

## SOEP Survey Papers

Series D – Variable Descriptions and Coding

# SOEP-Core v34 – HPATHL: Household- Related Meta-Dataset

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

Prof. Dr. Stefan Liebig, DIW Berlin and Freie Universität Berlin

Dr. David Richter, DIW Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Dr. Sabine Zinn, DIW Berlin

Please cite this paper as follows:

SOEP Group, 2019. SOEP-Core v34 – HPATHL: Household-Related Meta-Dataset. SOEP Survey Papers 760: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



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

© 2019 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-Core v34 – HPATHL: Household-Related Meta-Dataset

SOEP Group

2019

The file `hpathl` is part of a collection, which is released with doi:[10.5684/soep.v34](https://doi.org/10.5684/soep.v34).

## 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 – . . . . .	3
	syear – . . . . .	3
	cid – . . . . .	4
	hsample – . . . . .	4
<b>3</b>	<b>Survey History</b>	<b>5</b>
	hnetto – . . . . .	5
	hiyear – . . . . .	5
<b>4</b>	<b>Basic Demographic Information</b>	<b>6</b>
	sampreg – . . . . .	6
	hpop – . . . . .	6
<b>5</b>	<b>Weighting</b>	<b>6</b>
	hrgroup – . . . . .	6
	hbleib – . . . . .	6
	hhrf – . . . . .	6
	hhrf0 – . . . . .	7
	hhrf1 – . . . . .	7

## 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 –

---

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 –

---

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
... (4 rows omitted)	61650
2003	13756
2004	13117

2005	12836
2006	16481
2007	13410
2008	12651
2009	16642
2010	21479
2011	25896
2012	22331
2013	31231
2014	19699
2015	28293
2016	43343
2017	28242

Survey year (Reference Year for survey instruments)

### cid -

---

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 -

---

1	117156
2	31415
3	52043
4	9348
5	12945
6	77965
7	18554
8	14883
9	6499
10	24959
11	11210
12	16553
13	17206
14	6034
15	22258
16	12069
17	16585
18	12490
19	5390
20	2807
-1	0
-2	0
-3	0
-4	0
-5	0
-6	0
-8	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 -

---

1	356864
2	49853
3	81644
-1	0
-2	0
-3	8
-4	0
-5	0
-6	0
-8	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 HBRUTTO-Files.

#### hiyear -

---

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
... (6 rows omitted)	77273
2005	12810
2006	14012
2007	13378
2008	12600
2009	13170
2010	17355
2011	19232

2012	19223
2013	20861
2014	19572
2015	19267
2016	21065
2017	22148
2018	1624
-2	83874

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

## 4 Basic Demographic Information

### sampreg -

---

1	328776
2	86655
-1	0
-2	72930
-3	8
-4	0
-5	0
-6	0
-8	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 -

---

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

### hrgroup -

---

Random Groups (Total SOEP population is randomly divided in eight groups)

### hbleib -

---

inverse probability weights

### hhrf -

---

standard household weight



**hhf0** -

---

household weight for first wave of new samples

**hhf1** -

---

household weights without first wave of new samples