1551 2025

SOEP Survey Papers
Series D – Variable Descriptions and Coding

SOEP-Core v40.1 – HPATHL: Household-Related Meta-Dataset

SOEP Group

DIW BERLIN

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/soepsurveypapers

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, 2025. SOEP-Core v40.1 – HPATHL: Household-Related Meta-Dataset. SOEP Survey Papers 1551: 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)
Anton-Wilhelm-Amo-Straße 58
10117 Berlin
Germany

soeppapers@diw.de

DIW BERLIN The German Socio Economic Panel at DIW Berlin

SOEP-Core v40.1 – HPATHL: Household-Related Meta-Dataset

SOEP Group

2025

The file hpathl is part of a collection, which is released with doi:10.5684/soep.core.v40.1o.

Contents

1 General Information 3

2 Primary Key, Foreign Keys and Sample Information 3
hid – Current Wave HH Number . . . 3
syear – Survey Year . . . 3
cid – Original Houshold Number, Case ID . . . 4
hsample – Sample Member . . . 4

3 Survey History 5
hnetto – Aktueller Befragungsstatus . . . 5
hiyear – Interviewjahr (InterviewYear) . . . 5

4 Basic Demographic Information 6
sampreg – Aktuelle Stichprobenregion . . . 6
hpop – Aktuelle Populationszugehoerigkeit . . . 6

5 Weighting 7
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 houshold 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
...	(10 rows omitted)	143901
2009		16642
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

Survey year (Reference Year for survey instruments)

cid – Original Houshold 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)	125039
2	B 1984 Migration (until 1983, West)	32428
3	C 1990 Initial Sample (East)	57151
4	D 1994/5 Migration (1984-1994, West)	10172
5	E 1998 Refreshment	13276
6	F 2000 Refreshment	88143
7	G 2002 High Income	21673
8	H 2006 Refreshment	17938
9	I 2009 Innovation Sample	6499
10	J 2011 Refreshment	34736
11	K 2012 Refreshment	16586
12	L1 2010 Birth Cohort (2007-2010)	22589
13	L2 2010 Family Type (Low-Income, Single-Parent, Large Families)	25539
14	L3 2011 Family Type (Single-Parent, Large Families)	9627
15	M1 IAB-SOEP 2013 Migration (1995-2011)	29832
...	(9 rows omitted)	134408
25	M7 IAB-SOEP 2020 Migration	21865
26	M8 IAB-SOEP 2020 Migration Professionals	16070
27	R 2022 Refreshment	36945
28	M8b IAB-SOEP 2022 Refreshment	23017
29	M8c IAB-SOEP 2023 Refreshment	16500
30	M9 IAB-BAMF-SOEP 2023 Refugee	33766
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	0

-4	Inadmissable 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	468837
2	HH In Brutto Only	99275
3	unsuccessful, initial gross sample	225679
-1	No answer	0
-2	Does not apply	0
-3	Implausible value	8
-4	Inadmissable 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
...	(12 rows omitted)	160598
2011		19232
2012		19223
2013		20861
2014		19572
2015		19267
2016		21065
2017		22148
2018		24109
2019		24588
2020		23501
2021		14371
2022		23876
2023		1212
-2		200517
<NA>		78754

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

4 Basic Demographic Information

sampreg – Aktuelle Stichprobenregion

1	West Germany	577921
2	East Germany	142846
-1	No answer	0
-2	Does not apply	73024
-3	Implausible value	8
-4	Inadmissable 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	454518
2	Private HH, Foreign HH-Head	94148
3	Institutionalized HH, German HH-Head	4134

4	Institutionalized HH, Foreign HH-Head	10024
-1	No answer	0
-2	Does not apply	230975
-3	Implausible value	0
-4	Inadmissable 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	228379
1	70361
2	70182
3	70747
4	68266
5	71005
6	72496
7	71167
8	71196

weighted – Reason the household has no weight

1	Yes	460737
2	No: HH not completed (syear < 2021)	63056
3	No: HH not completed (syear >= 2021), see variable complete	36081
4	No: Sampled using non-random selection methods	4015
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	339
9	No: Out of sample: households without refugees	8
10	No: Improper interviews	96
-1	No answer	0
-2	Does not apply	225687
-3	Implausible value	0
-4	Inadmissable 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 houshold being realized completely, partial, or has missing instruments

1	Yes: HH matrix, HH interview, all PE interviews available	24900
2	Yes: HH matrix, HH interview, at least one PE interview missing	27845
3	No: HH matrix, no HH interview, at least one PE interview available	950
4	No: HH matrix, HH interview, no PE interview available	1017
5	No: HH matrix, no HH interview, no PE interview available	2595
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	31296
-1	No answer	0
-2	Does not apply	84298
-3	Implausible value	0
-4	Inadmissable 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