

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP-Core v32 – PPFAD

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

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

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

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

Editors:

Dr. Jan Goebel, DIW Berlin

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

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

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

Please cite this paper as follows:

SOEP Group, 2017. SOEP-Core v32 – PPFAD. SOEP Survey Papers 415: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

© 2017 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

soepapers@diw.de

SOEP-Core v32 – PPFAD

SOEP Group

2017

Contents

1	General Information	3
2	Primary Key, Foreign Keys and Sample Information	3
	persnr – Never Changing Person ID	3
	psample – Sample Member	3
	hhnr – Original Household Number	3
	\$hhnr – Current Wave Household Number [generic]	4
3	Survey History	4
	eintritt – Year First Contacted, Netto=10-99	4
	erstbefr – Year First Surveyed, Netto=10-99	4
	austritt – Year Of Last Contact, Netto=10-99	5
	letztbef – Year Of Last Survey, Netto=10-99	6
	\$netto – Current Wave Survey Status [generic]	7
	\$netold – Current Wave Survey Status [generic]	8
	\$casemat – Case Match, Combined Panel Households [generic]	9
4	Basic Demographic Information	9
	sex – Sex	9
	gebjahr – Year Of Birth	9
	gebmonat – Month of Birth	10
	gebmoval – Month Of Birth, Data Source	11
	todjahr – Year Died, 4 Digits	11
	todinfo – Year Died, Information Source	12
	germborn – Born In Germany	13
	corigin – Country Of Origin	14
	immiyear – Year Of Immigration To Germany	16
	migback – Migration background	17
	miginfo – Information source of MIGBACK	18
	loc1989 – Where did you live in 1989?	19
	\$sampreg – Current Wave Sample Region [generic]	20
	\$pop – Sample Membership [generic]	20

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 identifies 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)	18359
2	[2] B 1984 Migration (until 1983, West)	7532
3	[3] C 1990 Initial Sample (East)	8908
4	[4] D 1994/5 Migration (1984-1994, West)	2278
5	[5] E 1998 Refreshment	2940
6	[6] F 2000 Refreshment	17255
7	[7] G 2002 High Income	4000
8	[8] H 2006 Refreshment	3774
9	[9] I 2009 Innovation Sample	3702
10	[10] J 2011 Refreshment	7385
11	[11] K 2012 Refreshment	3484
12	[12] L1 2010 Birth Cohort (2007-2010)	8654
13	[13] L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	9745
14	[14] L3 2011 Family Type (Single-Parent, Large Families)	3848
15	[15] M1 2013 Migration (1995-2011)	8928
16	[16] M2 2015 Migration (2009-2013)	3048
-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 acquires 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 necessary 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
... (2 rows omitted)	15617
2001	898
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

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
... (3 rows omitted)	12741
2002	3312
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	2060
-2	33759

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

2000	1109
2001	3122
2002	3032
2003	2108
2004	1881
2005	1655
2006	2442
2007	2527
2008	2434
2009	2395
2010	6847
2011	4417
2012	4066
2013	5015
2014	5342
2015	50225

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	758
1994	724
1995	702
1996	757
1997	901
1998	1111
... (3 rows omitted)	5505
2002	2017
2003	1391
2004	1512
2005	1669
2006	2091
2007	1844
2008	1791
2009	2458
2010	4989
2011	4044
2012	3051
2013	4321

2014	3927
2015	27183
-2	33759

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 Successful 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, 17 Years	280
17	[17] Youth Biography First Time Surveyed, 17 Years	0
18	[18] Individual Questionnaire And Child und 17 Years	8
19	[19] Individual Questionnaire Without Household Interview	0
20	[20] Children in Successfully Interviewed Households (_Kind)	3928
21	[21] Children With Mother-Child Questionnaire_I, 0-1 Year	0
22	[22] Children With Mother-Child Questionnaire_II, 2-3 Year	0
23	[23] Children With Mother-Child Questionnaire_III, 5-6 Year	0
24	[24] Parents with children 7-8 year	0
25	[25] Parents with children 9-10 year	0
...	(13 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	97588
-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 (questionnaires) 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 necessary 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	97588
-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 households: \$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 differentiation and identifies additionally private and institutionalized households. [This information can be related to a specific variable and is not necessary 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	113837
-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 necessary generic.]

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

4 Basic Demographic Information

sex – Sex

1	[1] Male	56186
2	[2] Female	57630
-1	[-1] No Answer	24
-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)

gebjahr – Year Of Birth

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
... (97 rows omitted)	97393
2002	1261
2003	1274
2004	1289
2005	1263
2006	1189
2007	1603
2008	1516
2009	1337
2010	1330
2011	699
2012	663
2013	602
2014	497
2015	185
-1	1427

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

gebmonat – Month of Birth

1	[1] January	7268
2	[2] February	6648
3	[3] March	7363
4	[4] April	6624
5	[5] May	6941
6	[6] June	6553
7	[7] July	6971
8	[8] August	6772
9	[9] September	6857
10	[10] October	6563
11	[11] November	6045
12	[12] December	6272
-1	[-1] No Answer	15719
-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 preferred 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)	1524
2	[2] Ppfad, carry forward	0
3	[3] \$kind, Info from mother	11437
4	[4] Info From Sp	26769
5	[5] Info From \$lela	28471
6	[6] Info From bioage01 (mother)	7294
7	[7] Info from \$PAGE17	5382
-1	[-1] No Answer	15719
-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	145
1993	165
1994	140
1995	192
1996	185
1997	161
1998	183
... (4 rows omitted)	901
2003	290
2004	321
2005	353
2006	325
2007	304
2008	337
2009	189
2010	201
2011	163
2012	179
2013	214
2014	212
2015	186
-1	31
-2	107690

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)	4308
2	[2] survey about died person (\$v)	21
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]	111
-1	[-1] No Answer	31
-2	[-2] Does not apply	107690
-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

1	[1] Immigrant Before 49 Or German Born	87446
2	[2] Immigrant > 1984	17201
-1	[-1] No Answer	2390
-2	[-2] Does not apply	6803
-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, MIGINFO) and thus give information on the migration backgrounds of all persons who have ever been a part of a SOEP household (i.e., the population from PPFAD). Information for these variables is collected primarily 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 16-17-year-old questionnaire in use since 2000 (\$JUGEND).

GERMBORN specifies whether a person was born in Germany or in another country. Persons who immigrated to Germany before 1950 are considered as being born in Germany (1949 was the founding year of the Federal Republic of Germany; see also IMMIYEAR). Although most of the respondents gave information on their country of birth, for some cases, additional indicators were used to minimize the portion of missing values. These indicators were used in the following order:

1. Mothers' country of birth, their immigration history, and their place of residency at the time of the respondents' birth were taken into account to approximate the respondents' most probable place of birth. For instance, when a respondent was born after his/her mother immigrated to Germany, the respondent is considered to have been born in Germany.
2. The respondents' immigration year (IMMIYEAR), country of origin (CORIGIN), and citizenship (NATION\$\$ from PGEN, \$P or \$JUGEND, or \$PNAT from \$BRUTTO) were further taken into account to identify a respondents' country of birth. For instance, an immigration year implies another country of birth than Germany.
3. Minor children living in a household consisting only of persons born in Germany were considered to also have been born in Germany.
4. When sample D was launched, the country of birth was obtained from the address protocol (variable LPHERKFT). This information was used to fill in the missing values in GERMBORN.

If the country of birth is still missing after this procedure, GERMBORN is coded "-1". Permanent non-respondents are assigned the missing value "-2". A high share of these are sample G cases that were not asked to state their country of birth when the sample was launched (around 900 cases).

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)

corigin – Country Of Origin

1	[1] Germany	87446
2	[2] Turkey	2625
3	[3] Ex-Yugoslavia	906
4	[4] Greece	810
5	[5] Italy	1131
6	[6] Spain	587
7	[7] Ex-Gdr (Country of Origin Only)	0
10	[10] Austria	181
11	[11] France	129
12	[12] Benelux	13
13	[13] Denmark	24
14	[14] Great Britain	81
15	[15] Sweden	29
16	[16] Norway	6
17	[17] Finland	20
...	(160 rows omitted)	10348
178	[178] Rwanda	1
179	[179] Malawi	0
180	[180] Bessarabia	0

181	[181] Myanmar	0
182	[182] Fiji	0
183	[183] Niger	1
222	[222] Eastern Europe	155
333	[333] Other Unspecified Foreign Country	50
444	[444] Unspecified EU Member Country	0
-1	[-1] No Answer	2494
-2	[-2] Does not apply	6803
-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 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 (\$JUGEND). For those respondents who were not born in Germany and whose country of birth could not be determined, additional indicators were used to minimize the number of missing values. These indicators were used in the following order:

1. Respondents’ country of origin was considered to be the same country as that of their first, second, or previous citizenship if this was not German (NATION\$\$ from PGEN, \$P or \$JUGEND, or \$PNAT from \$BRUTTO).
2. Since 1996, first-time respondents’ have been asked to state whether they are a member of a broader group of immigrants such as Ethnic Germans from Eastern Europe or whether they are EEA or EU citizens. Respondents citing Eastern Europe as their country of origin were coded “222 - Unspecified Eastern European country,” whereas respondents who stated that they were from an EU country or a precursor country were coded “444 – Unspecified country within EU”.
3. Respondents who stated being from another region than Germany (including East and West Germany before 1989) are coded “333” (see also variables GP10803 to JP108B03).
4. 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.
5. When sample D was launched, the country of birth was obtained from the address protocol (variable LPHERKFT). This information was used to fill in further missing values in CORIGIN.

If the country of birth was still missing after this procedure, CORIGIN was coded “-1”. 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. Permanent non-respondents were assigned the missing value “-2”. A high share of these are sample G cases that were not asked to state their country of birth when the sample was launched (around 900 cases).

For persons who were born in another country than Germany, IMMIYEAR designates the year of immigration to Germany.

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

1950	19
1951	17
1952	11
1953	10
1954	12
1955	4
1956	19
1957	33
1958	38
1959	39
1960	85
1961	98
1962	119
1963	108
1964	149
... (38 rows omitted)	11348
2003	355
2004	348
2005	275
2006	242
2007	246
2008	241
2009	340
2010	404
2011	550
2012	557
2013	451
2014	85
2015	24
-1	3364
-2	94249

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). The 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 (\$JUGEND). Since sample M, information on all of a respondent’s stays in Germany is collected (up to 15 moves between countries, see MIGSPELL). 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.

For those respondents who were not born in Germany and whose year of immigration could not be determined, additional indicators were used to minimize the portion of missing values. These indicators were used in the following order:

1. When a respondent entered the SOEP for the first time because he/she 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 respondent was born before the mother immigrated to Germany.

If the year of immigration was still missing after this procedure, IMMIYEAR was coded "-1". Persons born in Germany were coded "-2" (see GERMBORN) as were permanent non-respondents. A high share of the latter are sample G cases that were not asked to state their country of birth when the sample was launched (around 900 cases).

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

1	[1] no migration background	72619
2	[2] direct migration background	17201
3	[3] indirect migration background	14827
4	[4] migration background, not differentiable	2412
-1	[-1] No Answer	6781
-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. The variable MIGBACK is useful in identifying the latter. It combines information on respondents' country of birth (see GERMBORN) and parental information such as their place of birth and their citizenship. The information for this variable comes predominantly from PPFAD (GERMBORN), PGEN (NATION\$\$), and the relevant biographical data sets (BIOPAREN, BIOIMMIG). The variables were also updated using information from the wave-specific individual questionnaires (\$P or \$PAUSL), 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 place of birth (see GERMBORN): Being born in another country 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". However, parental information is not available for all respondents. Whenever this information was missing (see MIGINFO) but the respondent was born in Germany and when further indicators also

suggested that there was no migration background (e.g. NATION\$\$), the respondent was considered having “1 - no migration background” (for more information, see MIGINFO). Since some of these respondents may be the descendants of immigrants, MIGBACK may slightly underestimate the number of persons having a “3 - indirect migration background”. Whenever information on a respondent’s country of birth was missing (see GERMBORN), the respondent’s first, second, and previous citizenships were taken into account. Having had non-German citizenship at some point in time was considered an indicator of a migration background. In a similar manner, BIIMGRP from BIOIMMIG was taken into account. If the migration background could not be differentiated further, however, respondents were assigned the code “4 – Migration background, not differentiable”. Given that citizenship is also available for household members who never participated in an interview (e.g., \$PNAT, AK07A, EK03A), MIGBACK contains fewer missing entries than, for example, GERMBORN. If the migration background is still missing after this procedure, MIGBACK is coded “-1”. 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 the sources 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 – Information source of MIGBACK

1	[1] direct personal w/o parental info	27728
2	[2] proxy personal w/o parental info	1294
3	[3] direct personal with parental info	51994
4	[4] proxy personal with parental info	26043
5	[5] only parental info and place of birth	0
-1	[-1] No Answer	6781
-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 provides information about the usage of proxy information in the generation process of MIGBACK due to missing values in respondents’ and their parents’ migration histories in the SOEP. Overall, MIGINFO can take on four different codes: either direct or proxy information is available on respondents, and either parental information is available or not.

With proxy information, we are referring to information on the respondent reported either by the parents (AK07A or EK03A) or by the interviewer (\$PNAT from \$PBRUTTO or the respective Infratest information), whereas we consider information on the country of birth (GERMBORN) or information on a respondents’ citizenship (NATION\$\$ from PGEN, \$P or \$JUGEND) to be direct respondent information. Please note that information on GERMBORN has partially been derived from other indicators (see GERMBORN). Whenever this is the case, MIGINFO reports proxy and not direct information.

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 (see GERMBORN), country of citizenship (NATION\$\$ from PGEN, \$P or \$JUGEND), or proxy information on the migration history (AK07A or EK03A or \$PNAT from \$PBRUTTO or the respective Infratest information).

MIGBACK information is considered to be particularly reliable for cases coded "3" on MIGINFO, in contrast to the other cases of missing parental information (1 and 2 in MIGINFO). The cases coded "4" in MIGINFO refer primarily to children in SOEP households whose information was reported by their parents.

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)

loc1989 - Where did you live in 1989?

1	[1] East Germany (DDR) incl. East Berlin	14420
2	[2] West Germany (FRG) incl. West Berlin	47841
3	[3] Abroad (Ausland)	8352
-1	[-1] No Answer	12471
-2	[-2] Does not apply	30756
-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

After asking this information from all respondents in 2003 (variable TP121 in file TP), a corresponding question was included in the biography questionnaire in Wave U (2004) [ub16 in file ULELA; uj58 in file UJUGEND] which will collect this time-independent information from all future first-time respondents. For all respondents interviewed up to 2013, the following information was used as input to generate LOC1989:

- Information on place and date of last school attendance (see variables BSSCHEND and BSSCHWO in file BIOSOC; variables \$B38 and \$B3701 in file \$LELA)
- Sample affiliation (variable PSAMPLE in file PPFAD)
- Year moved to current address (variable VRMOVEIN in file BIORESID; variable \$B68 in file \$LELA)
- Sample region (variables \$SAMPREG in file PPFAD)
- Year of first immigration to Germany (variable IMMIYEAR in file PPFAD)

In cases of inconsistent information from these various sources, the data collected in 2003 via variable TP121 and the information from the biography questionnaire collected since 2004 is considered superior. Persons without any individual information and aged less than 18 years in 1989 have been assigned parental information, if at all available.

Note: -2 means "does not apply; born after 1989" (see documentation on the biographical data in the section on PPFAD). With release v29, the generation process was updated. Before, people who never answered any questionnaire in the SOEP (or did answer a questionnaire but the household was not completed (N=3)), were also given the code -2, resulting in 5,341 cases with -2 with year of birth <1989. This was changed now, persons who never answered any question and were no information from other sources could be used, now have the code -1.

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: Ingrid Tucci (Tel. 030-89789-465)

\$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	94209
-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 necessary 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	97588
-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 HPPAD. See also the description of \$NETTO. [This information can be related to a specific variable and is not necessary generic.]

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