

# 1654<sup>2026</sup>

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

## SOEP-Core v41 – HPATHL: Household-Related Meta-Dataset

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

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 Group, 2026. SOEP-Core v41 – HPATHL: Household-Related Meta-Dataset. SOEP Survey Papers 1654: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



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

© 2026 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Anton-Wilhelm-Amo-Straße 58

10117 Berlin

Germany

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

# SOEP-Core v41 – HPATHL: Household-Related Meta-Dataset

SOEP Group

2026

## 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>
	hid – Current Wave HH Number . . . . .	3
	syear – Survey Year . . . . .	3
	cid – Original Household Number, Case ID . . . . .	4
	hsample – Sample Member . . . . .	4
<b>3</b>	<b>Survey History</b>	<b>5</b>
	hnetto – Aktueller Befragungsstatus . . . . .	5
	hiyear – Interviewjahr (InterviewYear) . . . . .	5
<b>4</b>	<b>Basic Demographic Information</b>	<b>6</b>
	sampreg – Aktuelle Stichprobenregion . . . . .	6
	hpop – Aktuelle Populationszugehoerigkeit . . . . .	6
<b>5</b>	<b>Weighting</b>	<b>7</b>
	hbleib – Inverse Staying Probability . . . . .	7
	hhrf – Weighting factor . . . . .	7
	hhrf0 – Weighting factor for new samples (wave 1 of new sample) . . . . .	7
	hhrf1 – Weighting factor without new samples (wave 1) . . . . .	7
	hrgroup – Random Groups . . . . .	7
	weighted – Reason the household has no weight . . . . .	8
	complete – Indicator for a household being realized completely, partial, or has missing instruments . . . . .	8

## 1 General Information

The path datasets should be the building block of any analysis. Path Files indicate the total population at the household and individual level (over time) and provide all IDs necessary to access further files at different levels (Krause/Glass/Reher 2019a,b). Path-Files are delivered in three data formats – in long-format [ H|P-PATHL] (as the most comprehensive version including weighting variables), in wide-format [ H|P-PFAD] (the traditional version), and in a short-version [ H|P-PATH] as a reduced population file (indicating the total of population of households and individuals) Household Level [ HID|SYEAR] {Navigation File: H-PATH-L (long-format)} Individual Level [ PID|SYEAR] {Navigation File: P-PATH-L (long-format)} Household Level [ HID] {Population File: H-PATH} Individual Level [ PID] {Population File: P-PATH} Household Level [ HID (HHNRAKT)] {Path File: HPFAD (wide-format)} Individual Level [ PID (PERSNR)] {Path File: PPFAD (wide-format)}. The constituting SOEP population considers three levels – cases, households, and individuals. Due to the SOEP sampling and survey process, these levels follow an implicit hierarchy. All samples refer to primary source households – indicated by the household id at the time when the survey starts – the (fixed) Case ID [ CID]. New Households may emerge from these original households during the longitudinal survey process by split-offs of family members – all (current) households are therefore indicated by a (variable) Household ID [ HID]. IDs for individuals living in the households are derived from the households, where they were living when they were surveyed for the first time – the (fixed) Personal ID [ PID].

## 2 Primary Key, Foreign Keys and Sample Information

### hid – Current Wave HH Number

---

The members and configuration of a household with a given ID (HID) can change between years. Depending on the respective research issue it may be necessary to check alterations using PPATHL, where persons are assigned to a household in each year.

### syear – Survey Year

---

1984	10363
1985	6142
1986	5878
1987	5512
1988	5403
1989	5213
1990	7270
1991	7335
1992	7359
1993	7334
1994	7431
1995	7717
1996	7529
1997	7423
1998	9403
... (11 rows omitted)	160543
2010	21479
2011	25896

2012	22331
2013	31231
2014	19699
2015	28293
2016	43343
2017	28242
2018	29228
2019	45577
2020	57341
2021	22723
2022	71807
2023	78754
2024	114062

Survey year (Reference Year for survey instruments)

### cid – Original Household Number, Case ID

---

Case Id - Household Source Identifier (fixed). The fixed household source id point to the first SOEP ancestor households to which current households can be traced back to. This variable is relevant for sampling information and calculating the weights.

### hsample – Sample Member

---

1	A 1984 Initial Sample (West)	125898
2	B 1984 Migration (until 1983, West)	32532
3	C 1990 Initial Sample (East)	57722
4	D 1994/5 Migration (1984-1994, West)	10264
5	E 1998 Refreshment	13312
6	F 2000 Refreshment	89275
7	G 2002 High Income	22060
8	H 2006 Refreshment	18272
9	I 2009 Innovation Sample	6499
10	J 2011 Refreshment	35886
11	K 2012 Refreshment	17213
12	L1 2010 Birth Cohort (2007-2010)	23393
13	L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	26822
14	L3 2011 Family Type (Single-Parent, Large Families)	10191
15	M1 IAB-SOEP 2013 Migration (1995-2011)	30705
...	(12 rows omitted)	223219
28	M8b IAB-SOEP 2022 Refreshment	25345
29	M8c IAB-SOEP 2023 Refreshment	17477
30	M9 IAB-BAMF-SOEP 2023 Refugee	38550
31	S 2024 Refreshment	42509
32	M8d IAB-SOEP 2024 Refreshment	7430
33	M10 IAB-SOEP 2024 Refreshment	33287
-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
-9	Missing due to a terminated interview	0

Unchanging sample membership at first contact with SOEP (at the point of selection of the original household into the sample).

### 3 Survey History

#### hnetto – Aktueller Befragungsstatus

---

1	Successful HH Interview	492743
2	HH In Brutto Only	113995
3	unsuccessful, initial gross sample	301115
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	8
-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
-9	Missing due to a terminated interview	0

Wave-specific survey status: for HNETTO=1 there is a HH interview in the respective year, i.e. an entry in the file HL. For HNETTO=2 there is only an entry in the file HBRUTTO, but the HH could not be interviewed (the reason for which is given by the variable HERGS in the file HBRUTTO). HNETTO=3 indicates unrealized Households of new samples in HBRUTT-Files.

#### hiyear – Interviewjahr (InterviewYear)

---

1984	5921
1985	6109
1986	5838
1987	5472
1988	5365
1989	5171
1990	7225
1991	7275
1992	7264
1993	7226
1994	7373
1995	7498

1996	7447
1997	7391
1998	8330
... (14 rows omitted)	199053
2013	20861
2014	19572
2015	19267
2016	21065
2017	22148
2018	24109
2019	24588
2020	23501
2021	14371
2022	23876
2023	20689
2024	38651
2025	1
-2	320440
-8	14764

Interview Year (indicates household interviews realized also outside of standard SYEAR)

## 4 Basic Demographic Information

### sampreg – Aktuelle Stichprobenregion

---

1	West Germany	727027
2	East Germany	170493
-1	No answer	722
-2	Does not apply	9611
-3	Implausible value	8
-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
-9	Missing due to a terminated interview	0

The variable SAMPREG tells whether a HH was located in West or East Germany in the respective year. This variable approximates also former regional differences within Berlin according to zip-codes.

### hpop – Aktuelle Populationszugehoerigkeit

---

1	Private HH, German HH-Head	476648
2	Private HH, Foreign HH-Head	108209
3	Institutionalized HH, German HH-Head	4469

4	Institutionalized HH, Foreign HH-Head	12100
-1	No answer	0
-2	Does not apply	306435
-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
-9	Missing due to a terminated interview	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 POP in PPATH/L.

## 5 Weighting

### **hbleib** – Inverse Staying Probability

---

inverse probability weights

### **hhrf** – Weighting factor

---

standard household weight

### **hhrf0** – Weighting factor for new samples (wave 1 of new sample)

---

household weight for first wave of new samples

### **hhrf1** – Weighting factor without new samples (wave 1)

---

household weights without first wave of new samples

### **hrgroup** – Random Groups

---

0	303899
1	75212
2	74968
3	75578
4	73052
5	75803
6	77374
7	76031
8	75944

**weighted** – Reason the household has no weight

---

1	Yes	484412
2	No: HH not completed (syear < 2021)	63056
3	No: HH not completed (syear >= 2021), see variable complete	50986
4	No: Sampled using non-random selection methods	4055
5	No: Weighted from 1995 onwards	93
6	No: Out of sample: below income threshold	436
7	No: Transferred to Innovation Sample	3251
8	No: Not eligible in first wave	344
9	No: Out of sample: households without refugees	9
10	No: Improper interviews	96
-1	No answer	0
-2	Does not apply	301123
-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
-9	Missing due to a terminated interview	0

Indicator for a household having a weight and if not giving a reason for it

**complete** – Indicator for a household being realized completely, partial, or has missing instruments

---

1	Yes: HH matrix, HH interview, all PE interviews available	36060
2	Yes: HH matrix, HH interview, at least one PE interview missing	40384
3	No: HH matrix, no HH interview, at least one PE interview available	1203
4	No: HH matrix, HH interview, no PE interview available	1224
5	No: HH matrix, no HH interview, no PE interview available	3670
6	No: No HH matrix, HH interview, at least one PE interview available	300
7	No: No HH matrix, no HH interview, at least one PE interview available	67
8	No: No HH matrix, HH interview, no PE interview available	16
9	No: No HH matrix, no HH interview, no PE interview available	44688
-1	No answer	0
-2	Does not apply	159734
-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	620515
-9	Missing due to a terminated interview	0

Vollständige oder teilweise Realisierung eines Haushaltes und fehlende Instrumente