

DIW Berlin / SOEP (Ed.)

**Research Report**

## SOEP-IS 2022 - PPATHL: Person-related meta-dataset (long)

SOEP Survey Papers, No. 1366

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* DIW Berlin / SOEP (Ed.) (2024) : SOEP-IS 2022 - PPATHL: Person-related meta-dataset (long), SOEP Survey Papers, No. 1366, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/307897>

**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.*



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

# 1366<sup>2024</sup>

**SOEP** Survey Papers  
Series D – Variable Descriptions and Coding

## SOEP-IS 2022 – PPATHL: Person-Related Meta-Dataset (Long)

SOEP-IS Group

Running since 1984, the German Socio-Economic Panel (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. Carina Cornesse, DIW Berlin and University of Bremen  
Dr. Jan Goebel, DIW Berlin  
Prof. Dr. Cornelia Kristen, University of Bamberg and DIW Berlin  
Prof. Dr. Philipp Lersch, DIW Berlin and Humboldt-Universität zu Berlin  
Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin  
Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin  
Prof. Dr. Sabine Zinn, DIW Berlin and Humboldt-Universität zu Berlin

Please cite this paper as follows:

SOEP-IS Group, 2024. SOEP-IS 2022 – PPATHL: Person-Related Meta-Dataset (Long). SOEP Survey Papers 1366: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  
© 2024 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-IS 2022 – PPATHL: Person-Related Meta-Dataset (Long)

SOEP-IS Group

2024

The file ppathl is part of a collection, which is released with doi:[10.5684/soep.is.2022](https://doi.org/10.5684/soep.is.2022).

## Contents

<b>1</b>	<b>Primary Key, Foreign Keys and Sample Information</b>	<b>3</b>
	pid – Never Changing Person ID . . . . .	3
	syear – Survey Year . . . . .	3
	hid – Current Wave HH Number . . . . .	4
	cid – Original Houshold Number, Case ID . . . . .	5
	rv_id – ID SUF pension insurance . . . . .	6
	sample1 – Subsample Identifier . . . . .	6
<b>2</b>	<b>Survey History</b>	<b>7</b>
	eintritt – Year First Contacted, Netto=10-99 . . . . .	7
	erstbefr – Year First Surveyed, Netto=10-99 . . . . .	8
	austritt – Year Of Last Contact, Netto=10-99 . . . . .	9
	letztbef – Year Of Last Survey, Netto=10-99 . . . . .	9
	netto – Current survey status . . . . .	10
	casemat – Case-Match, combined panel households . . . . .	11
<b>3</b>	<b>Basic Demographic Information</b>	<b>11</b>
	sex – Gender . . . . .	11
	gebjahr – Birth Year, 4-digit . . . . .	12
	gebmonat – Month Of Birth . . . . .	12
	todjahr – Year Died, 4 Digits . . . . .	13
	todinfo – Year Died, Information Source . . . . .	14
	germborn – Born in Germany . . . . .	14
	germborninfo – Germborn: Quality of information . . . . .	15
	corigin – Country Born In . . . . .	15
	corigininfo – Corigin: Quality of information . . . . .	16
	immiyear – Year Moved to Germany . . . . .	16
	immiyearinfo – Immiyear: Quality of information . . . . .	17
	migback – Migration background . . . . .	17
	miginfo – Migback: Quality of information . . . . .	18
	loc1989 – Where did you live in 1989? . . . . .	18
	locinfo – Loc1989: Source / Quality of information . . . . .	19
	sampreg – Current place of residence (FRG/GDR - with respect to borders from 1989) . . . . .	19
	pop – Sample Membership . . . . .	19
	parid – Partner Person Number . . . . .	20
	partner – Status Of Partnership . . . . .	21
<b>4</b>	<b>Weighting</b>	<b>22</b>
	pbleib – Inverse Staying Probability . . . . .	22
	phrf – PRELIMINARY: Weighting factor . . . . .	22

## 1 Primary Key, Foreign Keys and Sample Information

### pid – Never Changing Person ID

---

1153802	15
1153803	12
1153804	10
1156002	9
1156102	5
1161602	13
1161603	12
1168705	5
1168706	2
1168707	2
1168802	3
1168803	3
1168804	3
1172602	19
1172603	17
... (19173 rows omitted)	118948
55582402	1
55584102	1
55584402	1
55662302	1
55662303	1
55662304	1
55664302	1
-1 No Answer	0
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

A person can be uniquely identified with variable PID. Together with SYEAR primary key in this file.

### syear – Survey Year

---

1998	936
1999	963
2000	976
2001	990
2002	990
2003	1007
2004	1017
2005	1015

2006	1032
2007	1034
2008	1015
2009	4442
2010	4720
2011	4107
2012	5849
... (2 rows omitted)	19167
2015	10528
2016	11668
2017	11078
2018	10082
2019	9080
2020	9826
2022	7563
-1 No Answer	0
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

Together with PID primary key in this file.

#### hid – Current Wave HH Number

---

115380	57
115606	19
115614	10
116033	4
116092	16
116165	45
116262	2
116874	49
116882	14
117153	3
117269	68
117668	1
117722	9
118028	68
118036	2
... (8583 rows omitted)	118681
5566720	1
5566730	1
5566740	1
5566750	1
5566760	1

5566770	1
5566780	1
-1 No Answer	0
-2 Does not apply	30
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The household the person belongs to in the corresponding year (SYEAR).

#### **cid** – Original Houshold Number, Case ID

---

200034	46
200093	60
200140	72
200158	15
200247	61
200328	132
200417	24
200506	33
200581	17
200670	69
200689	46
200808	36
200816	22
200832	98
200840	101
... (7399 rows omitted)	118064
4905917	20
4905925	28
4905933	25
4905941	35
4905950	28
4905968	28
4905976	25
-1 No Answer	0
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

Case Id - Household Source Identifier. The fixed household source id points to the first SOEP ancestor household for this person. The person does not necessarily need to have ever lived



in a household with this number. This is relevant for sampling information and calculating the weights.

#### rv\_id – ID SUF pension insurance

---

80000013	19
80000021	19
80000030	7
80000048	24
80000056	23
80000064	23
80000072	23
80000080	24
80000099	24
80000102	24
80000110	24
80000129	24
80000137	24
80000145	24
80000153	24
... (2549 rows omitted)	25456
80025733	6
80025741	6
80025750	6
80025768	6
80025776	5
80025784	5
80025792	6
-1 No Answer	0
-2 Does not apply	93259
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question this year not part of Survey program	0

#### sample1 – Subsample Identifier

---

1 A German HH Head, West, 1984	0
2 B1 Turkish HH Head, 1984	0
3 B2 Yugoslav HH Head, 1984	0
4 B3 Greek HH Head, 1984	0
5 B4 Italian HH Head, 1984	0
6 B5 Spanish HH Head, 1984	0
7 C German HH Head, East, 1990	0
8 D1 Immigrant W. Germany (1984-1994), 1994/95	0
9 D2 Immigrant W. Germany (1984-1994), 1994/95	0

10	E1 Refreshment (PAPI), 1998	10469
11	E2 Refreshment (CAPI), 1998	9770
12	F1 Refreshment, 2000	0
13	F2 Refreshment, Additional Foreigner Addresses, 2000	0
15	G2 High Income, 2002	0
16	H Refreshment, 2006	0
...	(21 rows omitted)	98846
60	L2/3 FiD Screening, Low Income, 2010/11	0
61	L2/3 FiD Screening, Single Parent, 2010/11	0
62	L2/3 FiD Screening, Families with Several Children, 2010/11	0
63	L1 FiD Birth Cohort 2007, 2010/11	0
64	L1 FiD Birth Cohort 2008, 2010/11	0
65	L1 FiD Birth Cohort 2009, 2010/11	0
66	L1 FiD Birth Cohort 2010, 2010/11	0
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of survey	0

The sample membership never changes.

## 2 Survey History

**eintritt** – Year First Contacted, Netto=10-99

---

1998	16455
1999	406
2000	184
2001	384
2002	213
2003	383
2004	289
2005	195
2006	213
2007	265
2008	104
2009	23883
2010	2200
2011	604
2012	15968
...	(2 rows omitted)
2015	1113
2016	10867
2017	946
2018	587

2019	394
2020	3465
2022	140
-1 No Answer	0
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The year a person joined the SOEP.

**erstbefr** – Year First Surveyed, Netto=10-99

---

1998	12888
1999	537
2000	343
2001	375
2002	345
2003	486
2004	172
2005	353
2006	294
2007	463
2008	188
2009	17514
2010	2169
2011	976
2012	11996
... (2 rows omitted)	26778
2015	1838
2016	8234
2017	1164
2018	1123
2019	740
2020	2320
2022	564
-1 No Answer	0
-2 Does not apply	27225
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The year of a person's first interview.

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

---

1999	6
2000	39
2001	76
2002	61
2003	74
2004	132
2005	77
2006	163
2007	262
2008	197
2009	160
2010	1564
2011	2392
2012	4334
2013	3518
2014	3311
2015	4814
2016	3860
2017	5420
2018	6336
2019	7154
2020	17352
2022	57783
-1 No Answer	0
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The last year of a person's SOEP appearance.

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

---

1998	48
1999	48
2000	68
2001	106
2002	82
2003	116
2004	115
2005	163
2006	298
2007	162

2008	145
2009	1855
2010	1605
2011	4829
2012	2796
... (2 rows omitted)	6942
2015	3516
2016	5337
2017	5080
2018	5879
2019	6784
2020	25726
2022	20160
-1 No Answer	0
-2 Does not apply	27225
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The year of a person's most recent interview.

#### netto – Current survey status

---

10 Interviewee With Successful Interview (_P)	66541
12 Individual Questionnaire And Person Biography	2714
15 Individual Questionnaire And Experiments, Test	679
16 Individual Questionnaire, First Time Surveyed, Age 17	125
17 Youth Biography First Time Surveyed, Age 17	88
19 Individual Questionnaire Without Household Interview	5
20 Children in Successfully Interviewed Households (_Kind)	19810
21 Children in Kidlong	108
22 Children (only) in BioAgeL	73
23 Children With Mother-Child Questionnaire_III, Age 5-6	52
24 Children age 7-8, with parental questionnaire	36
29 Youth from refugee sample, age 16-17	0
30 Persons In Successfully Interviewed HH Without Individual Interview	27107
31 Successful Gap Interview (_LUECKE)	91
32 Successfully Completed Biography Questionnaires	8
... (22 rows omitted)	1498
91 Moved abroad	17
92 Moved abroad (abroad)	0
93 Moved abroad (exit)	0
94 Person Gap with advices	0
97 advice to dead person (exit)	0
98 advice to dead person (_VP)	72
99 Has Died	61

-1	No Answer	0
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

This variable indicates available information and files for the entire SOEP individuals. Netto-codes 10-19 (and 29) define the respondents population of PGEN, the codes 20-28 indicate children, 30-39 unit-non-responses in partially realized households, and the codes 90-99 describe permanent (or temporary) dropouts. Further differentiations point to the survey instruments (questionnaires). The Codes 10-39 describe the population in realized (and partially realized households).

#### casemat – Case-Match, combined panel households

---

0	CASE With HH Details	4
863637		1
4030805		1
4095826		1
-1	No Answer	0
-2	Does not apply	119078
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

It is possible that Individuals from different original households (CID) move together in one common household. Then people with identical values for (HID) in one wave may have different values for CID. Only for those persons moving together CASEMAT contains the HID of the other household members. This information is not relevant when linking person and household data based on the current household number HID.

### 3 Basic Demographic Information

#### sex – Gender

---

1	Male	57946
2	Female	61048
-1	No Answer	91
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0

-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

Respondent's (last) sex, plausibility longitudinally validated.

#### gebjahr – Birth Year, 4-digit

---

1907	4
1908	9
1909	2
1910	4
1913	5
1914	2
1915	15
1917	23
1918	10
1919	58
1920	39
1921	104
1922	58
1923	107
1924	183
... (91 rows omitted)	116516
2016	456
2017	386
2018	182
2019	179
2020	95
2021	25
2022	16
-1 No Answer	607
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

Respondent's year of birth, plausibility longitudinally validated.

#### gebmonat – Month Of Birth

---

1	January	9049
2	February	8526

3	March	10161
4	April	9268
5	May	9367
6	June	8782
7	July	9737
8	August	9234
9	September	9029
10	October	9280
11	November	8373
12	December	8219
-1	No Answer	10060
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

The month of birth of a respondent is created with information from net and gross datasets (i.a. information from questionnaires header or household protocol). For about 96% of the respondents, there was no discrepancy between this information. In 3.3% of the respondents, the mode of the month of birth is used when there is differing information. For the rest, the most plausible/newest information was used (for less than 1%). For about 20% of respondents there was no information about the month of birth at all, but for about 1000 persons it was possible to use indirect information, such as the move-in date of a newborn child.

#### **todjahr** – Year Died, 4 Digits

---

2011		46
2012		298
2013		177
2014		180
2015		294
2016		280
2017		398
2018		367
2019		495
2020		261
2022		852
-1	No Answer	0
-2	Does not apply	115437
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0



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.

#### **todinfo** – Year Died, Information Source

---

1	From Annual Survey (pbr_exit)	3648
2	survey about died person (\$v)	0
3	survey about parents (\$lela)	0
4	Infratest drop-out study 1992	0
5	Infratest drop-out study 2001	0
6	Infratest drop-out study 2007	0
7	Infratest drop-out study 2008/9	0
8	Modul Family changes [P]	0
-1	No Answer	0
-2	Does not apply	115437
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

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

#### **germborn** – Born in Germany

---

1	born in Germany or immigr.<1950	106482
2	not born in Germany	11853
-1	No Answer	750
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

GERMBORN and CORIGIN give information on the country of birth, with the exception of persons who immigrated to Germany before 1950 who are considered to have been born in Germany (the Federal Republic of Germany was founded in 1949).

**germborninfo** – Germborn: Quality of information

---

1	consistent information	87454
2	inconsistent information	683
3	no answer	30948
-1	No Answer	0
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

GERMBORNINFO indicates the quality of information given in GERMBORN.

**corigin** – Country Born In

---

4	Afghanistan	58
8	Albania	60
10	Antarctica	0
12	Algeria	2
16	American Samoa	0
20	Andorra	0
24	Angola	25
28	Antigua and Barbuda	0
31	Azerbaijan	32
32	Argentina	171
36	Australia	17
40	Austria	281
44	Bahamas	0
48	Bahrain	0
50	Bangladesh	17
...	(228 rows omitted)	115437
860	Uzbekistan	21
862	Venezuela	0
876	Wallis and Futuna	0
882	Samoa	0
887	Yemen	0
894	Zambia	0
900	Kosovo	83
-1	No Answer	2881
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0

-8 Question this year not part of Survey program 0

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. Respondents who were born in Germany were assigned the code (1) (see GERMBORN).

#### corigininfo – Corigin: Quality of information

---

1	consistent information	7
2	inconsistent information	10363
3	no answer	2233
4	filter germborn	106482
-1	No Answer	0
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

CORIGININFO indicates the quality of information given in CORIGIN.

#### immiyear – Year Moved to Germany

---

1950	2
1951	3
1952	14
1953	6
1954	11
1955	17
1956	37
1957	58
1958	16
1959	40
1960	38
1961	43
1962	16
1963	32
1964	14
... (50 rows omitted)	9710
2015	121
2016	30
2017	29
2018	12
2019	11

2020	3
2022	2
-1 No Answer	2338
-2 Does not apply	106482
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

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, PPATH or PPATHL) and who were not born in Germany (see GERMBORN).

#### immiyearinfo – Immiyear: Quality of information

---

1 consistent information	0
2 inconsistent information	10388
3 no answer	2215
4 filter germborn	106482
-1 No Answer	0
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question this year not part of Survey program	0

IMMIYEARINFO indicates the quality of information given in IMMIYEAR.

#### migback – Migration background

---

1 no migration background	94403
2 direct migration background	11871
3 indirect migration background	12085
-1 No Answer	726
-2 Does not apply	0
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question this year not part of Survey program	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, PPATH or PP-ATHL). In comparison to GERMBORN, the variable MIGBACK is useful to identify immigrants' descendants by combining information on respondents' country of birth (see GERMBORN) and (grand-)parental information such as their country of birth and their citizenship. 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".

#### miginfo – Migback: Quality of information

---

1	No (grand-)parental information	119085
2	At least 1 (grand-)parental information available	0
-1	No Answer	0
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

MIGINFO indicates the quality of information given in MIGBACK. MIGINFO provides information about the usage of (grand-)parents' migration histories in the SOEP.

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

---

1	East Germany (DDR) incl. East Berlin	16202
2	West Germany (FRG) incl. West Berlin	53981
3	Abroad (Ausland)	6901
-1	No Answer	8696
-2	Does not apply	33305
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

The variable LOC1989 in the meta-file PPFAD / PPATH 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 PPATH) were coded as "(-2) does not apply" on LOC1989.

**locinfo** – Loc1989: Source / Quality of information

---

0	Respondent born after 1989	33305
1	Direct information	75483
2	Indirect information	1601
-1	No Answer	8696
-2	Does not apply	0
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	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".

**sampreg** – Current place of residence (FRG/GDR - with respect to borders from 1989)

---

1	West-Germany	99268
2	East-Germany	19664
-1	No Answer	0
-2	Does not apply	153
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

Place of residence in East- or West-Germany (with regard to borders of 1989) in corresponding year (SYEAR). (For Berlin East-West-assignments are approximated by zip-codes.)

**pop** – Sample Membership

---

1	Private HH, German HH-Head	95501
2	Private HH, Foreign HH-Head	6055
3	Institutional. HH, Collective accommodation, German HH-Head	250
4	Institutional. HH, Collective accommodation, Foreign HH-Head	20
5	Not Compl. Private HH, German HH-Head	15705
6	Not Compl. Private HH, Foreign HH-Head	1463
7	Not Compl. Institutional. HH, Collective accommodation, German HH-Head	53
8	Not Compl. Institutional. HH, Collective accommodation, Foreign HH-Head	8

-1	No Answer	0
-2	Does not apply	30
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

POP was derived from WUM2 in HBRUTTO as well as PNAT\_V\* and STELL\_V\* (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 HPOP in HPATHL. See also the description of NETTO.

#### parid – Partner Person Number

---

1153802	14
1156002	9
1156102	3
1161603	12
1168705	5
1168802	1
1172602	18
1177202	3
1180202	19
1187302	3
1187402	18
1188402	13
1192002	4
1193902	1
1198802	11
... (9324 rows omitted)	46052
49059502	7
49059603	6
49059705	1
55581502	1
55582402	1
55584102	1
55664302	1
-1 No Answer	0
-2 Does not apply	72881
-3 Answer improbable	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

Partner indicators have the purpose of defining couples in SOEP households and thus to make possible analyses on the dyadic level. Persons without spouse and (cohabitating) partner receive a missing code “-2” (=does not apply). Also, the variable PARTNER is coded 0, 3, 4, 5 in these cases. In couples, partner is the value of the unchanging person ID number (=PID) of the partner. The assignment of the partner ID within households is based on four sources of information: A question in the person-file, that asks (unmarried) respondents to identify their partner in the household (bhppnr in 2017) (plk0001 in pl), the household matrix reported by the head of household at the beginning of the interview (bhstell in 2017) (stell\_v1 stell\_v2 stell\_h in pbrutto), the partnership biography in the lifehistory calendar reported by new respondents (see also, biomars), and self-reports on marital status and life events, such as marriage, move in with partner, separation, etc. In unclear cases, due to temporal non-response for instance, we also consider longitudinal information from previous and prospective waves. Moreover, PARID is self-consistent between two individuals. For analyses of partner relationships, this information can be used to link all persons with their respective partners, and all information on both partners can also be stored in a common dataset.

#### partner – Status Of Partnership

---

0	No partner	72410
1	Spouse, registered partner	39285
2	Partner	6658
3	Probably spouse, registered partner	102
4	Probably partner	135
5	not clear	37
-1	No Answer	419
-2	Does not apply	39
-3	Answer improbable	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question this year not part of Survey program	0

The variable PARTNER generated in the context of the partner identifier (PARID) to describe whether a person in a SOEP household has a partner in that household, and if so, the type of relationship existing between the partners. Relationships with persons outside the SOEP household are not covered by this variable. Code 0 is assigned to all single persons living in households and those with partners outside the household. Codes 1 to 4 describe relationships. To assign Codes 1 and 2, the partnership has to be definable from the perspective of both partners unanimously. If conflicting information exists between partners, the codes 3 or 4 are assigned. If it is unclear whether an individual has no partner or whether she forms a couple with one other household member, we assign the code 5. Registered partnerships (civil unions) for same-sex couples were introduced in Germany in 2001. Though, registered partnerships are legally not equal to marriage, they are listed in the same category.



## 4 Weighting

### pbleib – Inverse Staying Probability

---

0	69591
1	2118
1.00999999046326	1527
1.01	389
1.01999998092651	1081
1.02	36
1.02999997138977	1689
1.03	294
1.03999996185303	1940
1.04	373
1.04999995231628	1825
1.05	272
1.05999994277954	1680
1.06	644
1.07	101
... (461 rows omitted)	35506
5.08	1
5.17	1
5.19	2
5.22	1
5.27	1
5.59	1
5.94	2
6.1	1
6.23	3
6.65	1
6.8	1
6.87	1
7.46	1
7.47	1
7.72	1

inverse probability weights

### phrf – PRELIMINARY: Weighting factor

---

0	59018
91.3394822599804	1
116.474569906374	1
123.965483202547	1
131.615977841709	1
144.1321913998	1
149.637431023341	1
188.825645078472	1
189.157231162776	1

208.857712651777	1
214.592550418881	1
217.442815081181	1
221.709988887707	1
223.560939134865	1
224.035627932505	1
... (58253 rows omitted)	60038
270622.683133398	1
274244.473157692	1
284608.887648221	1
294547.305402483	1
304325.94623773	1
304422.697315345	1
304502.615586373	1
310427.337691357	1
315232.846108875	1
318267.180312637	1
328179.315322337	1
366153.039934138	1
372757.670483056	1
394464.574097253	1
535886.192438231	1

standard individual weights