

SOEP-IS Group

**Research Report**

## SOEP-IS 2023 - PPATHL: Person-related meta-dataset (long)

SOEP Survey Papers, No. 1493

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* SOEP-IS Group (2025) : SOEP-IS 2023 - PPATHL: Person-related meta-dataset (long), SOEP Survey Papers, No. 1493, Deutsches Institut für Wirtschaftsforschung (DIW), German Socio-Economic Panel (SOEP), Berlin

This Version is available at:

<https://hdl.handle.net/10419/316311>

**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/>

# 1493<sup>2025</sup>

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

## SOEP-IS 2023 – 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. Jan Goebel, DIW Berlin  
Dr. Christian Hunkler, DIW Berlin  
Prof. Dr. Philipp Lersch, DIW Berlin and Humboldt-Universität zu Berlin  
Dr. Levent Neyse, DIW Berlin and Berlin Social Science Center (WZB)  
Prof. Dr. Carsten Schröder, 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, 2025. SOEP-IS 2023 – PPATHL: Person-Related Meta-Dataset (Long). SOEP Survey Papers 1493: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.  
© 2025 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 2023 – PPATHL: Person-Related Meta-Dataset (Long)

SOEP-IS Group

2025

The file ppathl is part of a collection, which is released with doi:10.5684/soep.is.2023.

## Contents

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

## 1 Primary Key, Foreign Keys and Sample Information

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

---

-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.

### **hid** – Current Wave HH Number

---

-1	No Answer	0
-2	Does not apply	31
-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).

### **pid** – Never Changing Person ID

---

-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
... (3 rows omitted)	29695
2016	11668
2017	11078
2018	10082
2019	9080
2020	9827
2022	7560
2023	12114
-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.

**rv\_id** – ID SUF pension insurance

---

-1 No Answer	0
-2 Does not apply	103256
-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	10650
11	E2 Refreshment (CAPI), 1998	9940
12	F1 Refreshment, 2000	0
13	F2 Refreshment, Additional Foreigner Addresses, 2000	0
15	G2 High Income, 2002	0
16	H Refreshment, 2006	0
...	(22 rows omitted)	110607
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	16654
1999	411
2000	186
2001	390
2002	217
2003	388
2004	297
2005	201



2006	218
2007	278
2008	106
2009	24612
2010	2283
2011	632
2012	16805
... (3 rows omitted)	42620
2016	11856
2017	1062
2018	693
2019	499
2020	5053
2022	268
2023	5468
-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	13040
1999	541
2000	350
2001	381
2002	352
2003	493
2004	174
2005	360
2006	298
2007	474
2008	189
2009	18052
2010	2239
2011	1010
2012	12624
... (3 rows omitted)	29987
2016	8964
2017	1248
2018	1211
2019	809
2020	3189

2022	719
2023	2881
-1 No Answer	0
-2 Does not apply	31612
-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
... (2 rows omitted)	8121
2016	3860
2017	5420
2018	6330
2019	7146
2020	17344
2022	8092
2023	61829
-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	107
2002	82
2003	116
2004	115
2005	163
2006	298
2007	162
2008	147
2009	1860
2010	1611
2011	4835
2012	2805
... (3 rows omitted)	10500
2016	5378
2017	5072
2018	5879
2019	6654
2020	22410
2022	7119
2023	24108
-1 No Answer	0
-2 Does not apply	31612
-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)	71375
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)	21057
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	33009
31	Successful Gap Interview (_LUECKE)	91
32	Successfully Completed Biography Questionnaires	8
...	(22 rows omitted)	1627
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	131190
-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

#### parid – Partner Person Number

---

-1	No Answer	0
-2	Does not apply	81916
-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, the household matrix reported by the head of household at the beginning of the interview (stell\_v1 stell\_v2 stell\_h in pbrutto), 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	81303
1	Spouse, registered partner	41805
2	Partner	7149
3	Probably spouse, registered partner	138
4	Probably partner	139
5	not clear	205
-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.

#### sex – Gender

---

1	Male	63761
2	Female	67045
-1	No Answer	391
-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	108
1924	183
... (92 rows omitted)	128121
2017	489
2018	278
2019	275
2020	190
2021	94
2022	94

2023	30
-1 No Answer	1002
-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.

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

---

1 West-Germany	109185
2 East-Germany	21845
-1 No Answer	0
-2 Does not apply	167
-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	103512
2 Private HH, Foreign HH-Head	6407
3 Institutional. HH, Collective accommodation, German HH-Head	374
4 Institutional. HH, Collective accommodation, Foreign HH-Head	59
5 Not Compl. Private HH, German HH-Head	18996
6 Not Compl. Private HH, Foreign HH-Head	1719
7 Not Compl. Institutional. HH, Collective accommodation, German HH-Head	89
8 Not Compl. Institutional. HH, Collective accommodation, Foreign HH-Head	10
-1 No Answer	0
-2 Does not apply	31
-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.

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

---

2011	46
2012	298
2013	177
2014	180
2015	294
2016	280
2017	398
2018	367
2019	495
2020	229
2022	806
2023	554
-1 No Answer	0
-2 Does not apply	127073
-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)	4124
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	127073
-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.

### immiyear – Year Moved to Germany

---

1950	3
1951	4
1952	14
1953	6
1954	14
1955	19
1956	37
1957	61
1958	17
1959	43
1960	43
1961	47
1962	18
1963	37
1964	16
... (51 rows omitted)	10299
2016	31
2017	33
2018	12
2019	15
2020	4
2022	4
2023	3
-1 No Answer	3291
-2 Does not apply	117126
-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).

**germborn** – Born in Germany

---

1	born in Germany or immigr.<1950	117164
2	not born in Germany	12623
-1	No Answer	1410
-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).

**corigin** – Country Born In

---

4	Afghanistan	64
8	Albania	65
10	Antarctica	0
12	Algeria	2
16	American Samoa	0
20	Andorra	0
24	Angola	28
28	Antigua and Barbuda	0
31	Azerbaijan	35
32	Argentina	174
36	Australia	18
40	Austria	296
44	Bahamas	0
48	Bahrain	0
50	Bangladesh	17
...	(228 rows omitted)	126757
860	Uzbekistan	22
862	Venezuela	0
876	Wallis and Futuna	0
882	Samoa	0
887	Yemen	0
894	Zambia	0
900	Kosovo	87
-1	No Answer	3632
-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 CO-RIGIN 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).

### gebmonat – Month Of Birth

---

1	January	10081
2	February	9378
3	March	11149
4	April	10284
5	May	10380
6	June	9790
7	July	10744
8	August	10236
9	September	10030
10	October	10233
11	November	9276
12	December	9122
-1	No Answer	10494
-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.

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

---

1	East Germany (DDR) incl. East Berlin	17072
2	West Germany (FRG) incl. West Berlin	58529
3	Abroad (Ausland)	6814
-1	No Answer	11649
-2	Does not apply	37133

-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.

### migback – Migration background

---

1	no migration background	104260
2	direct migration background	12623
3	indirect migration background	12904
-1	No Answer	1410
-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 PPATHL). 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	131197
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.

#### germborninfo – Germborn: Quality of information

---

1	consistent information	93753
2	inconsistent information	1093
3	no answer	36351
-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.

#### corigininfo – Corigin: Quality of information

---

1	consistent information	7
2	inconsistent information	11153
3	no answer	2911
4	filter germborn	117126
-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.

#### immiyearinfo – Immiyear: Quality of information

---

1	consistent information	0
2	inconsistent information	11183

3	no answer	2888
4	filter germborn	117126
-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.

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

---

0	Respondent born after 1989	37133
1	Direct information	81043
2	Indirect information	1372
-1	No Answer	11649
-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".

## 4 Weighting

#### pbleib – Inverse Staying Probability

---

inverse probability weights

#### phrf – PRELIMINARY: Weighting factor

---

standard individual weights. The weighting for the 2023 survey year is based on the previous year's marginal distribution. The updated marginal distributions of the Federal Statistical Office of Germany for 2023 were not yet available to us at the time of the data release. We therefore ask you to apply the weighting for 2023 with caution. The newly drawn sample in 2023 had to be integrated into the weighting with the information on the marginal distributions of the previous year, which can lead to additional inaccuracies.