

DIW Berlin / SOEP (Ed.)

**Research Report**

## SOEP 2014 - Documentation of the person-related meta-dataset PPFAD for SOEP v31.1

SOEP Survey Papers, No. 308

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* DIW Berlin / SOEP (Ed.) (2016) : SOEP 2014 - Documentation of the person-related meta-dataset PPFAD for SOEP v31.1, SOEP Survey Papers, No. 308, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/130178>

**Standard-Nutzungsbedingungen:**

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

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

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

**Terms of use:**

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

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

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



<http://creativecommons.org/licenses/by-sa/4.0/>

## SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel study at DIW Berlin

2016

# SOEP 2014 – Documentation of the Person-related Meta-dataset PPFAD for SOEP v31.1

SOEP Group

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/soepsurveyspapers>

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, 2016. SOEP 2014 – Documentation of the Person-related Meta-dataset PPFAD for SOEP v31.1. SOEP Survey Papers 308: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



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

© 2016 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

[soeppapers@diw.de](mailto:soeppapers@diw.de)

# SOEP 2014 – Documentation of the Person-related Meta-dataset PPFAD for SOEP v31.1

SOEP Group

2016

## Contents

<b>1</b>	<b>General Information</b>	<b>3</b>
<b>2</b>	<b>Primary Key, Foreign Keys and Sample Information</b>	<b>3</b>
	persnr – Never Changing Person ID . . . . .	3
	psample – Sample Member . . . . .	3
	hhnr – Original Household Number . . . . .	4
	\$hhnr – Current Wave Household Number [generic] . . . . .	4
<b>3</b>	<b>Survey History</b>	<b>4</b>
	eintritt – Year First Contacted, Netto=10-99 . . . . .	4
	erstbefr – Year First Surveyed, Netto=10-99 . . . . .	5
	austritt – Year Of Last Contact, Netto=10-99 . . . . .	5
	letztbefr – 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
<b>4</b>	<b>Basic Demographic Information</b>	<b>9</b>
	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 . . . . .	12
	todinfo – Year Died, Information Source . . . . .	13
	corigin – Country Of Origin . . . . .	13
	germborn – Born In Germany . . . . .	14
	migback – Migration background . . . . .	15
	miginfo – Information source of MIGBACK . . . . .	15
	immiyear – Year Of Immigration To Germany . . . . .	16
	loc1989 – Where did you live in 1989? . . . . .	17
	\$sampreg – Current Wave Sample Region [generic] . . . . .	18
	\$pop – Sample Membership [generic] . . . . .	18

## 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)	18277
2	[2] B 1984 Migration (until 1983, West)	7518
3	[3] C 1990 Initial Sample (East)	8869
4	[4] D 1994/5 Migration (1984-1994, West)	2271
5	[5] E 1998 Refreshment	2936
6	[6] F 2000 Refreshment	17157
7	[7] G 2002 High Income	3968
8	[8] H 2006 Refreshment	3750
9	[9] I 2009 Innovation Sample	3702
10	[10] J 2011 Refreshment	7276
11	[11] K 2012 Refreshment	3418
12	[12] L1 2010 Birth Cohort (2007-2010)	8573
13	[13] L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	9642
14	[14] L3 2011 Family Type (Single-Parent, Large Families)	3809
15	[15] M1 2013 Migration (1995-2011)	8745
-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
1999	561
2000	15056
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

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
... (2 rows omitted)	11860
2001	881
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
-2	31890

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	1110
2001	3122
2002	3033
2003	2108
2004	1881
2005	1655
2006	2442
2007	2528
2008	2437
2009	2395
2010	6848
2011	4422
2012	4069
2013	5022
2014	51616

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	759
1994	724
1995	702
1996	757
1997	901
1998	1112
... (2 rows omitted)	3675
2001	1831
2002	2017
2003	1394
2004	1513
2005	1670
2006	2092
2007	1846
2008	1794
2009	2461

2010	4992
2011	4055
2012	3101
2013	5248
2014	28042
-2	31890

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, 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 Succesfully 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] Died	0
-1	[-1] No Answer	0
-2	[-2] Does not apply	93659
-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	93659
-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	109908
-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	54205
2	[2] Female	55685
-1	[-1] No Answer	21
-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
... (96 rows omitted)	93544
2001	1248
2002	1222
2003	1237
2004	1236
2005	1227
2006	1144
2007	1535
2008	1472
2009	1276
2010	1269
2011	628
2012	568
2013	504
2014	159
-1	1330

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	6949
2	[2] February	6408
3	[3] March	7035
4	[4] April	6368
5	[5] May	6653
6	[6] June	6297
7	[7] July	6689
8	[8] August	6489
9	[9] September	6566
10	[10] October	6307
11	[11] November	5797
12	[12] December	6016
-1	[-1] No Answer	32336
-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	0
-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)	1487
2	[2] Ppfad, carry forward	0
3	[3] \$kind, Info from mother	13617
4	[4] Info From Sp	26492
5	[5] Info From \$lela	26403
6	[6] Info From bioage01 (mother)	5957
7	[7] Info from \$PAGE17	3618
-1	[-1] No Answer	32336
-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	0
-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
... (3 rows omitted)	650
2002	250
2003	290
2004	321
2005	353
2006	325
2007	304
2008	337
2009	189
2010	201
2011	162
2012	184
2013	214
2014	203
-1	31
-2	103953

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)	4130
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]	100
-1	[-1] No Answer	31
-2	[-2] Does not apply	103953
-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)

**corigin** – Country Of Origin

---

1	[1] Germany	85998
2	[2] Turkey	2584
3	[3] Ex-Yugoslavia	969
4	[4] Greece	727
5	[5] Italy	1042
6	[6] Spain	534
7	[7] Ex-GDR (Country Of Origin Only)	0
10	[10] Austria	161
11	[11] France	105
12	[12] Benelux	13



13	[13] Denmark	24
14	[14] Great Britain	71
15	[15] Sweden	24
16	[16] Norway	5
17	[17] Finland	17
...	(159 rows omitted)	8369
177	[177] Bhutan	0
178	[178] Rwanda	1
179	[179] Malawi	0
180	[180] Bessarabia	0
181	[181] Myanmar	0
182	[182] Fiji	0
222	[222] Eastern Europe	163
333	[333] Other Foreign Country	51
444	[444] Unspecified EU Member Country	1
-1	[-1] No Answer	672
-2	[-2] Does not apply	5982
-3	[-3] Answer improbable	2398
-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

All SOEP samples also contain persons who immigrated to Germany from foreign countries. The variables CORIGIN, GERMBORN and IMMIYEAR, are given for all respondents in all SOEP samples who were interviewed at least once since 1984, based on information from the files \$P, \$PAUSL, \$PBRUTTO, BIOLELA (integrated biographical data files for Waves A to L) \$LELA (life course information on first-time respondents since Wave M) and \$JUGEND. For persons who were either not yet or no longer SOEP respondents when the immigration information was collected, the variables are set at “-1” (no response) if no adequately reliable decision on country of origin could be made based on the person-related responses. There is one exception to this: underage children, for whom country of origin is approximated based on the child’s year of birth and the biological mother’s year of immigration. Since the biographical questionnaire was first used with Sample G in the second wave (in the year 2003), the central basis for generating immigration variables is lacking for around 900 persons. For these individuals, the variables CORIGIN, GERMBORN and IMMIYEAR were set at “-3”.

For more information, contact: Elisabeth Liebau (+49 30 89789 259)

#### germborn – Born In Germany

---

1	[1] Immigrant Before 49 Or German Born	85998
2	[2] Immigrant > 1984	14966
-1	[-1] No Answer	2546
-2	[-2] Does not apply	5478
-3	[-3] Answer improbable	923
-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: Elisabeth Liebau (+49 30 89789 259)

**migback** – Migration background

---

1	[1] no migration background	72167
2	[2] direct migration background	14967
3	[3] indirect migration background	13833
4	[4] migration background, not differentiable	2275
-1	[-1] No Answer	6669
-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 provides information on migration background. It considers the nationality of the individual, his/her migration history as well as that of his/her parents for respondents born in Germany. In order to do so information on country of birth, citizenship and its change over time (in case of naturalization for example), and if available, parental information (direct information or proxy information) are combined. Direct/Indirect migration background indicates whether the person immigrated by him/herself (direct migration background) or if she/he is of migrant origin but born in Germany (indirect migration background). Respondents assigned the code 4 have a migration background. As the information on the region of birth is missing, it is not possible to know whether they migrated by themselves or if they were born in Germany.

Note that code 1 (“no migration background”) also includes persons of German nationality born in Germany whose parental information are not available. Some of them might be migrants’ offspring. Thus, the MIGBACK variable slightly underestimates the number of persons belonging to code 3, i.e. indirect migration background.

Note that in principle, an individual who gains a foreign nationality will be classified as being of migrant origin even if this was not the case before. Thus, the migration background variable as applied in SOEP takes a retrospective perspective and will be updated every year. Ethnic German repatriates (“Spät-/Aussiedler”) in this classification are coded “2” (direct migration background). Foreigners who were born in Germany and naturalized respondents who were born in Germany are coded “3” (indirect migration background).

In order to provide highest possible transparency we include a variable informing about the sources used to create the migration background variable. This information is provided in 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: Ingrid Tucci (Tel. 030-89789-465)*

**miginfo** – Information source of MIGBACK

---

1	[1] direct personal w/o parental info	37104
2	[2] proxy personal w/o parental info	992
3	[3] direct personal with parental info	39941
4	[4] proxy personal with parental info	25205
5	[5] only parental info and place of birth	0
-1	[-1] No Answer	6669
-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 MIGINFO variable informs on the information used to generate MIGBACK. There are two distinctions between used information. Firstly, the personal information. Either it stems from the interviewee or it is reported by someone else. Secondly, there is a further differentiation between persons for which we have parental information and those whose parents were not interviewed.

Direct information stem from the respondent (NATION\$\$, GERMBORN or variables about citizenship acquisition in \$P). A person's migration background is based on proxy information if at least one used information is either reported by the parents (AK07A or EK03A) or by the interviewer (\$PNAT out of \$PBRUTTO or the respective Infratest information). 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)

#### immiyear – Year Of Immigration To Germany

---

1949	13
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	117
1963	108
... (39 rows omitted)	11329
2003	348
2004	343
2005	265
2006	229
2007	242
2008	219
2009	211
2010	155
2011	103
2012	49
2013	19
2014	8
-1	821
-2	92547
-3	2400

See Variable CORIGIN.

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

For more information, contact: Elisabeth Liebau (+49 30 89789 259), Ingrid Tucci (+49 30 89789 465)

#### loc1989 – Where did you live in 1989?

1	[1] East Germany (DDR) incl. East Berlin	14235
2	[2] West Germany (FRG) incl. West Berlin	47129
3	[3] Abroad (Ausland)	6964
-1	[-1] No Answer	12590
-2	[-2] Does not apply	28993
-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	90280
-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	93659
-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 necessary generic.]

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