – Sexual Orientation:Source of information

0	[0] insufficient information	61400
1	[1] pointer	18010
2	[2] civil-self	155
3	[3] pointer, civil-self	85
4	[4] civil-hh	7
5	[5] pointer, civil-hh	16790
6	[6] civil-self, civil-hh	2
7	[7] pointer, civil-self, civil-hh	7475
8	[8] BIO	151
9	[9] pointer, bio	34
10	[10] civil-self, bio	22
11	[11] pointer, civil-self, bio	5
12	[12] civil-hh, bio	0

13	[13] pointer, civil-hh, bio	239
14	[14] civil-self, civil-hh, bio	0
...	(8 rows omitted)	9286
23	[23] pointer, civil-self, civil-hh, self-rep	11371
24	[24] bio, self-rep	191
25	[25] pointer, bio, self-rep	38
26	[26] civil-self, bio, self-rep	31
27	[27] pointer, civil-self, bio, self-rep	9
28	[28] civil-hh, bio, self-rep	0
29	[29] pointer, civil-hh, bio, self-rep	3
30	[30] civil-self, civil-hh, bio, self-rep	0
31	[31] pointer, civ-self, civ-hh, bio, self-r.	847
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

This integer variable indicates which sources of information coincide with the value of SEXOR for the respective respondent. Its digits in binary representation are to be interpreted as binary flags, according to the following scheme: 1=Pointer, 2=Marital status, 4=Relation to head of household, 8=Biography, 16=Self-reported. If SEXORINFO has the value 5=1x1+0x2+1x4, this means that partnership pointers and relationship to head of household variables indicate the sexual orientation which is coded in SEXOR. Similarly, a value of 16 indicates that the inference was drawn from the direct question about sexual orientation. The variable is labeled accordingly.

gebjahr – Year Of Birth

? What are your birth month and birth year? // Year *(from: soep-core/soep-core-2016-ll/1:lgeb [4312])*

? Your sex and date of birth: // Year 19... *(from: soep-core/soep-core-2016-pb-m12-erst/1:lgeb [9173])*

? Your date of birth: // Year *(from: soep-core/soep-core-2016-pb-m34/PA2:lgeb[7058])*

1882	1
1888	1
1892	7
1893	2
1894	3
1895	6
1896	14
1897	12
1898	15
1899	29
1900	36
1901	32
1902	54
1903	43
1904	57

...	(98 rows omitted)	105943
2003		1517
2004		1597
2005		1562
2006		1536
2007		1939
2008		1870
2009		1650
2010		1700
2011		1049
2012		979
2013		961
2014		836
2015		772
2016		393
-1		1535

Respondent's year of birth, plausibility longitudinally validated.
For more information, contact: Peter Krause (Tel. 030-89789-690)

gebmonat – Month Of Birth

[?] What are your birth month and birth year? // Month *(from: soep-core/soep-core-2016-ll/1:lgebmo[4312])*

[?] Your sex and date of birth: // Month *(from: soep-core/soep-core-2016-pb-m12-erst/1:lgebmo[9173])*

[?] Your date of birth: // Month *(from: soep-core/soep-core-2016-pb-m34/PA2:lgebmo[7058])*

1	[1] January	8593
2	[2] February	7129
3	[3] March	7842
4	[4] April	7034
5	[5] May	7408
6	[6] June	6928
7	[7] July	7425
8	[8] August	7167
9	[9] September	7273
10	[10] October	6964
11	[11] November	6356
12	[12] December	6598
-1	[-1] No Answer	22190
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	1
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	17243
-6	[-6] Version of questionnaire with modified filtering	0

The month of birth

- was asked starting with wave W (2006) in the mother-child-questionnaire, resulting in file BIOAGE01

- was asked starting with wave T (2003) in supplementary biography questionnaire, resulting in file $LELA (file not available with the SOEP data distribution)
- was asked in wave S individual questionnaire, resulting in file SP
- is recorded for all children within the file TKIND (wave T, 2003)
- can be approximately derived for newborn children from the month of moving into the household, stored in file $PBRUTTO
- can be reported by parents in the personal questionnaire (which might simultaneously establish a link to the child), stored in file $P

whereas the former information is prefered over the latter. This means the generated information (from TKIND, $PBRUTTO or $P) will only be utilized if no further, questionnaire based information for the month of birth is available. The generated month of birth could only be constructed for people who were born while their parents were members of the SOEP. Several adjustments and tests of the generated data have been done which showed that – in the cases in which the generated data was also collected by SP, $LELA or $KIND – the data generation is almost always congruent with the collected data and therefore has proven to be reliable. The used source of information is stored in GEBMOVAL.

While this provides the relevant information for most of the current panel members, the information remains missing for some persons including temporary dropouts or people who exited in a previous wave. For some of them the month of birth could be reconstructed. This reconstruction remains an approximation and might differ from the true month of birth in individual cases.

For more information, contact: Christian Schmitt (Tel. 030-89789-603)

gebmoval – Month Of Birth, Data Source

1	[1] Generated from gebmonth (parents)	1550
2	[2] Ppfad, carry forward	0
3	[3] $kind, Info from mother	6732
4	[4] Info From Sp	31912
5	[5] Info From $lela	28471
6	[6] Info From bioage01 (mother)	11510
7	[7] Info from $PAGE17	6542
-1	[-1] No Answer	22190
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	1
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	17243
-6	[-6] Version of questionnaire with modified filtering	0

Indicates the data source for the month of birth (GEBMONAT).
For more information, contact: Christian Schmitt (Tel. 030-89789-603)

todjahr – Year Died, 4 Digits

1984	7
1985	76
1986	101
1987	111
1988	114

1989	120
1990	92
1991	152
1992	144
1993	164
1994	140
1995	192
1996	185
1997	161
1998	183
... (5 rows omitted)	1190
2004	321
2005	352
2006	323
2007	304
2008	339
2009	194
2010	201
2011	162
2012	178
2013	216
2014	212
2015	199
2016	203
-1	31
-2	119784

The variable TODJAHR contains the four-digit year of death for persons whose death could be firmly established or a missing value code:

- (-2): persons, for whom it is unknown whether they are deceased (that is, both persons still living up to that wave, and persons whose exact whereabouts is unknown and have dropped out of SOEP)
- (-1): persons, for whom the fact of death is known, but the year of death is unknown.

For more information, contact: Martin Kroh (+49 30 89789 678)

todinfo – Year Died, Information Source

1	[1] From Annual Survey (pbr_exit)	4510
2	[2] survey about died person ($v)	18
3	[3] survey about parents ($lela)	2
4	[4] Infratest drop-out study 1992	5
5	[5] Infratest drop-out study 2001	718
6	[6] Infratest drop-out study 2007	5
7	[7] Infratest drop-out study 2008/9	949
8	[8] Modul Family changes [P]	129
-1	[-1] No Answer	31
-2	[-2] Does not apply	119784
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0

-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

For all persons who have been identified as deceased over the course of SOEP, the variable TODINFO gives the source of this information.

53 persons were identified as deceased in the Infratest Field Organization Study (Follow-up study of drop-outs between 1984 and 1992) carried out from April – June 1992.

In the framework of the Infratest Field Organization Study (follow-up study of drop-outs) of 2001, a total of over 700 persons were identified as deceased. Among them were several with multiple entries for year of death, that is, persons who were already identified as deceased in the standard wave-to-wave follow-up procedure (stored in the file PBR_EXIT) or in the Infratest Field Organization Study of 1992. A generally very high level of correspondence was found between the information given in the standard follow-up procedure and the point of death established ex-post in the Infratest Field Organization Studies. For ten persons, the year of dropping out of SOEP was used to impute the missing year of death. In the third of those follow-up studies which has been conducted in 2007, another 21 individuals were identified as deceased between 2001 and 2005. For 18 of those persons a valid year of death could be investigated, for the remaining three observations for which the exact year of death is unknown, TODJAHR has been set to the standard missing code "-1".

When the data from the Infratest Field Organization Study contradicted the data from PBR_EXIT (help for old friends: YPBRUTTO in earlier releases), the data from the Field Organization Study was used.

For more information, contact: Martin Kroh (+49 30 89789 678)

germborn – Born in Germany

? Where is your birthplace located today? *(from: soep-core/soep-core-2016-ll/3:lgebd[4314])*

? Did you come due to flight or expulsion after the second world war until 1950 on the today's territory of Germany? *(from: soep-core/soep-core-2016-ll/4:lflucht[4315])*

1	[1] born in Germany or immigr.<1950	95992
2	[2] not born in Germany	30159
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

The SOEP data comprises a sizeable number of immigrants to Germany and their descendants. Several user-friendly variables identify these groups (GERMBORN, CORIGIN, IMMIYEAR, MIGBACK) and thus give information on the migration background of all persons who have ever been a part of a SOEP household (i.e., the population from PPFAD). In addition, MIGINFO and the new variables GERMBORNINFO, CORIGININFO, IMMIYEAR-INFO indicate the quality of information given in GERMBORN, CORIGIN, IMMIYEAR and MIGBACK, respectively. Information for these variables is collected primarily from the wave-specific individual questionnaires ($P, $PAUSL, $MIG or $REFUGEES) or the variations of the "biography / life history" questionnaires (integrated biographical data files for Waves A to L in BIOLELA or life-course information on first-time respondents since Wave M in $LELA) and from the additional 16-17-year-old questionnaire in use since 2000 ($JUGEND).

GERMBORN specifies whether a person was born in Germany or in another country. Per-

sons who immigrated to Germany before 1950 are considered as being born in Germany (the Federal Republic of Germany was founded in 1949; see also IMMIYEAR). In v33, all relevant information (see documentation on the biographical data) available on persons who have ever been a part of a SOEP household (i.e., the population from PPFAD) was combined into a working dataset and compared to code GERMBORN. The vast majority of persons who have ever been part of a SOEP household gave consistent information on their country of birth. For this PPFAD population, GERMBORN could easily be coded according to the respondents' answers. For some cases (37% of the PPFAD population) "no direct information" or "inconsistent information" on the person's country of birth was available (see GERMBORNINFO) and additional indicators were used to code the GERMBORN values. In this process, information on a respondent's citizenship and their parents' migration biography were additionally used. We coded the values on GERMBORN in the following order (with descending priority):

1. First, mothers' immigration history and their place of residence at the time of the respondents' birth were taken into account to determine the respondents' probable country of birth. For instance, when a respondent was born after or in the year of their mother's immigration to Germany, the respondent is considered to have been born in Germany. A more detailed differentiation by month of birth and mother's immigration month is not possible due to missing information. When a mother's immigration year was missing, the father's immigration history was used to code a respondent's country of birth.

2. In the next step, GERMBORN was coded for the remaining "inconsistent information" cases. Respondents' information on their country of birth, their citizenship, and parental information was taken into account to identify a respondents' country of birth. For instance, a respondent who reported being born in Germany more often than being born abroad (country of birth), who had German citizenship (citizenship), and whose parents were both born in Germany (parental information) was considered to have been born in Germany.

3. In a last step, GERMBORN was coded for the remaining "no information" cases. Respondents' citizenship and parental information was used to approximate their most likely country of birth. By definition, information on their country of birth was missing. For instance, respondents with German citizenship whose parents were both born in Germany were coded as being born in Germany.

In comparison to the GERMBORN coding from 2015 (v32), all PPFAD cases in 2016 (v33) have a value on GERMBORN. Previously (v32) missing cases were coded in this version (v33), because more longitudinal and detailed citizenship and parental information was used to generate GERMBORN (and GERMBORNINFO). For a few PPFAD cases, the new generation procedure led to a change of the GERMBORN value. To provide the highest level of transparency possible, we include a variable for the quality of information used to create the GERMBORN variable: GERMBORNINFO. For persons who according to GERMBORN were not born in Germany, the variables CORIGIN and IMMIYEAR designate the country of origin and the year of immigration to Germany, respectively. More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

germborninfo – Germborn: Quality of information

1	[1] consistent information	79754
2	[2] inconsistent information	2079
3	[3] no information	44318
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

GERMBORNINFO indicates the quality of information given in GERMBORN. In v33, all relevant information available on persons who have ever been a part of a SOEP household (i.e., the population from PPFAD) was combined into a working dataset and compared to code GERMBORN. When information in this working dataset consistently indicated that a person was born either in Germany or abroad, GERMBORNINFO was coded with a (1) for "consistent information". When inconsistent or no direct information on a SOEP person was available, GERMBORNINFO was coded with a (2) indicating "inconsistent information" or with a (3) indicating "no information". GERMBORNINFO is thus an indicator of the quality of information given in GERMBORN. More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.
For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

birthregion – Birth place: German Federal Land

1	[1] Schleswig-Holstein	622
2	[2] Hamburg	377
3	[3] Lower Saxony	2014
4	[4] Bremen	179
5	[5] North Rhine-Westphalia	4241
6	[6] Hesse	1387
7	[7] Rhineland-Palatinate	958
8	[8] Baden-Wuerttemberg	2475
9	[9] Bavaria	3060
10	[10] Saarland	224
11	[11] Berlin	875
12	[12] Brandenburg	941
13	[13] Mecklenburg-West Pomerania	599
14	[14] Saxony	1826
15	[15] Saxony-Anhalt	1103
16	[16] Thuringia	1034
-1	[-1] No Answer	0
-2	[-2] Does not apply	104236
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

BITHREGION contains information about the German Federal State ("Bundesland") a person was born. In 2012 the SOEP asked all current respondents about the place of birth: 2. Where were you born? If there are other towns or cities with the same name, or if the town is very small, please state the nearest city. Please write the name of the town in the left blank and any additional information in the right blank. For example, write "Düsseldorf," "Frankfurt an der Oder," or "Frankfurt am Main" in the left blank, and in the case of "Roßdorf bei Schmalkalden," write "Roßdorf" in the left blank and "bei Schmalkalden" in the right blank. Since then this question has been part of the biography questionnaire and a variable *birthregion* is povided in dataset *ppfad*, which has to be updated each year for new respondents. The answers is given in clear text and coded by Kantar at the level of municipalities for german cities or villages (including the geocodes for the city center). For places outside Germany, Kantar provides only the geocodes, if possible. However, the reponses could not all be assigned to a unique municipality, therefore multiple municipality codes are provided by Kantar (up to 19 in 2012). For the variable *birthregion* in *ppfad* only those answers are used, where a unique assignment of a German Federal State ("Bundesland"), based on the possible municipality codes, was possible. For persons born in a SOEP household (the household was responding in this year) the code of the respective Federal State of this year is used.
For more information, contact: Jan Goebel (Tel. +49 30-89789-377)

corigin – Country Born In

? Where is your birthplace located today? // Outside of germany: *(from: soep-core/soep-core-2016-ll/3:lgebal[4314])*

? Which country were you born in? // Other country of birth (specify): *(from: soep-core/soep-core-2016-pb-m34/2:lgebal[12183])*

1	[1] Germany	95992
2	[2] Turkey	3117
3	[3] Ex-Yugoslavia	578
4	[4] Greece	955
5	[5] Italy	1334
6	[6] Spain	678
7	[7] Ex-GDR (only as country of origin)	0
10	[10] Austria	214
11	[11] France	158
12	[12] Benelux	13
13	[13] Denmark	28
14	[14] Great Britain	99
15	[15] Sweden	30
16	[16] Norway	5
17	[17] Finland	21
...	(161 rows omitted)	22539
179	[179] Malawi	0
180	[180] Bessarabia	0
181	[181] Myanmar	0
182	[182] Fiji	0
183	[183] Niger	4
222	[222] Eastern Europe	149
333	[333] Other Unspecified Foreign Country	0
444	[444] EU-Member State (unspecif.)	0
999	[999] ethnic minority	1

-1	[-1] No Answer	236
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

CORIGIN contains information on the country of birth for all persons who have ever been a part of a SOEP household (i.e., the population from PPFAD). Respondents who were born in Germany were assigned the code (1) (see GERMBORN). Persons who were not born in Germany were assigned another country of birth than Germany depending on the information given in the wave-specific individual questionnaires ($P, $PAUSL, $MIG or $REFUGEES) or the variations of the "biography / life history" questionnaires (integrated biographical data files for Waves A to L in BIOLELA or life course information on first-time respondents since Wave M in $LELA), and from the additional questionnaire for 16-17-year-olds in use since 2000 ($JUGEND). In addition, information from $PBRUTTO or the ELECTRONIC HOUSHOLD PROTOCOL for M1 was used (both answered by the household head).

The vast majority of respondents who have ever been a part of a SOEP household (i.e., the population from PPFAD) gave consistent information on their country of birth (see CORIGININFO). For 34% of the foreign-born population no direct or inconsistent information on the person's country of birth was available. For those respondents who were not born in Germany and whose country of birth could not be determined (CORIGININFO value (2) and (3)), additional indicators were used to code their country of origin (CORIGIN). The generation process was conducted in the following order (with descending priority):

1. Respondents' country of origin was considered to be the same as their country of citizenship if this was not German. The citizenship variable was constructed on the basis of all information given on first, second, and previous citizenships as well as naturalizations, and includes the countries of citizenship a respondent reported most frequently and/or first. Since citizenship information is collected annually for all persons who lived in a SOEP household, it is based on much more detailed information than the "(2) inconsistent information" collected for the country of origin. Respondents whose information on country of origin is "(2) inconsistent" answered only three questions on their country of origin on average (from 2 to 5 answers).

2. Mothers' country of birth was considered to be the respondents' most probable place of birth if the respondent was born before the mother immigrated to Germany (see also GERMBORN coding). If information on mothers' country of birth and the respondents' citizenship was missing, fathers' country of birth was used to code CORIGIN.

3. For the few cases without citizenship and parental information (around 150 cases), respondents' most recently mentioned country of origin was used.

4. For the few cases without citizenship, parental information and any information on their country of origin (CORIGININFO value (3)), respondents' legal status was used when it indicated that a person moved to Germany from an "Eastern European" country, resulting in the coding of around 150 cases to "(222) Eastern European" on CORIGIN.

If the country of birth was still missing after this procedure, CORIGIN was coded "(-1) don't know". CORIGIN includes a few more missing values than GERMBORN due to cases in

which it was not possible to determine a country of birth other than Germany. To provide the highest level of transparency possible, we include a variable for the quality of information used to create the country of birth variable: CORIGININFO. More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

corigininfo – Corigin: Quality of information

1	[1] consistent information	19830
2	[2] inconsistent information	281
3	[3] no information	10048
4	[4] filter germborn	95992
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

CORIGININFO indicates the quality of information given in CORIGIN. In v33, all relevant information available on persons who have ever been a part of a SOEP household (i.e., the population from PPFAD) was compiled into a working dataset and compared to code CORIGIN. CORIGININFO indicates whether "(1) consistent", "(2) inconsistent" or "(3) no information" was available on a respondent's country of birth after these comparisons. CORIGININFO is thus an indicator for the quality of information given in CORIGIN. The filtering of CORIGIN via GERMBORN was taken into account by implementing a separate category, "(4) Filter GERMBORN" on CORIGININFO for the persons who were considered being born in Germany on GERMBORN (for more information, see documentation on the biographical data). More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

immiyear – Year Of Immigration To Germany

? When did you move to Germany? // Year *(from: soep-core/soep-core-2016-ll/5:lzug01[4316])*

1950	12
1951	13
1952	11
1953	9
1954	9
1955	6
1956	12
1957	33
1958	37
1959	38
1960	86
1961	101
1962	120

1963		105
1964		147
…	(39 rows omitted)	11893
2004		361
2005		278
2006		250
2007		251
2008		268
2009		380
2010		436
2011		618
2012		685
2013		1363
2014		2264
2015		5045
2016		502
-1		4826
-2		95992

IMMIYEAR contains information on the year of immigration to Germany for all persons who have ever been a part of a SOEP household (i.e., the population from PPFAD) and who were not born in Germany (see GERMBORN). All relevant information on this variable was collected from the wave-specific individual questionnaires ($P or $PAUSL) or the variations of the "biography / life history" questionnaires (integrated biographical data files for Waves A to L in BIOLELA or life course information on first-time respondents since Wave M in $LELA), and from the additional questionnaire for 16-17-year-olds in use since 2000 ($JU-GEND). Since sample M, information on all of a respondent's stays in Germany has been collected (up to 15 moves between countries, see MIGSPELL in this SOEP Survey Paper). For all cases in which a respondent had more than one stay in Germany, IMMIYEAR contains the respondent's last year of immigration to Germany. In addition, information from the ELECTRONIC HOUSEHOLD PROTOCOL for M1 was used, which was only answered by the household head.

The vast majority of the persons who have ever been a part of a SOEP household (i.e., the population from PPFAD) gave consistent information on their year of immigration (see IM-MIYEARINFO). Over the course of the SOEP survey, only very few cases gave inconsistent information with regard to their year of immigration (around 50 cases on IMMIYEARINFO). For these cases, their latest year of immigration was used in IMMIYEAR. The respondent's year of birth was used as their year of immigration if they mentioned a year of immigration that was before their year of birth (8 cases).

For those respondents who were not born in Germany and whose year of immigration could not be determined (IMMIYEARINFO value (3)), additional indicators were used to mini-mize the portion of missing values. These indicators were used in the following order (with descending priority):

1. When a respondent entered the SOEP for the first time because they had just moved into the household from abroad (see $PZUG from $PBRUTTO), the household entry year was considered to be the same as the immigration year.

2. Mother's year of immigration was used as a proxy for the respondent when the respon-dent was born before the mother immigrated to Germany. If a mother's year of immi-gration was missing, the father's year of immigration was used to code IMMIYEAR.

If the year of immigration was still missing after this procedure, IMMIYEAR was coded "(-1) don't know". IMMIYEAR includes more missing values than GERMBORN and CORIGIN due to cases in which it was not possible to determine a respondent's year of immigration. However, users should be aware that the wording of questions on the year of immigration vary rather drastically over the course of the SOEP survey (see documentation on the biographical data in the section PPFAD). To provide the highest level of transparency possible, we include a variable for the quality of information used to create the year of immigration variable: IMMIYEARINFO. More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

immiyearinfo – Immiyear: Quality of information

1	[1] consistent information	19916
2	[2] inconsistent information	49
3	[3] no information	10194
4	[4] filter germborn	95992
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

IMMIYEARINFO indicates the quality of information given in IMMIYEAR. In v33, all relevant information available on persons who have ever been a part of a SOEP household (i.e., the population from PPFAD) was compiled into a working dataset and compared to code IMMIYEAR. IMMIYEARINFO indicates whether "(1) consistent", "(2) inconsistent" or "(3) no information" was available on a respondent's country of birth after these comparisons. IMMIYEARINFO is thus an indicator for the quality of information given in IMMIYEAR. The filtering of IMMIYEAR via GERMBORN was taken into account by implementing a separate category "(4) Filter GERMBORN" on IMMIYEARINFO for individuals who were considered to have been born in Germany on GERMBORN (for more information, see GERMBORN). When information in this working dataset consistently indicated a specific year of immigration, IMMIYEARINFO was coded "(1) consistent information" and the respective year of immigration was stated in IMMIYEAR. More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

migback – Migration background

? Was your father / mother born in Germany? // Yes *(from: soep-core/soep-core-2016-ll/29:lv03a [4339])*

? Was your father / mother born in Germany? // Yes *(from: soep-core/soep-core-2016-ll/29:lm03a [4339])*

? Was your father / mother born in Germany? // [Father] Yes *(from: soep-core/soep-core-2016-pb-m12-erst/116:lv03a[8969])*

? Was your father / mother born in Germany? // [Mother] Yes *(from: soep-core/soep-core-2016-pb-m12-erst/116:lm03a[8969])*

1	[1] no migration background	81134
2	[2] direct migration background	30159
3	[3] indirect migration background	14858
4	[4] migration background, not differentiable	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

MIGBACK contains information on respondents' migration background for all persons who have ever been a part of a SOEP household (i.e., the population from PPFAD). In comparison to GERMBORN, the variable MIGBACK is useful to identify immigrants' descendants by combining information on respondents' country of birth (see GERMBORN) and parental information such as their country of birth and their citizenship. The information for this variable comes predominantly from PPFAD (GERMBORN), auxiliary citizenship variables and the relevant biographical data sets (BIOPAREN, BIOIMMIG). The variables were also updated using information from the wave-specific individual questionnaires ($P, $PAUSL, $MIG or $REFUGEES), the variations of the "biography / life history" questionnaires (integrated biographical data files for Waves A to L in BIOLELA or life course information on first-time respondents since Wave M in $LELA), and the additional questionnaire for 16-17-year-olds in use since 2000 ($JUGEND).

Respondents were assigned to the MIGBACK categories based on country of birth (see GERMBORN): Being born in another country than Germany indicates, by definition, a direct migration background (2), while respondents born in Germany may have either no (1) or an indirect (3) migration background. Respondents whose parents had no migration background were assigned the code "(1) no migration background", while respondents whose father or mother had a migration background were assigned the code "(3) indirect migration background".

In comparison to the MIGBACK coding from 2015 (v32), all PPFAD cases in 2016 (v33) have a value on MIGBACK. Previously (v32) missing cases or cases where the migration background could not be differentiated further were coded in this version (v33), because more information on respondent´s country of birth was provided (see GERMBORN). Please note that any updates in related variables may also lead to an update of the MIGBACK variable. For instance, a respondent who never stated his or her citizenship but later states having German citizenship will be classified as having a migration background of some form. This retrospective perspective may lead to updates of the migration background variable with every new wave. To provide the highest level of transparency possible, we include a variable for quality of information used to create the migration background variable: MIGINFO. More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

miginfo – Migback: Quality of information

1	[1] direct personal w/o parental info	43064
2	[2] proxy personal w/o parental info	83087
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0

-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

MIGINFO indicates the quality of information given in MIGBACK. MIGINFO provides information about the usage of parents' migration histories in the SOEP. Overall, MIGINFO can take on two different codes: "(1) No parental information" or "(2) Parental information available". The parental information refers to any information on the migration background of the respondents' mother or father or both. This includes information on the country of birth and auxiliary citizenship variables.

Please note that the MIGINFO coding from 2015 (v32) is further differentiated between the availability of direct and proxy information on respondents. We changed the MIGINFO coding due to the introduction of the GERMBORNINFO variable. The quality of information given in MIGBACK can thus only be assessed by combining the GERMBORNINFO and MIGINFO variables. MIGBACK information is considered to be highly reliable in cases coded (2) "Parental information available" on MIGINFO and (1) "Consistent information" on GERMBORNINFO (around 41% of the PPFAD cases). In contrast, the quality of information given on MIGBACK is considered relatively uncertain in cases where parental information ((1) "No parental information" on MIGINFO) and respondents' information was missing ((3) "No information" on GERMBORNINFO)).

In a few cases, "(1) no parental information" (see MIGINFO) was available but we were nonetheless able to identify respondents with an "(2) indirect migration background" (see MIGBACK). In these cases, respondents were born in Germany but further variables suggested that there was a migration background (e.g., ethnic Germans). MIGBACK may slightly underestimate the number of persons having an "(3) indirect migration background", since some of the respondents born in Germany with missing parental information and for whom no further indicators were available may be the descendants of immigrants.

More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

refback – Refugee Experience

1	[1] without evidence of refugee experience	114419
2	[2] with evidence of direct refugee experience	9376
3	[3] with evidence of indirect refugee experience	2356
-1	[-1] No Answer	0
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

For more information, contact: Peter Krause (Tel. 030-89789-690)

refinfo – refback: Source of information

1	[1] Direct refugee experience	9408
2	[2] Indirect information by mother	2288
3	[3] Indirect information by partner of mother	0

4	[4] Indirect information by household head	173
5	[5] Indirect information by partner of household head	0
9	[9] no information	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	114282
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

For more information, contact: Peter Krause (Tel. 030-89789-690)

loc1989 – Where did you live in 1989?

? Where did you live before German reunification, i.e. before 1989? *(from: soep-core/soep-core-2016-ll/20:lwied[4331])*

? Where did you live before German reunification, that is, before 1989? *(from: soep-core/soep-core-2016-pb-m12-erst/41:lwied[3725])*

1	[1] East Germany (DDR) incl. East Berlin	14642
2	[2] West Germany (FRG) incl. West Berlin	47191
3	[3] Abroad (Ausland)	8351
-1	[-1] No Answer	17740
-2	[-2] Does not apply	38227
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

The variable LOC1989 in the meta-file PPFAD provides information about a person's residence prior to German reunification, distinguishing among "(1) German Democratic Republic [GDR]", "(2) Federal Republic of Germany [FRG] (including West Berlin)", and "(3) abroad". Respondents born after 1989 (GEBJAHR in PPFAD) were coded as "(-2) does not apply" on LOC1989. This information has been generated for all individuals who were ever a member of a SOEP household (the population of PPFAD). LOC1989 combines information from two main sources: In 2003, the individual questionnaire included information on the place of residence before German reunification (TP). Since 2004, this question has been included in the biography questionnaires ($LELA). Along with these sources, the following indicators were used to code the variable LOC1989 (with descending priority):

1. $HHNR in PPFAD: Place of residence in the former FRG before German reunification
2. IMMIYEAR in PPFAD: Respondents who first immigrated to Germany after 1989 were coded as living "(3) abroad" in 1989
3. IMMIYEAR, CORIGIN in PPFAD: Respondents who immigrated to Germany before 1990 were assumed to have been living in the "(2) Federal Republic of Germany [FRG] (including West Berlin)" in 1989
4. PSAMPLE in PPFAD: Respondent's sample affiliation in 1990, differentiating between members of the former West samples (A, B) and the former East sample (C)
5. $SAMPREG in PPFAD & BRMOVEIN and SYEAR in BIORESID: Respondents who moved into their current dwelling in the former FRG or GDR before 1989
6. GSAMPREG in PPFAD: Respondent living in the West or East sample region in 1990

The vast majority of information given in LOC1989 is based on information from these sources. For the remaining respondents, indirect information is derived from the following proxies to code their place of residence in 1989:

1. $PZUG in $PBRUTTO: New entrants to the SOEP who previously lived in East Germany or abroad
2. BSSCHEND and BSSCHWO in BIOSOC: Place and year of the last school attended
3. LPGRUPPE in LPBRUTTO: Place of birth that was asked in 1995
4. $P: Country of origin GDR
5. KPNAT in KPBRUTTO: Citizens of (former) GDR
6. $P: Place of residence in 1984
7. BIOPAREN in PPFAD: Parental residence in 1989 for individuals younger than 18 in 1989

More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

locinfo – Loc1989: Source / Quality of information

0	[0] Respondent born after 1989	38227
1	[1] Direct information	68268
2	[2] Indirect information	1916
-1	[-1] No Answer	17740
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

The variable LOCINFO indicates the quality of information given in LOC1989, differentiating between direct and indirect information. LOCINFO provides information about the use of proxy information in the process of generating LOC1989 due to missing values in respondents' and their parents' residence in 1989 in the SOEP. Overall, LOCINFO can take on three different codes: either "(1) direct" or "(2) indirect information" is available on respondents or they were "(0) born after 1989". More detailed information on this variable can be found in the documentation on the biographical data in the section on PPFAD.

For more information, contact: Diana Schacht (Tel. +49 30-89789-465)

$sampreg – Current Wave Sample Region [generic]

1	[1] West-Germany	13500
2	[2] East-Germany	6131
-1	[-1] No Answer	0
-2	[-2] Does not apply	106520
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: since g

Current place of residence in East- or West-Germany with regard to borders of 1990. (West-Berlin is understood as a part of West-Germany.) As people can move, not all individuals, which are belonging to sample „German West" (PSAMPLE=1) have value for „West-German" in this variable. [This information can be related to a specific variable and is not necessarily generic.]

$pop – Sample Membership [generic]

1	[1] Private HH, German HH-Head	11401
2	[2] Private HH, Foreign HH-Head	4775
3	[3] Institutional. HH, German HH-Head	21
4	[4] Institutional. HH, Foreign HH-Head	55
5	[5] Not Compl. Private HH, German HH-Head	0
6	[6] Not Compl. Private HH, Foreign HH-Head	0
7	[7] Not Compl. Institutional. HH, German HH-Head	0
8	[8] Not Compl. Institutional. HH, Foreign HH-Head	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	109899
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

$POP was derived from $WUM2 (see private household, institutional household in $HBRUTTO) as well as $PNAT and $STELL (nationality and relationship to head of household in $PBRUTTO). Missing values were imputed based on the person's history. Thus, the only admissible missing value is –2, meaning not applicable.

This variable is therefore particularly important, as it enters into the determination of cross-sectional weights. The variable corresponds with $SHPOP in HPFAD. See also the description of $NETTO. [This information can be related to a specific variable and is not necessarily generic.]

For more information, contact: Peter Krause (Tel. 030-89789-690)