

SOEPcampus June 2019

Introduction to the German Socio-Economic Panel (SOEP) study – Part I

SOEP Group, Sandra Bohmann

Introduction to the SOEP I

1. A Short Introduction to Panel Data
2. What is the SOEP?
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

A Short Introduction to Panel Data

- Panel data: repeated survey of multiple, identical units over multiple, identical points in time

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
2010	✓	✓	✓	✓	✓
2011	✓	✓	✓	✓	✓
2012	✓	✓	✓	✓	✓
2013	✓	✓	✓	✓	✓
2014	✓	✓	✓	✓	✓
2015	✓	✓	✓	✓	✓
2016	✓	✓	✓	✓	✓

1

A Short Introduction to Panel Data

- Panel data: repeated survey of multiple identical units over multiple, identical points in time

Cross-section:
all units, one year

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
2010	✓	✓	✓	✓	✓
2011	✓	✓	✓	✓	✓
2012	✓	✓	✓	✓	✓
2013	✓	✓	✓	✓	✓
2014	✓	✓	✓	✓	✓
2015	✓	✓	✓	✓	✓
2016	✓	✓	✓	✓	✓

1

A Short Introduction to Panel Data

- Panel data: repeated survey of multiple, identical units over multiple, identical points in time

Longitudinal/time series:
one unit, all years

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
2010	✓	✓	✓	✓	✓
2011	✓	✓	✓	✓	✓
2012	✓	✓	✓	✓	✓
2013	✓	✓	✓	✓	✓
2014	✓	✓	✓	✓	✓
2015	✓	✓	✓	✓	✓
2016	✓	✓	✓	✓	✓

- Panel data: repeated survey of multiple identical units over multiple, identical points in

Pooled cross-section:
several units, several
years

	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
2010			✓		
2011	✓				
2012		✓			
2013			✓		
2014				✓	
2015	✓				
2016					✓

1

A Short Introduction to Panel Data

Complete information analysis with an unbalanced panel

t1	t2	t3	
			dropouts
			successfully interviewed in all waves
			new subjects/ households in sample

Information from different life stages

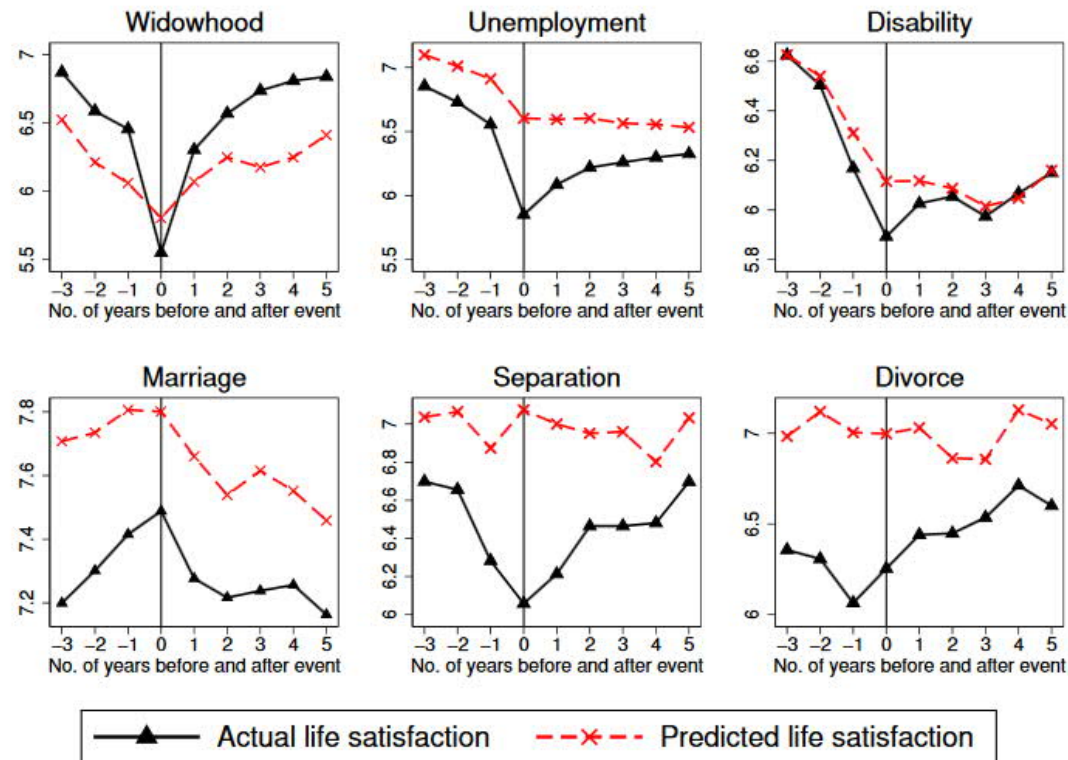
- Identification of transition phases and important life events and their (causal) effects
- Measurement of phenomena as e.g. “time in poverty”
- Control of time-constant extraneous variables (approaching causality)
- Analysis of trends

(Mis-)Predicted Subjective Well-Being Following Life Events

(Odermatt and Stutzer, 2017 – Working Paper No. 2017 – 12, Center for Research in Economics, Management and the Arts)

Advantages of Panel Data: Example 1

Figure 1: Patterns in mean actual and predicted life satisfaction around life events



Note: Sample sizes are presented in Table 1.
Data source: SOEP.

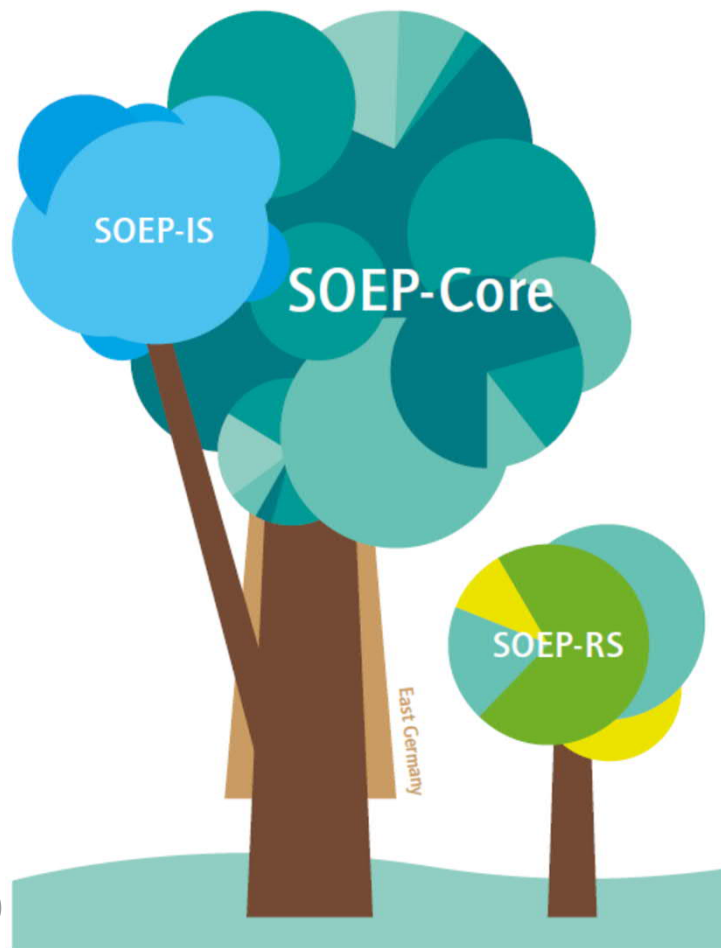
- Selection into sample
→ Weights
- Distortion due to panel attrition
→ Weights
- Limited generalizability due to changes in the population
→ Refreshment samples
- Trade-off: Repetition of identical questions vs. adaptation to new topics (limited survey time)
→ Rotating modules

Introduction to the SOEP - Part I

1. A Short Introduction to Panel Data
- 2. What is the SOEP?**
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

Introduction to the SOEP - Part I

1. A Short Introduction to Panel Data
- 2. What is the SOEP?**
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services



SOEPcore

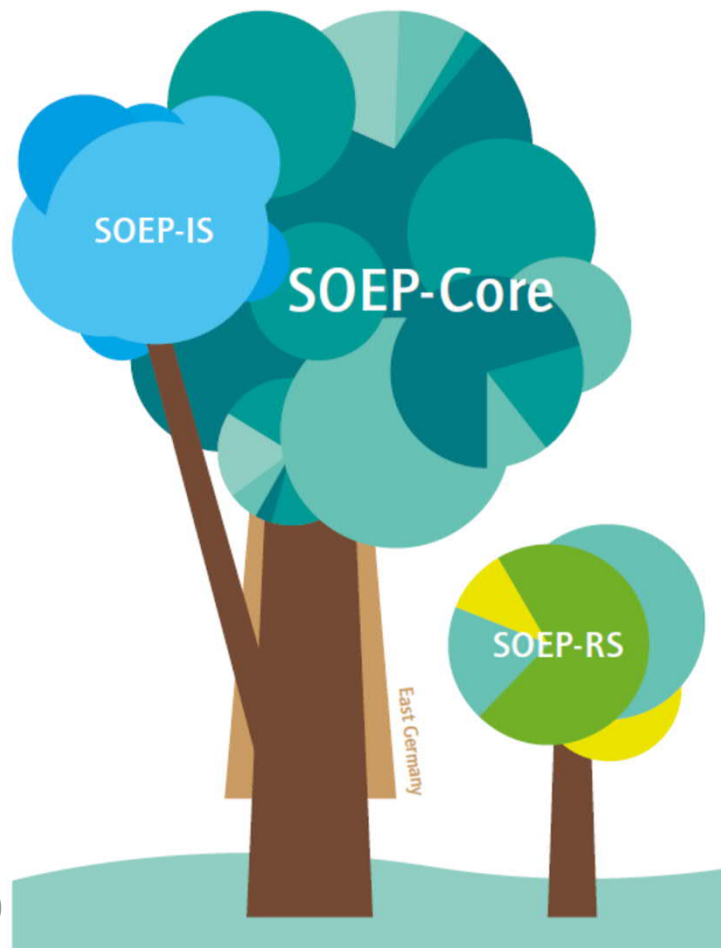
- 1984 - today
- 18 subsamples, regular refreshments

SOEP IS - Innovation Sample

- From 2012 onwards
- Subsamples E and I
- Opportunity for external users to propose survey modules

SOEP RS - Related Studies

- Question+ data structure of SOEP
 - Berlin Aging Study II (BASE II)
 - Families in Germany (FiD)
 - PIAAC-L
 - BRISE



SOEP Core

- 1984 - today
- 18 subsamples, regular refreshments

SOEP IS - Innovation Sample

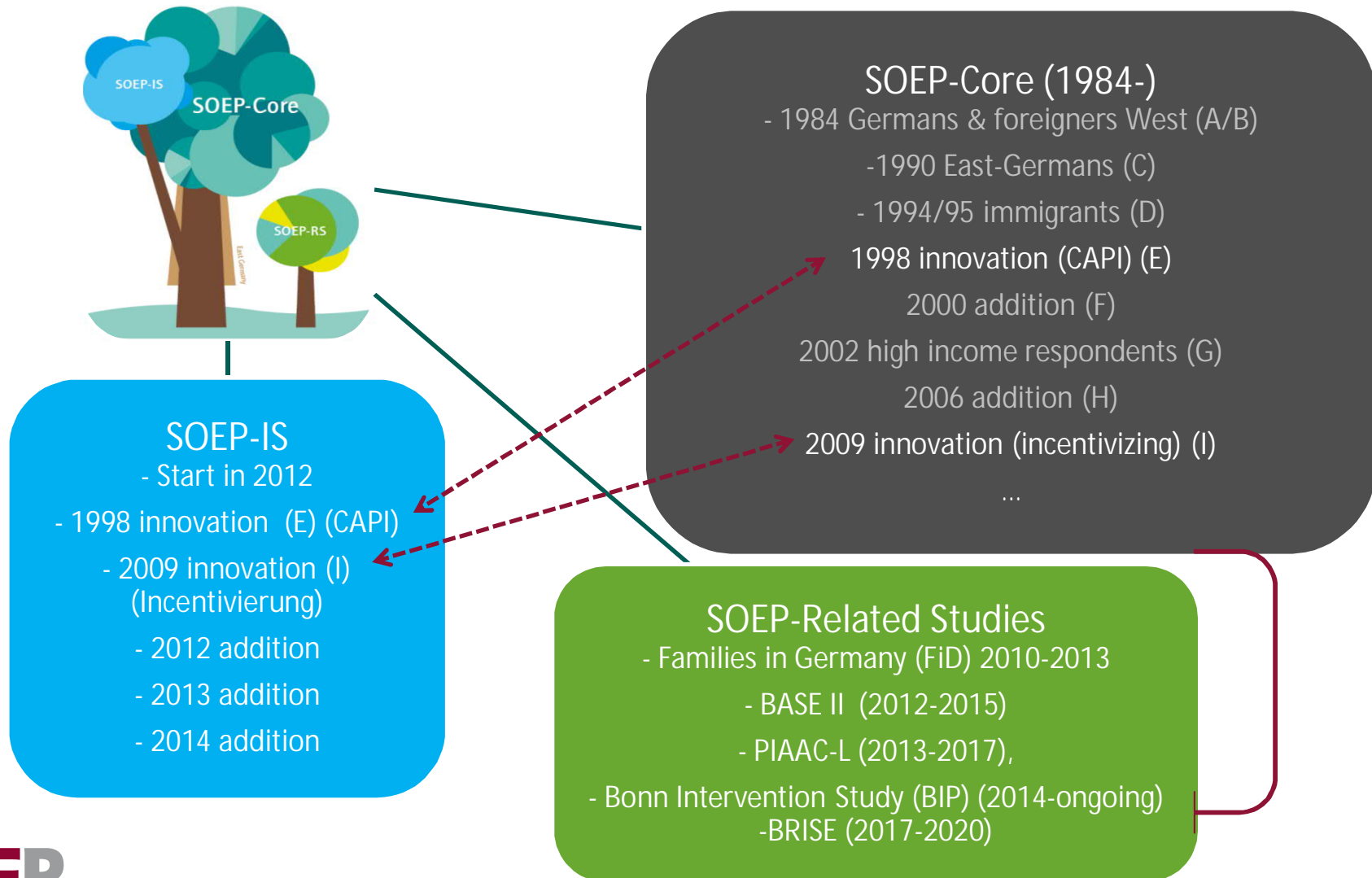
- From 2012 onwards
- Subsamples E and I
- Opportunity for external users to propose survey modules

SOEP RS - Related Studies

- Question+ data structure of SOEP
 - Berlin Aging Study II (BASE II)
 - Families in Germany (FiD)
 - PIAAC-L
 - BRISE

2

SOEP IS and Related Studies



Household Panel Study: Representative annual panel survey of private households in Germany

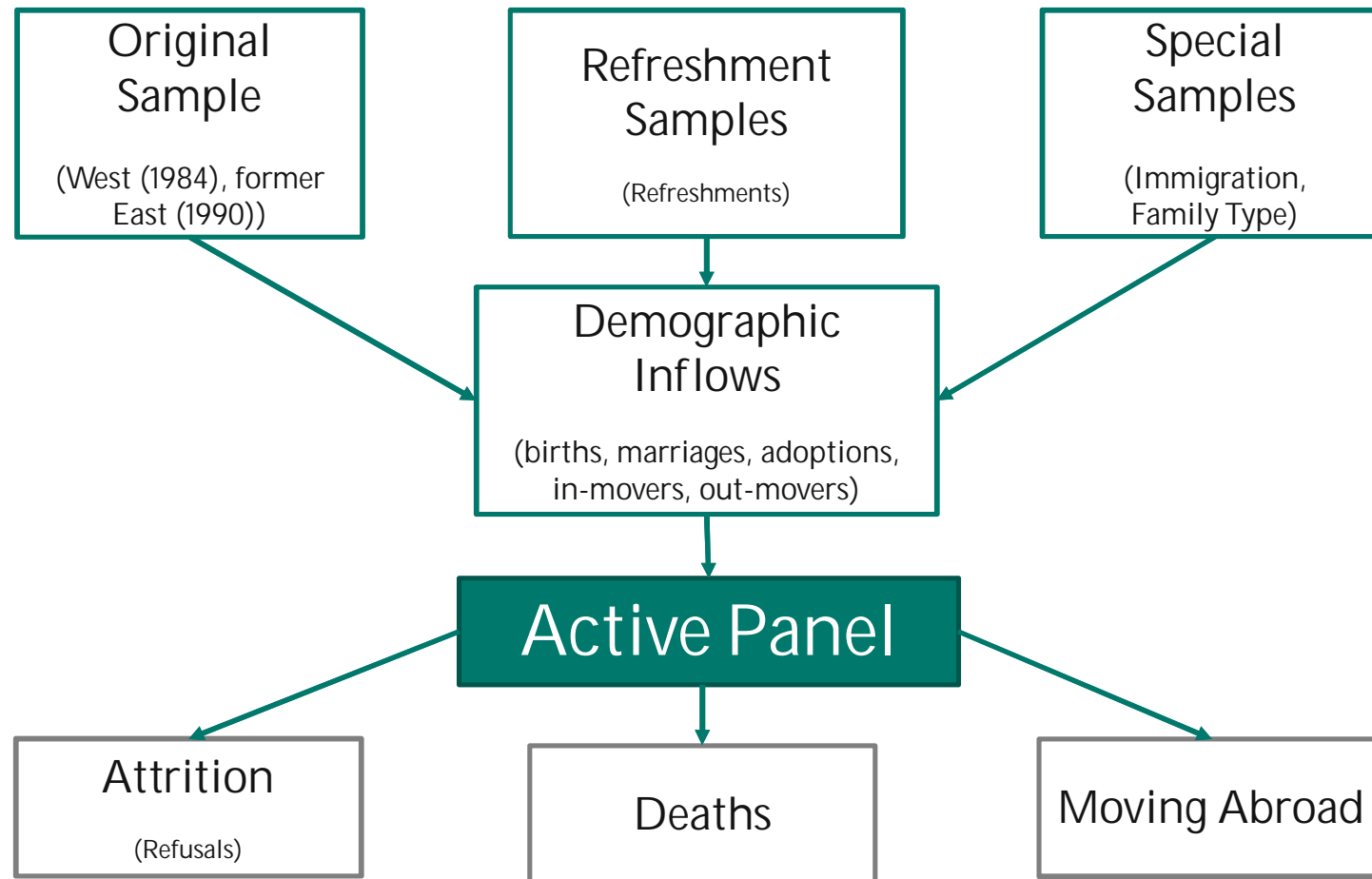
- Since 1984
- In 2018 about 15.000 households
- Information for up to three generations due to genealogical design
- Life-Course perspective with a focus on well-being
- Individual information on all individuals living in the sampled household
- Theory-based multidisciplinary mix of subjective and objective data on households, individuals and families.
- Provision of micro-data for research in social and behavioral sciences as well as economics for universities and independent research centers in Germany and abroad
- Possibility of regional comparison and availability local context indicators (only at SOEP Research Data Center at DIW Berlin)

Introduction to the SOEP - Part I

1. A Short Introduction to Panel Data
- 2. What is the SOEP?**
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

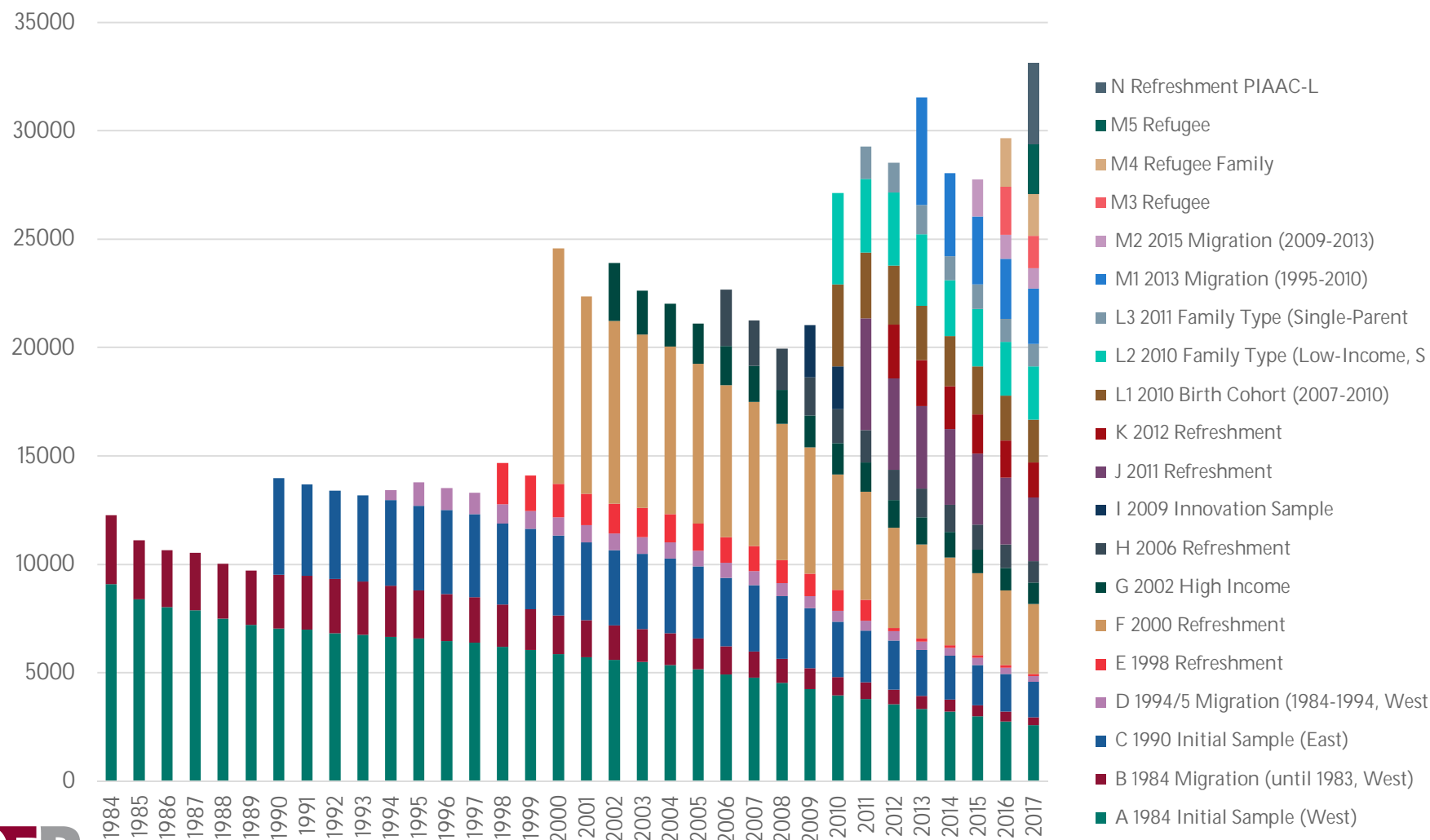
2b

Composition of SOEP Core



2b

Sample Development (Individuals) by Subsamples



Household Panel Study with information on individuals within the households

- Information on the household is gathered from the household head
- Personal interviews with all household members above the age of 18 (since 1984)
 - Since 2001: All individuals age 17 with retrospective questions on youth
 - Since 2014: Personal interviews with children living the household aged 11-12
 - Since 2016: Personal interviews with children living in the household aged 13-14
- Proxy information for younger children living in the household
- Proxy information on individuals who dropped out of the survey due to death or moving abroad
- Temporary dropouts will be re-interviewed with an additional special short questionnaire

Introduction to the SOEP - Part I

1. A Short Introduction to Panel Data
- 2. What is the SOEP?**
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

Multi-topic study with a multidisciplinary perspective, covering questions concerning:

- Demography and population
- Education and qualification
- Occupation and employment
- Income, taxes and social security
- Housing
- Health
- Attitudes, values and personality
- Subjective wellbeing (Life satisfaction).
- Migration and integration

+ Additional topics with multi-annual survey rhythms

2c

Topics with Multi-Annual Survey Rhythm*

	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Consumption and savings																														
Health																														
Globalization & trans-nationalisation																														
Grip strength																														
Reciprocity																														
Big Five																														
Risk aversion																														
Trust																														
Health (SF12 BMI)																														
Transport and energy consumption																														
Work Environment																														
Expectations about the future																														
Family and social networks																														
Time use and preferences																														
Labor market & subj. indicators																														
Training & occup. qualification																														
Household finance und wealth																														
Social security und retirement																														
Neighborhood																														

*Complete list of topics is available in the SOEPCompanion (2019):
Companion.soep.de

Subjective

&

Objective

Information are
gathered on a temporal dimension spanning



2c

Subjective and Objective Information

Subjective

&

Objective

Health and Illness

158. How would you describe your current health?

Very good ☐
 Good ☐
 Satisfactory ☐
 Poor ☐
 Bad ☐

159. How many hours do you sleep on average on a normal day during the working week?
 How many hours on a normal weekend day?
 Please give only whole hours.

Normal working day hours Normal weekend day hours

160. Has a doctor ever diagnosed you to have one or more of the following illnesses?

Sleep disorder ☐
 Diabetes ☐
 Asthma ☐
 Cardiac disease (also cardiac insufficiency, weak heart) ☐
 Cancer ☐
 Stroke ☐
 Migraine ☐
 High blood pressure ☐
 Depression ☐
 Dementia ☐
 Joint diseases (including arthritis, rheumatism) ☐
 Chronic back trouble ☐
 Burnout ☐
 Other illness ☐
 No illness diagnosed ☐

161. Do you have a health problem that limits you in normal everyday life?

Yes, severely ☐ ☐
 Yes, somewhat ☐ ☐
 No, not at all ☐

162. Have you had this health problem for more than half a year?

Yes ☐ No ☐

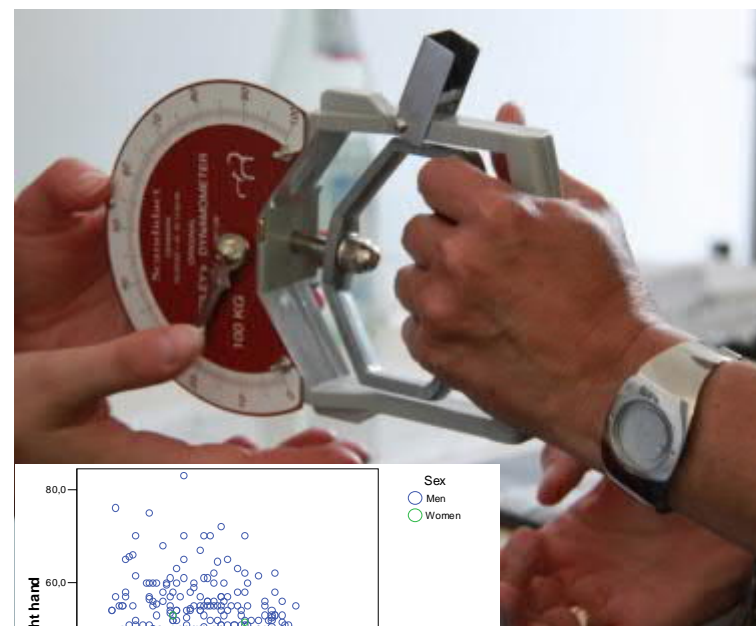
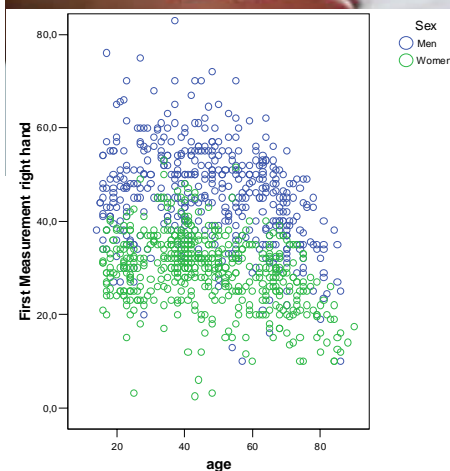


Foto: C. Kurka



Information regarding the present:

- current employment status or current life satisfaction

Information regarding the past:

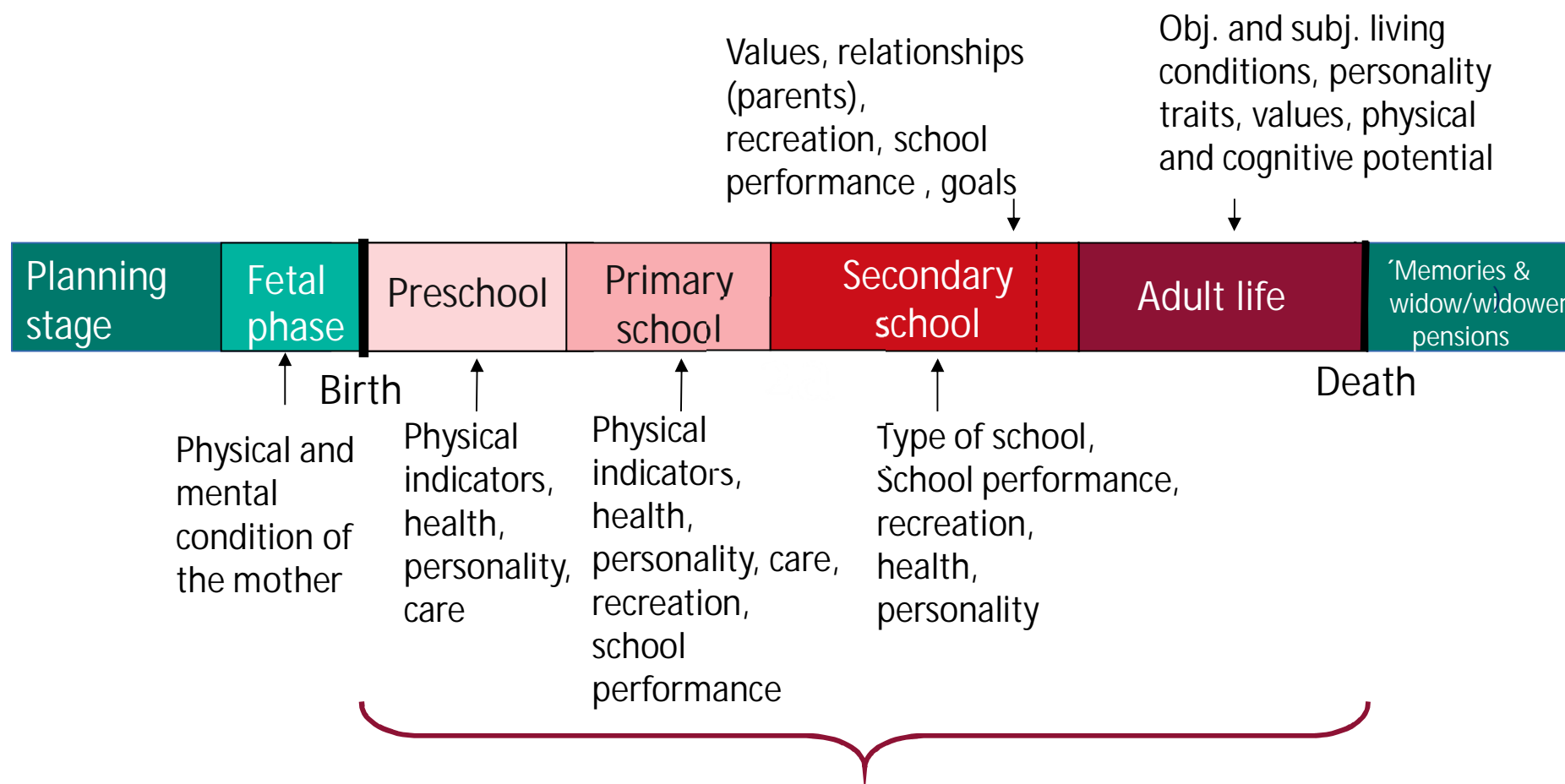
- Retrospective questions about events
 - How often have you changed jobs in the recent years?
- Retrospectively assessed history of events since age of 15
 - Employment or family status
- Monthly calendar of income and employment
 - Employment status from January to December of the past year
 - Monthly changes in education or family status since the last

Information regarding the future (expectations):

- Expected life satisfaction in 5 years
- Expectations regarding (re-)employment, retirement, expectations on schooling

2c

Life Course Perspective

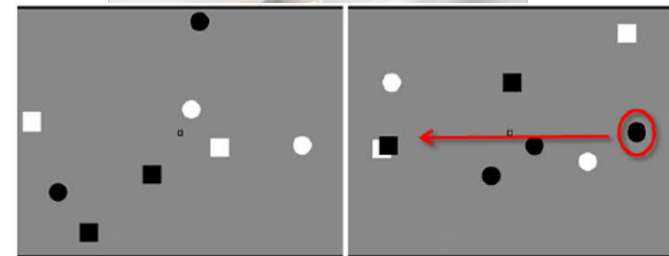


Information from "cradle to stretcher"
 → Ideally for all individuals in the household

Subjective and Objective Information

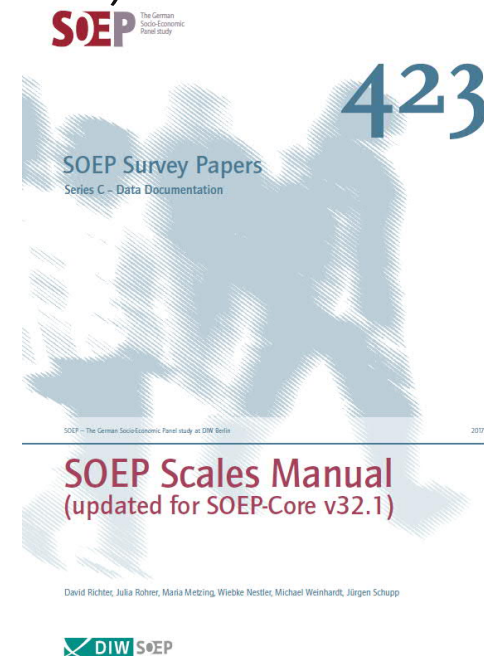
Extension of survey instruments besides standardized questionnaires:

- Objective measure of physical characteristics
- Instruments measuring cognitive potential
- Behavioral experiments
 - Charitable giving experiment
 - Meeting Chancellor Merkel Intervention
- Innovative modules in SOEP-IS:
 - Jelly-Bear Test
 - Saliva Collection for Epigenetics
 - Variation of the “invisible gorilla”
 - Use of Smartphone Apps for real-life data collection



Personality Traits

- Big Five Inventory (2005, 2009, 2013)
- Risk Aversion (2004, 2006, 2008, 2010, 2012, 2014, 2016)
- Locus of Control (1994, 1995, 1996, 2005, 2010, 2015)
- Reciprocity (2005, 2010, 2015)
- Impulsiveness & Patience (2008, 2013)
- Trust and Fairness (2003, 2008, 2013)
- Personal Networks (2006, 2011)
- Self-esteem (2010, 2015)
- Loneliness (2013)
- Anomy Indicators (1991, 1996, 2008)
- Life Goals (1990, 1992, 1995, 2004, 2008, 2012, 2016)



- 90 seconds test
- Part of the individual questionnaire
- Only for CAPI-respondents

Test	2006	2012	2016
Animal Naming Task	x	x [*]	x ^{**}
Symbol-Digit-Test	x	x	x
Vocabulary test		x	x

* only for those who answered already in 2006,

** only for those who answered already in 2012

Hier haben wir ein kleines Lösungsbeispiel:

\div)	+	⊥	⌈	∨	($\overline{\cdot}$	⊥
1	2	3	4	5	6	7	8	9

Aufgabe:

Zeichen:

) \div ⊥

Zahl?

? ? ?

Lösung:

Zeichen:

) \div ⊥

Zahl?

2 1 4

Zu dem Zeichen in Feld 1 ist "2" einzugeben.
Zu dem Zeichen in Feld 2 ist "1" einzugeben.
Zu dem Zeichen in Feld 3 ist "4" einzugeben.

Welches ist das richtige Wort?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
kurehen	krusienen	kasseren	kursieren	kustieren	

„Denksport Jugend“ (DJ) for 16-17 year olds

- Annually since 2006
- Incentivized intelligence-structure test (I-S-T 2000 R) (Amthauer et al. 2001)
 - ca. 30 minutes
 - 90% participation rate

Dimensions

- Verbal-cognitive potential (word analogies)
- Numeric-cognitive potential (numeric sequences)
- Figural-cognitive potential (matrices and pictures)
- Logic thinking (= sum index)

Matrix 1:

Matrix 2:

Matrix 3:

Matrix 4:

Beispiel:

Wald : Bäume = Wiese : ?

Gräser ... ☒ Heu ... ☐ Futter ... ☐ Grün ... ☐ Weide ... ☐

Hier sind Oberbegriffe und Unterbegriffe gegenüber gestellt. Für den „Wald“ sind „Bäume“, was für die „Wiese“ die „Gräser“ sind. Kreuzen Sie also Gräser an.

Ein weiteres Beispiel:

dunkel : hell = nass : ?

Regen ... ☐ Tag ... ☐ feucht ... ☐ Wind ... ☐ trocken ... ☒

Da „dunkel“ das Gegenteil von „hell“ ist, muss zu „nass“ auch das Gegenteil gefunden werden. Hier wird also **trocken** angekreuzt.

Beispiel 1: 6 + - : 2 + - : 3 = 5

Hier muss man die Zahlen 6 und 2 zusammenzählen, dann vom Zwischenergebnis 8 die Zahl 3 abziehen, um das vorgegebene Ergebnis 5 zu erhalten:

(6 + 2 = 8 ; 8 - 3 = 5)

Lösung: 6 + - : 2 + - : 3 = 5

Beispiel 2: 7 + - : 2 + - : 4 = 10

Hier sind die Zahlen 7 und 2 malzunehmen, das Zwischenergebnis ist 14, von dem muss man die Zahl 4 abziehen, um das Ergebnis 10 zu erhalten:

(7 · 2 = 14 ; 14 - 4 = 10)

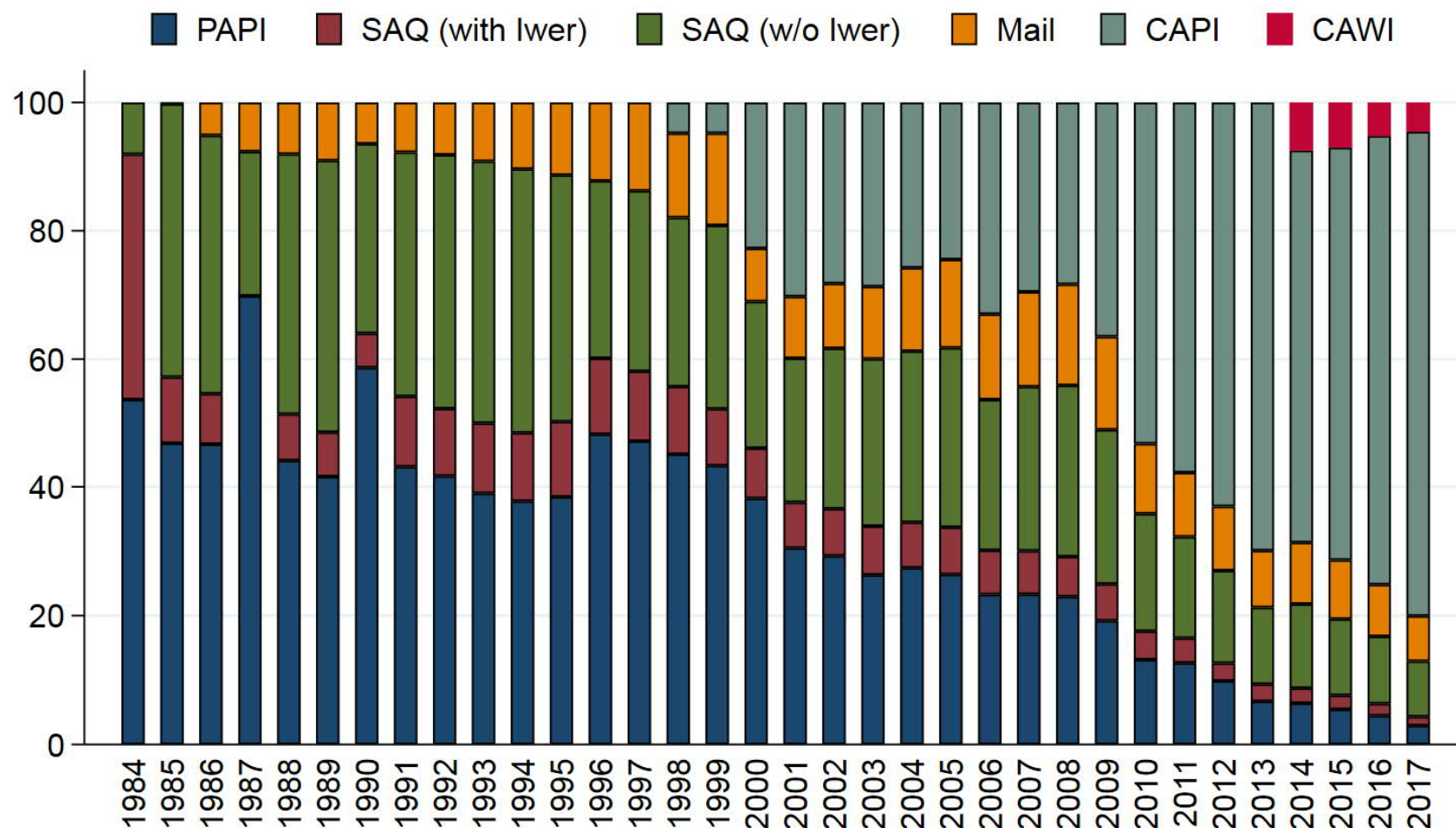
Lösung: 7 + - : 2 + - : 4 = 10

Introduction to the SOEP - Part I


1. A Short Introduction to Panel Data
2. What is the SOEP?
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

3


Development of the Survey Method



A A A
Intranet Deutsch Sitemap Newsletter Contact Legal Details Data Protection DIW Berlin



About SOEP
Research Data Center SOEP
News and Events
Publications with SOEP data



Documentation


Questionnaires & Fieldwork Documents

[Research Data Center SOEP](#) > [Data and Documentation](#) > [SOEP-Core](#) > [Questionnaires & Fieldwork Documents](#) >

Data and Documentation
SOEP-Core
Questionnaires & Fieldwork Documents
Generated Variables
Changes in the Dataset
Versions / DOI
Survey Methods
SOEP-IS
SOEP Cross-country
Linking Possibilities
SOEP in the College


All printed questionnaires are available in quotable form as → [SOEP Survey Papers](#).

2017	2016	2015	2014	2013
2012	2011	2010	2009	2008
2007	2006	2005	2004	2003
2002	2001	2000	1999	1998
1997	1996	1995	1994	1993
1992	1991	1990	1989	1988
1987	1986	1985	1984	Additional



Dieses Dokument auf Deutsch

Your contact person

→ **Florian Griese**



Sozio-ökonomisches Panel
DIW Berlin
Mohrenstraße 58
10117 Berlin
Tel.: +49 30 89789-359
Fax: +49 30 89789-115

 E-mail


SOEP hotline

SOEP

37

SOEPCampus

https://www.diw.de/en/diw_02.c.238114.en/questionnaires_fieldwork_documents.html



Questionnaires

2017

■ Samples A-L3

Standard questionnaires

- ↓ **Individual Questionnaire** | PDF, 1 MB (English and German)
- ↓ **Household Questionnaire** | PDF, 0.9 MB (English and German)

Biographical questionnaires

- ↓ **Youth Questionnaire** | PDF, 0.64 MB (German only)
- ↓ **Supplementary Biography Questionnaire** | PDF, 1.13 MB (German only)
- ↓ **Catch-Up Individual Questionnaire** | PDF, 515.85 KB (German only)

Mother-child instruments

- ↓ **Mother-Child-Questionnaire (Newborns)** | PDF, 0.65 MB (German only)
- ↓ **Mother-Child-Questionnaire (2-3 year olds)** | PDF, 0.7 MB (German only)
- ↓ **Mother-Child-Questionnaire (5-6 year olds)** | PDF, 0.6 MB (German only)
- ↓ **Parents Questionnaire (7-8 year olds)** | PDF, 0.58 MB (German only)
- ↓ **Mother-Child-Questionnaire (9-10 year olds)** | PDF, 0.7 MB (German only)

For a translation of the Mother-Child instruments, please refer to the generated questionnaires with reference to variables (see below)

Pupils/ early youth

- ↓ **Pre-Teen (11-12 year olds)** | PDF, 0.87 MB (German only)
- ↓ **Early Youth (13-14 year olds)** | PDF, 0.87 MB (German only)

Deceased person

- ↓ **The deceased Individual** | PDF, 446.06 KB (German only)

Target-group-specific
questionnaires
(migrants & refugees)

■ Samples M1-M2

- ↗ **Household Questionnaire** (English and German)
- ↗ **Individual with Biography** (English and German)
- ↗ **Individual Questionnaire (continued sample)** (English and German)
- ↗ **Youth Questionnaire** (German only)

Metadata-based questionnaires with reference to variables

With reference to variables

■ Samples A-L3

- ↓ **Individual (PAPI)** | PDF, 367.12 KB
- ↓ **Household (PAPI)** | PDF, 242.01 KB
- ↓ **Catch-Up Individual** | PDF, 173.77 KB
- ↓ **Biography** | PDF, 251.15 KB

- Standard instruments (annually)
 - Household questionnaire (household head)
 - Time-variant household information
 - Individual questionnaire (every person in household aged > 17/18)
 - Time-variant individual information
- Biographical instruments (once, during first interview)
 - Supplementary Biography Questionnaire (aged 18 and above, since 1989)
 - retrospective questions regarding important life events and biographical data (e.g. place of birth, nationality, nationality of parents)

3

Target-group-specific Instruments

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Starting with birth-cohort
Newborns																		2002/2003
2-3 years (Muki2)																		2002
5-6 years (Muki3)																		2002
7-8 years (FID)																		2002
9-10 years (Muki5)																		2002
11-12 years* (pre-teen)																		2002
13-14 years* (early youth)																		2002
17 years** (youth)																		1983

* Personally interviewed, with parental consent

** personally interviewed


- „Luecke“ (from 2006 onwards)
 - Temporary unit non-response
- The deceased person (from 2009 onwards)
 - Family relations, cause of death, testament
- Your life abroad (2006, 2007, 2012, 2013, 2014)
 - Reasons for moving abroad, reasons to return
- IAB-SOEP-migration sample (since 2013 and 2015) (M1-M2)
- IAB-BAMF-SOEP-survey of refugees (since 2016 and 2017) (M3-M5)

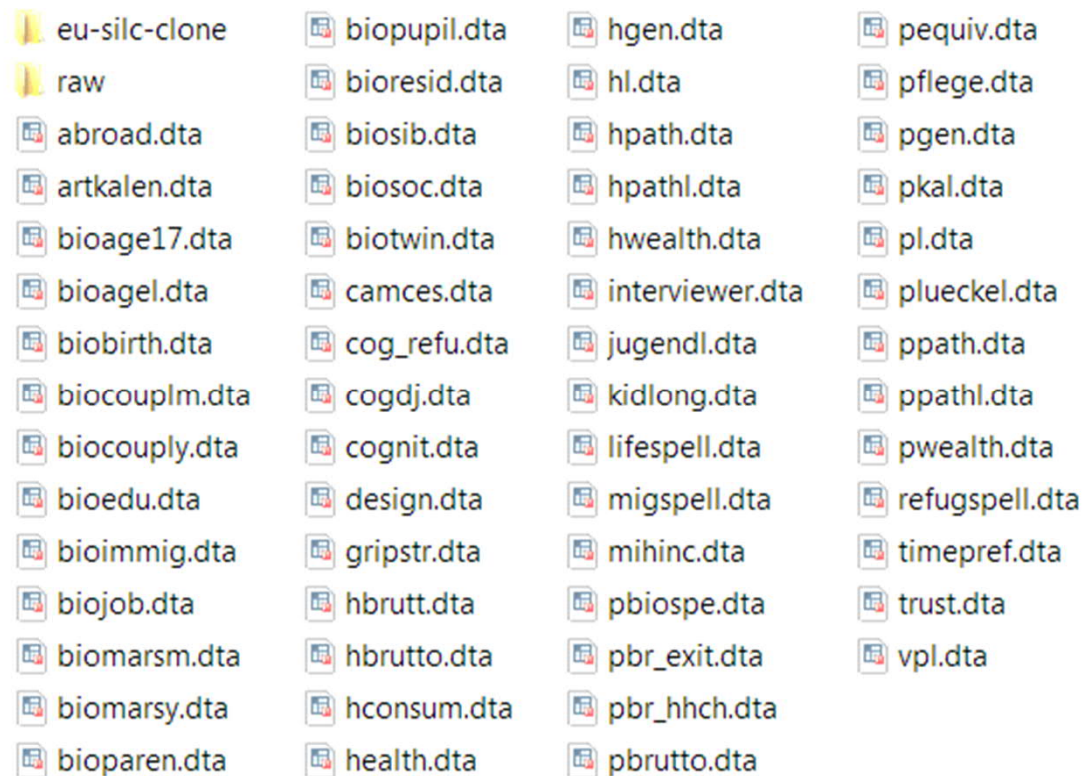
Introduction to the SOEP - Part I

1. A Short Introduction to Panel Data
2. What is the SOEP?
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

4

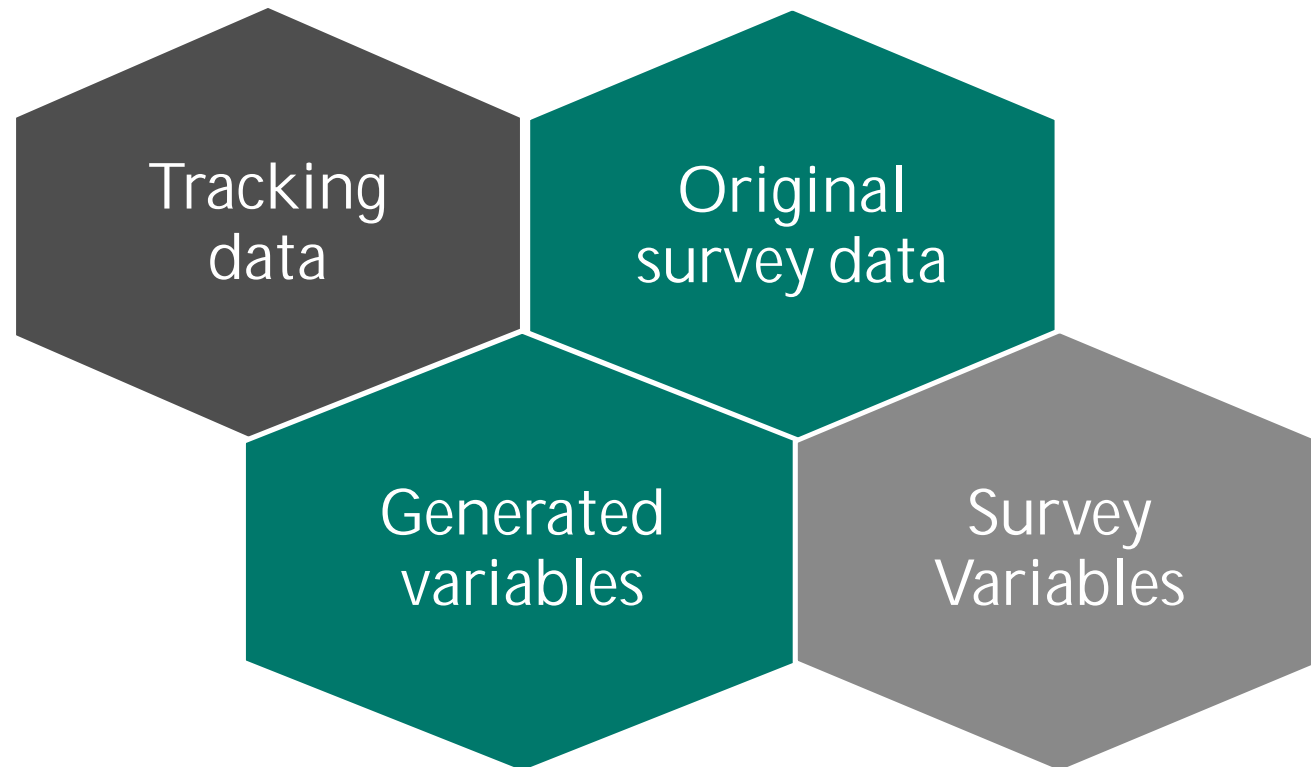
When you get the data


 Version number: corresponds to survey waves



4

We provide 4 Types of Data



4

Naming Conventions for datasets

General naming conventions for datasets in SOEPcore

The SOEP provides „speaking“ names for datasets, which contain hints with regards to:

Survey unit

Type of data

Underlying survey Instruments

Wave (for Wide Format in raw)

4

Naming Conventions for datasets

General naming conventions for datasets in SOEPCore

Subject of inquiry

p: person
h: household
kid: children
jugend: youth
vp: deceased person

Type of data

Original:
Generated: *gen, bio*, *kalen, equiv,
*spell, kal,
Tracking: *brutto, *path, pbr_exit,
Survey: *hrf, design, exit, pbr_hhch,

Underlying survey instruments

luecke, abroad,
cog*, gripstr, timepref, trust, ...

4

Biography Data Sets

- BIOAGE17 information from youth questionnaires (age 17)
- BIOAGEL information from mother-child questionnaires (age 0 – 10)
- BIOPUPIL information from pupil and pre-teen questionnaires (age 12-14)
- BIOEDU educational biographies (since 2010)
- BIOIMMIG migrant information
- BIOJOB first job
- BIOBIRTH births biographies for mothers
- BIOPAREN information on mothers and fathers
- BIOSOC socialization info– life course
- BIORESID residence biographies
- BIOTWIN information on twins, triplets and quadruplets (since 2006)
- BIOSIB sibling information (since 2010)
- BIOCOUPL(M/Y) Relationship Biographies (month/year)
- BIOMARS(M/Y) Marriage Biographies (month/year)

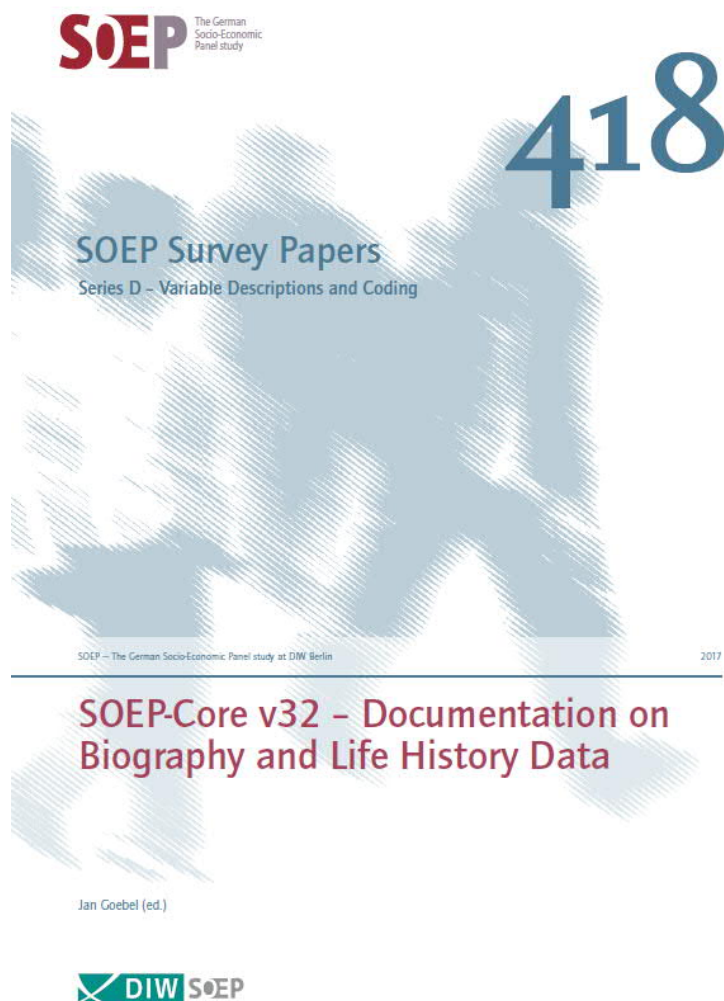
*Some are updated
annually,
others contain time-
constant information*

Cumulative data sets

- PWEALTH assets person (every 5 years since 2002)
- HWEALTH assets households (every 5 years since 2002)
- HEALTH health person (biannually since 2002)
- PFLEGE in-need-of-care person (annually 1985 to present)
- GRIPSTR gripping force test (biannually since 2006)

Spell data sets

- PBIOSPE employment biography– 15+
- ARTKALEN labor status calendar
- EINKALEN income calendar
- BIOMARS(Y/M) marriage events (month)
- BIOCOUPL(Y/M) marriage events (year)
- SOZKALEN social welfare calendar (household level)
- REFUGSPELL information on flight history and reasons
- MIGSPELL information on migration history and reasons



4

Conventions for variable names in SOEPcore

1st digit: **D**ataset-identifier

P = Person, H = Household,
L = Biography, J = Youth

3rd digit: **T**opic-identifier

a = demography and population
b = work and employment
c = income, taxes and social security
d = family and social networks
e = health and care
f = habitation, equipment and
services of private households
g = education and qualification
h = attitudes, values and personality
i = time use and ecological conduct
j = integration, migration and
transnationalization
k = survey methodology

2nd digit: **V**ariable-identifier

L = numeric variable,
A = alphanumeric variable
C = numeric variable containing
original information for recoded
variables

4th-7th digit: run-on number

Examples:

ple0010	person	numeric	health & care	0010	(birthyear)
pli0044	person	numeric	time use	0044	(time spent on childcare weekday)

4

The path files and the netto variable

ppath

contains basic information on all individuals, who ever lived in a SOEP household including ...

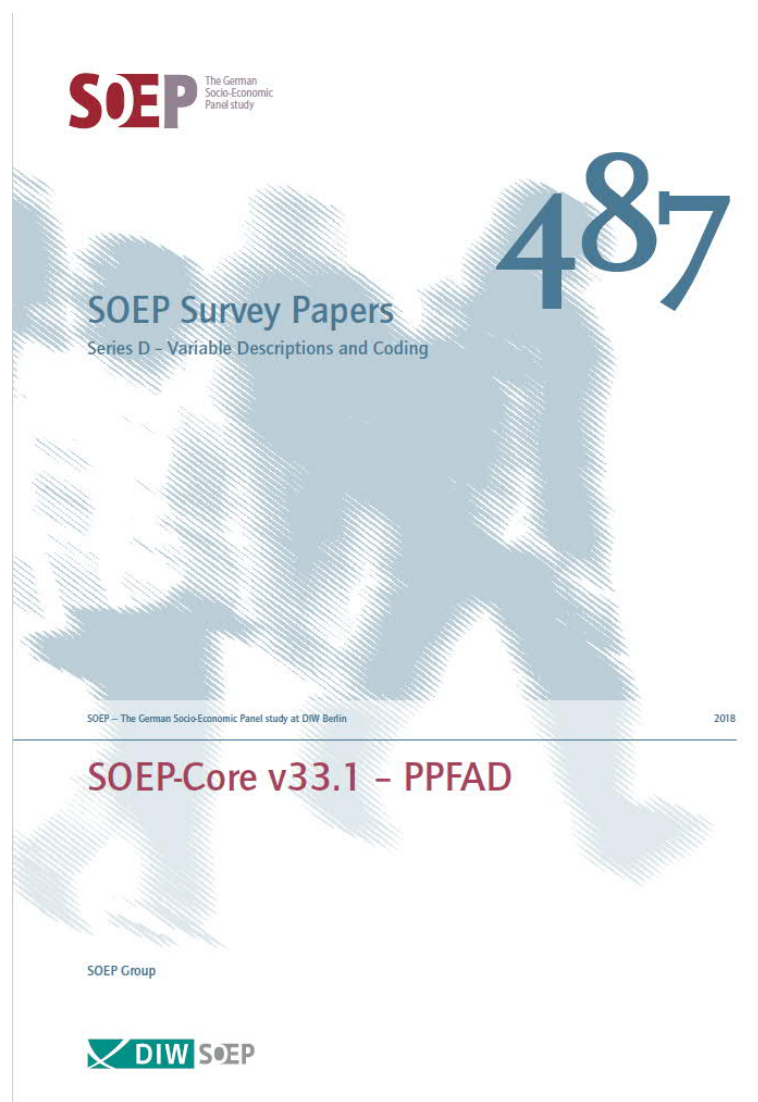
- identifiers (`pid`, `hid`, `cid`)
- sex, year & month of birth, `migback`
- sample (`psample`)
- survey status for each year (`netto`)
- household type (`pop`)
- sample region (`psamp`)
- ...

*we recommend using
the netto variable to
identify your study
population*

hpath

contains time consistent and time variant information on household level including

- current and root household number (`hid`, `cid`)
- sample regions (`psamp`)
- response status
- ...



Name	Label
cid	Case-ID, Original Household Number
pid	Never Changing Person ID
sex	Gender
gebjahr	Birth Year, 4-digit
eintritt	Year First Contacted, Netto=10-99
erstbefr	Year First Surveyed, Netto=10-99
austritt	Year Of Last Contact, Netto=10-99
psample	Sample Member
letztbef	Year Of Last Survey, Netto=10-99
todjahr	Year Died, 4 Digits
todinfo	Year Died, Information Source
immiyear	Year Moved to Germany
germborn	Born in Germany
corigin	Country Born In
gebmonat	Month Of Birth
gebmoval	Month Of Birth, Data Source
loc1989	Where did you live in 1989?
migback	Migration background
miginfo	Migback: Quality of information
germborninfo	Germborn: Quality of information
corigininfo	Corigin: Quality of information
immiyearinfo	Immiyear: Quality of information
sexor	Sexual Orientation
sexorinfo	Sexual Orientation:Source of information
locinfo	Loc1989: Source / Quality of information
birthregion	Birth place: German Federal Land
arefback	Refugee Experience
arefinfo	arefback: Source of Information

4

Current Wave Survey Status (netto)

\$netto - Current Wave Survey Status [generic]

respondents	10	[10] Interviewee With Successful Interview (_P)	11957
	12	[12] Individual Questionnaire And Person Biography	0
	13	[13] Individual Questionnaire And Youth Biography	0
	14	[14] Individual Questionnaire And Other Questionnaires	0
	15	[15] Individual Questionnaire And Experiments, Test	0
	16	[16] Individual Questionnaire, First Time Surveyed, 17 Years	280
	17	[17] Youth Biography First Time Surveyed, 17 Years	0
	18	[18] Individual Questionnaire And Child und 17 Years	8
	19	[19] Individual Questionnaire Without Household Interview	0
children	20	[20] Children in Successfully Interviewed Households (_Kind)	3928
	21	[21] Children With Mother-Child Questionnaire_I, 0-1 Year	0
	22	[22] Children With Mother-Child Questionnaire_II, 2-3 Year	0
	23	[23] Children With Mother-Child Questionnaire_III, 5-6 Year	0
	24	[24] Parents with children 7-8 year	0
	25	[25] Parents with children 9-10 year	0
Etc.	...	(13 rows omitted)	79
	89	[89] Repatriate - (was drop out [90])	0
	90	[90] Individual Dropouts PBR_EXIT	0
	91	[91] Moved abroad	0
	92	[92] Moved abroad (abroad)	0
	93	[93] Moved abroad (exit)	0
	94	[94] Person Gap with advices	0

4

Conventions for Missing Values in SOEP

Code	Meaning
-1	No reply
-2	Does not apply
-3	Not valid
-4	Invalid multiple reply
-5	Not available version of questionnaire
-6	Questionnaire with different filtering
-8	Question not part of questionnaire

Note: In SOEP no SYSTEM MISINGS („.“ in Stata)

Combining information from different datasets can be relevant for various types of research questions.

Examples:

- How does the birth of a third child affect the household income?
 - Household income from hgen or pequiv – information in births from biobirth
- What is the correlation between length of unemployment, life satisfaction and trust in democratic institutions?
 - Employment spells from artkalen – life satisfaction and trust in democratic institutions from p

If the relevant information is part of different datasets (and different levels of inquiry), they need to be merged via **identifiers**.

To combine information from **different datasets and levels**

- PID person, invariant over time
- HID current housold ID, variant over time
- CID root household ID, constant over time
- SPELLNR spell number, in spell data
- INTNR number of interviewer in data set INTVIEW
- SYEAR survey year
- Regional identifiers e.g. federal states, official district codes

For other questions, it might be necessary to combine information from different individuals.

Examples:

- How does the divorce of a child affect mental health of the parents?
- How does entering unemployment affect the grades of the children?

Information from different persons can be combined via **pointers**

- Certain datasets contain pointers to partners, parents or children (in addition to the identifiers for the anchor person)

How to combine the data depends on the question, e.g.:

- Analysis from the child's perspective: Does the employment status of the mother affect the child's education success?
→ persnr = child
- Analysis from the mother's/parents' perspective: Does the birth of a child affect labor supply from women?
→ persnr = (potential) mother
- Analysis from women's perspective: Does the partner affect the labor supply of women?
→ persnr = woman with partner

- Survey Unit which determines the dependent variable is the anchor person which keeps HID and PID
- Pointers are used to merge information on other individuals to the anchor person

4

Pointers in the SOEP

Anchor Person	Merged Person	Dataset	Variable
Person	Partner	pgen ppfadl	pgpartz parid
Mother	Child	biobirth	kidnrXX
Child < 17	Mother	kidl	k_pmum
Child < 17	Partner of the mother	kidl	k_pmump
Child > 17 / Adults	Mother	bioparen	mnr
Child > 17 / Adults	Partner of the mother	bioparen	fnr
Person	Sibling	biosib	sibpnrXX
Person	Twin, Triplet, Quadruplet	biotwin	pnrtwin /pnrtrip /pnrquad

4

Combining information with Pointers – An Example

How is mothers' labor market participation affected by income changes of their partners?

1. Create a dataset of females with children containing the pointer to partners:

- DepVar: Mothers' labour market participation → PID with mothers

```
use pid hid syear netto parid using ppath1.dta
merge 1:1 pid hid syear using biobirth.dta , keepus(kidnr01)
keep(3) nogen
merge 1:1 pid hid syear using pgen.dta , keepus(pglfs pgstib)
keep (1 3) nogen
save $temp\mothers.dta
```

2. Prepare dataset for "partners" containing the relevant information and rename PID for the partner dataset

```
use pid hid syear netto pglabgro using pgen.dta
rename pid parid
save $temp\partners.dta
```

3. Merge datasets keeping only merged cases

```
use $temp\mothers.dta, clear
merge 1:1 parid syear using $temp\partners.dta ,
keepus(pglabgro) keep (3)
```

4

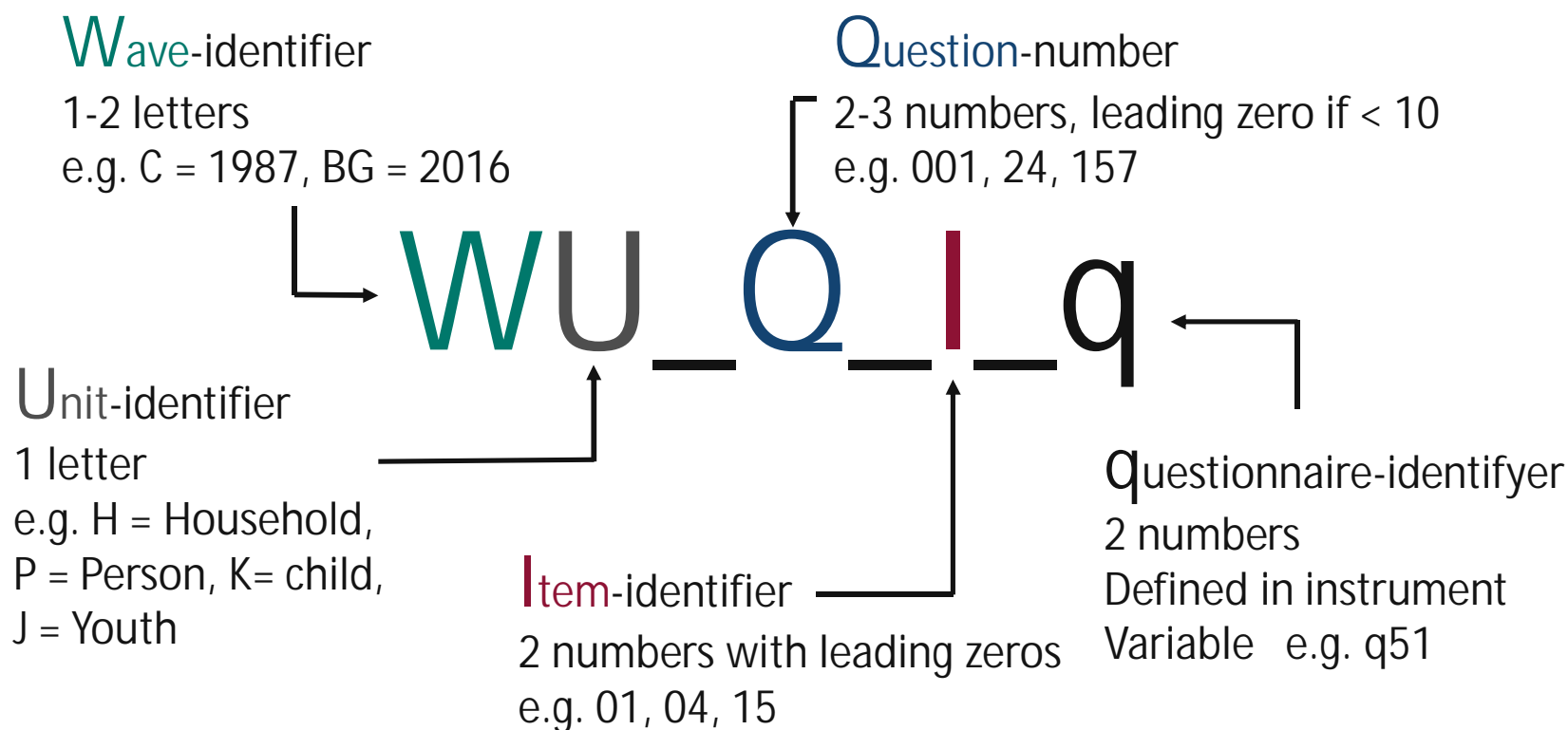
The raw folder

- Equivalent information but in **wave-specific data sets**
- \$ = wave token

wave token	year	wave nr.
a	1984	1
b	1985	2
...
bg	2016	33
bh	2017	34

4

Variable names in wide – v.34 onwards



Examples:

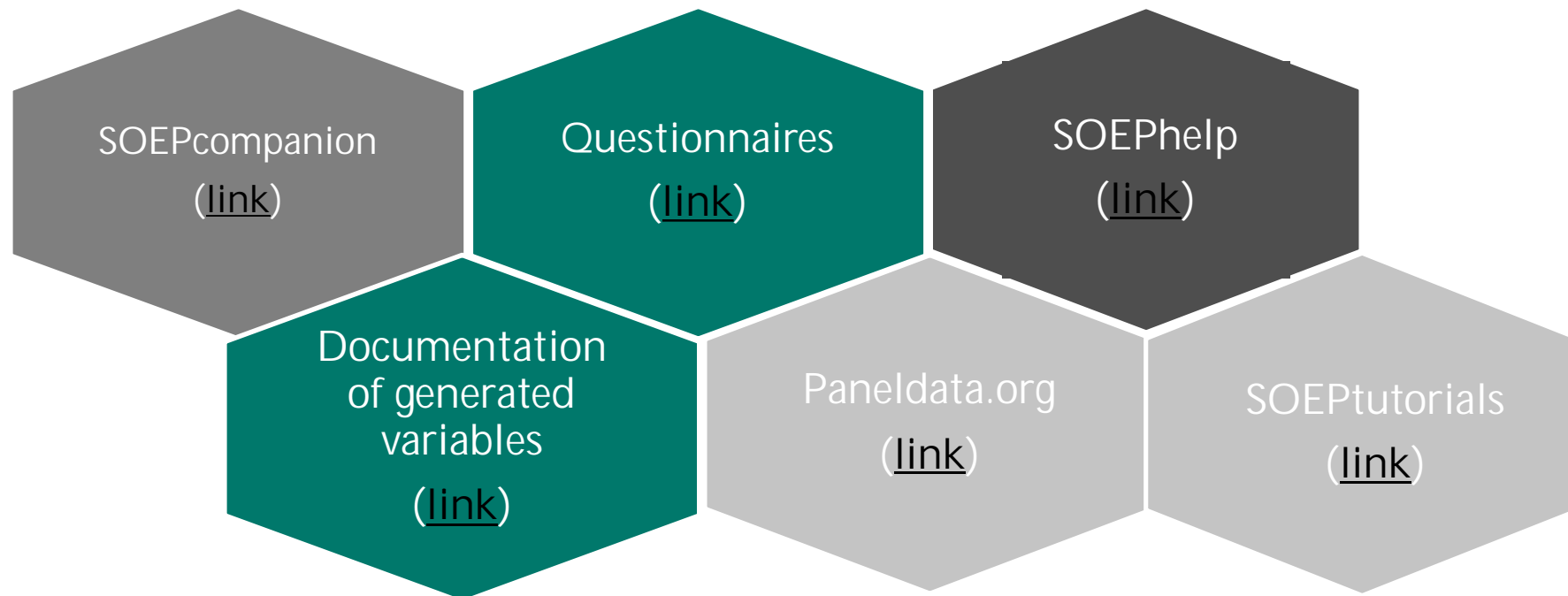
AP_06	wave A (1984)	person	question 6		
BEK_78_10	wave BE (2014)	child	question 78	item 10	
BHP_109_01_q75	wave BH (2017)	person	question 109	item 1	questionnaire 57 *

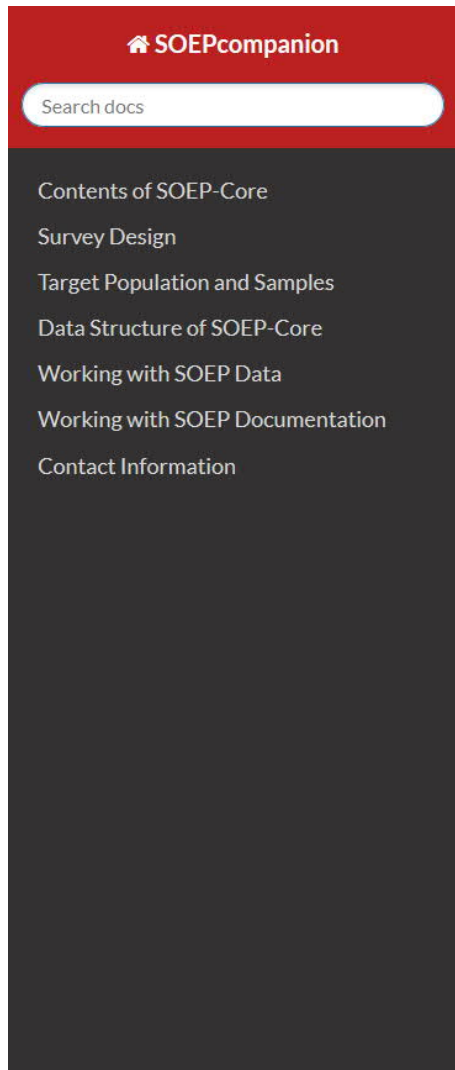
* Personal Biography Q.(M3-M4 Re-Interviewed; CAPI)

Introduction to the SOEP - Part I

1. A Short Introduction to Panel Data
2. What is the SOEP?
 - a) Basic Information on SOEPcore
 - b) Composition and Sample Development of SOEPcore
 - c) Topics covered in SOEPcore
3. Survey Methods and Instruments
4. Data Structure
5. Documentation and Online Services

Important sources for working with the SOEP





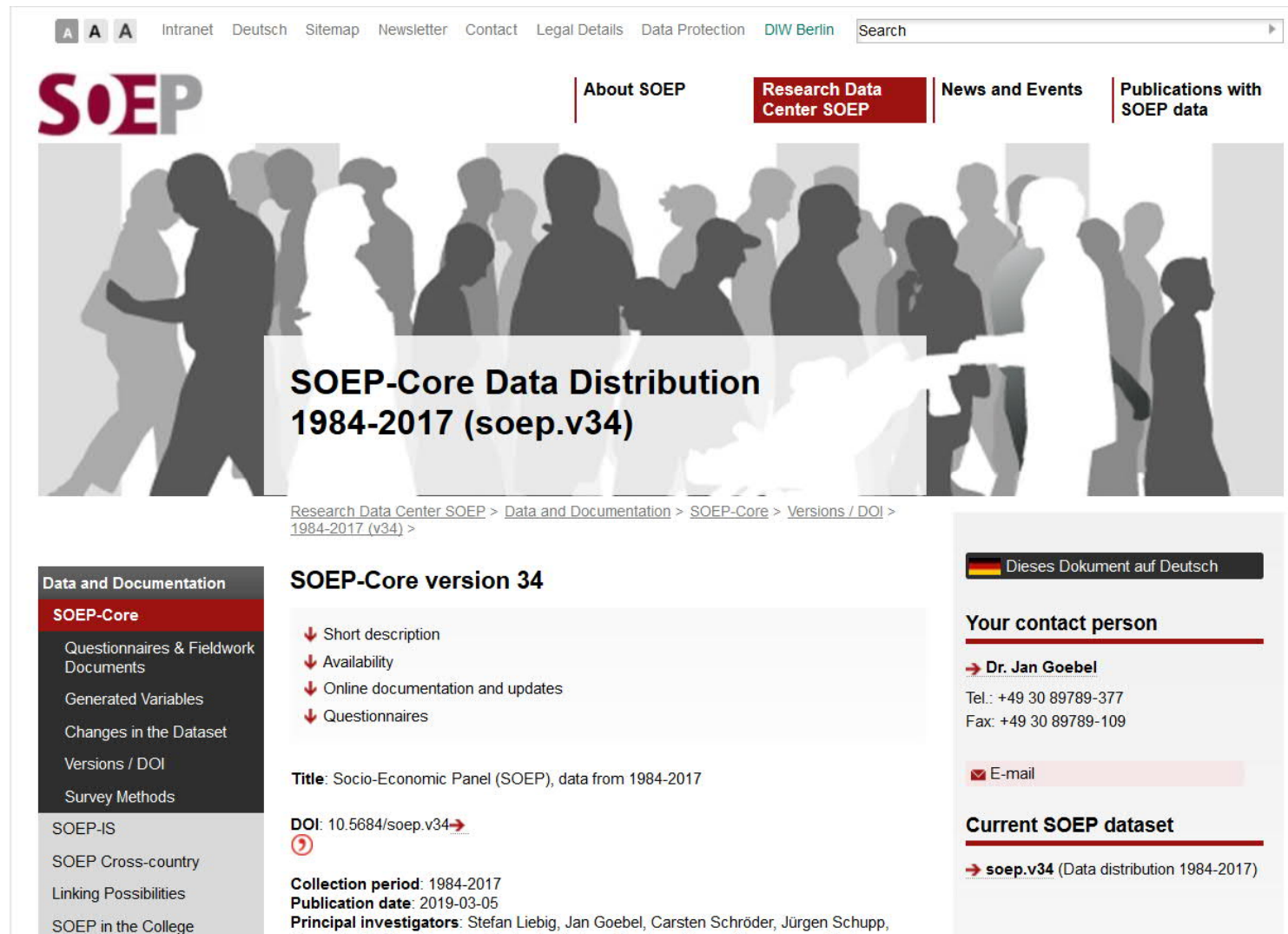
[Docs](#) » Welcome to the SOEPcompanion!

[View page source](#)

Welcome to the SOEPcompanion!

SOEP-Core is THE centerpiece of the wide-ranging representative longitudinal study of private households located at the German Institute for Economic Research, DIW Berlin. SOEP-Core was started in 1984 and in 1990—just after German reunification—we enlarged the area covered by the SOEP study by adding a representative sample from East Germany. This feature makes the SOEP unique among other household panel surveys worldwide. Each year since 1984, around 15,000 households and about 30,000 individuals have been surveyed by the SOEP's fieldwork organization, Kantar Public Germany. The data provide information on every member of every household taking part in the survey. Respondents include Germans living in the states of both the former East and West Germany, foreign citizens residing in Germany, recent immigrants, and a new sample of refugees added in 2016. Some of the many topics include household composition, education, occupational biographies, employment, earnings, health, and satisfaction indicators.

The SOEPcompanion provides data users with the content of the currently released SOEP-Core data (v34) and will introduce into different versions of SOEP-Core data structures. It will also provide data users with a lot of retrievals in Stata as well as multiple instructions on how to use our various documentation services. This collection of information in the SOEPcompanion is intended to become a yearly updated main reference guide, and a practical companion in basic understanding and implementation of the SOEP.



SOEP

[Intranet](#) [Deutsch](#) [Sitemap](#) [Newsletter](#) [Contact](#) [Legal Details](#) [Data Protection](#) [DIW Berlin](#)

[About SOEP](#) [Research Data Center SOEP](#) [News and Events](#) [Publications with SOEP data](#)

SOEP-Core Data Distribution 1984-2017 (soep.v34)

[Research Data Center SOEP](#) > [Data and Documentation](#) > [SOEP-Core](#) > [Versions / DOI](#) > [1984-2017 \(v34\)](#) >

Data and Documentation

- SOEP-Core**
 - Questionnaires & Fieldwork Documents
 - Generated Variables
 - Changes in the Dataset
 - Versions / DOI
 - Survey Methods
- SOEP-IS
- SOEP Cross-country
- Linking Possibilities
- SOEP in the College

SOEP-Core version 34

- ↓ Short description
- ↓ Availability
- ↓ Online documentation and updates
- ↓ Questionnaires

Title: Socio-Economic Panel (SOEP), data from 1984-2017

DOI: 10.5684/soep.v34 →

Collection period: 1984-2017
Publication date: 2019-03-05
Principal investigators: Stefan Liebig, Jan Goebel, Carsten Schröder, Jürgen Schupp,

[Dieses Dokument auf Deutsch](#)

Your contact person

→ **Dr. Jan Goebel**
 Tel.: +49 30 89789-377
 Fax: +49 30 89789-109

☒ E-mail

Current SOEP dataset

→ **soep.v34** (Data distribution 1984-2017)

A A A Intranet Deutsch Sitemap Newsletter Contact Legal Details Data Protection DIW Berlin Search

SOEP

About SOEP Research Data Center SOEP News and Events Publications with SOEP data

Documentation

Questionnaires & Fieldwork Documents

[Research Data Center SOEP](#) > [Data and Documentation](#) > [SOEP-Core](#) > [Questionnaires & Fieldwork Documents](#) >

Data and Documentation

- SOEP-Core**
 - Questionnaires & Fieldwork Documents**
 - Generated Variables
 - Changes in the Dataset
 - Versions / DOI
 - Survey Methods
- SOEP-IS
- SOEP Cross-country
- Linking Possibilities
- SOEP in the College


All printed questionnaires are available in quotable form as → **SOEP Survey Papers**.

2017	2016	2015	2014	2013
2012	2011	2010	2009	2008
2007	2006	2005	2004	2003
2002	2001	2000	1999	1998
1997	1996	1995	1994	1993
1992	1991	1990	1989	1988
1987	1986	1985	1984	Additional

Dieses Dokument auf Deutsch

Your contact person

→ **Florian Gries**

 Sozio-oekonomisches Panel
DIW Berlin
Mohrenstraße 58
10117 Berlin
Tel.: +49 30 89789-359
Fax: +49 30 89789-115

☒ E-mail

SOEPHotline



[Research Data Center SOEP](#) > [Data and Documentation](#) > [SOEP-Core](#) > [Generated Variables](#) >

Data and Documentation
SOEP-Core
Questionnaires & Fieldwork Documents
Generated Variables
Changes in the Dataset
Versions / DOI
Survey Methods
SOEP-IS
SOEP Cross-country
Linking Possibilities
SOEP in the College Classroom
SOEP as Reference Data
Data Access
Services
Getting Started

- ↓ Generated variables (wave specific v33.1)
- ↓ Generated variables not updated
- ↓ Previous versions

Generated variables (wave specific v33.1)

SOEP Survey Paper 483 (Series D):

↓ SOEP-Core v33.1 – \$PGEN | PDF, 258.92 KB

(Documentation of Person-related Status and Generated Variables)

SOEP Survey Paper 487 (Series D):

↓ SOEP-Core v33.1 – PPFAD | PDF, 174.07 KB

(Documentation of the Person-related Meta-dataset)

SOEP Survey Paper 482 (Series D):


↓ SOEP-Core v33.1 – \$HGEN | PDF, 199.96 KB

(Documentation of Household-related Status and Generated Variables)

SOEP Survey Paper 484 (Series D):

↓ SOEP-Core v33.1 – HPFAD | PDF, 107.66 KB

(Documentation of the Household-related Meta-dataset)

 Dieses Dokument auf Deutsch

SOEPHotline



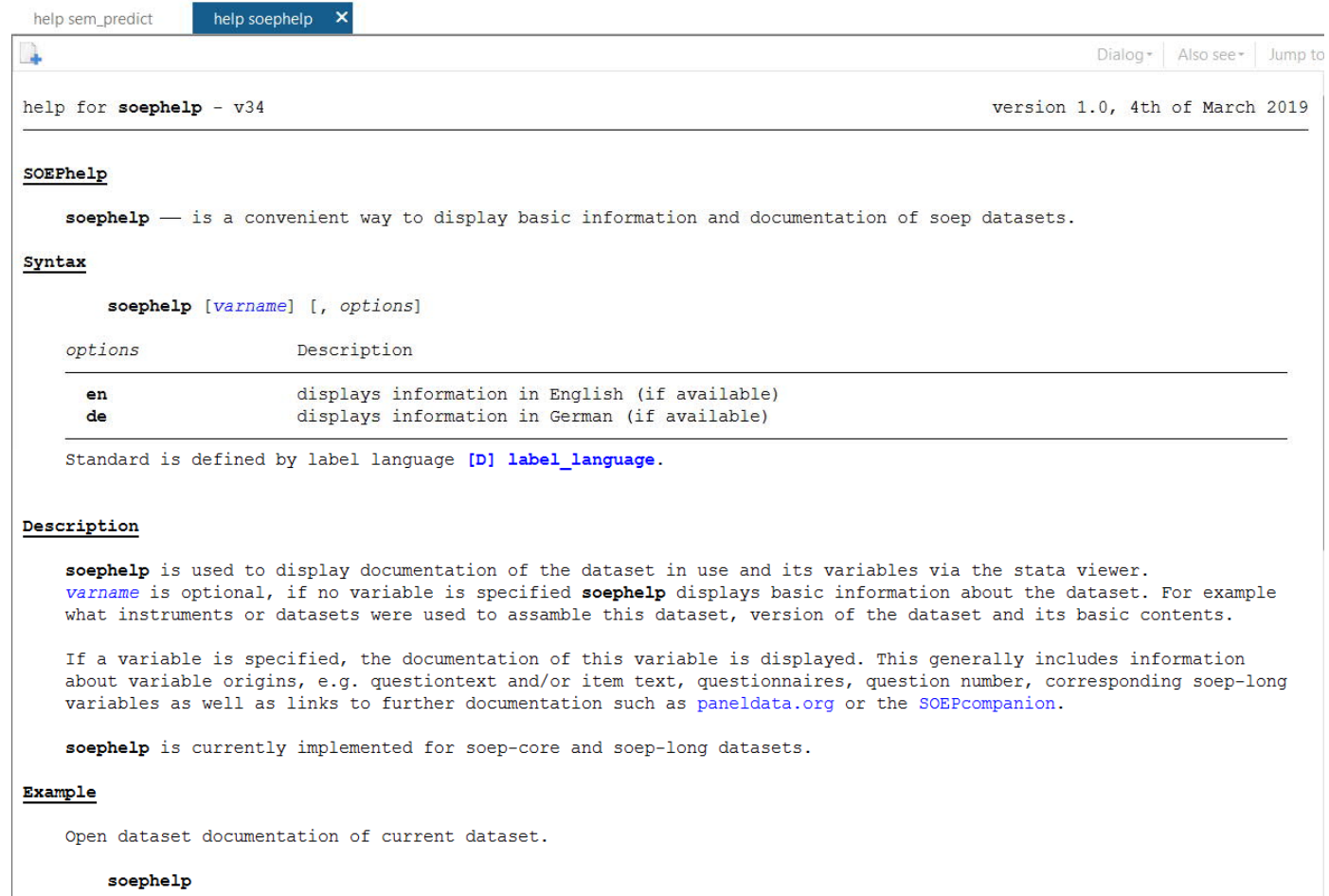
Contact:
Alisa Fränkel
Philipp Kaminsky

Tel.: +49 30 89789-292
Fax.: +49 30 89789-109
✉ soepmail@diw.de

SOEPcompanion

The ideal guide for working with SOEP-Core

```
net install soephelp, replace from(http://companion.soep.de/SOEPhelp/)
```



```
help for soephelp - v34 version 1.0, 4th of March 2019
```

SOEPhelp

soephelp — is a convenient way to display basic information and documentation of soep datasets.

Syntax

```
soephelp [varname] [, options]
```

<i>options</i>	Description
en	displays information in English (if available)
de	displays information in German (if available)

Standard is defined by label language [\[D\] label_language](#).

Description

soephelp is used to display documentation of the dataset in use and its variables via the stata viewer. [varname](#) is optional, if no variable is specified **soephelp** displays basic information about the dataset. For example what instruments or datasets were used to assemble this dataset, version of the dataset and its basic contents.

If a variable is specified, the documentation of this variable is displayed. This generally includes information about variable origins, e.g. questiontext and/or item text, questionnaires, question number, corresponding soep-long variables as well as links to further documentation such as paneldata.org or the SOEPcompanion.

soephelp is currently implemented for soep-core and soep-long datasets.

Example

Open dataset documentation of current dataset.

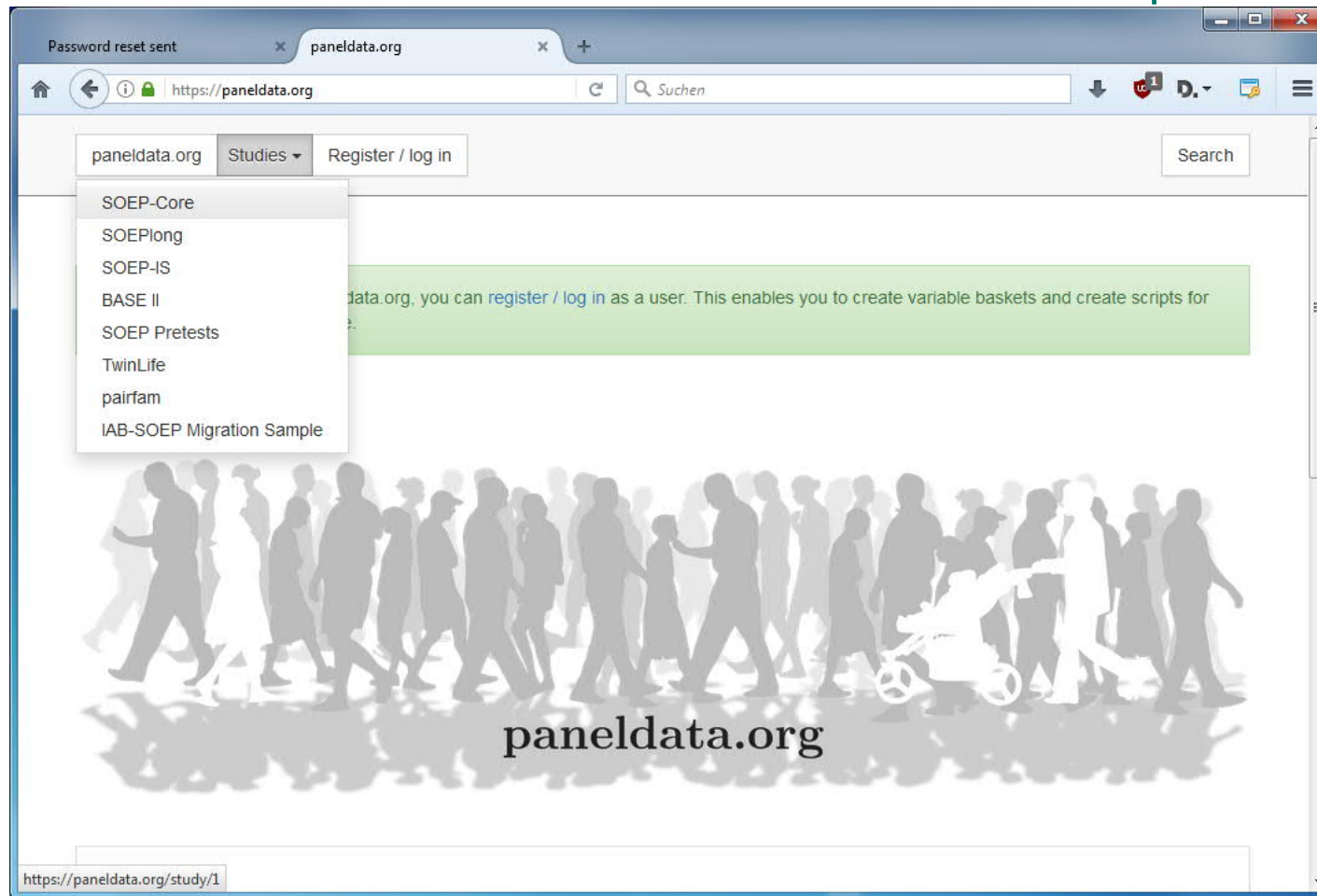
```
soephelp
```


In our media-center and on youtube

The screenshot shows a YouTube search results page for 'soep tutorials'. The search bar at the top contains 'soep tutorials'. The results are filtered by 'DIW SOEP'. The first four results are listed below:

Video Title	Views / Time	Description
SOEPTutorials: Data Structure and Naming Conventions	SOEP Studie • 9 Aufrufe • vor 2 Wochen 8:04	This tutorial introduces the general structure of the Socio-Economic Panel Study and explains the naming conventions for ...
SOEPTutorials: Identifiers and Pointers	SOEP Studie • 23 Aufrufe • vor 2 Wochen 4:28	This tutorial introduces the different identifiers used for individuals, households, and interviewers, and explains how you can use ...
SOEPTutorials: Panel Composition and Sample Development	SOEP Studie • 24 Aufrufe • vor 2 Wochen 5:40	This tutorial explains who we survey each year. It illustrates the household-centered survey design and the methodologies used to ...
SOEPTutorials: Questionnaires	SOEP Studie • 5 Aufrufe • vor 2 Wochen	This tutorial introduces the different questionnaires currently used in the SOEP survey. It illustrates how we design the ...

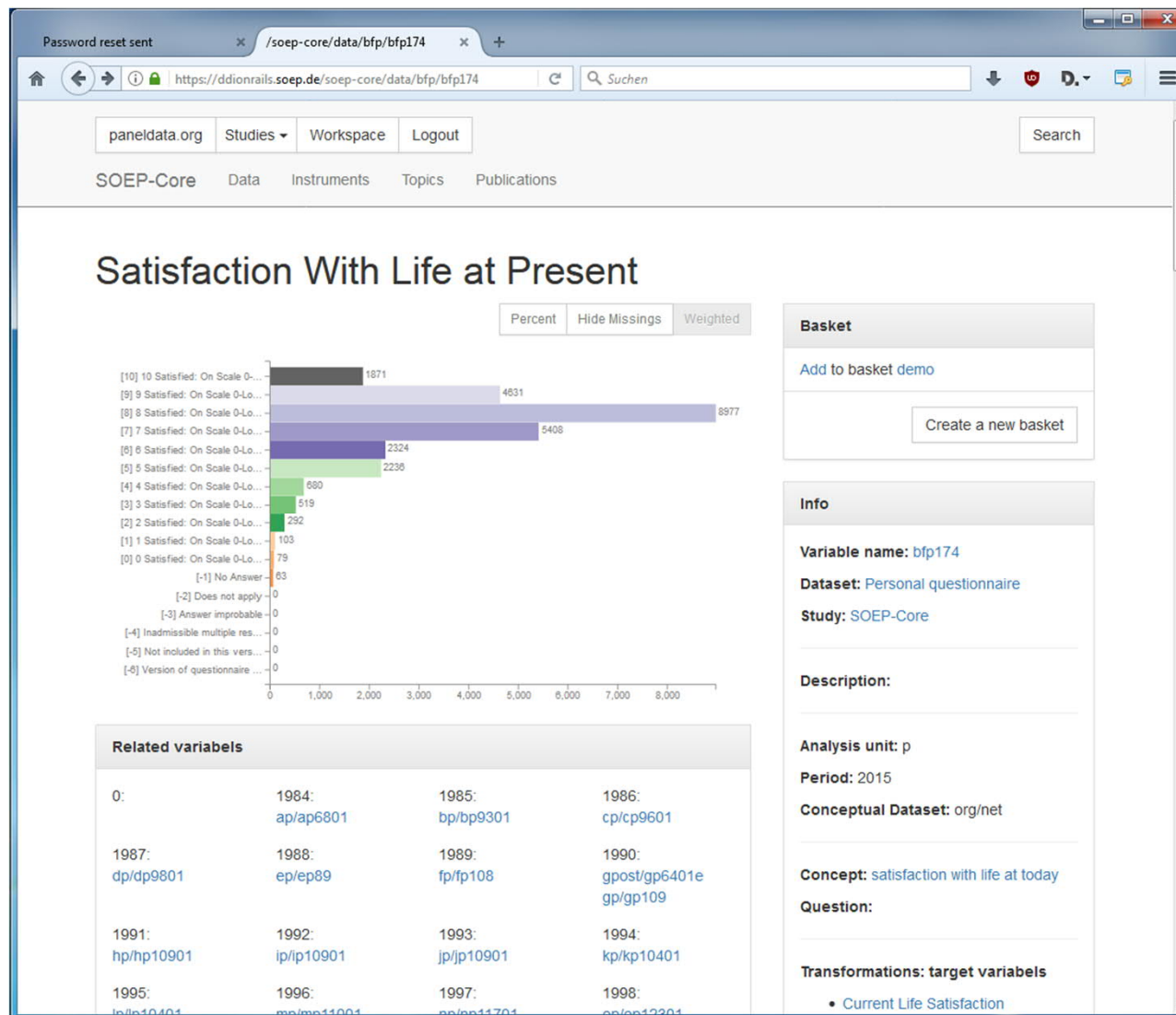
Overview of all studies of the SOEP and other panels



All data sets and increasingly also all questionnaires

The screenshot shows the Paneldata.org website interface. At the top, there's a navigation bar with links for 'paneldata.org', 'Studies', 'Register / log in', and a 'Search' button. Below this, a secondary navigation bar lists 'SOEP-Core', 'Data', 'Instruments', 'Topics', and 'Publications'. The main content area is titled 'Datasets' and includes a 'Show 10 entries' dropdown and a search box. A table lists various datasets with columns for Name, Label, Conceptual, Period, and Analysis unit. The table is paginated, showing 'Showing 1 to 10 of 414 entries' and a page navigation bar with 'Previous', '1', '2', '3', '4', '5', '...', '42', and 'Next'.

Name	Label	Conceptual	Period	Analysis unit
abroad	Questionnaire for people moved abroad	org/net	0	p
ah	Household questionnaire	org/net	1984	h
ahbrutto	Gross Household Data	org/gross	1984	h
ahgen	Generated Household Data	gen	1984	h
akind	Data on children (from HH-Questionnaire)	org/net	1984	p
ap	Personal questionnaire	org/net	1984	p
apausl	Migrant specific questions in the Personal Questionnaire	org/net	1984	p
apbrutto	Gross Individual Data	org/gross	1984	p
apequiv	Cross-national Equivalent File	gen	1984	p
apgen	Generated Individual Data	gen	1984	p



5

Working with the Basket

The screenshot shows a web browser window with the URL `https://ddionrails.soep.de/workspace/baskets/1`. The page title is "Basket: demo". The interface includes a navigation bar with "paneldata.org", "Studies", "Workspace", and "Logout" buttons. A search bar is also present. Below the navigation bar, there is a section for "Info" with the following details:

- Title: demo
- Study: [soep-core](#)

To the right of the info section, there is a "List of scripts" button and a "CREATE A NEW SCRIPT" button. The main area displays a table of variables selected for the basket, organized by year from 1984 to 1993. Each variable is represented by a shopping cart icon and a label. The variables are:

Year	Variable
1984	pzule1
1985	ap6801, bp9301
1986	cp9601
1987	dp9801
1988	ep89
1989	fp108
1990	gp6401e, gp109
1991	hp10901
1992	ip10901
1993	jp10901

On the left side of the variable list, there are buttons for "Add all" and "Remove all". Below the variable list, there is a section for "satisfaction with life at today".

5

Script Generator for Cross-section With Stata, SPSS and R

demo/script-1 (script in paneldata) | Password reset sent | 90% | Suchen

paneldata.org | Studies | Workspace | Logout | Search

demo | No more links at the moment...

Script: demo

Configure basket

Name
script-1

Label

Script generator
soep-stata

Output path
out/

Analysis Unit
Individual

Private households
Private households only

Sample composition
balanced

Age group
All adult respondents

Update script

raw script

```

*** LOCAL VARIABLES ***

global MY_PATH_IN  "data/"
global MY_PATH_OUT  "out/"
global MY_FILE_OUT  ${MY_PATH_OUT}new.dta
global MY_LOG_FILE  ${MY_PATH_OUT}new.log
capture log close
log using "${MY_LOG_FILE}", text replace
set more off

*** NOT PROCESSED ***

*** PFAD ***

use hhnr persnr sex gebjahr psample hhnr hnetto hpop chnr cnetto cpop using "${MY_PATH_IN}pp

*** BALANCED VS UNBALANCED ***

keep if ( (hnetto >= 10 & hnetto < 20) & (cnetto >= 10 & cnetto < 20) )

*** PRIATVE VS ALL HOUSEHOLDS ***

keep if ( (hpop == 1 | hpop == 2) | (cpop == 1 | cpop == 2) )

*** SORT PFAD ***

sort hhnrakt
save "${MY_PATH_OUT}pfad.dta", replace
clear

*** HRF ***

use "${MY_PATH_IN}phrf.dta"
sort persnr
save "${MY_PATH_OUT}phrf.dta", replace

```

Thank you for your attention.



DIW Berlin — Deutsches Institut
für Wirtschaftsforschung e.V.
Mohrenstraße 58, 10117 Berlin
www.diw.de

Editor