

Kara, Selin; Zimmermann, Stefan

Research Report

SOEPcompanion (v38)

SOEP Survey Papers, No. 1261

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Kara, Selin; Zimmermann, Stefan (2023) : SOEPcompanion (v38), SOEP Survey Papers, No. 1261, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/273538>

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/>

1261²⁰²³

SOEP Survey Papers
Series F – General Issues and Teaching Materials

SOEPcompanion (v38)

Selin Kara, Stefan Zimmermann, and 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. Carina Cornesse, DIW Berlin and University of Bremen

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Cornelia Kristen, University of Bamberg and DIW Berlin

Prof. Dr. Philipp Lersch, DIW Berlin and Humboldt-Universität zu Berlin

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

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

Prof. Dr. Sabine Zinn, DIW Berlin and Humboldt-Universität zu Berlin

Please cite this paper as follows:

Selin Kara, Stefan Zimmermann, and SOEP Group, 2023. SOEPcompanion (v38). SOEP Survey Papers 1261: Series F – General Issues and Teaching Materials. Berlin: DIW Berlin/SOEP



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

© 2023 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

soeppapers@diw.de

SOEPcompanion (v38)

Selin Kara, Stefan Zimmermann, and SOEP Group

2023

SOEPcompanion

Release 2023

Selin Kara, Stefan Zimmermann, SOEP Group

Jul 24, 2023

CONTENTS

1	Preface	1
2	Topics of SOEP-Core	2
2.1	Demography and Population	3
2.2	Work and Employment	4
2.3	Income, Taxes, and Social Security	15
2.4	Family and Social Networks	29
2.5	Health and Care	37
2.6	Home, Amenities, and Contributions of Private HH	43
2.7	Education and Qualification	50
2.8	Attitudes, Values, and Personality	61
2.9	Time Use and Environmental Behavior	66
2.10	Integration, Migration, Transnationalization	71
2.11	Survey Methodology	74
3	Survey Design	75
3.1	SOEP Questionnaires	75
3.1.1	Overview of the Questionnaires	77
3.1.2	Household Questionnaire	77
3.1.3	Individual Questionnaire	79
3.1.4	Biography Questionnaire	82
3.1.5	Mother and Child Instruments	83
3.1.6	Youth Instruments	85
3.1.7	Additional Instruments	88
3.2	Scales Manual	90
3.2.1	Affective Well-Being	90
3.2.2	Anomie	92
3.2.3	Basic Social Justice Orientations Scale	94
3.2.4	Cognitive Competencies	96
3.2.5	Conspiracy Mentality	99
3.2.6	Effort-Reward Imbalance Model	100
3.2.7	Impulsiveness & Patience	103
3.2.8	Life Goals	104
3.2.9	Life Satisfaction	108
3.2.10	Locus of Control	118
3.2.11	Loneliness	120
3.2.12	Narcissistic Admiration and Rivalry Questionnaire Short Scale (NARQ-S)	121
3.2.13	Optimism/Pessimism – Attitudes toward the Future	123
3.2.14	Parenting Goals	124
3.2.15	Parenting Role	125

3.2.16	Parenting Style	127
3.2.17	Patient Health Questionnaire – 4 (PHQ-4)	129
3.2.18	Personality – Big Five	131
3.2.19	Reciprocity	139
3.2.20	Risk Aversion	142
3.2.21	Self Esteem	145
3.2.22	Sources of Social Inequality	146
3.2.23	Strengths and Difficulties Questionnaire (SDQ)	147
3.2.24	Supportive Parenting	152
3.2.25	Temperament	154
3.2.26	Tendency to Forgive	155
3.2.27	Trust, Trustworthiness, Fairness	156
3.2.28	Vineland Adaptive Behavior Scales	158
3.3	Survey Concepts and Modes	161
3.4	Panel Care	162
4	Target Population and Samples	164
4.1	The SOEP Samples in Detail	165
4.1.1	Sample-Specific Questionnaires	168
4.2	Eligibility and Follow-up	175
4.3	Development of Sample Sizes	176
5	Data Structure of SOEP-Core	179
5.1	Data Editions of SOEP-Core	179
5.1.1	Teaching, International, and EU Edition	180
5.1.2	Add-ons: Area Types and Planning Regions	180
5.1.3	Remote Edition	180
5.1.4	Onsite Edition	181
5.2	Principles of Data Analysis	181
5.2.1	Cross-Sectional Data Structure (CS)	182
5.2.2	Data Structure in “Wide” Format (wide)	182
5.2.3	Data Structure in “Long” Format (long)	182
5.2.4	Data Structure in Spell Format (spell)	183
5.3	Data Distribution File	183
5.3.1	Core Datasets	186
5.3.2	Raw Datasets	186
5.3.3	eu-silc-like-panel	188
5.4	Datasets SOEP-Core	189
5.4.1	Tracking Data	190
5.4.2	Original Data	192
5.4.3	Survey Data	194
5.4.4	Generated Data	195
5.4.5	Spell Data	199
5.5	Data Processing	201
5.6	Dataset Identifiers	201
5.6.1	Partner Identifier	202
5.6.2	Family Identifier	203
5.6.3	Interviewer Identifier	205
5.7	Versioning and Harmonization	205
5.8	Missing Conventions	206
6	Working with SOEP Data	208
6.1	Working with Tracking Data (PPATHL)	208
6.2	Generating a Cross-Sectional Dataset	220

6.3	Syntax Generator on paneldata.org	228
6.4	Generating a Longitudinal Dataset	235
6.5	Working with harmonized Variables	249
6.6	Longitudinal Data Analysis	264
6.6.1	Clean and inspect the data	265
6.6.2	Univariate inspection & analysis	268
6.6.3	Simple cross sectional analyses	273
6.7	Working with Migration Data (BIOIMMIG)	276
6.8	Fixed Effects Estimation	285
6.9	Working with SOEP Regional Data	299
6.10	Working with spatial data in R	306
6.10.1	Prerequisites	306
6.10.2	Reading data	308
6.10.3	Transformations	312
6.10.4	Plotting Spatial Data	313
6.10.5	Frequently Used Operations	314
6.10.6	Complete Example	321
6.10.7	Appendix	325
6.11	How to Use SOEP IGEL	326
6.11.1	IGEL Workstation	326
6.11.2	Logging in	326
6.11.3	Working with SOEP DATA	331
6.11.4	Importing Scripts or External Data	332
6.11.5	Instructions for exporting from Hauser to user	333
6.11.6	Data transfer from Moran to Hauser	335
6.12	Working with SOEP data in csv format	336
6.13	How to Merge SOEP Data in Stata	339
6.13.1	1:1 merge - one-to-one on key variables	340
6.13.2	1:m merge - one-to-many on key variables	343
6.13.3	m:1 merge – many-to-one on key variables	346
6.13.4	joinby	348
7	Working with SOEP Documentation	350
7.1	Variable Search with Questionnaires	350
7.2	Variable Search with paneldata.org	352
7.3	Topic Search with paneldata.org	360
7.4	Documentation on Generated Data	367
7.5	Working with SOEPhelp	373
7.5.1	Working with SOEPhelp in R	373
7.5.2	Working with SOEPhelp in STATA	375
7.6	Working with Metadata-based Questionnaires	382
8	Contact Information	384

PREFACE

SOEP-Core is the centerpiece of the Socio-Economic Panel, a wide-ranging representative longitudinal study of private households in Germany, based at the German Institute for Economic Research, DIW Berlin. SOEP-Core was started in 1984, and in 1990—shortly after German reunification—it was enlarged to include a representative sample from East Germany. This feature makes the SOEP unique among household panel surveys worldwide. Every year since 1984, individuals in households have been surveyed by the SOEP’s fieldwork organization, infas Institut für angewandte Sozialwissenschaften GmbH. The data provide information on every member of every household taking part in the survey. Respondents include Germans living in 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 describes the current version of the SOEP-Core data (v38) and introduces users to the different SOEP-Core data structures. It also provides applications in Stata as well as instructions on how to use our various documentation services. We plan to revise the information in the SOEPcompanion annually to continue providing users a comprehensive, up-to-date introductory understanding of the SOEP.

We know that starting to use any new dataset is difficult, and this is especially true of panel data given their complexity. We hope that this introduction will help. We always welcome any feedback or tips on how to improve our documentation.

- Recommendation of our most recent version of a general short description of SOEP study: [The German Socio-Economic Panel Study \(SOEP\)](#)
- To the information system for efficient working with complex datasets: paneldata.org

TOPICS OF SOEP-CORE

The topics of the SOEP questionnaires and the various modules they contain can be grouped into 11 areas. Some of the modules deal with aspects of life that tend to change from one year to the next, and are therefore repeated annually, while other modules are repeated at intervals of several years. How often a module is repeated is stated in the “module” column of our topic tables. Some SOEP modules are also adapted in different ways to the different questionnaires. The questions in the “Big Five” personality traits module, for instance, are formulated differently in the mother-child questionnaires than they are in the individual questionnaire.

Overview of Modules in Different SOEP Questionnaires

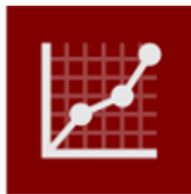
	<i>Individual</i>	<i>Youth</i>	<i>Mother-Child A</i>	<i>Mother-Child B</i>	<i>Mother-Child C</i>	<i>Parents D</i>	<i>Mother-Child E</i>
Affective well-being	x	x					
Big Five personality traits	x	x		x	x		x
Birth history		x	x	x	x	x	x
Childcare			x	x	x	x	x
Educational aspirations		x				x	x
Health of child			x	x	x		x
Height and weight of child			x	x	x		
Height and weight	x	x					
Language ability German / native language	x	x					
Leisure and activities (with child)				x	x		
Life satisfaction	x	x					
Language use				x			x
Locus of control	x	x					
Family background	x	x					
Parental interest in school performance		x					x
Allowance		x					x
Political orientation	x	x					
Risk aversion	x	x					
State of health	x	x					
Strengths and difficulties					x		x
Temperament			x	x			

The Scales Manual briefly describes the theoretical background and development of all of the scales used in the Socio-Economic Panel (SOEP) study. It also provides the relevant citations as well as the items belonging to the scales and the answer format, including the verbal anchors.

Note: The topic tables list the modules in a questionnaire, not the question items. The modules listed do not represent all of the variables in the SOEP, nor do they refer to all questionnaires. Specific information can also be found in our generated datasets.

2.1 Demography and Population

The demography and population topic provides information including the birth date and sex of each household member (including children) and of each interviewer; the place and history of births in the household; household size; and relationships among household members.



Demography and Population

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Birth date		birthd, birthm, birthy_v2	[2022]
	Country of Origin, Acquisition Date German Citizenship	[2002-2020], [2013], [2002-2012,2014-2020]	plj0024_h, plj0024_v1, plj0024_v2	[2022, 2023]
	Country of Origin, Birthcountry Parents	[2012-2018,2020]	plj0175	[2022, 2023]
	Country of Origin, Citizenship	[1984-1993,1996-2020], [1984-1995]	plj0014_v1, plj0014_v2	[2022, 2023]
	Country of Origin, German Citizenship	[1996-2020]	plj0014_v3	
	Country of Origin, Second Citizenship	[2000-2020]	plj0022, plj0023_v1	[2022, 2023]
	Country of Origin, Second/Other Citizenship	[2020]	plj0023_v2	
	Gender Identification		pla0047_v2	[2022, 2023]
	Gender Identification, Birth Register	[2021]	psex3	
	Gender Identification, Other		psubjsexso	[2022, 2023]
<i>Youth Questionnaire</i>	Gender Identification, Self-Identification		pla0047_v2, psubjsexso	[2022]
	Birth history, Birth date	[2000-2020], [2003-2020]	jl0233, jl0234	

continues on next page

Table 1 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Birth history, Born in Germany	[2000-2020], [2000-2005, 2011-2020], [2006-2010]	j10235_h, j10235_v1, j10235_v2	
	Birth history, Country of Birth	[2000-2020]	j10238	
	Birth history, Year of Migration	[2000-2018]	j10239	
	Country of Origin, Acquisition Date German Citizenship	[2012-2018]	j10419	
	Country of Origin, German Citizenship	[2006-2018]	j10241, j10244	
	Country of Origin, Nationality	[2006-2018]	j10245	
	Country of Origin, Residence status	[2000-2013], [2014-2018]	j10246, j10445	
	Country of Origin, Second Citizenship	[2006-2018]	j10242, j10243	
	Country of Origin, Status upon Migration	[2000-2018]	j10240	
<i>Household Questionnaire</i>	Birth date		birthd, hgeb, hgebmo	[2022, 2023]
<i>Mother and Child Instruments</i>	Birth history, Birth date	[2003-2020]	birthm, birthy	[2022]

2.2 Work and Employment

The work and employment modules provide information on diverse employment-related topics including the respondent's first job, further training, changes in working conditions following parenthood, part-time work, and unemployment. Modules cover not only objective information such as working hours but also subjective perceptions of working conditions and feelings about work.



Work and Employment

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Average Weekly Hours		plb0726, plb0727	[2022]
	Care period (Pflegezeit)	[2011-2019], [2015-2019, 2021], [2011-2014, 2020]	plb0020_h, plb0020_v1, plb0020_v2	[2022, 2023]

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Change of job	[1985-2020], [1985-1993], [1994-2020, 2021]	plb0031_h, plb0031_v1, plb0031_v2	[2022, 2023]
	Change of job, Early Retirement	[2011-2020, 2021]	plb0480	[2022, 2023]
	Change of job, Interruption due to Children	[2010-2020, 2021]	plb0295	[2022, 2023]
	Change of job, Month Change in Company	[1985-1993]	plb0033_v5	
	Change of job, Month First Job	[1985-1993]	plb0033_v1	
	Change of job, Month Position taken up	[1994-2020, 2021]	plb0033_v7	[2022, 2023]
	Change of job, Month Self Employed	[1985-1993]	plb0033_v4	
	Change of job, Month Taken Over	[1991-1993]	plb0033_v6	
	Change of job, Month Work resumed	[1985-1993]	plb0033_v2	
	Change of job, Month new Employer	[1985-1993]	plb0033_v3	
	Change of job, Number of Changes	[2012-2020, 2021]	plb0478, plb0479	[2022, 2023]
	Change of job, Start Current Position non-response	[2004-2020]	plb0034	
	Change of job, Start new Position	[1994-2020, 2021]	plb0032	[2022, 2023]
	Change of job, Type of Change	[1994-2020], [1994-2004], [2005-2020, 2021]	plb0284_h, plb0284_v1, plb0284_v2	[2022, 2023]
	Changes in workplace tools / technologies	[2015-2018]	plb0595	
	Changes in workplace tools / technologies, Influences	[2015-2018]	plb0596, plb0598, plb0600	plb0597, plb0599,
	Commuter Module, Distance	(irregular) [1985-2019, 2021]	plb0158	[2023]
	Commuter Module, Frequency	[2015, 2017, 2019, 2021]	plb0591	
	Commuter Module, No Commute	[1985, 1990], [1993, 1995, 1997-1999], [2000-2013, 2015, 2017, 2019, 2021]	plb0159_v1, plb0159_v2, plb0159_v3	[2023]
	Commuter Module, Second Residence	[2015, 2017, 2019, 2021]	plb0589, plb0590	
	Commuter Module, Time	[2015, 2017, 2019, 2021]	plb0592	[2023]

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Contract to Provide Specific Services (Werkvertrag)	[2013,2015]	plb0482	
	Contractual working hours, Actual working hours	[1984-2020], [1984-1990], [1990-2019]	plb0186_h, plb0186_v1, plb0186_v2	
	Contractual working hours, Desired working hours	[1985-1995,1997-2019], [1985-1995,1997-1999,2016], [2000-2015,2017-2019]	plb0241_h, plb0241_v1, plb0241_v2	
	Contractual working hours, Not fixed	[1987-1989], [1984-1986,1990-2021]	plb0185_v1, plb0185_v2	[2022, 2023]
	Contractual working hours, Working days	(irregular) [1990-2021]	plb0209, plb0210	[2022, 2023]
	Current job, Civil Servant	[1984-2021]	plb0065	[2022, 2023]
	Current job, Civil Service	[1984-2021]	plb0040	
	Current job, Employment Agency	[2001-2021]	plb0041	
	Current job, Fixed-Term Employment	[1985-2020], [1985-1999], [2000-2005], [2006-2021]	plb0037_h, plb0037_v1, plb0037_v2, plb0037_v3	[2022, 2023]
	Current job, Industrial Worker	[1984-2021]	plb0058	[2022, 2023]
	Current job, Month current Position	[1984-1988,1990-1993,1996-2021]	plb0035	[2022, 2023]
	Current job, Number of Employees	[1984-1990], [1990], [1991-1998], [1999-2004], [2005-2015]	plb0049_v1, plb0049_v2, plb0049_v3, plb0049_v4, plb0049_v5	
	Current job, Number of Employees Self Employed	[1984-2020], [2014-2020]	plb0057_h2, plb0057_v8	
	Current job, Occupational Status	[2019], [2014-2020]	plb0568_v1, plb0568_v2	[2022, 2023]
	Current job, Salaried employees	[1984-1990], [2018, 2021]	plb0064_v1, plb0064_v2	
	Current job, Self Employed	[1984-2020], [2014-2020]	plb0057_h1, plb0057_v1, plb0057_v2, plb0057_v3, plb0057_v4, plb0057_v5, plb0057_v6, plb0057_v7, plb0057_v9	[2022, 2023]
	Current job, Trainee/Intern	[1984-1999], [1990]	plb0063_v1, plb0063_v2	

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Current job, Working at current Employer	[1984-1988,1990-2020], [1984-1988,1990-1998], [1999-2021]	plb0036_h, plb0036_v1, plb0036_v2	[2022, 2023]
	Digitalisation in the workplace	[2020]	perlarb1, perlarb2, perlarb3, perlarb4, perlarb5, perlarb6, perlarb7, perlarb8, perlarb9, perlarb10, perlarb11, perlarb12, pdigarb1, pdigarb2, pdigarb3, pdigarb4, pdigarb5, pdigarb6, pdigarb7, pdigarb8, pdigisys1, pdigisys2, pdigisys3, pdigisys4, pdigisys5, parbsitm1, parbsitm2, parbsitm3, parbsitm4, parbsitm5, parbsitm6, parbsitm7, parbsitm8	
	Employment / education calendar, Full-Time Employment	[1984], [1985-1989,1991-2021]	pab0001_v1, pab0001_v2, pab0001_v3	
	Employment / education calendar, Further Education	[2000-2020], [2020]	pab0010_v1, pab0010_v2	
	Employment / education calendar, Homemaker	[1984-2021]	pab0008	
	Employment / education calendar, In School	[1984-2021]	pab0013	
	Employment / education calendar, Maternity Leave	[1996-2021]	pab0006	
	Employment / education calendar, Military/Civil Service	[1984-2021]	pab0007	
	Employment / education calendar, Mini-Job	[2005-2021]	pab0011	
	Employment / education calendar, Other	[1984-2021]	pab0012	
	Employment / education calendar, Part-Time Employment	[1985-2021]	pab0002	
	Employment / education calendar, Retired	[1984-2021]	pab0005	
	Employment / education calendar, Short-Time Work	[2010-2014,2017-2018]	pab0003_v1	
	Employment / education calendar, Unemployed	[1984-2021]	pab0004	

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Employment / education calendar, Vocational Training	[2020]	pab0003_v2	
	Employment status	[1984-2020], [1984], [1985-1990], [1990], [1991-1995], [1996-1998], [1999], [2000-2001], [2002-2015], [2016-2021]	plb0022_h, plb0022_v1, plb0022_v2, plb0022_v3, plb0022_v4, plb0022_v5, plb0022_v6, plb0022_v7, plb0022_v8, plb0022_v9	
	Employment, October 2014	[2015]	plb0574	
	Entitlement to paid breaks	[2015-2018]	plb0601, plb0602, plb0603	
	Evening and weekend work, Evening	[1990], [2012], [2013], (irregular) [1995-2019]	plb0205_v1, plb0205_v2, plb0205_v3, plb0205_v4	[2023]
	Evening and weekend work, Night	[1990], [2012], [2013], (irregular) [1995-2019]	plb0206_v1, plb0206_v2, plb0206_v3, plb0206_v4	[2023]
	Evening and weekend work, Saturday	(irregular) [2005-2019]	plb0218	[2023]
	Evening and weekend work, Sunday	(irregular) [2005-2019]	plb0219	[2023]
	Financial compensation for overtime	[1984-1986, 1988-2014, 2018, 2020], [1984-1986, 1988-1995], [1996], [1997-2014, 2018, 2020]	plb0195_h, plb0195_v1, plb0195_v2, plb0195_v3	[2022]
	Gross / net income, October 2014	[2015]	plb0584, plb0585	
	Income from Internet activities	[2020]	pintver, pintverx, pintvstd1, pintvstd2, pintvstdno, pintvv1, pintvv2, pintvvno, pintet, pintetx, pintetstd1, pintetstd2, pintetstdno, pintetv1, pintetv2, pintetvno, pintba, pintbax, pintbastd1, pintbastd2, pintbastdno	

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Industry sector, occupational classification	[2013-2020], [1990-1993], [2020], [1999-2021]	p_isco08, p_nace, plb0072_v1, plb0072_v2, plb0072_v3, plb0073_h, plb0073_v1, plb0073_v2, plb0073_v3, plb0073_v4, plb0073_v5	[2022, 2023]
	Job search	[1985-2020]	plb0358_h	
	Job search, Active Search	[1989-2020]	plb0362	[2022, 2023]
	Job search, Applied on Speculation	[1989-1998]	plb0358_v8	
	Job search, Friends / Acquaintances	[1985-1998], [1985-1988]	plb0358_v3, plb0358_v5	
	Job search, Job Centre	[1985-1998]	plb0358_v1	
	Job search, Learn about current Position	[1999-2002], [2003-2013], [2014], [2015-2021]	plb0358_v10, plb0358_v11, plb0358_v12, plb0358_v13	
	Job search, Newspaper	[1985-1998]	plb0358_v2	
	Job search, Offer within Company	[1985-1988]	plb0358_v4	
	Job search, Other	[1989-1998]	plb0358_v7	
	Job search, Private Agent	[1995-1998]	plb0358_v9	
	Job search, Self Employed	[1985-1998]	plb0358_v6	
	Job search, motives	(irregular) [1994-2017]	plb0111	
	Job search, preferences	(irregular) [1994-2017]	plb0426	
	Labor Market Experience		plb0737, plb0739, plb0741	plb0738, plb0740, [2022]
	Labor intensity	[2015], [2016-2018], [2015-2018]	plb0593_v1, plb0593_v2, plb0594	
	Leaving a job	[1985-2020], [2001-2020], [1985-2000]	plb0282_h, plb0282_v1, plb0282_v2	[2022, 2023]
	Leaving a job, Abandonment of own business	[1985-1998]	plb0304_v8	
	Leaving a job, Closure of operations	[1991-1998]	plb0304_v11	
	Leaving a job, Compensation	[1991-2020], [1991-2001], [1991-2020], [2002-2020]	plc0040, plc0041_h, plc0041_v1, plc0041_v2	[2022, 2023]
	Leaving a job, Early Retirement	[1987-1998]	plb0304_v10	
	Leaving a job, End Fixed-Term Contract	[1985-1998]	plb0304_v2	

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Leaving a job, End Vocational Training	[1985-1998]	plb0304_v3	
	Leaving a job, Exempted	[1991-1998]	plb0304_v12	
	Leaving a job, Month	[1985-2020]	plb0298, plb0299	[2022, 2023]
	Leaving a job, Months Worked	[1985-2020]	plb0302	[2022, 2023]
	Leaving a job, Mutually agreed dissolution	[1985-1990]	plb0304_v5	
	Leaving a job, Non-response	[2004-2020]	plb0300	[2022]
	Leaving a job, Other	[1985-1998]	plb0304_v9	
	Leaving a job, Own Resignation	[1985-1998]	plb0304_v4	
	Leaving a job, Perspective after Leaving	[1999], [2000-2020]	plb0305_v1, plb0305_v2	[2022, 2023]
	Leaving a job, Retirement	[1991-1998]	plb0304_v15	
	Leaving a job, Termination by Employer	[1985-1998]	plb0304_v1	
	Leaving a job, Transfer	[1985-1998]	plb0304_v7	
	Leaving a job, Transfer at own request	[1985-1998]	plb0304_v6	
	Leaving a job, Type of Leaving	[1985-2020], [1999-2000], [2001-2020]	plb0304_h, plb0304_v13, plb0304_v14	[2022, 2023]
	Leaving a job, Years Worked	[1985-2020]	plb0301	[2022, 2023]
	Maternity / parental leave	[1999-2000], [2001-2020]	plb0019_v1, plb0019_v2	[2022]
	Occupational expectations, non-employed	(irregular) [1999-2020]	plb0427, plb0429	plb0428, [2022]
	Overtime, October 2014	[2015]	plb0582, plb0583	
	Paid breaks, October 2014	[2015]	plb0575	
	Paid breaks, October 2015	[2015]	plb0576	
	Paid breaks, October 2016	[2015]	plb0577	
	Paid breaks, October 2017	[2015]	plb0578	
	Performance evaluation by superior	[2004,2008,2011,2016]	plb0098, plb0100, plb0102	plb0099, plb0101,

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Professional expectations	[1985,1987,1989-1994,1996,1998]	plb0432_v1, plb0433_v1, plb0434_v1, plb0435_v1, plb0436_v1, plb0437_v1, plb0438_v1, plb0439_v1, plb0440_v1, plb0441_v1, plb0442_v1	
	Professional expectations, next two years	(irregular) [1999-2018]	plb0432_v2, plb0433_v2, plb0434_v2, plb0435_v2, plb0436_v2, plb0437_v2, plb0438_v2, plb0439_v2, plb0440_v2, plb0441_v2, plb0442_v2	[2023]
	Registered unemployed	[1985-2020]	plb0021	[2022, 2023]
	Self-employment, reasons	[2010,2015]	plb0333, plb0334, plb0335, plb0336, plb0337, plb0338	
	Short-time compensation (Kurzarbeitergeld)	[1984-2001,2003-2005,2010-2011], [1984-2001,2003-2005], [1984], [2010-2011], [1985-2001,2003-2005,2010-2011]	plc0057_h, plc0057_v1, plc0057_v2, plc0058_v1, plc0058_v2	[2022, 2023]
	Side jobs	[1998-2007], [1998]	plb0382_h, plb0382_v1	
	Side jobs, Agriculture	[1999-2007]	plb0382_v2	
	Side jobs, Construction	[1999-2007]	plb0382_v3	
	Side jobs, Days	[1985-2016]	plb0396	
	Side jobs, Gross Income	[1995-2016], [1995-2001], [2002-2016]	plc0062_h, plc0062_v1, plc0062_v2	
	Side jobs, Helping Family Members out	[1986-2016]	plb0392	
	Side jobs, Hours per Month	[1985-2014]	plb0397	
	Side jobs, Hours per Week	[2015-2016]	plb0573	
	Side jobs, Industrial Sector	[1999-2007]	plb0382_v4	
	Side jobs, Iregual	[1985-2016]	plb0395	
	Side jobs, Months	[2000-2013]	plb0398	

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Side jobs, Occupational Classification ISCO08	[2013-2016]	p_isco08_sidejob, p_isco08_sidejob1, p_isco08_sidejob2, p_isco08_sidejob3	
	Side jobs, Occupational Classification ISCO88	[1991-2016]	p_isco88_sidejob, p_isco88_sidejob1, p_isco88_sidejob2, p_isco88_sidejob3	
	Side jobs, Other	[1985-2016]	plb0393	
	Side jobs, Regular	[1985-2016]	plb0394	
	Side jobs, Service Sector	[1999-2007]	plb0382_v5	
	Standby duty	[2011,2014-2019]	plb0212, plb0213, plb0214, plb0215	[2023]
	Start of working hours	(irregular) [2002-2019]	plb0180, plb0181, plb0182	
	Starting a new job, Acceptable Position	[1984-2020]	plb0423	[2022, 2023]
	Starting a new job, Active Job Search	[1994-1998], [1999-2020]	plb0424_v1, plb0424_v2	[2022, 2023]
	Starting a new job, Desired Employment Type	[1984-2020]	plb0240	[2022, 2023]
	Starting a new job, Expected Minimum Income	[1987-1989,1992-1994,1996-2001], [2002-2020]	plb0420_v1, plb0420_v2	[2022, 2023]
	Starting a new job, Intention	[1984-1993], [1994-2020]	plb0417_v1, plb0417_v2	[2022, 2023]
	Starting a new job, Non-response Salary	[1987-1989,1992-1994,1996-2020]	plb0421	[2022, 2023]
	Starting a new job, Number of Hours	[2007-2020]	plb0422	[2022, 2023]
	Starting a new job, Suitable Job	[1987-2020], [1987-2002], [2003-2020]	plb0419_h, plb0419_v1, plb0419_v2	
	Starting a new job, Timing	[1984-2020]	plb0418	[2022, 2023]
	Supervisory position	[2007,2009,2011,2013,20	plb0067, plb0068, plb0069	[2023]
	Use of professional skills in job	[1985-2007,2009]	plb0357	
	Vacation entitlement, Carried over Vacation	[2005,2010]	plb0275, plb0276	
	Vacation entitlement, Contracted Days	[2000,2005,2010]	plb0269	
	Vacation entitlement, Days on Vacation	[1985-1990,2000,2005,2010]	plb0265	
	Vacation entitlement, Expired Vacation	[2005,2010]	plb0273, plb0274	
	Vacation entitlement, Not specified	[2005,2010]	plb0270, plb0272	
	Work council (Betriebsrat)	[2001,2006,2011,2016,20	plb0050	[2022]

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Work from home	[1997,1999,2002,2009-2014,2020], (irregular) [1997-2020], [2012], [2013]	plb0095, plb0096_v1, plb0096_v2, plb0096_v3	
	Work from home, Possibility	[1997,1999,2009-2014]	plb0097	[2022]
	Work from home, Possibility in Contract	[2020]	plb0697	[2022]
	Work in black economy	[2015-2016], [2015]	plb0571, plb0572	
	Work time regulations	[2003,2005,2007,2009-2019]	plb0211	[2023]
	Work, last 7 days	[1999-2020]	plb0018	[2022, 2023]
	Working hours, October 2014	[2015]	plb0579, plb0579_h, plb0580, plb0581, plb0581_h	
	Working overtime	[1997-2020]	plb0193	
	Working overtime, Compensation Period	[2020], [2002-2020]	plb0194_v1, plb0194_v2	[2022, 2023]
	Working overtime, Compensation period	[2002-2020], [2018-2020], [2002-2017]	plb0220_h, plb0220_v1, plb0220_v2	[2022, 2023]
	Working overtime, Financial Compensation	[2015-2020]	plb0605	[2022, 2023]
	Working overtime, Hours Last Month	[1986,1988-2020]	plb0197	[2022, 2023]
	Working overtime, Last Month	[1986,1988-2020], [1986,1988-1996], [1997-2001], [2002-2020]	plb0196_h, plb0196_v1, plb0196_v2, plb0196_v3	[2022, 2023]
	Working overtime, Paid Hours Last Month	[2001-2020]	plb0198	[2022, 2023]
	Working overtime, Time taken off	[2013-2020]	plb0483, plb0484	[2022, 2023]
	Workload (effort-reward imbalance), Career Prospects	[2006,2011-2012,2016]	plb0134, plb0135	
	Workload (effort-reward imbalance), Interruptions	[2006,2011-2012,2016]	plb0120, plb0121	
	Workload (effort-reward imbalance), Job at risk	[2006,2011-2012,2016]	plb0128, plb0129	
	Workload (effort-reward imbalance), Poor Career Prospects	[2006,2011-2012,2016]	plb0124, plb0125	
	Workload (effort-reward imbalance), Poor Working Conditions	[2006,2011-2012,2016]	plb0126, plb0127	
	Workload (effort-reward imbalance), Problems Sleeping	[2006,2011-2012,2016]	plb0117	

continues on next page

Table 2 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
<i>Youth Questionnaire</i>	Workload (effort-reward imbalance), Recognition by Superiors	[2006,2011-2012,2016]	plb0130, plb0131	
	Workload (effort-reward imbalance), Recognition for Performance	[2006,2011-2012,2016]	plb0132, plb0133	
	Workload (effort-reward imbalance), Sacrifices for Career	[2006,2011-2012,2016]	plb0115	
	Workload (effort-reward imbalance), Salary	[2006,2011-2012,2016]	plb0136, plb0137	
	Workload (effort-reward imbalance), Thinking about Work	[2006,2011-2012,2016]	plb0113, plb0116	plb0114,
	Workload (effort-reward imbalance), Time Pressure	[2006,2011-2012,2016]	plb0112, plb0119	plb0118,
	Workload (effort-reward imbalance), Work Volume	[2006,2011-2012,2016]	plb0122, plb0123	
	Jobs and money, Employment Form	[2000-2020]	jl0014	[2022, 2023]
	Jobs and money, First Job	[2000-2020]	jl0017, jl0018	[2022]
	Jobs and money, Job Search	[2006-2020]	jl0386	[2022, 2023]
	Jobs and money, Own Earnings	[2000-2020]	jl0013	[2022, 2023]
	Jobs and money, Paid Work	[2006-2020]	jl0385	[2022, 2023]
	Jobs and money, Reason for Working	[2001-2020]	jl0019	[2022]
	Jobs and money, Savings	[2000-2020], [2000-2001], [2000-2020], [2002-2020], [2000-2020]	jl0023, jl0024_v1, jl0025	jl0024_h, jl0024_v2, [2022, 2023]
	Jobs and money, Unemployment	[2006-2020]	jl0387	[2022, 2023]

2.3 Income, Taxes, and Social Security

The income, taxes, and social security modules collect wide-ranging financial information from earnings and spending to public benefits, pensions, inheritances, taxes, and debts. They also cover assets such as real estate and other property.



Income, Taxes and Social Security

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Additional questions for employed people, 13th month payment prev. year	[1984-2020], [1984-2001], [1984-2020], [2002-2020]	plc0042, plc0043_h, plc0043_v1, plc0043_v2	[2022, 2023]
	Additional questions for employed people, 14th month payment prev. year	[1984-2020], [1984-2001], [1984-2020], [2002-2020]	plc0044, plc0045_h, plc0045_v1, plc0045_v2	[2022, 2023]
	Additional questions for employed people, Christmas Bonus prev. year	[1984-2020], [1984-2001], [1984-2020], [2002-2020]	plc0046, plc0047_h, plc0047_v1, plc0047_v2	[2022, 2023]
	Additional questions for employed people, No Bonus prev. year	[1984-2020]	plc0054	[2022, 2023]
	Additional questions for employed people, Other Bonus prev. year	[1984-2020], [1984-2001], [1984-2020], [2002-2020]	plc0052, plc0053_h, plc0053_v1, plc0053_v2	[2022, 2023]
	Additional questions for employed people, Profit-sharing Bonus prev. year	[1985-2020], [1985-2001], [1985-2020], [2002-2020]	plc0050, plc0051_h, plc0051_v1, plc0051_v2	[2022, 2023]
	Additional questions for employed people, Vacation Bonus prev. year	[1984-2020], [1984-2001], [1984-2020], [2002-2020]	plc0048, plc0049_h, plc0049_v1, plc0049_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Accident Insurance Retirement Pension	[1986-2001, 2003-2020], [1986-2001], [2003-2020]	plc0243_h, plc0243_v1, plc0243_v2	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Additional questions for retirees / pensioners, Accident Insurance Widow's Pension	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0286_h, plc0286_v1, plc0286_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Company Retirement Pension	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0240_h, plc0240_v1, plc0240_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Company Widow's Pension	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0283_h, plc0283_v1, plc0283_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Invalid Pension non-response	[2003-2020]	plc0251	
	Additional questions for retirees / pensioners, Orphan Benefit non-response	[2003-2020]	plc0290	
	Additional questions for retirees / pensioners, Other Retirement Pensions	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0249_h, plc0249_v1, plc0249_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Other Widow's Pensions	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0288_h, plc0288_v1, plc0288_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Private Retirement Pension	[2003-2020], [2018-2020]	plc0242_v1, plc0242_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Private Widow's Pension	[2003-2020]	plc0285	[2022, 2023]
	Additional questions for retirees / pensioners, Retirement Pension Civil Servants	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0236_h, plc0236_v1, plc0236_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Riester Pension	[2015-2020]	plc0516, plc0517	[2022, 2023]
	Additional questions for retirees / pensioners, Shareholder Company	[2019]	plc0572	

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Additional questions for retirees / pensioners, Supplementary Pension Civil Servants	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0238_h, plc0238_v1, plc0238_v2	[2022, 2023]
	Additional questions for retirees / pensioners, Supplementary Widow's Pension Civil Servants	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0281_h, plc0281_v1, plc0281_v2	[2022, 2023]
	Additional questions for retirees / pensioners, War Victims Pension	[1986-2001,2003-2016], [1986-2001], [2003-2016]	plc0245_h, plc0245_v1, plc0245_v2	
	Additional questions for retirees / pensioners, War Victims Widow's pension	[1986-2001,2003-2016], [1986-2001], [2003-2016]	plc0247_h, plc0247_v1, plc0247_v2	
	Additional questions for retirees / pensioners, Widow's pension	[1986-2001,2003-2020], [2003-2020]	plc0268_h, plc0268_v1, plc0268_v2, plc0268_v3	[2022, 2023]
	Additional questions for retirees / pensioners, Widow's pension Civil Servants	[1986-2001,2003-2020], [1986-2001], [2003-2020]	plc0279_h, plc0279_v1, plc0279_v2	[2022, 2023]
	Asset balance	[2002]	plc0340	
	Asset balance, Building Loan Contract (Bausparvertrag)	[2007,2012], [2017,2019], [2007,2012], [2017,2019]	plc0317_v1, plc0317_v2, plc0318_v1, plc0318_v2	
	Asset balance, Building Society Savings	[2007,2012,2017,2019]	plc0315, plc0319	plc0316,
	Asset balance, Cash Surrender	[2002]	plc0327, plc0336, plc0338	plc0335, plc0337,
	Asset balance, Enterprise	[2002]	plc0341, plc0365, plc0367, plc0369	plc0364, plc0366, plc0368,
	Asset balance, Financial Assets	[2002], [2002,2007,2012], [2002,2007,2012,2017,2019]	plc0314, plc0328, plc0330, plc0332, plc0334	plc0326, plc0329, plc0331, plc0333,
	Asset balance, Financial Burden	[2002], [2002,2007,2012]	plc0408, plc0411, plc0413, plc0415, plc0417, plc0419, plc0420	plc0409, plc0412, plc0414, plc0416, plc0418,

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Asset balance, Life Insurance	[2002,2007,2012,2017,2019]	plc0363	
	Asset balance, Non self used Property	[2002,2007,2012,2017,2019], [2002,2007,2012], [2002,2007,2012,2017,2019]	plc0356, plc0358, plc0360, plc0362	plc0357, plc0359, plc0361,
	Asset balance, Other Property	[2002,2007,2012,2017,2019]	plc0354	
	Asset balance, Property non-response	[2007,2012,2017,2019]	plc0355	
	Asset balance, Remaining Debt	[2002]	plc0410, plc0422, plc0424, plc0425	plc0421, plc0423,
	Asset balance, Residential property	[2002]	plc0339, plc0343, plc0345, plc0347, plc0349	plc0342, plc0344, plc0346, plc0348,
	Asset balance, Residential property Usage	[2002,2007,2012,2017,2019]	plc0350, plc0352	plc0351,
	Asset balance, Tangible Assets	[2002,2007,2012,2017,2019]	plc0370, plc0372, plc0374	plc0371, plc0373,
	Asset balance, Undeveloped Land	[2002,2007,2012]	plc0353	
	Asset development	[2019]	plc0570i01, plc0570i02, plc0570i03, plc0570i04, plc0570i05, plc0570i06, plc0570i07, plc0570i08	
	Benefits and bonuses from employer	(irregular) [2006-2020]	plc0026, plc0028, plc0030, plc0032, plc0034, plc0036, plc0038, plc0039	plc0027, plc0029, plc0031, plc0033, plc0035, plc0037,
	Financial advantages from use of company car	[2016-2018,2020]	plc0532	[2022]
	Financial support received, Children	[2009-2013]	plj0156, plj0158, plj0159	plj0157,
	Financial support received, No Payments	[2009-2013]	plj0172	
	Financial support received, Other Relatives	[2009-2013]	plj0164, plj0166, plj0167	plj0165,

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Financial support received, Parents	[2009-2013]	plj0152, plj0153, plj0154, plj0155	
	Financial support received, Spouse	[2009-2013]	plj0160, plj0161, plj0162, plj0163	
	Financial support received, Unrelated Persons	[2009-2013]	plj0168, plj0169, plj0170, plj0171	
	Financial support to relatives or others, Children	[1984-1991,1993,1995-2020]	plj0135, plj0136_h, plj0136_v1, plj0136_v2, plj0137_h, plj0137_v1, plj0137_v2, plj0137_v3	[2022, 2023]
	Financial support to relatives or others, No Payments	[1984-1991,1993,1995-2020]	plj0151	[2022, 2023]
	Financial support to relatives or others, Other Relatives	[1984-1991,1993,1995-2020]	plj0143, plj0144_h, plj0144_v1, plj0144_v2, plj0145_h, plj0145_v1, plj0145_v2, plj0145_v3	[2022, 2023]
	Financial support to relatives or others, Parents	[1984-1991,1993,1995-2020]	plj0131, plj0132_h, plj0132_v1, plj0132_v2, plj0133_h, plj0133_v1, plj0133_v2, plj0133_v3	[2022, 2023]
	Financial support to relatives or others, Spouse	[1984-1991,1993,1995-2020]	plj0139, plj0140_h, plj0140_v1, plj0140_v2, plj0142_h, plj0142_v1, plj0142_v2, plj0142_v3	[2022, 2023]
	Financial support to relatives or others, Unrelated Persons	[1984-1991,1993,1995-2020]	plj0147, plj0148_h, plj0148_v1, plj0148_v2, plj0149_h, plj0149_v1, plj0149_v2, plj0149_v3	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Gross / net income, collective wage agreements	[1984-2020]	plc0013_h, plc0013_v1, plc0013_v2, plc0014_h, plc0014_v1, plc0014_v2, plc0502_v1, plc0502_v2, plc0507, plc0508, plc0509	[2023]
	Income, Alimony	[2010-2015], [2012], [2012,2018-2020]	plc0181, plc0182, plc0183_v1, plc0183_v2, plc0184, plc0188_v1, plc0188_v2, plc0190_v1, plc0190_v2, plc0494, plc0496	[2022, 2023]
	Income, Alimony Months	[2010-2017], [2010- 2013,2015], [2012,2018- 2020]	plc0189_v1, plc0189_v2, plc0495	[2022, 2023]
	Income, Child Support	[2010-2015]	plc0177, plc0178, plc0488, plc0490	
	Income, Child Support Months	[2010-2013,2015]	plc0489	
	Income, Gross Selfem- ployed prev year	[1990-2020], [1995-2020]	plb0474_h, plb0474_v1, plb0474_v2, plc0073_v1, plc0073_v2, plc0075_h, plc0075_v1, plc0075_v2	[2022, 2023]
	Income, Gross prev year	[1990-2020], [1990-2001], [1995-2020], [2017-2020], [2002-2020], [1995-2020]	plb0471_h, plb0471_v1, plb0471_v2, plc0015_h, plc0015_v1, plc0015_v2	[2022, 2023]
	Income, Maternity benefit	[2019-2020]	plc0126_v1, plc0126_v2, plc0152_v1, plc0152_v2, plc0153_h, plc0153_v1, plc0153_v2, plc0155_h, plc0155_v1, plc0155_v2	[2022, 2023]
	Income, Maternity benefit Months	[1995-2020]	plc0154	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Income, Months of Second Job Income	[1995-2020]	plc0065	[2022, 2023]
	Income, Months of Self-Employed Income	[1995-2020]	plc0074	[2022, 2023]
	Income, Months of Wages	[2000-2020], [2000-2001], [2002-2020], [1995-2020]	plc0016, plc0017_h, plc0017_v1, plc0017_v2	[2022, 2023]
	Income, No Other Income	[1995-2020], [2001-2020]	plc0116, plc0117	[2022, 2023]
	Income, Retirement pension	[2017-2020], [1995-2020], [1995-2001], [2017-2020], [1995-2020], [2002-2020], [1995-2020]	plc0232_v1, plc0232_v2, plc0233_h, plc0233_v1, plc0233_v2, plc0234_v1, plc0234_v2	[2022, 2023]
	Income, Retirement pension Months	[1995-2020]	plc0235	[2022, 2023]
	Income, Second Job	[2015,2019-2020]	plc0515	
	Income, Second Job prev year	[1990-2020], [1990-2001], [2002-2020]	plb0477_h, plb0477_v1, plb0477_v2	[2022, 2023]
	Income, Self-Employment	[2015,2019-2020]	plc0514	
	Income, Sideline Job prev year	[2019-2020], [1995-2020]	plc0064_v1, plc0064_v2	[2022, 2023]
	Income, Student loans	[1995-2020], [1995-2001], [2017-2020], [1990-2020], [1990-2001], [2002-2020], [1995-2020], [2002-2020]	plc0168_h, plc0168_v1, plc0168_v2, plc0169_v1, plc0169_v2, plc0171_h, plc0171_v1, plc0171_v2	[2022, 2023]
	Income, Student loans Months	[1995-2020]	plc0170	[2022, 2023]
	Income, Support from outside the household	[1990-2020], [1990-2001], [2019-2020], [1995-2020], [1995-2001], [2019-2020], [2002-2020], [1995-2020], [2002-2020], [1995-2020]	plc0198_h, plc0198_v1, plc0198_v2, plc0202_v1, plc0202_v2, plc0203_h, plc0203_v1, plc0203_v2, plc0204_v1, plc0204_v2	[2022, 2023]
	Income, Support from outside the household Months	[2018], [1995-2020]	plc0205_v1, plc0205_v2	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Income, Unemployment benefit	[2017-2020]	plc0130_v1, plc0130_v2, plc0131_h, plc0131_v1, plc0131_v2, plc0132_v1, plc0132_v2, plc0132_v3, plc0135_v1, plc0135_v2, plc0136, plc0137_h, plc0137_v1, plc0137_v2	[2022]
	Income, Unemployment benefit II	[2018-2020]	plc0138_v1, plc0138_v2, plc0139	[2022, 2023]
	Income, Wages	[2015,2019-2020]	plc0513	
	Income, Widow's pension	[2019-2020], [1995-2020], [1995-2001], [2019-2020], [1995-2020], [2002-2020], [1995-2020]	plc0273_v1, plc0273_v2, plc0274_h, plc0274_v1, plc0274_v2, plc0275_v1, plc0275_v2	[2022, 2023]
	Income, Widow's pension Months	[1995-2020]	plc0276	[2022, 2023]
	Inheritances	[2001,2019], [2017]	plc0375_v1, plc0375_v2	
	Inheritances, First Inheritance, amount	[2001], [2017], [2001,2019], [2017]	plc0383_v1, plc0383_v2, plc0384_v1, plc0384_v2	
	Inheritances, First Inheritance, last 15 years	[2017]	plc0376_v2, plc0377_v2, plc0378_v2, plc0379_v2, plc0380_v2, plc0381_v2, plc0382_v2	
	Inheritances, First Inheritance, once/ever	[2001,2019]	plc0376_v1, plc0377_v1, plc0378_v1, plc0379_v1, plc0380_v1, plc0381_v1, plc0382_v1, plc0383_h	
	Inheritances, From Whom	[2001]	plc0385, plc0395, plc0405	
	Inheritances, Future	[2001]	plc0406, plc0407	

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Inheritances, Second Inheritance, last 15 years	[2017]	plc0386_v2, plc0387_v2, plc0388_v2, plc0389_v2, plc0390_v2, plc0391_v2, plc0392_v2, plc0393_v2, plc0394_v2	
	Inheritances, Second Inheritance, once/ever	[2001,2019], [2001,2019]	[2001], plc0386_v1, plc0387_v1, plc0388_v1, plc0389_v1, plc0390_v1, plc0391_v1, plc0392_v1, plc0393_v1, plc0394_v1	
	Inheritances, Third Inheritance, amount	[2001], [2017], [2019], [2001,2019], [2017]	plc0403_v1, plc0403_v2, plc0403_v3, plc0404_v1, plc0404_v2	
	Inheritances, Third Inheritance, last 15 years	[2017]	plc0396_v2, plc0397_v2, plc0398_v2, plc0399_v2, plc0400_v2, plc0401_v2, plc0402_v2	
	Inheritances, Third Inheritance, once/ever	[2001,2019]	plc0396_v1, plc0397_v1, plc0398_v1, plc0399_v1, plc0400_v1, plc0401_v1, plc0402_v1, plc0403_h	
	Pension entitlements, company	[2013,2018], [2013], [2018], [2013,2018]	plc0432, plc0433, plc0434_v1, plc0434_v2, plc0435, plc0441, plc0442, plc0443, plc0444_v1, plc0444_v2, plc0445	
	Pension payments	[2013,2018], [2013], [2018]	plc0437, plc0438, plc0439_v1, plc0439_v2	

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Riester / Ruerup pension plans	(irregular) [2004-2020]	plc0313_h, plc0313_v1, plc0313_v2, plc0430, plc0431	[2022]
	Social security, Don't know	[2002,2007,2012,2017]	plc0009	
	Social security, Financial Security	[2002,2007,2012,2017]	plc0111, plc0112, plc0113, plc0114	[2022]
	Social security, Minimum Household Income	[1992,2002,2007,2012,2017], [1992], [2002,2007,2012,2017]	plc0001_h, plc0001_v1, plc0001_v2	[2022]
	Wage tax classification	[1991,1993,2004,2016], [2004,2016]	plc0091_h, plc0091_v1, plc0091_v2, plc0091_v3, plc0091_v4, plc0091_v5, plc0091_v6, plc0091_v7, plc0091_v8, plc0091_v9	
<i>Household Questionnaire</i>	Alimony	[2010]	hlc0091, hlc0092, hld0004, hld0005	
	Credit burden	[2005-2016], [2005-2011], [2011-2016]	hlc0115_h, hlc0115_v1, hlc0115_v2	
	Expenditures on Food, Month	(irregular) [1998-2020], [1998,2000-2001], (irregular) [2003-2020]	hlf0436_h, hlf0436_v1, hlf0436_v2	[2022, 2023]
	Expenditures on Food, Week	(irregular) [1998-2020], [1998,2000-2001], (irregular) [2003-2020]	hlf0435_h, hlf0435_v1, hlf0435_v2	[2022, 2023]
	Good/Low Income, Good Household Income	[1992,1997,2007,2018], [1992,1997], [2007,2018]	hlc0022_h, hlc0022_v1, hlc0022_v2	
	Good/Low Income, Insufficient Household Income	[1992,1997,2007,2018], [1992,1997], [2007,2018]	hlc0020_h, hlc0020_v1, hlc0020_v2	
	Good/Low Income, Just Sufficient Household Income	[1992,1997,2007,2018], [1992,1997], [2007,2018]	hlc0021_h, hlc0021_v1, hlc0021_v2	
	Good/Low Income, Poor Household Income	[1992,1997,2007,2018], [1992,1997], [2007,2018]	hlc0019_h, hlc0019_v1, hlc0019_v2	
	Good/Low Income, Very Good Household Income	[1992,1997,2007,2018], [1992,1997], [2007,2018]	hlc0023_h, hlc0023_v1, hlc0023_v2	
	Good/Low Income, Very Poor Household Income	[1992,1997,2007,2018], [1992,1997], [2007,2018]	hlc0018_h, hlc0018_v1, hlc0018_v2	

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Household income / expenses, Basic financial security in old age prev. Year	[2005-2020]	hlc0061_h, hlc0061_v1, hlc0061_v2, hlc0062, hlc0063, hlc0071	[2022, 2023]
	Household income / expenses, Child Allowance prev. year	[1984-2000], [1985-2020], [1985-2001], [2001-2020], [2002-2020]	hlc0040, hlc0041, hlc0042_h, hlc0042_v1, hlc0042_v2	[2022, 2023]
	Household income / expenses, Child Allowance today	[1995-2020], [2000-2009], [1995-2020], [1995-2001], [2010-2020], [2002-2020]	hlc0044_h, hlc0044_v1, hlc0044_v2, hlc0045_h, hlc0045_v1, hlc0045_v2	[2022, 2023]
	Household income / expenses, Child Benifit	[1984-2020], [1985-1990], [1991-1995], [1996-2020]	hlc0039_h, hlc0039_v1, hlc0039_v2, hlc0039_v3	[2022, 2023]
	Household income / expenses, Child Care Subsidy	[2009-2020], [2011-2020], [2010-2020]	hlc0046_h, hlc0046_v1, hlc0046_v2, hlc0046_v3, hlc0046_v4, hlc0047_h, hlc0047_v1, hlc0047_v2, hlc0124, hlc0125	[2022, 2023]
	Household income / expenses, Child Care Subsidy prev. Year	[2009-2020]	hlc0049_h, hlc0049_v1, hlc0049_v2, hlc0050_h, hlc0050_v1, hlc0050_v2, hlc0051_h, hlc0051_v1, hlc0051_v2, hlc0121, hlc0122, hlc0123	[2022, 2023]
	Household income / expenses, Compulsory Long Term Care Insurance	[2001-2020], [2001], [1996-2020], [1996], [1997-1999], [1997-2000], [1997-1998], [2000-2009], [1996-2020], [1996-2001], [2002-2020], [2010-2020], [2002-2020]	hlc0079_h, hlc0079_v1, hlc0079_v2, hlc0085_h, hlc0085_v1, hlc0085_v2, hlc0085_v3, hlc0085_v4, hlc0085_v5, hlc0085_v6, hlc0090_h, hlc0090_v1, hlc0090_v2	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Household income / expenses, Compulsory Long Term Care Insurance prev. Year	[2001-2020]	hlc0078	[2022, 2023]
	Household income / expenses, Family Members Support	[2001-2020]	hlc0077	[2022, 2023]
	Household income / expenses, Help with living costs	[1984-2009], [2009], [1995-2020], [1995-2020], [1995-2001], [1995-1998,2010-2020], [2002-2020]	[1984,1991-1985-1990], [1999-2009], [1995-2001], [1995-2001], hlc0066_h, hlc0066_v1, hlc0066_v2, hlc0067_h, hlc0067_v1, hlc0067_v2, hlc0068_h, hlc0068_v1, hlc0068_v2	[2022, 2023]
	Household income / expenses, Help with living costs prev. Year	[1984-2020], [1985-1990], [1992-2009], [1984-1991,2001-2020], [1984-1991,2001], [1984,1991,2010-2020], [2002-2020]	hlc0055_h, hlc0055_v1, hlc0055_v2, hlc0055_v3, hlc0059_h, hlc0059_v1, hlc0059_v2	[2022, 2023]
	Household income / expenses, Housing assistance	[1984-2020], [1985-1990], [1995-2020], [1999-2009], [1995-2020], [1995-2001], [1984,1991-2020], [1995-1998,2010-2020], [2002-2020]	hlc0080_h, hlc0080_v1, hlc0080_v2, hlc0083_h, hlc0083_v1, hlc0083_v2, hlc0084_h, hlc0084_v1, hlc0084_v2	[2022, 2023]
	Household income / expenses, Housing assistance prev. Year	[1984-2020], [1984-2001], [1984-2020], [2002-2020]	hlc0081, hlc0082_h, hlc0082_v1, hlc0082_v2	[2022, 2023]
	Household income / expenses, Income Bracket	[1999-2000], [2001-2002], [2003-2020]	hlc0006_v1, hlc0006_v2, hlc0006_v3	
	Household income / expenses, Monthly Household Income	[1984-2020], [1984-2001], [2002-2020]	hlc0005_h, hlc0005_v1, hlc0005_v2	[2022, 2023]
	Household income / expenses, Reduction of earning capacity	[2005-2020], [2005-2009], [2010-2020]	hlc0070_h, hlc0070_v1, hlc0070_v2	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Household income / expenses, Special Circumstances Assistance	[1984-2009], [1984-1991,2001], [2002-2009]	hlc0056_h, hlc0056_v1, hlc0056_v2, hlc0056_v3, hlc0058, hlc0060_h, hlc0060_v1, hlc0060_v2	
	Household income / expenses, Subsistence Support prev. year	[1984-1991,2001-2020]	hlc0057	[2022, 2023]
	Household income / expenses, Unemployment Subsidy II	[2005-2020], [2005-2009], [2010-2020], [2005-2020]	hlc0064_h, hlc0064_v1, hlc0064_v2, hlc0065	[2022, 2023]
	Household income / expenses, Unemployment Subsidy II prev. year	[2006-2020]	hlc0052, hlc0053, hlc0054	[2022, 2023]
	Household income / expenses; Number of Children	[1995-2020]	hlc0043	[2022, 2023]
	Income and expenses from renting / leasing	[1984-1990,1992-2020], [1984-1990,1992-2001], [1984-1990,1992-2020], [2002-2020]	hlc0007, hlc0008_h, hlc0008_v1, hlc0008_v2	[2022, 2023]
	Income and expenses from renting / leasing, Maintenance costs	[1984-1990,1992-2020], [1984-1990,1992-2001], [2002-2020], [2016-2020]	hlc0111_h, hlc0111_v1, hlc0111_v2, hlc0176	[2022, 2023]
	Income and expenses from renting / leasing, Redemption payments	[1985-1990,1992-2020], [1985-1990,1992-2001], [2002-2020], [2016-2020]	hlc0112_h, hlc0112_v1, hlc0112_v2, hlc0177	[2022, 2023]
	Income and expenses from renting / leasing, Tax Deduction	[2005-2020]	hlc0009, hlc0010	[2022, 2023]
	Inheritance, present, lottery prize	[2016-2020]	hlc0178, hlc0179, hlc0180, hlc0181, hlc0182, hlc0183	[2022, 2023]
	Investments, Building Society	[1984-2020]	hlc0105	[2022, 2023]
	Investments, Combined Savings	[1990]	hlc0097	
	Investments, Fixed Interest Securities	[1984-2020]	hlc0107	[2022, 2023]
	Investments, Interest and Dividend Income	[1984-2001], [2002-2020], [1985-2020]	hlc0013_v1, hlc0013_v2, hlc0014	[2022, 2023]
	Investments, Life Insurance	[1984-2020]	hlc0106	[2022, 2023]
	Investments, No Securities	[1984-2020]	hlc0093	[2022, 2023]

continues on next page

Table 3 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Investments, Non-response	[2003-2020], [2016-2020]	hlc0096, hlc0184	[2022, 2023]
	Investments, Operating Assets	[1984-2020]	hlc0104	[2022, 2023]
	Investments, Other Securities	[2001-2020]	hlc0108	[2022, 2023]
	Investments, Savings Account	[1984-2020]	hlc0098	[2022, 2023]
	Investments, Tax-deductible Loan	[2005-2020]	hlc0094, hlc0095	[2022, 2023]
	Ratio between income and expenditures	[2010-2013], [2016-2018], [2010]	hlc0024_v1, hlc0024_v2, hlc0030	
	Ratio between income and expenditures, Cap shortfall	[2010-2013], [2016]	hlc0029_v1, hlc0029_v2	
	Ratio between income and expenditures, Expenditure Surplus	[2010-2013], [2016-2018], [2010-2013], [2016]	hlc0027_v1, hlc0027_v2, hlc0028_v1, hlc0028_v2	
	Ratio between income and expenditures, Income Surplus	[2010-2013], [2016-2018], [2010-2013], [2016]	hlc0025_v1, hlc0025_v2, hlc0026_v1, hlc0026_v2	
	Repayments of loans	[1997-2020], [1997-2011], [1997-2020], [1997-2001], [2002-2011], [2011-2020]	hlc0113_h, hlc0113_v1, hlc0113_v2, hlc0114_h, hlc0114_v1, hlc0114_v2	[2022, 2023]
	Savings	[1992-2020], [2002-2014]	hlc0119_h, hlc0119_v1, hlc0119_v2, hlc0119_v3, hlc0119_v4, hlc0120_h, hlc0120_v1, hlc0120_v2, hlc0120_v3, hlc0120_v4	[2022, 2023]
	Savings, Deficit	[2021]	hfehlb1, hfehlb3, hfehlb5, hfehlb7	hfehlb2, hfehlb4, hfehlb6,

2.4 Family and Social Networks

As a household study, the SOEP offers rich information on family and social relationships and how these connections change in different stages of life. The modules dealing with family and social networks cover the entire life cycle beginning with pregnancy and childbirth and continuing through parenthood, family formation, friendships, marriage, divorce, and death, and also provide a wealth of additional information on important life events.



Family and Social Networks

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Childcare	[2003-2020]	suppartn	
	Circle of friends, sociodemographics	[2011,2016]	pld0104, pld0105, pld0106	
	Circle of friends, sociodemographics, age	[2006,2011-2012,2016]	pld0095, pld0096, pld0097	
	Circle of friends, sociodemographics, education	[2006,2011,2016]	pld0101, pld0102, pld0103	
	Circle of friends, sociodemographics, labor force status	[2006,2011-2012,2016]	pld0098, pld0099, pld0100	
	Circle of friends, sociodemographics, relations	(unregelmaessig) [1988-2016], [2012]	pld0089_h, pld0089_v2, pld0090_v1, pld0091_h, pld0091_v2	pld0089_v1, pld0090_h, pld0090_v2, pld0091_v1,
	Circle of friends, sociodemographics, sex	(unregelmaessig) [1988-2016]	pld0092, pld0093, pld0094	
	Circle of friends, trustworthy persons	[2006,2011,2016,2019], [2006,2011,2016,2019], (unregelmaessig) [1991-2019], [1991,1996], [2001], [2006,2011,2016,2019], (unregelmaessig) [1991-2019], [1991,1996], [2001], [2006,2011,2016,2019]	pld0062_v1, pld0064_v1, pld0066_v1, pld0068_v1, pld0070_v1, pld0072_v1, plf0049_h, plf0049_v2, plf0050_h, plf0050_v2, plf0050_v3	pld0063_v1, pld0065_v1, pld0067, pld0069_v1, pld0071_v1, pld0073, plf0049_v1, plf0049_v3, plf0050_v1,

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Circle of friends, trustworthy persons (M3-M5)	[2017-2019]	pld0062_v2, pld0063_v2, pld0064_v2, pld0065_v2, pld0066_v2, pld0068_v2, pld0069_v2, pld0070_v2, pld0071_v2, pld0072_v2	
	Circle of friends, trustworthy persons: conflicts	[2006,2011,2013,2016,2017-2018], [2006,2011,2013,2016,2017-2018]	pld0077, pld0078, pld0079, pld0080, pld0081, pld0082	
	Circle of friends, trustworthy persons: help	[2006,2011,2016,2017-2018]	pld0074, pld0075, pld0076	
	Circle of friends, trustworthy persons: unpleasant truths	[2006,2011,2016,2017-2018], [2006,2011,2016,2017-2018], [2006,2011,2016,2017-2018], [2006,2011,2016,2017-2018], [2011,2016,2019], [2017-2018], [2011,2016,2019], [2017-2018], [2006,2011,2016,2017-2018]	pld0083_v1, pld0083_v2, pld0084_v1, pld0084_v2, pld0085_v1, pld0085_v2, pld0086_v1, pld0086_v2, pld0087_v1, pld0087_v2, pld0088	
	Family changes	[1991,1996,2001], [1985-2020], [2003-2020]	pld0012, pld0013, pld0014, pld0038, pld0039, pld0040, pld0159, pld0160	[2022, 2023]
	Family changes, child-birth	[1999-2020]	pld0152, pld0153, pld0154	[2022, 2023]
	Family changes, death	[1999-2020]	pld0146, pld0147, pld0148, pld0161, pld0162, pld0163, pld0164, pld0165, pld0166, pld0167, pld0168, pld0169, pld0170, pld0171	[2022, 2023]
	Family changes, divorce	[1999-2020]	pld0140, pld0141, pld0142	[2022, 2023]
	Family changes, marriage	[1999-2020]	pld0134, pld0135, pld0136	[2022, 2023]
	Family changes, moving in	[1999-2020]	pld0137, pld0138, pld0139	[2022, 2023]
	Family changes, moving out (child)	[1999-2020]	pld0149, pld0150, pld0151	[2022, 2023]
	Family changes, other	[1985-1995,1999-2008,2010-2020]	pld0155, pld0156, pld0158	[2022, 2023]
	Family changes, separation	[1999-2020]	pld0143, pld0144, pld0145	[2022, 2023]
	Family network, aunt	[2006,2011,2016]	pld0115, pld0116	
	Family network, children	[1991,1996,2001,2003-2020]	pld0025, pld0026, pld0027, pld0028, pld0301i01, pld0301i02, pld0302	

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Family network, distance	[1991], [1991], [1996,2001], [2006,2011,2016]	plj0117_v1, plj0117_v2, plj0117_v3, plj0118_v1, plj0118_v2, plj0118_v3, plj0119_v1, plj0119_v2, plj0119_v3, plj0120, plj0121, plj0122_v1, plj0122_v2, plj0122_v3, plj0123_v1, plj0123_v2, plj0123_v3, plj0124_h, plj0124_v1, plj0124_v2, plj0124_v3, plj0125_v1, plj0125_v2, plj0125_v3, plj0126, plj0127_v1, plj0127_v2, plj0127_v3, plj0128, plj0129, plj0130_v1, plj0130_v2, plj0130_v3	
	Family network, grandchildren	[1991,1996,2001,2006,2011,2016]	pld0033, pld0034	
	Family network, grandparents	[2006,2011,2016]	pld0110, pld0111, pld0112, pld0113, pld0114	
	Family network, other relatives	[1991,1996,2001,2006,2011,2016]	pld0035, pld0036	
	Family network, parents	[1991,1996,2001,2006,2011,2016]	pld0023, pld0024	
	Family network, siblings	[1991,1996,2001,2006,2011,2016] [1991,1996,2001,2006,2011,2016] (unregelmässig) [1991-2016]	pld0029, pld0030, pld0031, pld0032	
	Family network, spouse	[1996,2001,2006,2011,2016] [1996,2001], [2006,2011,2016], [1991,1996,2001,2006,2011,2016]	pld0020, pld0021_h, pld0021_v1, pld0021_v2, pld0022, pld0107	
	Family network, stepparents	[2006,2011,2016]	pld0108, pld0109	
	Family network, uncle	[2006,2011,2016]	pld0117, pld0118	
	Friends	(unregelmässig) [2003-2020]	pld0047	[2022]
	Leisure and activities (with child)	[2003-2020]	tvhrs, tvyn	
	Marital / partnership status	[2019]	pld0131_v2, pld0131_v3, [2022] pld0132_v1, pld0132_v2, pld0133, pld0299, pld0300, plk0001_v2, plk0001_v3	
	Pregnancy and childbirth	[2003-2020]	pregplan	
	Sexual orientation	[2016]	pld0298_v1, pld0298_v2, pld0298_v3	
<i>Youth Questionnaire</i>	Allowance (Pocket money)	[2000-2020], [2002-2020]	, jl0022_h, jl0022_v2	[2022, 2023]

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Allowance (Pocket money, Deutschmark)	[2000-2001]	j10021_v1, j10022_v1	
	Childhood and parental home	[2000-2018]	j10273, j10279	
	Childhood and parental home (parent's education, ISCO-08)	[2013-2018]	j_isco08_jobfather, j_isco08_jobmother	
	Childhood and parental home (parent's education, ISCO-88)	[2000-2017]	j_isco88_jobfather, j_isco88_jobmother	
	Childhood and parental home (parent's education, KldB 2010)	[2013-2018]	j_kldb2010_jobfather, j_kldb2010_jobmother	
	Childhood and parental home (parent's education, KldB 92)	[2000-2017]	j_kldb92_jobfather, j_kldb92_jobmother	
	Childhood and parental home, father	[2000-2018], [2014-2017], [2014-2018], [2015-2018]	j10307, j10309, j10311, j10313_v1, j10313_v2, j10315, j10327_h, j10327_v1, j10327_v2, j10506, j10508, j10510, j10512, j10514, j10516, j10518, j10520, j10522	
	Childhood and parental home, mother	[2000-2018], [2014-2017], [2014-2018], [2015-2018]	j10304, j10308, j10310, j10312, j10314_v1, j10314_v2, j10316, j10328_h, j10328_v1, j10328_v2, j10507, j10509, j10511, j10513, j10515, j10517, j10519, j10521, j10523	
	Childhood and parental home, siblings	[2004-2012]	j10274, j10275, j10276, j10277, j10278, j10446, j10447, j10454, j10455, j10456, j10457, j10458, j10459, j10460, j10461, j10462, j10463, j10464, j10465, j10466, j10467, j10468, j10469, j10470, j10471, j10472, j10473, j10474, j10475, j10476, j10477, j10478, j10479, j10480, j10481, j10482, j10483, j10484, j10485, j10486, j10487, j10488, j10489, j10490, j10491, j10492, j10493, j10494, j10495, j11406, j11407, j11408, j11409, j11410, j11411	
	Parental interest in child's performance in school	[2000-2018]	j10167, j10168, j10169, j10170, j10171, j10172, j10173, j10174	[2022]

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
<i>Household Questionnaire</i>	Relationship to family members	[2001-2018]	j10026, j10027, j10028, j10029, j10030, j10031, j10032, j10033, j10034, j10040, j10041, j10043, j10044, j10045, j10046, j10047, j10048, j10049, j10050, j10051, j10052, j10053, j10054, j10055, j11043	[2022]
	Relationship to family members, conflicts	[2001-2018]	j10035_v1, j10035_v2, j10036_v1, j10036_v2, j10037_v1, j10037_v2, j10038_v1, j10038_v2, j10039_v1, j10039_v2	[2022]
	Care Provision, Outside Household		hle0025, hle0039, hle0040, hle0041, hle0042, hle0043, hle0044, hle0026, hle0027, hle0028, hle0029, hle0030, hle0031, hle0032, hle0033, hle0034, hle0035, hle0036, hle0037, hle0038	[2022]
	Child, COVID-19		kd_covquat_1, kc_covimpf_1, kd_covquat_2, kc_covimpf_2, kd_covquat_3, kc_covimpf_3, hcovidquat_4, hcovimpf1_4, kd_covquat_5, kc_covimpf_5, kd_covquat_6, kc_covimpf_6, kd_covquat_7, kc_covimpf_7, kd_covquat_8, kc_covimpf_8, kd_covquat_9, kc_covimpf_9, kd_covquat_10, kc_covimpf_10	[2022]

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Child, Citizenship		k_originde_1, k_origin2_1, k_origin_a_1, k_origin_b_1, k_oorigin_a_1, k_oorigin_b_1, k_originde_2, k_origin2_2, k_origin_a_2, k_origin_b_2, k_oorigin_a_2, k_oorigin_b_2, k_originde_3, k_origin2_3, k_origin_a_3, k_origin_b_3, k_oorigin_a_3, k_oorigin_b_3, k_originde_4, k_origin2_4, k_origin_a_4, k_origin_b_4, k_oorigin_a_4, k_oorigin_b_4, k_originde_5, k_origin2_5, k_origin_a_5, k_origin_b_5, k_oorigin_a_5, k_oorigin_b_5, k_originde_6, k_origin2_6, k_origin_a_6, k_origin_b_6, k_oorigin_a_6, k_oorigin_b_6, k_originde_7, k_origin2_7, k_origin_a_7, k_origin_b_7, k_oorigin_a_7, k_oorigin_b_7, k_originde_8, k_origin2_8, k_origin_a_8, k_origin_b_8, k_oorigin_a_8, k_oorigin_b_8, k_originde_9, k_origin2_9, k_origin_a_9, k_origin_b_9, k_oorigin_a_9, k_oorigin_b_9, k_originde_10, k_origin2_10, k_origin_a_10, k_origin_b_10, k_oorigin_a_10, k_oorigin_b_10	[2022]

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Childcare, Hours	Weekly	kc_wselbst_1, kc_wpart_1, kc_weltern_1, kc_wgeltern_1, kc_wgesch_1, kc_wverw_1, kc_wtagm_1, kc_wbezb_1, kc_wfreund_1, kc_wnone_1, kc_wselbsth_1, kc_wparth_1, kc_welternh_1, kc_wgelternh_1, kc_wgeschh_1, kc_wverwh_1, kc_wtagmh_1, kc_wbezbh_1, kc_wfreundh_1, kc_wselbst_2, kc_wpart_2, kc_weltern_2, kc_wgeltern_2, kc_wgesch_2, kc_wverw_2, kc_wtagm_2, kc_wbezb_2, kc_wfreund_2, kc_wnone_2, kc_wselbsth_2, kc_wparth_2, kc_welternh_2, kc_wgelternh_2, kc_wgeschh_2, kc_wverwh_2, kc_wtagmh_2, kc_wbezbh_2, kc_wfreundh_2, kc_wselbst_3, kc_wpart_3, kc_weltern_3, kc_wgeltern_3, kc_wgesch_3, kc_wverw_3, kc_wtagm_3, kc_wbezb_3, kc_wfreund_3, kc_wnone_3, kc_wselbsth_3, kc_wparth_3, kc_welternh_3, kc_wgelternh_3, kc_wgeschh_3, kc_wverwh_3, kc_wtagmh_3, kc_wbezbh_3, kc_wfreundh_3, kc_wselbst_4, kc_wpart_4, kc_weltern_4, kc_wgeltern_4, kc_wgesch_4, kc_wverw_4, kc_wtagm_4, kc_wbezb_4, kc_wfreund_4, kc_wnone_4, kc_wselbsth_4, kc_wparth_4, kc_welternh_4, kc_wgelternh_4, kc_wgeschh_4, kc_wverwh_4, kc_wtagmh_4, kc_wbezbh_4, kc_wfreundh_4, kc_wselbst_5, kc_wpart_5, kc_weltern_5, kc_wgeltern_5, kc_wgesch_5, kc_wverw_5, kc_wtagm_5, kc_wbezb_5, kc_wfreund_5, kc_wnone_5, kc_wselbsth_5, kc_wparth_5, kc_welternh_5, kc_wgelternh_5, kc_wgeschh_5, kc_wverwh_5, kc_wtagmh_5, kc_wbezbh_5, kc_wfreundh_5, kc_wselbst_6, kc_wpart_6, kc_weltern_6, kc_wgeltern_6, kc_wgesch_6,	[2022]
2.4. Family and Social Networks			kc_wverw_6, kc_wtagm_6, kc_wbezb_6, kc_wfreund_6, kc_wnone_6, kc_wselbsth_6, kc_wparth_6, kc_welternh_6, kc_wgelternh_6	35

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
<i>Mother and Child Instruments</i>	Household members		pre_vorname1, pre_vorname2, pre_vorname3, pre_vorname4, pre_vorname5, pre_vorname6, pre_vorname7, pre_vorname8, pre_vorname9, pre_vorname10	[2023]
	Allowance (Pocket money)	[2003-2020]	allow, allowpm, allowpw	
	Attitude toward parental role	[2003-2020]	bepar1, bepar10, bepar2, bepar3, bepar4, bepar5, bepar6, bepar8, bepar9	
	Attitude towards maternal role	[2003-2020]	change1, change2, change3, change4, change5, change6, change7, change8, health	
	Breastfeeding	[2003-2020]	breastf, breastfc, breastfm	
	Childcare	[2003-2020]	care10h, care11h, care12h, care19, care1h, care3h, care4h, care5h, care6h, care7h, care8h, care9h, maincare	
	Eating behavior (child)	[2003-2020]	eatsat1, eatsat2, eatsat3, eatson1, eatson2, eatson3, eatweek1, eatweek3	
	Frequency of leisure and other activities (child)	[2003-2020]	freqact1, freqact10, freqact11, freqact12, freqact13, freqact14, freqact15, freqact16, freqact17, freqact18, freqact19, freqact2, freqact20, freqact3, freqact4, freqact5, freqact6, freqact7, freqact8, freqact9	
	Friends (child)	[2003-2020]	frndadlt, frndchld	
	Language use	[2003-2020]	language	[2023]
	Leisure and activities (with child)	[2003-2020]	activ1, activ2, activ3, activ4, activ5, activ6, activ7, activ8, activ9	
	Parental interest in child's performance in school	[2003-2020]	conscho1, conscho2, conscho3, conscho4, conscho5, conscho6, conscho7	
	Parenting goals	[2003-2020]	edgoal1, edgoal10, edgoal11, edgoal12, edgoal13, edgoal14, edgoal15, edgoal16, edgoal17, edgoal18, edgoal2, edgoal3, edgoal4, edgoal5, edgoal6, edgoal7, edgoal8, edgoal9	
	Parenting style	[2003-2020]	edbeh1, edbeh10, edbeh11, edbeh12, edbeh13, edbeh14, edbeh15, edbeh16, edbeh17, edbeh18, edbeh2, edbeh3, edbeh4, edbeh5, edbeh6, edbeh7, edbeh8, edbeh9	
	Pregnancy and child-birth	[2003-2020]	birthpw, delivpl, nchild	

continues on next page

Table 4 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Relationship to other parent or child	[2003-2020]	biochild	

2.5 Health and Care

The modules on health and care cover doctor visits, sports and fitness, alcohol consumption, health insurance, health status, and grip strength, both on respondents themselves and on other individuals in the household, such as children and deceased household members.



Health and Care

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Additional private insurance	(irregular) [1999-2020]	ple0128_h , ple0128_v1 , ple0128_v2 , ple0130 , ple0132 , ple0134	[2022] ple0129 , ple0131 , ple0133 ,
	Alcohol consumption	[2006,2008,2010]	ple0090 , ple0092 , ple0177 , ple0178	ple0091 , ple0093 ,
	COVID-19, Diagnosed		ple0197 , ple0199	ple0198 , [2022]
	COVID-19, Vaccination		ple0201 , ple0202i02 , ple0202i03 , ple0203i01 , ple0203i02 , ple0203i03	ple0202i01 , [2022]
	Cannabis consumption		ple0196i01 , ple0196i02 , ple0196i03	[2022]
	Disabilities in everyday life (SF-12)	(irregular) [2002-2020]	ple0004 , ple0005	[2022]
	Electronic cigarette: liquid	[2020]	ple0195	[2022]
	Family doctor		ple0216	[2023]

continues on next page

Table 5 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Flu, Vaccination		pgrippim, pgrippim2, pgrippim3	[2023]
	Health insurance	[1999-2020]	ple0097, ple0099_h, ple0099_v1, ple0099_v2, ple0099_v3, ple0099_v4, ple0099_v5, ple0104_h, ple0104_v1, ple0104_v2, ple0104_v3, ple0104_v4, ple0104_v5, ple0104_v6, ple0104_v7, ple0160	[2022]
	Health insurance debts	[2017]	plc0567	
	Health insurance, private	[1984-1986], [1999-2020]	(irregular) ple0098_v1, ple0098_v2, ple0098_v3, ple0098_v4, ple0098_v5	[2022]
	Health restrictions	[2011-2013,2015-2020], [2012-2013,2015-2020]	ple0009, ple0162	[2023]
	Height and weight	(irregular) [2002-2020]	ple0006, ple0007	[2022]
	Hospital stays	[1984-1989,1991-1992,1994-2020]	ple0053, ple0055, ple0056	[2022, 2023]
	Illness	[2011,2013,2015,2017,2019], [2009,2011,2013,2015,2017,2019], [2009,2011,2013,2015,2017,2019]	ple0011, ple0012, ple0013, ple0014, ple0015, ple0016, ple0017, ple0018, ple0019, ple0020, ple0021, ple0022, ple0023, ple0024, ple0189	[2023]
	Individual health services	[2016,2018,2020]	ple0186	
	Nutritional awareness	[2016,2018,2020]	ple0179, ple0180, ple0181, ple0182	[2022, 2023]
	Private supplementary care insurance	[2016,2018,2020]	ple0183, ple0184, ple0185	
	Qualification for additional benefits	[1999-2011]	ple0121	
	Reduced ability to work	[1984-1989,1991-1992,1994-2020]	ple0040, ple0041	[2022, 2023]

continues on next page

Table 5 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Sickness notifications to employer	[1985-1989,1991-1992,1994-2020]	plb0024_h, plb0024_v1, plb0024_v2, plb0024_v3, ple0044_h, ple0044_v1, ple0044_v2, ple0046, ple0048, ple0049, ple0050, ple0051, ple0052, ple0174, ple0175	[2022, 2023]
	Smoking	[1998], [1998,2001], (irregular) [2002-2020], (irregular) [2002-2018], (irregular) [2006-2020], (irregular) [2004-2020], (irregular) [2002-2018], [2016,2018,2020]	ple0080_v1, ple0080_v2, ple0080_v3, ple0081_h, ple0081_v1, ple0081_v2, ple0082, ple0083, ple0084, ple0085, ple0086_v1, ple0086_v2, ple0086_v3, ple0086_v4, ple0089, ple0176	[2022]
	State of health	[1992,1994-2020]	ple0008	[2022, 2023]
	Stress and exhaustion (SF-12)	(irregular) [2002-2020], (irregular) [1984-2020]	ple0026, ple0027, ple0028, ple0029, ple0030, ple0031, ple0032, ple0033, ple0034, ple0035, ple0036	[2022]
	Visits to the doctor	[1988-1989,1991-1992,1995-2020], [1984-1989,1991-1992,1994-2020]	ple0072, ple0073	[2022, 2023]
	Visits to the doctor, Delay		ple0217, ple0218, ple0219i01, ple0219i02, ple0219i03, ple0219i04	[2023]
	Visits to the doctor, Dentist		ple0062	[2023]
	Visits to the doctor, Specialist		ple0060	[2023]
<i>Youth Questionnaire</i>	Height and weight	[2006-2020]	j10219, j10220	[2022, 2023]
	State of health	[2006-2020]	j10218	[2022, 2023]

continues on next page

Table 5 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
<i>Household Questionnaire</i>	COVID-19, Test Result	Positive [2021]	hkcovarzta_1, hkmnt1a_1, hkmnt2a_1, hkmnt3a_1, hkmnt4a_1, hk- covkra_1, hkco- stesta_1, hkco- varztb_2, hkmnt1b_2, hkmnt2b_2, hkmnt3b_2, hkmnt4b_2, hk- covkrb_2, hkco- stestb_2, hkco- varztc_3, hkmnt1c_3, hkmnt2c_3, hkmnt3c_3, hkmnt4c_3, hk- covkrc_3, hkco- stestc_3, hkco- varztd_4, hkmnt1d_4, hkmnt2d_4, hkmnt3d_4, hkmnt4d_4, hk- covkrd_4, hkco- stestd_4, hkco- varzte_5, hkmnt1e_5, hkmnt2e_5, hkmnt3e_5, hkmnt4e_5, hk- covkre_5, hkco- steste_5, hkco- varztf_6, hkmnt1f_6, hkmnt2f_6, hkmnt3f_6, hkmnt4f_6, hk- covkrf_6, hkco- stestf_6, hkco- varztg_7, hkmnt1g_7, hkmnt2g_7, hkmnt3g_7, hkmnt4g_7, hk- covkrg_7, hkco- stestg_7, hkco- varzth_8, hkmnt1h_8, hkmnt2h_8, hkmnt3h_8, hkmnt4h_8, hk- covkrh_8, hkco- stesth_8, hkco- varzti_9, hkmnt1i_9, hkmnt2i_9, hkmnt3i_9, hkmnt4i_9, hkco- covarztj_9, hkmnt1j_10, hkmnt2j_10, hkmnt3i_10	

Table 5 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	COVID-19, Quarantine	[2021]	hkcovhqa_1, hkcovhqb1a_1, hkcovhqb2a_1, hkcovhqb2b_2, hkcovhqb1b_2, hkcovhqb2c_3, hkcovhqb1c_3, hkcovhqb2d_4, hkcovhqb1d_4, hkcovhqb2e_5, hkcovhqb1e_5, hkcovhqb2f_6, hkcovhqb1f_6, hkcovhqb2g_7, hkcovhqb1g_7, hkcovhqb2h_8, hkcovhqb1h_8, hkcovhqb2i_9, hkcovhqb1i_9, hkcovhqb2j_10, hkcovhqb1j_10, hkcovhqb2k_10	

continues on next page

Table 5 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	COVID-19, Symptoms	[2021]	hktmntgr1a_1, hktmntgr2a_1, hktmntgr3a_1, hktmntgr4a_1, hktmntgr5a_1, hktmntgr6a_1, hktmntgr7a_1, hktmntgr8a_1, hkcovsb1_1, hkcovsb2_1, hkcovsb3_1, hkcovse1_1, hkcovse2_1, hkcovse3_1, hktmntgr1b_2, hktmntgr2b_2, hktmntgr3b_2, hktmntgr4b_2, hktmntgr5b_2, hktmntgr6b_2, hktmntgr7b_2, hktmntgr8b_2, hkcovsb1b_2, hkcovsb2b_2, hkcovsb3b_2, hkcovse1b_2, hkcovse2b_2, hkcovse3b_2, hktmntgr1c_3, hktmntgr2c_3, hktmntgr3c_3, hktmntgr4c_3, hktmntgr5c_3, hktmntgr6c_3, hktmntgr7c_3, hktmntgr8c_3, hkcovsb1_3, hkcovsb2_3, hkcovsb3_3, hkcovse1_3, hkcovse2_3, hkcovse3_3, hktmntgr1d_4, hktmntgr2d_4, hktmntgr3d_4, hktmntgr4d_4, hktmntgr5d_4, hktmntgr6d_4, hktmntgr7d_4, hktmntgr8d_4, hkcovsb1_4, hkcovsb2_4, hkcovsb3_4, hkcovse1_4, hkcovse2_4, hkcovse3_4, hktmntgr1e_5, hktmntgr2e_5, hktmntgr3e_5, hktmntgr4e_5, hktmntgr5e_5, hktmntgr6e_5, hktmntgr7e_5, hktmntgr8e_5, hkcovsb1_5	

Table 5 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Satisfaction with availability of care	[1997,2002,2008]	hlf0318	
<i>Mother and Child Instruments</i>	Health of child	[2003-2020]	chhealth, lstmedex, medaid3mb	
	Health of child, disorders	[2003-2020]	disord, disord1, disord2, disord3, disord4, disord5, disord6, disord7, disord8, disord9	
	Health of child, hospital stays	[2003-2020]	hospital12m, hospital3mb	
	Health of child, illnesses	[2003-2020]	ill0, ill10, ill11, ill12, ill13, ill14, ill15, ill2, ill31, ill32, ill4, ill5, ill6, ill7, ill8, ill9, illno	
	Height and weight of child	[2003-2020]	height, weight, weightb	
	Physical and mental health of mother	[2003-2020]	feeling1, feeling2, feeling3, feeling4	

2.6 Home, Amenities, and Contributions of Private HH

The housing, amenities, and household expenses modules provide wide-ranging information on everyday life including the type of dwelling and whether it is a rental property or owner-occupied; expenditures on personal hygiene, transportation, and vacations; and the division of household labor.



Home, Amenities and Contributions of Private HH

Questionnaire	Module	Years	Variables	Preview
<i>Household Questionnaire</i>	Childcare costs	[2010-2013,2015,2017-2019], [2010-2012], [2013,2015,2017-2019]	ks_cost_h, ks_cost_v2	ks_cost_v1,

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Childcare provider	[1987,1995,1997,2002,2002,2005,2007]	kd_insta_h, kd_insta_v1, kd_insta_v2, kd_insta_v3, kd_insta_v4, kd_insta_v5, kd_insta_v6, kd_insta_v7	
	Childcare situation	[1987,1997,1999-2002,2004-2020]	kc_care_h, kc_care_v1, kc_care_v2, kc_care_v3, kc_care_v4, kc_care_v5, kc_care_v6, kc_care_v7	
	Dependence on childcare hours	[2002]	kd_rely	
	Leisure activities, children	(unregelmaessig) [2006-2020]	ka06_art, ka06_mus, ka06_non, ka06_oth, ka06_spo, ka16_art, ka16_ctr, ka16_mus, ka16_non, ka16_org, ka16_sar, ka16_smu, ka16_sot, ka16_spo, ka16_ssp, ka16_sth, ka16_yth	
	Leisure costs, children	(unregelmaessig) [2002-2019], [2002,2005,2007], [2010-2013,2015,2017,2019], [2017-2018], [2010-2013,2015,2017-2019], [2010-2012], [2013,2015,2017,2019], [2017-2018]	kk_amtp_h, kk_amtp_v1, kk_amtp_v2, kk_amtp_v3, kk_cost_h, kk_cost_v1, kk_cost_v2, kk_cost_v3	
	Lunch, childcare	(unregelmaessig) [1997-2019], [1997,2002,2005,2007], [2010-2013,2015,2017,2019], [2017-2018]	kd_lunch_h, kd_lunch_v1, kd_lunch_v2, kd_lunch_v3	
	Lunch, school	[2010-2013,2015,2017-2019]	ks_lunch	
	School attendance by child	[1984-2020], [1984-1994], [1990], [1991], [1995-2020], [2017], [2016-2019]	ks_gen_h, ks_gen_v1, ks_gen_v2, ks_gen_v3, ks_gen_v4, ks_gen_v5, ks_spe	
	School provider and costs	(unregelmaessig) [1987-2019], [1987,1996], [2010-2012], [2013,2015,2017-2019]	ks_amtp_h, ks_amtp_v1, ks_amtp_v2, ks_amtp_v3	

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
<i>Household Questionnaire</i>	Change in residential situation	[1991-2020], [1991-1998], [2015-2020], [1991-2020], [1999-2020]	hlf0106, hlf0107_v1, hlf0523	hlf0107_h, hlf0107_v2, [2022, 2023]
	Changes in home fixtures and furnishings since last year	[2004,2006,2008,2010-2013]	hlc0116, hlc0117, hlf0159, hlf0164, hlf0165_h, hlf0165_v1, hlf0165_v2, hlf0166, hlf0167, hlf0223, hlf0224, hlf0225, hlf0226, hlf0227, hlf0228, hlf0229, hlf0230, hlf0231, hlf0232, hlf0233, hlf0234, hlf0235, hlf0236, hlf0237, hlf0238, hlf0244, hlf0245, hlf0246, hlf0247, hlf0248, hlf0249, hlf0250, hlf0251, hlf0252	
	Changes in home fixtures and furnishings since last year: Internet	[2000,2002,2004,2006], (unregelmaessig) [2000-2013], [2000], [2002,2004,2006,2008,2013]	hlf0169_v1, hlf0169_v2, hlf0169_v3, hlf0169_v4, hlf0169_v5, hlf0169_v6, hlf0169_v7, hlf0170_h, hlf0170_v1, hlf0170_v2	
	Changes in home fixtures and furnishings since last year: car	(unregelmaessig) [2000-2013], [2000], [2002,2004,2006,2008,2013], [2010-2013], [2010-2011]	hlf0209_h, hlf0209_v1, hlf0209_v2, hlf0210, hlf0211	
	Changes in home fixtures and furnishings since last year: cell phone	(unregelmaessig) [2000-2020], (unregelmaessig) [2000-2013], [2010-2013]	hlf0241_h, hlf0241_v1, hlf0241_v2, hlf0241_v3, hlf0241_v4, hlf0241_v5, hlf0241_v6, hlf0241_v7, hlf0241_v8, hlf0242, hlf0243	
	Changes in home fixtures and furnishings since last year: kitchen appliances	(unregelmaessig) [1998-2013]	hlf0214, hlf0215, hlf0216, hlf0217, hlf0218, hlf0219, hlf0220, hlf0221, hlf0222	
	Changes in home fixtures and furnishings since last year: motorcycle, moped	(unregelmaessig) [2000-2013], [2000], [2002,2004,2006,2008,2013], [2010-2013]	hlf0212_h, hlf0212_v1, hlf0212_v2, hlf0213	
	Changes in home fixtures and furnishings since last year: phone	(unregelmaessig) [1990-2020], (unregelmaessig) [1990-2013], [2013,2015], [2014], [2016-2020], (unregelmaessig) [1998-2013]	hlf0239_h, hlf0239_v1, hlf0239_v2, hlf0239_v3, hlf0239_v4, hlf0240	
	Cleaning or household help	[1991,1994,1999-2020], [2010-2020]	hlf0261, hlf0262	[2022, 2023]
	Comparison of old and new home	[1985-2013,2015,2017,2019-2020]	hlf0126, hlf0127, hlf0128, hlf0129, hlf0130, hlf0131, hlf0132	[2022, 2023]

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Consumption Module	[2010-2013]	hlf0172	
	Consumption Module: (unregelmässig)	[1990-2013], [1991]	hlf0163_h, hlf0163_v1, hlf0163_v2	
	Cars			
	Consumption Module: Clothes and Shoes	[2010-2013]	hlf0379, hlf0380, hlf0381, hlf0382	[2023]
	Consumption Module: Cosmetics	[2010-2013]	hlf0383, hlf0384, hlf0385, hlf0386	
	Consumption Module: Culture	[2010-2013]	hlf0399, hlf0400, hlf0401, hlf0402	
	Consumption Module: Education	[2010-2013]	hlf0395, hlf0396, hlf0397, hlf0398	
	Consumption Module: Food and Drinks	[2010-2013]	hlf0371, hlf0372, hlf0373, hlf0374, hlf0375, hlf0376, hlf0377, hlf0378	
	Consumption Module: Furniture	[2010-2013]	hlf0427, hlf0428, hlf0429, hlf0430	
	Consumption Module: Health	[2010-2013]	hlf0387, hlf0388, hlf0389, hlf0390	
	Consumption Module: Hobby	[2010-2013]	hlf0403, hlf0404, hlf0405, hlf0406	
	Consumption Module: Holiday	[2010-2013]	hlf0407, hlf0408, hlf0409, hlf0410	[2023]
	Consumption Module: Insurance	[2010-2013]	hlf0411, hlf0412, hlf0413, hlf0414, hlf0415, hlf0416, hlf0417, hlf0418	
	Consumption Module: Internet	[2010-2013]	hlf0168, hlf0171	
	Consumption Module: Other	[2010-2013]	hlf0431, hlf0432, hlf0433, hlf0434	
	Consumption Module: Repair	[2010-2013]	hlf0419, hlf0420, hlf0421, hlf0422	
	Consumption Module: Telecommunication	[2010-2013]	hlf0391, hlf0392, hlf0393, hlf0394	
	Consumption Module: Transportation	[2010-2013]	hlf0423, hlf0424, hlf0425, hlf0426	
	Costs of comparable rental homes	[1984-2002, 2005-2014]	hlf0094	
	Costs of home ownership	[1986-2014, 2016-2020]	hlf0084, hlf0090_h, hlf0090_v1, hlf0090_v2, hlf0601, hlf0602, hlf0603, hlf0604, hlf0605	[2022, 2023]
	Dwelling / building type	[1986-2020], [1986-1990, 1991-2008, 2010-2018], [2009], [2016-2020]	hlf0155_h, hlf0155_v1, hlf0155_v2, hlf0596	
	Financial burden of home ownership	[2016]	hlf0606	
	Financial burden of home rental	[2016]	hlf0611	

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Government-subsidized housing	[1984-2020], [1998-2015], [1986-2002,2008-2020]	hlf0011_h, hlf0011_v1, hlf0011_v2, hlf0011_v3, hlf0011_v4, hlf0073	[2022, 2023]
	Groceries: Organic		hli0150	[2023]
	Hereditary lease interest	[1984-2013,2015-2020], [1986-2020], [1986-1987], [1988-1992], [1993-2020]	hlf0016, hlf0154_h, hlf0154_v1, hlf0154_v2, hlf0154_v3	[2022, 2023]
	Home fixtures and furnishings	[1991], [2015-2020]	hlf0023, hlf0024, hlf0025, hlf0026, hlf0027, hlf0028, hlf0029, hlf0030, hlf0031, hlf0032, hlf0033, hlf0034, hlf0035, hlf0036, hlf0037, hlf0529, hlf0530, hlf0531	[2022, 2023]
	Home ownership / rental	[1984-2020], [2003-2020]	hlf0001_h, hlf0001_v1, hlf0001_v2, hlf0001_v3, hlf0006, hlf0015	[2022, 2023]
	Home ownership / rental, ownership acquisition	[1984-2020], [1984-1990,1999-2001], [1991-1998], [1991-1997]	hlf0007_h, hlf0007_v1, hlf0007_v2, hlf0007_v3	
	Home ownership / rental, ownership change		hlf0703	[2023]
	Home ownership / rental, ownership transfer	[1999-2020]	hlf0007_v4, hlf0009	[2022, 2023]
	Homeowner	[1990-2020], [1990-2002,2005-2012], [2003-2004], [2013-2020]	hlf0013_h, hlf0013_v1, hlf0013_v2, hlf0013_v3	
	Loans, mortgages, building loan agreements	[1985-2020], [1985-1990, 1991-1998], [1999-2020], [1984-2020], [1984-1990, 1991-2001], [2002-2020]	hlf0087_h, hlf0087_v1, hlf0087_v2, hlf0088_h, hlf0088_v1, hlf0088_v2	
	Material deprivation	(unregelmässig) [2001-2015], [2016-2019]	hlf0175, hlf0177, hlf0178_h, hlf0178_v1, hlf0178_v2, hlf0178_v3, hlf0178_v4, hlf0178_v5, hlf0179, hlf0180, hlf0181, hlf0183, hlf0185, hlf0186, hlf0187, hlf0188, hlf0189, hlf0190, hlf0191, hlf0192, hlf0193, hlf0194, hlf0195, hlf0444, hlf0613, hlf0622	[2023]
	Modernization costs	[2016-2020]	hlf0599	
	Monthly rent, heating, other expenses	[2016-2020]	hlf0607, hlf0608, hlf0610	[2022, 2023]

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Monthly rent, heating, other expenses (Deutschmark)	[2002-2014,2016-2020]	hlf0081_v2	[2022, 2023]
	Monthly rent, heating, other expenses: electricity	[2010-2014,2016-2020]	hlf0078, hlf0079	[2022, 2023]
	Monthly rent, heating, other expenses: heating	[1986-2014,2016-2020], [1986-1990,1997-2001], [2002-2014,2016-2020]	hlf0069_h, hlf0069_v5	[2022, 2023]
	Monthly rent, heating, other expenses: heating and hot water	[1990,1996], [1991-1995]	hlf0069_v2, hlf0069_v4	
	Monthly rent, heating, other expenses: other	[1991-2014,2016-2020], [1991-2001], [1996-2014,2016-2020]	hlf0081_h, hlf0082	
	Monthly rent, heating, other expenses: rent	[1984-2020], [1984-2001]	hlf0074_h, hlf0074_v1	
	Monthly rent, heating, other expenses: rent (Deutschmark)	[2002-2020]	hlf0074_v2	[2022, 2023]
	Name and birth of children	[1984-2020], [2017,2020]	hlk0044_v1, hlk0044_v2	
	Number of books in household	[2001,2006,2011,2016]	hlf0197	
	Persons in household in need of care	[1985-2020], [2015-2020], [2016-2020]	hlf0291, hlf0292, hlf0300, hlf0301, hlf0302, hlf0303, hlf0304, hlf0315_h, hlf0315_v1, hlf0315_v2, hlf0315_v3, hlf0317_h, hlf0317_v1, hlf0317_v2, hlf0317_v3, hlf0319, hlf0320, hlf0321, hlf0322, hlf0331, hlf0332, hlf0369, hlf0370_h, hlf0370_v1, hlf0370_v2, hlf0446, hlf0448, hlf0595, hlf0631	[2022, 2023]
	Pets	[2006,2011,2016], [2006,2011]	hlf0196, hlf0254, hlf0255, hlf0256, hlf0257, hlf0258, hlf0259	
	Photovoltaic and solar thermal system	[2015-2016,2020]	hlf0532, hlf0535, hlf0536, hlf0537, hlf0538, hlf0539	

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Reasons for moving	[1985-2013, 2015, 2017-2020]	hlf0108_h, hlf0108_v1, hlf0108_v10, hlf0108_v11, hlf0108_v12, hlf0108_v13, hlf0108_v14, hlf0108_v15, hlf0108_v2, hlf0108_v3, hlf0108_v4, hlf0108_v5, hlf0108_v6, hlf0108_v7, hlf0108_v8, hlf0108_v9, hlf0109, hlf0124, hlf0125	
	Reasons for moving, comparison of old and new home	[2015, 2017-2020], [2015, 2017], [2015, 2017, 2019-2020]	hlf0524, hlf0525, hlf0526	[2022, 2023]
	Residential area	(unregelmaessig) [1986-2019], [1994, 1999, 2004, 2009, 2014, 2016-2020], [1994, 1999, 2004, 2009, 2018-2020], [1986-1987], [1988-1992], [1986, 1994, 1999, 2004, 2014, 2019], [1993-2020]	hlf0148, hlf0149, hlf0150, hlf0151, hlf0152, hlf0153_h, hlf0153_v1, hlf0153_v2, hlf0153_v3, hlj0004_v1, hlj0004_v2	[2022, 2023]
	Residential area, distances	(unregelmaessig) [1986-2019]	hlf0135, hlf0136, hlf0137, hlf0138, hlf0139, hlf0140, hlf0141, hlf0142, hlf0143, hlf0144, hlf0145, hlf0146, hlf0147	
	Residential area, neighbors	(unregelmaessig) [1986-2019]	hld0001, hld0002, hld0003	
	Savings		hlf0722	[2023]
	Second Residence	[2011, 2016], [2011], [2011, 2016]	hlf0156, hlf0157, hlf0158	
	Size and condition of home, rooms	[1984-2020], [1984-1990, 1998-2020]	hlf0021_h, hlf0021_v1, hlf0021_v2, hlf0021_v3	[2022, 2023]
	Size and condition of home, size	[1984-2020]	hlf0018, hlf0019_h, hlf0019_v1, hlf0019_v2, hlf0019_v3, hlf0071_h, hlf0071_v1, hlf0071_v2, hlf0071_v3	[2022, 2023]
	Spending, current		hlf0428_v2, hlf0380, hlf0404_v4, hlf0424_v3, hlf0384_v3, hlf0408	[2023]
	Spending, future		hlf0714, hlf0715, hlf0716, hlf0717, hlf0718, hlf0719, hlf0720, hlf0721	[2023]
	Type of energy used in household		, hlf0591	[2023]

continues on next page

Table 6 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Type of energy used in household: biomass	[2015,2020]	hlf0582, hlf0583, hlf0584, hlf0585, hlf0586	[2023]
	Type of energy used in household: coal	[2015]	hlf0570, hlf0571, hlf0572_v1, hlf0572_v2, hlf0573, hlf0574, hlf0575	[2023]
	Type of energy used in household: district heat	[2015,2020]	hlf0540, hlf0541, hlf0542, hlf0543, hlf0544	[2023]
	Type of energy used in household: electricity	[2015]	hlf0557, hlf0558, hlf0559_v1, hlf0559_v2, hlf0560, hlf0561, hlf0562, hlf0563	[2023]
	Type of energy used in household: environmental heat	[2015,2020]	hlf0589, hlf0590	[2023]
	Type of energy used in household: gas	[2015]	hlf0545, hlf0546, hlf0547_v1, hlf0547_v2, hlf0548, hlf0549, hlf0550	[2023]
	Type of energy used in household: heating oil	[2015]	hlf0564, hlf0565, hlf0566_v1, hlf0566_v2, hlf0567, hlf0568, hlf0569	[2023]
	Type of energy used in household: liquid gas	[2015]	hlf0551, hlf0552, hlf0553_v1, hlf0553_v2, hlf0554, hlf0555, hlf0556	[2023]
	Type of energy used in household: solar	[2015,2020]	hlf0587, hlf0588	[2023]
	Type of energy used in household: wood/pellets	[2015]	hlf0576, hlf0577, hlf0578_v1, hlf0578_v2, hlf0579, hlf0580, hlf0581	[2023]

2.7 Education and Qualification

Education is one of the cornerstones of society today. The education, training, and qualifications modules provide extensive information on educational attainment and outcomes, the level of completed education and training, reasons for not completing education or training, educational goals, and much more, along with data on children's skill development, for instance, whether they are able to speak in full sentences or use scissors.



Education and Qualification

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Completed education and training, Current year	[1993-2020]	plg0074	[2022, 2023]
	Completed education and training, Degree Nonresponse	[2014-2020]	plg0268	[2022, 2023]
	Completed education and training, Degree Recognized in Germany	[2015-2020]	plg0284	[2022, 2023]
	Completed education and training, Education Completed	[2000-2020], [2005-2020]	plg0076, plg0077	[2022, 2023]
	Completed education and training, Field of study	[1985-2009]	p_field	
	Completed education and training, General school-leaving certificate	[1985-2020], [1991, 1985-2020]	plg0078_h, plg0078_v2	plg0078_v1, [2022, 2023]
	Completed education and training, Nonresponse	[2004-2020]	plg0075	
	Completed education and training, Since previous year	[1985-2020], [1993-2020]	plg0072, plg0073	[2022, 2023]
	Completed education and training, Type of college degree	[1985-2009]	p_degree	
	Completed education and training, University school-leaving certificate	[1985-2008], [2009-2013], [2014-2020]	plg0079_v1, plg0079_v4	plg0079_v3, [2022, 2023]
	Further education, course details and motives for participation, Course 1 Adjusting to new Demands	[1989,1993,2000,2004,2	plg0144	
	Further education, course details and motives for participation, Course 1 Applicability	[2004,2008]	plg0117	
	Further education, course details and motives for participation, Course 1 Career Change	[1989,1993,2000,2004,2	plg0135	
	Further education, course details and motives for participation, Course 1 Duration	[1989,1993]	plg0120_v1, plg0120_v2, plg0120_v3, plg0120_v4	

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Further education, course details and motives for participation, Course 1 During Working Hours	[1989,1993,2000,2004,2008]	plg0150	
	Further education, course details and motives for participation, Course 1 Financial Support	[1989,1993,2000,2004,2008]	plg0177_h, plg0177_v2, plg0177_v4, plg0177_v6	plg0177_v1, plg0177_v3, plg0177_v5,
	Further education, course details and motives for participation, Course 1 Hours of Instruction	[1989,1993,2000,2004,2008]	plg0129	
	Further education, course details and motives for participation, Course 1 Introduction New Job	[1989,1993,2000,2004,2008]	plg0138	
	Further education, course details and motives for participation, Course 1 No Own Costs	[2000,2004,2008]	plg0174	
	Further education, course details and motives for participation, Course 1 Organizer	[1989,1993,2000,2004,2008]	plg0155_h, plg0155_v10, plg0155_v3, plg0155_v5, plg0155_v7, plg0155_v9	plg0155_v1, plg0155_v2, plg0155_v4, plg0155_v6, plg0155_v8,
	Further education, course details and motives for participation, Course 1 Other	[1989,1993,2000,2004,2008]	plg0147	
	Further education, course details and motives for participation, Course 1 Own Costs	[1989,1993,2000,2004,2008]	plg0169_h, plg0169_v2	plg0169_v1,
	Further education, course details and motives for participation, Course 1 Participation Certificate	[1989,1993,2000,2004,2008]	plg0184	
	Further education, course details and motives for participation, Course 1 Qualification for Promotion	[1989,1993,2000,2004,2008]	plg0141	

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Further education, course details and motives for participation, Course 1 Start	[1989,1993,2000,2004,2 [1989,1993], [2000,2004,2008], [1989,1993,2000,2004,2	plg0108_h, plg0108_v1, plg0108_v2, plg0111	
	Further education, course details and motives for participation, Course 1 Telecourse	[1989,1993,2000,2004,2	plg0132	
	Further education, course details and motives for participation, Course 1 pay off	[2004,2008]	plg0114	
	Further education, course details and motives for participation, Course 2 Adjusting to new Demands	[1989,1993,2000,2004,2	plg0145	
	Further education, course details and motives for participation, Course 2 Applicability	[2004,2008]	plg0118	
	Further education, course details and motives for participation, Course 2 Career Change	[1989,1993,2000,2004,2	plg0136	
	Further education, course details and motives for participation, Course 2 Duration	[1989,1993]	plg0121_v1, plg0121_v2, plg0121_v3, plg0121_v4	
	Further education, course details and motives for participation, Course 2 During Working Hours	[1989,1993,2000,2004,2	plg0151	
	Further education, course details and motives for participation, Course 2 Financial Support	[2000,2004,2008], [2000], [2004,2008]	plg0182_h, plg0182_v1, plg0182_v2	
	Further education, course details and motives for participation, Course 2 Hours of Instruction	[1989,1993,2000,2004,2	plg0130	
	Further education, course details and motives for participation, Course 2 Introduction New Job	[1989,1993,2000,2004,2	plg0139	

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Further education, course details and motives for participation, Course 2 No Own Costs	[2000,2004,2008]	plg0175	
	Further education, course details and motives for participation, Course 2 Organizer	[2000,2004,2008]	plg0164	
	Further education, course details and motives for participation, Course 2 Other	[1989,1993,2000,2004,2	plg0148	
	Further education, course details and motives for participation, Course 2 Own Costs	[2000,2004,2008], [2000], [2004,2008]	plg0171_h, plg0171_v2	plg0171_v1,
	Further education, course details and motives for participation, Course 2 Participation Certificate	[1989,1993,2000,2004,2	plg0185	
	Further education, course details and motives for participation, Course 2 Qualification for Promotion	[1989,1993,2000,2004,2	plg0142	
	Further education, course details and motives for participation, Course 2 Start	[1989,1993,2000,2004,2 [1989,1993], [2000,2004,2008], [1989,1993,2000,2004,2	plg0109_h, plg0109_v1, plg0109_v2, plg0112	
	Further education, course details and motives for participation, Course 2 Telecourse	[1989,1993,2000,2004,2	plg0133	
	Further education, course details and motives for participation, Course 2 pay off	[2004,2008]	plg0115	
	Further education, course details and motives for participation, Course 3 Adjusting to new Demands	[1989,1993,2000,2004,2	plg0146	
	Further education, course details and motives for participation, Course 3 Applicability	[2004,2008]	plg0119	

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Further education, course details and motives for participation, Course 3 Career Change	[1989,1993,2000,2004,2008]	plg0137	
	Further education, course details and motives for participation, Course 3 Duration	[1989,1993]	plg0122_v1, plg0122_v2, plg0122_v3, plg0122_v4	
	Further education, course details and motives for participation, Course 3 During Working Hours	[1989,1993,2000,2004,2008]	plg0152	
	Further education, course details and motives for participation, Course 3 Financial Support	[2000,2004,2008], [2000], [2004,2008]	plg0183_h, plg0183_v2	plg0183_v1,
	Further education, course details and motives for participation, Course 3 Hours of Instruction	[1989,1993,2000,2004,2008]	plg0131	
	Further education, course details and motives for participation, Course 3 Introduction New Job	[1989,1993,2000,2004,2008]	plg0140	
	Further education, course details and motives for participation, Course 3 No Own Costs	[2000,2004,2008]	plg0176	
	Further education, course details and motives for participation, Course 3 Organizer	[2000,2004,2008]	plg0165	
	Further education, course details and motives for participation, Course 3 Other	[1989,1993,2000,2004,2008]	plg0149	
	Further education, course details and motives for participation, Course 3 Own Costs	[2000,2004,2008], [2000], [2004,2008]	plg0172_h, plg0172_v2	plg0172_v1,

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Further education, course details and motives for participation, Course 3 Participation Certificate	[1989,1993,2000,2004,2008]	plg0186	
	Further education, course details and motives for participation, Course 3 Qualification for Promotion	[1989,1993,2000,2004,2008]	plg0143	
	Further education, course details and motives for participation, Course 3 Start	[1989,1993,2000,2004,2008], [1989,1993,2000,2004,2008]	plg0110_h, plg0110_v1, plg0110_v2, plg0113	
	Further education, course details and motives for participation, Course 3 Telecourse	[1989,1993,2000,2004,2008]	plg0134	
	Further education, course details and motives for participation, Course 3 pay off	[2004,2008]	plg0116	
	Further education, course details and motives for participation, Course Subject / Content	[1989,1993]	plg0153	
	Further education, course details and motives for participation, Financial Support	[1989,1993]	plg0167, plg0168	
	Further education, course details and motives for participation, Initiative for taking Course	[1989,1993]	plg0166	
	Further education, course details and motives for participation, Pay Off	[1989]	plg0187, plg0188, plg0189	
	Further training measures, Further professional training	[2014-2015], [2016-2020]	plg0269_v1, plg0269_v2	[2022]
	Further training measures, Trainig measures prev year	[2014-2020]	plg0270, plg0271	[2022]
	Further training, financing	[2015-2018,2020]	plg0285, plg0286, plg0287, plg0288, plg0289, plg0290, plg0291	

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables		Preview
	Further training, motivation		plg0373i01, plg0375i01, plg0377i01, plg0374i02, plg0376i02, plg0373i03, plg0375i03, plg0377i03	plg0374i01, plg0376i01, plg0373i02, plg0375i02, plg0377i02, plg0374i03, plg0376i03,	[2022]
	Further training, online		plg0372i01, plg0372i03	plg0372i02,	[2022]
	Further training, reasons for not taking part	[2014]	plg0277, plg0278, plg0279, plg0280, plg0281		
	Further training, results		plg0378i01, plg0378i03, plg0379i02, plg0380i01, plg0380i03, plg0381i02, plg0381i03	plg0378i02, plg0379i01, plg0379i03, plg0380i02, plg0381i01,	[2022]
	Further training, suggested / provided by employer	[2014]	plg0274		
	Further training, suggested / provided by employere	[2014]	plg0273		
	Lifelong learning	[2014]	plg0266		
	Vocational training, Currently in education / training	[1984-2020], [2020]	plg0012_v1, plg0012_v2		[2022, 2023]
	Vocational training, General school	[1984-2015], [2016-2020]	plg0013_v1, plg0013_v3		
	Vocational training, Scholarship	[2007-2020]	plg0015_h, plg0015_v2, plg0015_v4	plg0015_v1, plg0015_v3,	
	Vocational training, University	[1984-1995], [1999-2008], [2009-2012], [2013-2020]	plg0014_v1, plg0014_v3, plg0014_v5, plg0014_v7	plg0014_v2, plg0014_v4, plg0014_v6,	[2022, 2023]
<i>Youth Questionnaire</i>	Education and career plans	[2013-2019], [2000-2017], [2013-2019], [2000-2017]	j_isco08_jobwish, j_isco88_jobwish, j_kldb2010_jobwish, j_kldb92_jobwish		
	Education and career plans, Apprenticeship	[2000-2020]	j10177, j10182, j10183, j10203		[2022]
	Education and career plans, Career Training	[2014-2020]	j10438, j10439		[2022]
	Education and career plans, Engineering school	[2013-2020]	j10440, j10441		[2022]
	Education and career plans, Exploring Skills	[2001-2020]	j10205		[2022, 2023]

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Education and career plans, Financial Independence	[2000-2020]	j10197, j10198	[2022]
	Education and career plans, Informed about Future Occupation	[2001-2020]	j10201	[2022]
	Education and career plans, No Particular Plans	[2001-2020]	j10204	[2022, 2023]
	Education and career plans, Occupational Foundation	[2000-2020]	j10179	[2022]
	Education and career plans, Occupational Integration	[2000-2020]	j10178, j10180, j10181	[2022]
	Education and career plans, Parents Suggestions	[2001-2020]	j10202	[2022, 2023]
	Education and career plans, Preferred Occupation	[2000-2020]	j10199	[2022]
	Education and career plans, Vocational School	[2000-2020]	j10184, j10185	[2022]
	Education and career plans, Volunteering	[2000-2020]	j10186, j10187	[2022]
	Educational aspirations, Apprenticeship	[2014-2020]	j10504	[2022]
	Educational aspirations, Aspired school-leaving qualification	[2000-2020], [2000-2001-2020], [2000-2020]	j10130_h, j10130_v1, j10130_v2, j10131	[2022]
	Educational aspirations, Career Training	[2000-2020]	j10193	[2022, 2023]
	Educational aspirations, Civil Servant Training	[2000-2020]	j10192	[2022, 2023]
	Educational aspirations, Completed Apprenticeship	[2000-2020]	j10189	[2022, 2023]
	Educational aspirations, Engineering school	[2000-2020]	j10194	[2022, 2023]
	Educational aspirations, Future Apprenticeship	[2003-2020], [2000-2020]	j10188, j10196	[2022, 2023]
	Educational aspirations, Trade and Technical School	[2000-2020]	j10191	[2022, 2023]

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Educational aspirations, University	[2000-2020]	j10195	[2022, 2023]
	Educational aspirations, Vocational School	[2000-2020]	j10190	[2022, 2023]
	School, attendance & homework		j10125_v4, j10125_v5	[2023]
	School, attendance & homework, Class Representative	[2000-2020]	j10139	[2022]
	School, attendance & homework, Course type	[2000-2020]	j10162, j10163	[2022, 2023]
	School, attendance & homework, Extracurricular activities	[2000-2020]	j10141, j10142, j10143, j10144, j10145, j10146	[2022]
	School, attendance & homework, First Foreign Language	[2000-2020], [2000-2005], [2006-2020]	j10132_h, j10132_v1, j10132_v2, j10132_v3	[2022]
	School, attendance & homework, Foreign Language		j11972	[2023]
	School, attendance & homework, Grade / Year	[2014-2020]	j10434	[2022, 2023]
	School, attendance & homework, Grades / Points	[2000-2020]	j10152, j10153, j10154, j10155, j10156, j10157	[2022, 2023]
	School, attendance & homework, Graduation		j11973	[2023]
	School, attendance & homework, Number classmates	[2000-2020], [2000-2018]	j10176_h, j10176_v2	j10176_v1,
	School, attendance & homework, Private School	[2001-2018]	j10138	
	School, attendance & homework, Satisfaction with grades	[2000-2020]	j10147, j10148, j10149, j10150	[2022, 2023]
	School, attendance & homework, School attendance abroad	[2000-2020]	j10137_h, j10137_v1, j10137_v2, j10435, j10436	[2022]
	School, attendance & homework, School recommendation	[2001-2018]	j10151	
	School, attendance & homework, School-leaving certificate	[2000-2020], [2000-2011], [2012-2020]	j10127_h, j10127_v2	j10127_v1, [2022]

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	School, attendance & homework, Second Foreign Language	[2000-2020], [2000], [2001-2005], [2006-2020]	j10133_h, j10133_v1, j10133_v2, j10133_v3	[2022]
	School, attendance & homework, Still in School	[2000-2020], [2000-2002], [2003-2005], [2006-2020]	j10125_h, j10125_v1, j10125_v2, j10125_v3	[2022]
	School, attendance & homework, Student Body President	[2000-2020]	j10140	[2022]
	School, attendance & homework, Year of leaving school	[2000-2020]	j10126	[2022]
	School, attendance & homework, Year repeated	[2000-2020]	j10164, j10165, j10166	[2022]
<i>Household Questionnaire</i>	Online learning materials	[2021]	hdigis1a_1, hdigis3a_1, hdigis2b_2, hdigis1c_3, hdigis3c_3, hdigis2d_4, hdigis1e_5, hdigis3e_5, hdigis2f_6, hdigis1g_7, hdigis3g_7, hdigis2h_8, hdigis1i_9, hdigis3i_9, hdigis2j_10, hdigis3j_10	hdigis2a_1, hdigis1b_2, hdigis3b_2, hdigis2c_3, hdigis1d_4, hdigis3d_4, hdigis2e_5, hdigis1f_6, hdigis3f_6, hdigis2g_7, hdigis1h_8, hdigis3h_8, hdigis2i_9, hdigis1j_10, idegrad1, idegrad2, idegrad3
<i>Mother and Child Instruments</i>	Educational aspirations, Ideal school completion	[2003-2020]	idegrad1, idegrad2, idegrad3	
	Educational aspirations, intermediate secondary	[2003-2020]	probgra2	
	Educational aspirations, lower secondary	[2003-2020]	probgra1	
	Educational aspirations, upper secondary	[2003-2020]	probgra3	
	School and homework	[2003-2020]	scolcon1, scolcon2, scolcon3, scolcon4, scolcon5, scolcon6, scolcon7	
	School and homework, Comprehensive school	[2003-2020]	curscol7	
	School and homework, Grammar secondary class	[2003-2020]	curscol6	

continues on next page

Table 7 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	School and homework, Intermediate secondary school	[2003-2020]	curscol5	
	School and homework, Last report mark	[2003-2020]	lamark , matmark , nomark	
	School and homework, Other school	[2003-2020]	curscol8	
	School and homework, Place	[2003-2020]	hwplace	
	School and homework, Primary school	[2003-2020]	curscol1	
	School and homework, Second general school	[2003-2020]	curscol4	
	School and homework, Special pedagogic concept	[2003-2020]	curscol2	
	School and homework, Special school	[2003-2020]	curscol3	
	School and homework, Support	[2003-2020]	hwsupprt	
	School enrollment	[2003-2020]	scenrolm , scenroln , scenroly	

2.8 Attitudes, Values, and Personality

The attitudes, values, and personality modules provide extensive information on respondents' personality traits, political orientations, concerns, satisfaction with different aspects of life, willingness to take risks, and much more.



Attitudes, Values and Personality

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	10,000-euro question	[2010,2017]	plh0134 , plh0135 , plh0136	
	Affective well-being	[2007-2020]	plh0184 , plh0185 , plh0186 , plh0187	[2022, 2023]
	Anomie	(irregular) [1990-2018]	plh0188 , plh0189 , plh0190 , plh0191	[2023]

continues on next page

Table 8 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Attitudes towards genders	[2019]	plh0395i01, plh0395i03, plh0395i05, plh0395i06	plh0395i02, plh0395i04,
	Attitudes towards refugees	[2016,2018,2020]	plj0433, plj0436, plj0439, plj0442, plj0443	plj0434, plj0437, plj0440, plj0441, [2023]
	Big Five personality traits	[2005,2009,2012-2013,2017,2019], [2009,2012-2013,2017,2019]	plh0212, plh0215, plh0218, plh0221, plh0224, plh0255	plh0213, plh0216, plh0219, plh0222, plh0225, plh0214, plh0217, plh0220, plh0223, plh0226, [2023]
	Bundestag election, Eligibles	[2014,2018]	plh0333	[2022]
	Bundestag election, Non-eligibles	[2021]	ppolpar1, ppolpar2, ppolpar3	
	Depressive traits	[2016,2019]	plh0339, plh0340, plh0342	plh0341, [2023]
	Discrimination	[2019]	plh0387i01, plh0387i03, plh0387i05, plh0387i07, plh0387i09, plh0387i11	plh0387i02, plh0387i04, plh0387i06, plh0387i08, plh0387i10,
	Donation of blood	[2010]	plh0131_v1, plh0132, plh0133	plh0131_v2,
	Donations	[2010,2015,2018,2020]	plh0129, plh0130	
	Donations of goods	[2010,2020]	plj0108, plj0111, plj0114, plj0115	plj0109, plj0112, plj0110, plj0113,
	Flourishing	[2015-2020]	plh0334	[2022]
	Goals in life (Kluckhohn)	(irregular) [1990-2016], [2013,2017-2019], [2016]	plh0104, plh0107, plh0110, plh0343_v1, plh0343_v2	plh0105, plh0108, plh0111, plh0106, plh0109, plh0112,
	Impulsivity, patience	[2008,2013,2018]	plh0253, plh0254	[2023]
	Income justice, general	[2005]	plh0116, plh0119, plh0122, plh0125, plh0126, plh0127	plh0117, plh0120, plh0123, plh0124,
	Inflation expectations		pli0196, pli0198, pli0201i01, pli0201i02, pli0201i03	pinfgrp, pli0197, pli0200, pim-mogrp, [2023]
	Life satisfaction	[1984-2020]	plh0182	[2022, 2023]

continues on next page

Table 8 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Locus of control	[1994-1996]	plh0369, plh0370, plh0371, plh0372, plh0373, plh0374, plh0375, plh0376, plh0377_v1, plh0378_v1, plh0379_v1, plh0380_v1, plh0381_v1, plh0382_v1, plh0383_v1, plh0384_v1, plh0385_v1, plh0386_v1	
	Locus of control, rephrased	[2005,2010,2015-2016,2020]	plh0377_v2, plh0378_v2, plh0379_v2, plh0380_v2, plh0381_v2, plh0382_v2, plh0383_v2, plh0384_v2, plh0385_v2, plh0386_v2	
	Loneliness	[2013,2016-2019], [2013,2016-2020]	plj0587, plj0588, plj0589	
	Lottery question	[2004,2009,2014]	plh0203	
	Money and account balance	[2016,2018]	plh0344, plh0345, plh0346	
	Optimism/pessimism	[1999,2005,2009,2016]	plh0244	
	Organisational and community membership	(irregular) [1985-2019], (irregular) [2001-2019], [2003,2007,2011]	plh0263_h, plh0263_v2, plh0264_h, plh0264_v1, plh0264_v2, plh0265, plh0266, plh0267	[2023]
	Organizational and community membership	[1985,1989,1993]	plh0263_v1	
	Policy objectives (Inglehart Index)	[1984-1986,1996,2006,2016]	plh0054, plh0056, plh0058, plh0061	
	Political Tendency, Left-Right	[2005,2009,2014,2016]	plh0004	
	Political influence	[2019]	plh0397i01, plh0397i02, plh0397i03, plh0397i04, plh0397i05	
	Political orientation	[1985-2020]	plh0007	[2022, 2023]
	Political orientation (Party Affiliation)	[1984-2020], [1984-1989], [1990], [1991], [1992], [1993], [1987-1988], [1994-2020]	plh0012_h, plh0012_v1, plh0012_v2, plh0012_v3, plh0012_v4, plh0012_v5, plh0012_v6, plh0013_v1	[2022, 2023]
	Political orientation (Party Preference)	[1984-2020]	plh0011_h, plh0011_v1, plh0011_v2, plh0013_h, plh0013_v2	[2022, 2023]
	Reciprocity	[2005,2010,2015-2020]	plh0206i01, plh0206i02, plh0206i03, plh0206i04, plh0206i05, plh0206i06	

continues on next page

Table 8 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Religious affiliation	(irregular) [1990-2020], [2013,2016- 2020]	plh0258_h, plh0258_v1, plh0258_v10, plh0258_v11, plh0258_v12, plh0258_v13, plh0258_v2, plh0258_v3, plh0258_v4, plh0258_v5, plh0258_v6, plh0258_v7, plh0258_v8, plh0258_v9	[2023]
	Religiousness		plm0560	[2023]
	Risk aversion in different domains	[2004,2009,2014]	plh0197, plh0198, plh0199, plh0200, plh0201, plh0202	
	Risk aversion in general	[2004,2006,2008- 2020], [2013], [2004,2006,2008- 2020]	plh0204_h, plh0204_v1, plh0204_v2	[2022, 2023]
	Satisfaction with various aspects	(irregular) [1989-2019], [2006,2011- 2013,2016], [1984-2020], [2008-2020], [1984-2020], [1984- 1990,1993- 2020], [1984- 2020], [2004- 2020], [1984- 2020], [1984- 1989,1991- 1994,1996- 2020], [1990,1997- 2020], [2006- 2020]	plh0164, plh0171, plh0172, plh0173, plh0174, plh0175, plh0176, plh0177, plh0178, plh0179, plh0180, plh0181	[2022, 2023]
	Self-esteem	[2010,2015- 2020]	plh0206i11	
	Social justice	[2019]	plh0396i01, plh0396i02, plh0396i03, plh0396i04	
	Social responsibility	[1997,2002,2017]	plh0016, plh0017, plh0018, plh0019, plh0020, plh0021, plh0022, plh0023, plh0024, plh0025, plh0026	
	Tendency to forgive	[2010,2015- 2016,2020]	plh0206i07, plh0206i08, plh0206i09, plh0206i10	
	Trust, trustworthiness and fairness	[2003,2008,2013,2017]	p1d0043, p1d0044, p1d0045, plh0192, plh0193, plh0194, plh0195, plh0196	[2023]
	Wage justice	[2015], [2017,2019], [2017-2019]	plh0138, plh0139, plh0140, plh0141, plh0337_v1, plh0337_v2, plh0338_v1, plh0338_v2	[2023]

continues on next page

Table 8 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
<i>Youth Questionnaire</i>	Well-being aspects	[1994,1998-1999]	plh0091_v2, plh0092_v2, plh0093_v2, plh0094_v2, plh0095_v2, plh0096_v2, plh0097_v2, plh0098_v2, plh0099_v2, plh0100_v2, plh0101, plh0102, plh0103	
	Well-being aspects, East Germany	[1990-1991]	plh0091_v1, plh0092_v1, plh0093_v1, plh0094_v1, plh0095_v1, plh0096_v1, plh0097_v1, plh0098_v1, plh0099_v1, plh0100_v1	
	Worries	[2009-2014,2019]	plh0032, plh0033, plh0034, plh0035, plh0038, plh0040, plh0042, plh0043, plh0046, plh0047, plh0335, plh0336	[2022, 2023]
	Affective well-being	[2007-2020]	jl0381, jl0382, jl0383, jl0384	[2022, 2023]
	Attitudes and opinions	[2000-2020]	jl0329, jl0330, jl0360, jl0364	[2022]
	Big Five personality traits	[2006-2020]	jl0365, jl0366, jl0367, jl0368, jl0369, jl0370, jl0371, jl0372, jl0373, jl0374, jl0375, jl0376, jl0377, jl0378, jl0379, jl0380	[2022, 2023]
	Future	[2000-2020]	jl0222, jl0223, jl0224, jl0225, jl0226, jl0227, jl0228, jl0229, jl0230, jl0231, jl0232	
	Life satisfaction	[2006-2020]	jl0392	[2022, 2023]
	Locus of control	[2006-2020]	jl0350_v1, jl0351_v1, jl0352_v1, jl0353_v1, jl0354_v1, jl0355_v1, jl0356_v1, jl0357_v1, jl0358_v1, jl0359_v1	
	Locus of control, rephrased	[2001-2005]	jl0350_v2, jl0351_v2, jl0352_v2, jl0353_v2, jl0354_v2, jl0355_v2, jl0356_v2, jl0357_v2, jl0358_v2, jl0359_v2	[2022]
	Political orientation	[2006-2020]	jl0388, jl0389, jl0390, jl0391	[2022]
	Risk aversion in general	[2006-2020]	jl0349	[2022, 2023]
	Social justice	[2019-2020]	jl1909, jl1910, jl1911, jl1912	[2022]
	Sources of social inequality	[2000-2020]	jl0337, jl0338, jl0339, jl0340, jl0341, jl0342, jl0343, jl0344, jl0345, jl0346, jl0347, jl0348	[2022]
	Trust	[2006-2020]	jl0361, jl0362, jl0363	[2022, 2023]
<i>Mother and Child Instruments</i>	Big Five personality traits	[2003-2020]	char10, char1a, char1b, char2, char3, char4, char5, char6, char7, char8, char9	

continues on next page

Table 8 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Strengths and difficulties questionnaire	[2003-2020]	behav1, behav10, behav11, behav12, behav13, behav14, behav15, behav16, behav17, behav18, behav2, behav3, behav4, behav5, behav6, behav7, behav8, behav9	
	Temperament	[2003-2020]	temp1, temp2, temp3, temp4, temp5, temp6, temp7	
	Vineland adaptive behavior scales (Movements)	[2003-2020]	mvmn1, mvmn3, mvmn4, mvmn5, mvmn6	
	Vineland adaptive behavior scales (Playing)	[2003-2020]	sclr2, sclr3, sclr4, sclr5, sclr6	
	Vineland adaptive behavior scales (Skills)	[2003-2020]	skill1, skill2, skill3, skill4, skill5	
	Vineland adaptive behavior scales (Speaking and Listening)	[2003-2020]	spch3, spch5, spch6, spch7, spch8	

2.9 Time Use and Environmental Behavior

The modules on time use and environmental behavior give information on time commitments, free time, and time planning as well as environmental awareness, for instance, the use of public transport and different energy sources, as well as what respondents think about renewable energies.



Time Use and Environmental Behavior

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Carbon footprint		pli0195	[2023]
	Computer usage: Private	[1997,2000-2001], [1997,2001], [2000], [1997,2001], [1997], [2001]	pli0066_h, pli0066_v1, pli0066_v2, pli0067_h, pli0067_v1, pli0067_v2	
	Computer usage: Private (Internet)	[2001]	pli0068, pli0069	

continues on next page

Table 9 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Computer usage: Work	[1997,1999,2001], [2000], [1997,2001], [1997], [2001]	pli0070_v1, pli0070_v2, pli0071_h, pli0071_v1, pli0071_v2	
	Computer usage: Work (Internet)	[2001]	pli0072, pli0073	
	Leisure activities (long)	[2019], (unregelmaessig) [1990-2019], [2013]	plh0390, plh0391, plh0392, plh0393, plh0394, pli0083, pli0084, pli0085_v1, pli0085_v2, pli0087, pli0088, pli0089, pli0090_v1, pli0090_v2, pli0090_v3, pli0091_h, pli0091_v1, pli0091_v2, pli0165, pli0178, pli0182	[2023]
	Leisure activities (long): Art and Music	(unregelmaessig) [1990-2019], (unregelmaessig) [2001-2017], (unregelmaessig) [1990-2019]	pli0093_h, pli0093_v1, pli0093_v2	[2023]
	Leisure activities (long): Politics	(unregelmaessig) [1984-2019], [1984], (unregelmaessig) [1985-2017], (unregelmaessig) [1990-2019]	pli0097_h, pli0097_v1, pli0097_v2, pli0097_v3	
	Leisure activities (long): Religion	(unregelmaessig) [1990-2019], (unregelmaessig) [1990-2017], (unregelmaessig) [1990-2019]	pli0098_h, pli0098_v1, pli0098_v2	
	Leisure activities (long): Socializing	(unregelmaessig) [1990-2019]	pli0079, pli0080, pli0081, pli0082, pli0094_v1, pli0094_v2, pli0095_v1, pli0095_v2	[2023]
	Leisure activities (long): Sports	(unregelmaessig) [1984-2019], [1984], (unregelmaessig) [1985-2017], (unregelmaessig) [1990-2019]	pli0092_h, pli0092_v1, pli0092_v2, pli0092_v3	[2023]
	Leisure activities (long): Voluntary work	(unregelmaessig) [1984-2019], [1984], (unregelmaessig) [1990-2019], (unregelmaessig) [1985-2017]	pli0096_h, pli0096_v1, pli0096_v2, pli0096_v3	[2023]
	Time use for different activities (Saturdays)	(unregelmaessig) [1990-2019], (unregelmaessig) [2001-2019], (unregelmaessig) [2003-2019], [2008- 2013,2015,2017,2019- 2020]	pli0003_h, pli0003_v1, pli0003_v2, pli0003_v3, pli0003_v4, pli0005, pli0036, pli0054, pli0055, pli0056, pli0060	[2022]

continues on next page

Table 9 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Time use for different activities (Saturdays): Childcare	(unregelmaessig) [1990-2019], (unregelmaessig) [1993-2019]	pli0019_h, pli0019_v2, pli0019_v4	pli0019_v1, pli0019_v3,
	Time use for different activities (Saturdays): Chores	(unregelmaessig) [1990-2019], (unregelmaessig) [1990-2019]	pli0012_h, pli0012_v2, pli0012_v3	pli0012_v1,
	Time use for different activities (Saturdays): Education	(unregelmaessig) [1990-2019], (unregelmaessig) [1990-2019]	pli0024_h, pli0024_v2, pli0024_v3	pli0024_v1,
	Time use for different activities (Saturdays): Repair work	(unregelmaessig) [1990-2019], (unregelmaessig) [1993-2019]	pli0031_h, pli0031_v2, pli0031_v4	pli0031_v1, pli0031_v3,
	Time use for different activities (Sundays)	[1984], (unregelmaessig) [1992-2019], (unregelmaessig) [1984-2019], (unregelmaessig) [1990-2019], (unregelmaessig) [2001-2019], (unregelmaessig) [2003-2019]	pli0002_v1, pli0007_v1, pli0007_v3, pli0007_v5, pli0010, pli0011, pli0057, pli0058	pli0002_v2, pli0007_v2, pli0007_v4,
	Time use for different activities (Sundays): Childcare	(unregelmaessig) [1985-2019], (unregelmaessig) [1992-2019]	pli0022_h, pli0022_v2, pli0022_v4	pli0022_v1, pli0022_v3,
	Time use for different activities (Sundays): Chores	(unregelmaessig) [1984-2019], (unregelmaessig) [1992-2019]	pli0016_h, pli0016_v2, pli0016_v4	pli0016_v1, pli0016_v3,
	Time use for different activities (Sundays): Education	(unregelmaessig) [1984-2019], (unregelmaessig) [1992-2019]	pli0028_h, pli0028_v2, pli0028_v4	pli0028_v1, pli0028_v3,
	Time use for different activities (Sundays): Repair work	[1984-1990], (unregelmaessig) [1992-2019]	pli0034_v1, pli0034_v3, pli0034_v4	pli0034_v2,
	Time use for different activities (weekdays)	[1984], [2003-2020], [2001-2020], [1992-2020], [1990-2020], [1984-2020], [2008-2013,2015,2017,2019-2020]	pli0001_v1, pli0038_v1, pli0038_v3, pli0040, pli0046, pli0051, pli0052, pli0059	pli0001_v2, pli0038_v2, pli0038_v4,
	Time use for different activities (weekdays): Childcare	[1985-2020], [1985-1991], [1990], [1992-2020]	pli0044_h, pli0044_v2, pli0044_v3	pli0044_v1, [2022, 2023]
	Time use for different activities (weekdays): Chores	[1984-2020], [1984-1991], [1990], [1992-2020]	pli0043_h, pli0043_v2, pli0043_v3	pli0043_v1, [2022, 2023]
	Time use for different activities (weekdays): Education	[1990], [1984-2020]	pli0047_v1, pli0047_v3	pli0047_v2, [2022]
	Time use for different activities (weekdays): Repair work	[1984-2020], [1984-1991], [1990], [1992-2020]	pli0049_h, pli0049_v2, pli0049_v3	pli0049_v1, [2022]

continues on next page

Table 9 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Trip to work	[1985,1993,1995], [1998, 2003], (unregel- maessig) [1985-2019]	plb0144, plb0145, plb0156_v1, plb0157_v1, plb0157_v2, plb0158	
	Trip to work: Car pas- senger	[1985], [1998,2003]	plb0175_v1, plb0175_v2	
	Trip to work: Cost	[1985,1993,1995]	plb0142, plb0143	
	Trip to work: Starting year	[1997,1999], [1997], [1999]	plb0146_h, plb0146_v1, plb0146_v2	
	Trip to work: Time	[1985,1990- 1993,1995,1998,2003]	plb0147, plb0148, plb0149, plb0150, plb0151, plb0152, plb0153, plb0154, plb0155	
	Trip to work: Way to work	[1998, 2003], [1985, 1990], [1993,1995,1997- 1999], [2000- 2013,2015,2017,2019]	plb0156_v2, plb0159_v1, [2023] plb0159_v2, plb0159_v3	
	Use of transportation for errands	[1998], [2003, 2018], [1998], [2003,2018], [1998], [2003,2018], [1998], [2003, 2018], [1998], [2003, 2018], [1998,2003,2018]	pli0133_v1, pli0133_v2, pli0134_v1, pli0134_v2, pli0135_v1, pli0135_v2, pli0136_v1, pli0136_v2, pli0137_v1, pli0137_v2, pli0138	
	Use of transportation for excursions	[1998], [2003,2018], [1998], [2003, 2018], [1998], [2003, 2018], [1998], [2003,2018], [1998], [2003,2018], [1998,2003,2018]	pli0139_v1, pli0139_v2, pli0140_v1, pli0140_v2, pli0141_v1, pli0141_v2, pli0142_v1, pli0142_v2, pli0143_v1, pli0143_v2, pli0144	
	Use of transportation in general	[1998,2003,2018]	pli0106	
	Use of transportation in general: Car sharing		pli0185	[2023]
	Use of transportation in general: Cruise		pli0194	[2023]
	Use of transportation in general: Licenses	[1985,1991,1998,2003], [1985,1991,1995,1998,2 [1985], [1991,1995,1998,2003]	pli0101, pli0102, pli0103, pli0104, pli0105_h, pli0105_v1, pli0105_v2	
	Use of transportation in general: Plane		pli0193i01, pli0193i02, [2023] pli0193i03, pli0193i04, pli0193i05, pli0193i06	
	Use of transportation in general: Public transportation		pli0186, pli0187i01, [2023] pli0187i02, pli0188, pli0189, pli0190i01, pli0190i02, pli0191, pli0192i01, pli0192i02	

continues on next page

Table 9 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Use of transportation in leisure time	[1998], [2003, 2018], [1998], [2003, 2018], [1998], [2003,2018], [1998], [2003, 2018], [1998], [2003, 2018], [1998], [2003,2018]	pli0145_v1, pli0145_v2, pli0146_v1, pli0146_v2, pli0147_v1, pli0147_v2, pli0148_v1, pli0148_v2, pli0149_v1, pli0149_v2, pli0150_v1, pli0150_v2	
	Use of transportation to take children places	[1998], [2003, 2018], [1998], [2003, 2018], [1998], [2003, 2018], [1998], [2003, 2018], [1998], [2003, 2018], [1998]	pli0151_v1, pli0151_v2, pli0152_v1, pli0152_v2, pli0153_v1, pli0153_v2, pli0154_v1, pli0154_v2, pli0155_v1, pli0155_v2, pli0156	
	Use of transportation to work	[1998], [2003,2018,2020], [1998], [2003,2018,2020], [1998], [2003,2018,2020], [1998], [2003,2018,2020], [1998], [2003,2018,2020], [1998,2003,2018,2020]	pli0127_v1, pli0127_v2, pli0128_v1, pli0128_v2, pli0129_v1, pli0129_v2, pli0130_v1, pli0130_v2, pli0131_v1, pli0131_v2, pli0132	
<i>Youth Questionnaire</i>	Leisure and hobbies	[2001-2018], [2006-2020]	j10058, j10064, j10065, j10066, j10067, j10072, j10073	[2022, 2023]
	Leisure and hobbies: Internet	[2006-2013]	j10060_v1, j10060_v2, j10060_v3, j10060_v4	[2022, 2023]
	Leisure and hobbies: Music	[2001-2018]	j10061, j10062, j10074, j10075, j10076, j10087, j10104	[2022]
	Leisure and hobbies: Socializing	[2001-2020], [2006-2020]	j10068, j10069, j10070, j10071	[2022, 2023]
	Leisure and hobbies: Sports	[2000-2020], [2000-2020]	j10063, j10105_h, j10105_v1, j10105_v2, j10109, j10112_v1, j10112_v2, j10116, j10117	[2022]
	Leisure and hobbies: Videogames	[2001-2020], [2001-2015], [2016-2020]	j10059_h, j10059_v1, j10059_v2	[2022, 2023]
<i>Household Questionnaire</i>	Private Vehicles	[2015]	hli0005, hli0077, hli0078, hli0079, hli0080, hli0081, hli0082, hli0083, hli0084, hli0085	[2023]
	Private Vehicles:Fuel consumption		hli0011_v2, hli0146, hli0017_v2, hli0147, hli0023_v2, hli0148, hli0029_v2, hli0149	[2023]
	Private Vehicles:Mileage		hli0010, hli0016, hli0022, hli0028	[2023]
	Traffic and energy: E-Bike	[2015,2020]	hli0135, hli0136	

continues on next page

Table 9 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Traffic and energy: Electricity provider change	[2015,2020]	hli0139, hli0140, hli0141, hli0142	
	Traffic and energy: Green electricity	[2015], [2015,2020]	hli0137, hli0138	[2023]
	Traffic and energy: Private Vehicles (Biodiesel)	[2015,2020]	hli0114, hli0115, hli0116, hli0117, hli0118, hli0119, hli0120	[2023]
	Traffic and energy: Private Vehicles (Diesel)	[2015,2020]	hli0100, hli0101, hli0102, hli0103, hli0104, hli0105, hli0106	[2023]
	Traffic and energy: Private Vehicles (E10)	[2015,2020]	hli0093, hli0094, hli0095, hli0096, hli0097, hli0098, hli0099	[2023]
	Traffic and energy: Private Vehicles (Electricity)	[2015,2020]	hli0121, hli0122, hli0123, hli0124, hli0125, hli0126, hli0127	[2023]
	Traffic and energy: Private Vehicles (Gas)	[2015,2020]	hli0107, hli0108, hli0109, hli0110, hli0111, hli0112, hli0113	[2023]
	Traffic and energy: Private Vehicles (Hydrogen)	[2015,2020]	hli0128, hli0129, hli0130, hli0131, hli0132, hli0133, hli0134	[2023]
	Traffic and energy: Private Vehicles (Normal Petrol/Super)	[2015,2020]	hli0086, hli0087, hli0088, hli0089, hli0090, hli0091, hli0092	[2023]

2.10 Integration, Migration, Transnationalization

Migration and forced migration are changing German society. The SOEP offers diverse migration samples and numerous specific migration questions that allow researchers to study migration-related questions in detail. The modules on integration, migration, and transnationalization provide data on migration histories, discrimination, inter-ethnic contact, education, cultural integration, transnational relations, identification with Germany, and the intention to stay in Germany.



Integration, Migration,
Transnationalization

Questionnaire	Module	Years	Variables	Preview
<i>Individual Questionnaire</i>	Applying for German citizenship	(unregelmaessig) [1998-2018]	plj0021	
	Circle of friends, percentage of migrants	[2013,2018]	plm0143	
	Contacts abroad, thoughts about moving abroad	[2009,2014,2019]	plj0089, plj0090, plj0091, plj0092, plj0104, plj0105	
	Country of origin	[2020]	plj0725	[2022, 2023]
	Disadvantage / discrimination based on ethnic origins (detailed)	[2019]	plh0387i01, plh0387i04, plh0387i06, plh0387i08, plh0387i10, plj0048_v1, plj0327, plj0328, plj0329, plj0330, plj0331, plj0332, plj0333, plj0334, plj0335, plj0336, plj0337, plj0338, plj0339	
	Discrimination, areas		plh0412, plh0414, plh0414	[2022]
	Foreign language skills	[2013]	plm0135	
	Integration indicators	(unregelmaessig) [1984-2018], (unregelmaessig) [1997-2019], [2020]	plj0078, plj0080_v1, plj0080_v2	[2022]
	Intention to stay	[1996-2011,2015-2020], [2013], [1996-2011,2013,2015-2020], [2020], [1984-2011,2013,2015-2020], [1996-2011,2013,2015-2020]	plj0085_v1, plj0085_v2, plj0086_v1, plj0086_v2, plj0087, plj0088	
	Language ability German	[2007-2011,2013-2020], [2010-2011,2013-2020]	plj0071, plj0072, plj0073	[2023]
	Language ability native language	[2007-2011,2013-2019], [2010-2011,2013-2019]	plj0074, plj0075, plj0076	[2023]
	Language use, media	[2014,2016], [2017-2020]	plj0226_v1, plj0226_v2	[2022]
	Language use, newspapers	(unregelmaessig) [1988-2012]	plj0070	
	Native language	[2007-2011,2013,2015-2019]	plj0009	
	Native language (family)	[2013,2015-2020], [2013], [2015-2020]	plm0136_h, plm0136_v2	plm0136_v1,
	Native language (friends)	[2013,2015-2020], [2013], [2015-2020]	plm0137_h, plm0137_v2	plm0137_v1,
	Native language (workplace)	[2013,2015-2020], [2013], [2015-2020]	plm0138_h, plm0138_v2	plm0138_v1,

continues on next page

Table 10 – continued from previous page

Questionnaire	Module	Years	Variables	Preview
	Regional attachment	[2009,2014,2019]	plj0043, plj0044, plj0045	
	Sense of home	(unregelmaessig) [1988-2012], [2014]	plj0083, plj0340	
	Translation help	[2013-2018]	p_buh1, p_buh10, p_buh2, p_buh3, p_buh4, p_buh5, p_buh6, p_buh7, p_buh8, p_buh9	
	Visited country of origin in last 2 years	[2014,2016,2018,2020]	plj0322, plj0323	[2022]
<i>Youth Questionnaire</i>	Visiting / being visited by Germans and foreigners at home	(unregelmaessig) [2007-2019]	plj0060, plj0061, plj0062, plj0063	
	Language ability German	[2006-2018], [2010-2018], [2014-2020]	jl0248, jl0442, jl0443, jl0444, jl1249	[2023]
	Language ability native language	[2006-2013], [2010-2013]	jl0251, jl1251	
<i>Biography Questionnaire</i>	German language courses, Federal Office for Migration and Refugees		plm0721_v2	[2023]
	German language courses, before moving to Germany		lm0131_v5	[2023]
	German language courses, since moving to Germany		lr3579	[2023]
	German language exam, before moving to Germany		lm1064i01_v4	[2023]
	German language exam, since moving to Germany		lb1450	[2023]
	Moving to Germany, COVID-19		lb1442	[2022]
	Moving to Germany, Obstacles		lr3577i01, lr3577i02, lr3577i03, lr3577i04, lr3577i05, lr3577i06, lr3577i07, lr3577i08, lr3577i09, lr3577i10, lr3577i11	[2022]

2.11 Survey Methodology

Survey methodology modules offer diverse variables on imputation, weighting, SOEP-Core fieldwork, identifiers, interview methods, survey modes, and information about the respondent's exit from the survey.



Survey Methodology

Questionnaire	Module	Variables
<i>Interviewer Questionnaire</i>		
	Identificators	hhnr , intid , syear , wave
	Interview information	typint , lenghtinth , lenghtintp , lenghtintj
	Demography	gender , birth , marital , educ , modbula , modggk , ista1 , ibstam1 , ibstav1 , imusp , irel
	Interviewer history	startint , endint , experience , firstintm , firstintd , lastintm , lastintd
	Employment	iberuf , ioed , istell
	Interviewer activity	meancontacthh , responserate , amountinth , amountintp , amountintj , papi , capi , cawi , mail
	Patience	iged
	Health	iges
	Risk aversion	irisk
	Life satisfaction	izule
	Incentives	ibbarhon , ibbeval
	Optimism	ibopt
	Motivation & Fulfillment	igru01 - igru07 , ierf01 - ierf14
	Assessment of Participation	itebe01 - itebe13
	Interviewer Training	ibseval01 - ibseval04 , ibschul , ibschul02
	Big Five personality traits	iego01 - iego22
	Attitudes and social interaction	ibez01 - ibez05 , iverh01 - iverh06
	Political orientation	ipol1 - ipol4
	Worries	isor01 - isor14 , isor21 - isor22
	Working hours	ibwsist01 - ibwsist05 , ibwssol01 - ibwssol03
	Interviewer and other studies	ibsozer01 - ibsozer08 , ibsozero , ibsozerso , ibef01 - ibef03
	Foreign language skills	ispr01 - ispr10 , ibspre01 - ibspre10
	Flags (conflicts)	genderconfl , birthconfl , maritalconfl , educconfl , startintconfl , ista1confl

Important documents regarding this Topic are available [here](#)

Last change: Jul 24, 2023

SURVEY DESIGN

3.1 SOEP Questionnaires

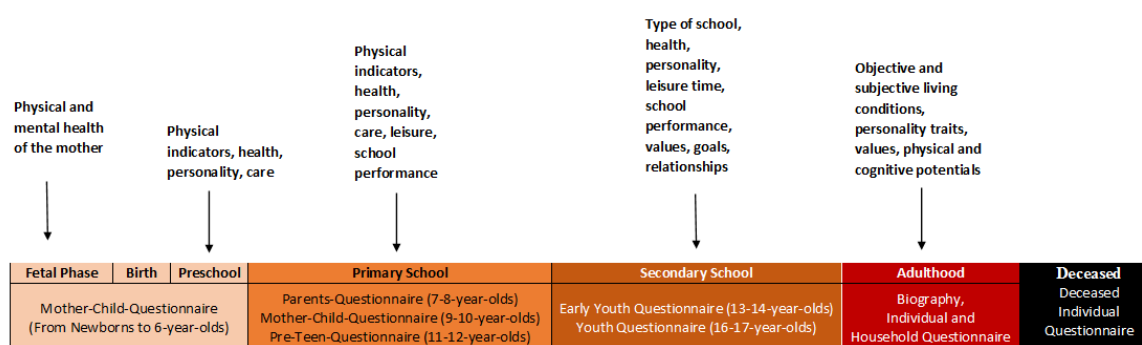
The interview methodology of the SOEP is based on a set of pre-tested questionnaires for households and individuals. Interviewers try to obtain face-to-face interviews with all members of a given survey household aged 16 and over. Thus, there are no proxy interviews for adult household members. Additionally, one person (the “head of household”) is asked to answer a household-related questionnaire covering information on housing, housing costs, and different sources of income (e.g., social transfers such as social assistance or housing allowances). This questionnaire also includes questions on children up to the age of 16 in the household, mainly concerning daycare, kindergarten, and school attendance.

The questions in the SOEP are largely identical for all participants of the survey to ensure comparability across the participants within a given year, but of course there are differences across years. There are a few exceptions to this rule, which are due to different requirements in the target population. Up to 1996, the questionnaires for the sample of foreigners (B) and the immigrant sample (D) covered additional measures of integration or information on re-migration behavior. Between 1990 and 1992, i.e., during the first years of the German reunification process, the questionnaire for the East German sample (C) also contained some additional specific variables. From 1996 to 2012, all questionnaires were uniform and completely integrated for all of the main SOEP samples. For the IAB-SOEP Migration Sample, which was launched in 2013, specific questions were added to the SOEP questionnaires. The same is true of the IAB-BAMF-SOEP Survey of Refugees, which was launched in 2016.

Another special questionnaire is used for first-time respondents since some questions do not have to be repeated every year. Each respondent is asked to fill out a biographical questionnaire covering information on the life course up to the first SOEP interview (e.g., marital history, social background, and employment biography).

Additional information not provided directly by the respondent can be obtained from the “address logs”, which are stored for every year in the \$PBRUTTO and \$HBRUTTO files. Every address log is filled in by the interviewer even in the case of non-response, thus providing very valuable information, e.g. for attrition analysis. For researchers interested in methodological issues, these data also contain information on the fieldwork process such as the number of contacts, reasons for drop-outs, and interview mode. For households that were contacted successfully, the address logs cover the size of the household, some regional information, survey status, etc. The individual data for all household members include the relationship to the household head, survey status of the individual, and some demographic information.

Life History



The SOEP questionnaires are designed so that people in a SOEP household can be analyzed from birth to adulthood and throughout the rest of their lives. In addition to the *Youth Questionnaire*, which was conducted for the first time in 2000/01, a series of questionnaires for specific cohorts of children living in SOEP households have been introduced since 2003. These have been completed annually since their year of introduction by mothers (in exceptional cases by fathers) with children of the appropriate age. In 2003, a questionnaire was developed for the mothers of newborn children, *Mother and Child Questionnaire (Newborns)*. The following instruments were developed in such a way that this starting cohort (born 2002/2003) can be followed up in their development and analyzed longitudinally. This was followed in 2005 by a questionnaire for mothers of 2-3-year-old children, *Mother and Child Questionnaire (2-3-year-olds)* and in 2008 by a questionnaire for 5-6-year-olds, *Mother and Child Questionnaire (5-6-year-olds)*. In 2010, the questionnaire for 7-8-year-old children, *Parents and Child Questionnaire (7-8-year-olds)*, completed by both mothers and fathers, was launched. In 2012, the questionnaire for 9-10-year-old children, *Mother and Child Questionnaire (9-10-year-olds)* was added as the last questionnaire to be answered by the mothers. This was followed by two youth instruments in which the children, aged 12, *Pre-Teen Questionnaire* and 14, *Early Youth Questionnaire*, answered questions about their own lives for the first time. These were introduced in 2014 and 2016, respectively. In 2018, the first cohort completed the entire battery of age-specific instruments and from then on, they will complete the annual questionnaires of the long-term SOEP study. Each person in a SOEP household receives the *Individual Questionnaire* as soon as they reach the age of 18, and the head of the household also receives the *Household Questionnaire*. If a respondent states in their interview that someone has died in the last year, regardless of whether the deceased person was part of a SOEP household, the *Deceased Individual Questionnaire* is given to the respondent providing the information.

3.1.1 Overview of the Questionnaires

Questionnaires \ Years	1984	1985	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	2016	2017	2018	2019	2020	2021
Household																																						
Individual																																						
Biography																																						
Catch-Up Individual																																						
Youth (16-17-year-olds)																																						
Mother-Child (Newborns)																																						
Mother-Child (2-3-year-olds)																																						
Mother-Child (5-6-year-olds)																																						
Parents (7-8-year-olds)																																						
Mother-Child (9-10-year-olds)																																						
Pre-Teen (11-12-year-olds)																																						
Early Youth (13-14-year-olds)																																						
Deceased Individual																																						
Cognitive Tests for Youth („Lust auf DJ“)																																						
Grip Strength Test																																						

3.1.2 Household Questionnaire

The household questionnaire in its basic form has been an important part of the SOEP surveys since 1984 and has been improved and expanded continuously. The data collected and the questionnaire itself have become so complex that the original topics are no longer sufficient. Between 1984 and 2016, the number of questions more than doubled from 46 to 97. The multitude of questions offer users many options for analysis. Each year, the number of questions varies because new innovative question modules are added or because some questions are not asked every year. An overview of the modules included at different intervals can be found in the section *Topics of SOEP-Core*. The questions provide diverse information about the respondents' households that is stored in several hundred variables. Child-specific questions asked in the household questionnaire are found in the separate dataset \$kind.

Availability: Since 1984

Dataset: \$h (CS), *hl* (long)

Respondent: Head of household

The following question modules are part of the core program of the Household Questionnaire:

- Change of living situation
- Neighborhood
- Building type
- Size and condition of dwelling
- Amenities
- Type of dwelling
- Loans, mortgages, building-society loans
- Hereditary lease interest
- Modernization costs
- Ownership costs

- Photovoltaic and solar thermal system
- Owner debt
- Government-subsidized housing
- Home ownership
- Rental and expenses
- Tenant debt
- Cleaning or household assistance
- Persons in need of care
- Names and birth dates of children
- Child's school attendance
- Childcare situation
- Income and expenses from renting/leasing
- Loan repayment
- Debt
- Inheritances, gifts, winnings
- Investments
- Income/expenses household
- Savings
- Material depreciation
- Number of books
- Pets
- Cause of moving

where applicable:

+ migration-specific modules for the IAB-SOEP Migration Sample

- distinguishing repayment of loans, debt, income / expenses between Germany and foreign country

or where applicable:

+ refugee-specific modules for the IAB-BAMF-SOEP Sample of Refugees

- Information on shared accommodations
- Location preferences

3.1.3 Individual Questionnaire

The individual questionnaire has been a standard instrument since the beginning of the SOEP. In order to enable analysis over time, the individual questionnaire has a large number of question modules that are asked every year. There are also questions that do not have to be asked every year, as short-term changes are unlikely. In order to be able to react to current social changes, new topics are added to the individual questionnaire and repeated at intervals of more than one year.

Availability: Since 1984

Dataset: \$p\$ (CS), *pl* (long)

Respondent: Persons over 18 years in the household

The following question modules are part of the core program of the Individual Questionnaire:

- Satisfaction with various live aspects
- Satisfaction with current life situation
- Feelings
- Flourishing
- Risk aversion
- Political orientation
- Worrying
- Life satisfaction overall
- Ethnic/national origins
- Vocational training
- Completed level of education
- Higher education
- Family situation
- Family changes
- State of health
- Disability or severe disability
- Visits to the doctor
- Hospital stays
- Sick leave
- Health insurance
- Wages and collective wage agreements
- Additional questions for employees
- Additional questions for retirees/pensioners
- Government transfers
- Calendar
- Time use
- Second jobs

- Income
- Work, last 7 days
- Maternity/ parental leave
- Care period (Pflegezeit)
- Registered unemployed
- Quitting a job
- Employment status
- Start of job
- Change of job
- Job search
- Current profession
- Current job
- Working hours
- Overtime
- Optimism
- Religion
- Organization and Association membership
- Personality traits (Big Five)
- Anomie
- Life goals
- Locus of control
- Reciprocity
- Trust and Fairness
- Narcissism
- Loneliness
- Impulsiveness and Patience
- Political Goals (Inglehart-Index)
- Attitude towards refugees
- Just society
- Discrimination
- Bundestag election
- Social responsibility
- Influence on public decisions
- Friends
- LGBT-Status
- Child wish

- Gender stereotypes
- Attitudes towards gender
- On-Call occupation
- Commuting
- Home-Office
- Short-Time work payment
- Work council
- Payment equity
- Workload
- Occupational expectations
- Depressive traits
- Smoking and drinking
- Integration indicators
- Free time
- Leisure activities
- Donation

where applicable:

+ migration specific modules for the IAB-SOEP-Migrationsample

- First Job in Germany
- Job before immigration
- Language proficiency before and since immigration
- Partnership during immigration
- Living situation since immigration
- Religion and faith of parents
- Satisfaction in various areas of life before and after immigration

or where applicable:

+ refugee specific modules for the IAB-BAMF-SOEP-Sample of Refugees

- Legal status
- Religion and faith
- Language proficiency
- Integration courses and government measures
- Special questions for interviewers concerning language
- Recognition of qualifications

Re-Interviewed

- Cultural and political participation
- Application for recognition

- Trauma screener
- OK (Judgement of different actions)
- Citizenship (inkl. connection with country of origin/ Germany)
- Disadvantages
- Location preferences
- Willingness to participate in a tandem program
- Satisfaction in various areas before and after fleeing

New respondents

- Obtaining help and knowledge about advice services
- Assessment of current situation in country of origin
- Government, democracy and woman's position

3.1.4 Biography Questionnaire

Availability: Since 1987

Dataset: \$lela (CS), *biol* (long)

Respondent: Supplementary, one-time data from the personal questionnaire of all persons aged 18 and over in the household.

Content:

- Nationality
- Country of Origin
- Childhood
- Parents
- Life course since the age of 15
- Education
- Occupation
- Partnership/marriage
- Information on children
- Siblings

where applicable:

+ migration specific modules for the IAB-SOEP-Migrationsample

- Travel to Germany
- Stays Abroad
- Citizenship
- Language proficiency
- Work before moving to Germany
- First job in Germany

- Relationship at the time of moving to Germany

or where applicable:

+ refugee-specific modules for the IAB-BAMF-SOEP Sample of Refugees

- Travel to Germany
- Questions concerning parents of respondent
- Lodging and living situation
- Language proficiency before moving to Germany

3.1.5 Mother and Child Instruments

Mother and Child Questionnaire (Newborns)

Mothers of newborn children answer questions dealing primarily with pregnancy, birth, breastfeeding, and the health of the newborn child. The questionnaire also asks to what extent the mother feels that her living situation changed after the birth of the child, how childcare is handled, and how mothers assess their baby's temperament (as a precursor to personality).

Availability: Since 2003

Dataset: \$muki (CS), *bioagel* (long)

Respondent: Mother in household (child age 0-1)

Content:

- Course of pregnancy
- Childbirth
- Health screening
- Well-being
- Childcare
- Living situation

Mother and Child Questionnaire (2-3-year-olds)

Mothers of 2-3-year-old children answer questions about their child's health and how long they have been breastfeeding. The questionnaire asks again about the childcare situation and the child's temperament and includes a short scale on personality (the dimensions of agreeableness, extraversion, openness, and conscientiousness from the "Big Five"; McCrae and Costa 1987). In addition, it asks what language is spoken with the child and what activities they or the main caregiver engages in with their child (e.g., going to the playground, reading or telling stories, visiting other families with children). Mothers are asked to assess their children's adaptive behavior in the areas of communication, everyday skills, social relationships, and motor skills. This is based on a translated version of the Vineland Adaptive Behavior Scale, which was reduced to 20 items for the SOEP to provide data on the child's stage of development in everyday life.

Availability: Since 2005

Dataset: \$muki2 (CS), *bioagel* (long)

Respondent: Mother in household (child age 2-3)

Content:

- Personality of the child
- Well-being
- Childcare
- Language skills
- Development
- Abilities

Mother and Child Questionnaire (5-6-year-olds)

Mothers of 5-6-year-old children complete this questionnaire in the survey year when their child will turn six. It has many of the same topics as in previous years: health, childcare, a more comprehensive battery of items on the personality (from this age on, the “Big Five” dimension of neuroticism is also included) and activities that they or the main caregiver engages in with their child. In addition, the questionnaire includes a shortened version of the Strength and Difficulties Questionnaire (SDQ), a frequently used instrument to measure the mental health of children and adolescents, reduced here to 17 items of the German SDQ.

Availability: Since 2008

Dataset: \$muki3 (CS), *bioagel* (long)

Respondent: Mother in household (child age 5-6)

Content:

- Personality of the child
- Activities with children
- Well-being
- Childcare

Parents and Child Questionnaire (7-8-year-olds)

This questionnaire on 7-8-year-old children is the only age-specific instrument that is completed by both parents, as long as they live together in the same household. In this age range, questions about school attendance (date of school enrolment) and parent’s aspirations for their children’s level of school completion become relevant for the first time. However, the focus is on parenting goals, parenting styles, and the roles of both parents. Parenting goals range between conformity and autonomy. Parenting styles are surveyed using 18 items, which can be divided into six scales: emotional warmth, inconsistent education, monitoring, negative communication, psychological control, strict control. The items were taken from the pairfam study, as were the 10 items on the role of parents, which can be divided into three scales: autonomy, hostile attributes, and willingness to make sacrifices.

Availability: Since 2012

Dataset: \$elt (CS), *bioagel* (long)

Respondent: Parents in household (child age 7-8)

Content:

- Hopes and expectations for children’s educational attainment
- Parental goals
- Parental styles
- Parental role

- Childcare

Mother and Child Questionnaire (9-10-year-olds)

In addition to questions on health and child care, which are asked in almost all age groups, mothers of 9-10-year-old children are asked for more detailed information about the children's school situation. They are asked what level of schooling they would like their children to complete and what level they think is realistic, what their children's most recent grades were in their three main subjects, whether someone helps the child with homework, and whether the child likes going to school. Since friends and leisure activities are gaining in importance in this age group, some questions deal with these topics. Questions about allowance money are asked for the first time in this age group.

Availability: Since 2012

Dataset: \$muki5 (CS), *bioangel* (long)

Respondent: Mother in household (child age 9-10)

Content:

- Hopes and expectations for children's educational attainment
- Education
- Parental involvement
- Leisure activities
- Family environment
- Social behavior of child
- Personality of Child
- Health of Child
- Supervision
- Allowance money

3.1.6 Youth Instruments

Pre-Teen Questionnaire

Young people complete a questionnaire for the first time themselves in the year they turn twelve. Here, as in the preceding questionnaires, the focus is on their school situation: what time their school day starts and ends on different days of the week, what type of school they attend, how many students are in their class, how many of their classmates are not from Germany, whether they feel discriminated against by their teacher, and what their grades were on their last report card in Math, German, and English. The questionnaire also asks how much time they spend on homework, where they do their homework, and who helps them with homework and studying. They are asked what level of schooling they would like to complete and what level they realistically expect to complete. Since friends play an important role at this age, pre-teens are asked how often they go to friends for support when they have problems. They are asked how many close friendships they have and how often their parents interfere in their choice of friends. They are asked about the educational aspirations of their three closest friends and three older siblings (if any). Several questions deal with their cultural capital and learning environment (e.g., books, musical instruments, and art in the household; whether they have a desk and a room of their own). They are asked about how they spend their free time, how much allowance money they get, and about their personality, willingness to take risks, and life satisfaction. Further questions deal with what languages are spoken with the child and who the child eats meals with.

Availability: Since 2014

Dataset: \$school (CS), *biopupil* (long)

Respondent: 11-12-year-olds in the household

Content:

- Attitude
- Personality
- School (schedule, educational attainment, extra-curricular activities)
- Recreational activities
- Social and family surroundings
- Living situation

Early Youth Questionnaire

The questionnaire for early youth is designed similarly to the pre-teen questionnaire to provide important data on developmental psychology. There are fewer questions about homework and the learning environment and more questions on involvement in extra-curricular activities at school (e.g., student council, after-school clubs) since such activities build social capital. Early youth are asked about the importance of various family members and friends in their lives and about their own educational aspirations as well as those of their three best friends. They are asked how late they are allowed to stay out on school nights and weekends, and what types of activities they have taken part in without their parents (e.g., vacation, doctor visits, shopping, drinking alcohol, smoking cigarettes). They are asked how much allowance they get, and whether they have any savings. Another new topic in this age group is interest in politics and political orientations.

Availability: Since 2015

Dataset: \$school2 (CS), *biopupil* (long)

Respondent: 13-14-year-olds in the household

Content:

- Self-perception
- School (schedule, educational attainment, extra-curricular activities)
- Recreational activities
- Friends
- Siblings
- Parents
- Allowance money
- Political party preferences
- Self-perceptions
- Willingness to take risks
- Life satisfaction
- Attitudes/opinions
- Future

Youth Questionnaire

In the SOEP, young people who turn 17 in the year of the survey are considered adult respondents. Like other first-time adult respondents, they receive a biography questionnaire and an individual questionnaire. Since part of the adult biography (e.g., employment history, relationships) does not yet apply to the young respondents, whereas other aspects such as relationships with parents, leisure activities, and school or vocational training play a greater role, a youth questionnaire was developed in 2000 to replace the biographical questionnaire in this age group. The content of this questionnaire corresponds in many respects to the adult biographical questionnaire so that the data can be used to supplement the information on parents (if parents do not live in the household; dataset: BIOPAREN). Health status, personality, willingness to take risks, locus of control, trust, time preferences, political preferences, knowledge of German, as well as information on the respondent's living situation, work situation, training, career plans, and educational aspirations are also covered in this questionnaire. For the period from 2000 to 2005, respondents in this age group completed the youth questionnaire and the individual questionnaire. Since 2006, they have only completed the youth questionnaire. The version used since then has been expanded to include a few additional indicators. A test was added to assess cognitive potential based on the I-S-T 2000R (Amthauer et al. 2001) using 20 subtasks each for the components of analogies, number series, and matrices (see Solga et al. 2005). The test measures fluid cognitive abilities, a strongly biologically determined dimension of cognitive abilities that is not influenced by education and is primarily based on reasoning, processing rate, and working memory capacity (Cattell 1971; Horn 1982). Although the format of the test differs from those usually used in surveys, young people's willingness to participate has been high (Schupp and Hermann 2009).

Availability: Since 2000

Dataset: \$jugend (CS), *jugendl* (long)

Respondent: 16-17-year-olds in the household

Content:

- Living
- Relationships
- Leisure and sports
- School (educational attainment, foreign languages, extra-curricular activities)
- Allowance money
- Education
- Career plans
- Future
- Background
- Childhood and Upbringing
- Attitudes/opinions
- Self-Perception
- Life satisfaction
- Political party preferences

Cognitive Tests for Youth

In 2006, a separate questionnaire with cognitive tests for adolescents was used for the first time in the SOEP. It was named “Lust auf DJ” (or “interest in DJ”) as a play on disc jockey, but DJ stands for “Denksport und Jugend”, or mind sports and youth. The questionnaire was created for young people between the ages of 16 and 17.

Availability: Since 2007

Dataset: *cogdj* (CS)

Respondent: 16-17-year-olds in the household as a supplement to the youth questionnaire

Content:

- Assignment of word pairs
- Complete equations
- Assign figures

3.1.7 Additional Instruments

Catch-Up Individual Questionnaire

The Catch-Up or “Gap” (German:Lücke) questionnaire is given to respondents who failed to respond in the previous year of the study. They are asked to provide important data about the year they missed.

Availability: Since 1987

Dataset: *pluecke* (CS), *plueckel* (long)

Respondent: SOEP respondents who are temporarily unavailable.

Content:

All data refer to the previous survey year

- Status of the respondent
- Occupational change
- Receipt of social benefits within the last year
- Completion of education
- Type of educational attainment
- Change of family status

Deceased Individual Questionnaire

In 2009, for the first time in SOEP-Core, information was collected on former SOEP participants who had died since the last survey in 2008. The Deceased Individual questionnaire thus completes the life history information in the SOEP. The primary aim is to obtain as much information as possible about the causes and circumstances of death of former SOEP respondents. As the questionnaire also collects information on individuals who have never participated in the SOEP survey, this can be used together with the causes and circumstances of death in socio-scientific analysis.

Availability: Since 2009

Dataset: *vp* (CS), *vpl* (long)

Respondent: SOEP respondents who lost a loved one.

Content:

- Relationship to the deceased
- Was the deceased a survey respondent?
- Domestic environment of the deceased
- Cause and place of death
- Last will and testament
- Health of the deceased
- Life satisfaction of the deceased
- Influence of bereavement on respondent's own life

Grip Strength Test

Availability: Since 2008

Dataset: *gripstr* (long)

Respondent: Persons over 17 years in the household

Content:

This test measures hand grip strength, which is useful in assessing respondents' physical condition.

Interviewer Questionnaire

We derive basic demographical and employment information on interviewers from personnel data of the fieldwork organization. Since 2000, Kantar Public regularly updates these information. Additionally, at irregular intervals, the SOEP interviewers complete a short version of the standard individual questionnaire themselves, which is called the interviewer questionnaire.

Availability: 2006, 2012, 2016

Dataset: *interviewer* (long)

Respondent: SOEP interviewers

Content:

- Basic Demography
- Occupational History
- Personality
- Motivation
- Interviewer Training
- Worries
- Language Skills

Last change: May 16, 2023

3.2 Scales Manual

Introduction

This manual briefly describes the theoretical background and development of all of the scales used in the Socio-Economic Panel (SOEP) study. It also provides the relevant citations as well as the items belonging to the scales and the answer format, including the verbal anchors.

The unique value of this manual lies in the presentation of each scale in the form of easy-to-understand tables listing variable names of the items in the scale in a wave-specific dataset (labels). This allows the individual items to be found and aggregated quickly.

In addition, the tables in this manual include mean values (M), standard deviations (SD), corrected item-total correlations (CITC), and information as to whether the item has to be recoded before aggregation (R). The number of valid cases in the particular survey years, as well as two measures of reliability, the internal consistency (Cronbach's Alpha) and test-retest correlations (if available), are also reported.

3.2.1 Affective Well-Being

Summary

Affective well-being describes the balance between positive and negative emotional experiences and is of interest for the research on well-being in psychology and economics as well as in the social sciences in general. The scale consists of four items and has been used in the SOEP since 2007.

Theoretical Background

In the psychological literature, subjective well-being is assumed to consist of two components: cognitive well-being and affective well-being (Schimmack et al., 2002). Here, affective well-being represents the emotional component of subjective well-being. In contrast to cognitive well-being, which is based on reflexive evaluation of subjective well-being, affective well-being hinges on the balance between positive and negative emotions (Sumner, 1996). The distinction between cognitive and affective well-being is important for the scientific investigation of subjective well-being and for policy considerations about public well-being. If the aim is to maximize subjective well-being on both the individual and the social level, and if the two are indeed separate components, it is crucial to measure their relative importance. Despite the importance of distinguishing between cognitive and affective well-being, their relationship has been researched little to date (Schimmack, 2009). Their relationship also highlights the need to examine factors such as unemployment that potentially influence subjective well-being (e.g., Schimmack et al., 2008). Conceptually, the relationship between affective well-being and other general personality characteristics is also of interest.

Scale Development

Theoretical considerations and results from the 2006 SOEP pretest led to the construction of a scale consisting of four items. One item deals with positive experiences ("happy"), while the other three items deal with negative experiences (annoyed, afraid, sad; Schimmack, 2009). In the pilot study, the measure constructed in this manner showed a high correlation with a longer measure of affective balance consisting of 10 items and produced results that are consistent with the previous results in the relevant psychological literature (Schimmack et al., 2008).

References

- Schimmack, U. 2009. *Measuring wellbeing in the SOEP*. *Schmollers Jahrbuch*, 129, 241-249.
- Schimmack, U., Diener, E., & Oishi, S. (2002). *Life-Satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources*. *Journal of Personality*, 70, 345-384.
- Schimmack, U., Schupp, J., & Wagner, G. G. (2008). *The influence of environment and personality on the affective and cognitive component of subjective well-being*. *Social Indicators Research*, 89, 41-60.
- Sumner, L. W. (1996): *Welfare, happiness, and ethics*. Oxford: University Press.

Items

I will now read to you a number of feelings. Please indicate for each feeling how often or rarely you experienced this feeling in the last four weeks (Ich lese Ihnen eine Reihe von Gefühlen vor. Geben Sie bitte jeweils an, wie häufig oder selten Sie dieses Gefühl in den letzten vier Wochen erlebt haben):

1. Angry (ärgerlich gefühlt)?
2. Worried (ängstlich gefühlt)?
3. Happy (glücklich gefühlt)?
4. Sad (traurig gefühlt)?

Scale: 1 (Very rarely / Sehr selten) to 5 (Very often / Sehr oft)

Test-Retest Correlations

In 2009, this scale was included in a retest taken by a subsample (N = 164 completed the scale) within 30 to 49 days after the initial test. Test-retest correlations for the items were (in scale order) .46, .49, .51, and .46; scale scores correlated .54.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2007	plh0184	20829	2.89	1.03	0.40	0.66
2007	plh0185	20796	1.96	1.01	0.49	0.66
2007	plh0186R	20815	2.53	0.87	0.30	0.66
2007	plh0187	20819	2.39	1.04	0.57	0.66
2008	plh0184	19631	2.82	0.99	0.41	0.68
2008	plh0185	19602	1.94	0.98	0.50	0.68
2008	plh0186R	19605	2.52	0.86	0.36	0.68
2008	plh0187	19620	2.38	1.03	0.57	0.68
2009	plh0184	20722	2.78	1.00	0.39	0.65
2009	plh0185	20666	1.93	0.98	0.49	0.65
2009	plh0186R	20691	2.55	0.88	0.29	0.65
2009	plh0187	20689	2.37	1.02	0.56	0.65
2010	plh0184	18859	2.74	0.96	0.38	0.66
2010	plh0185	18844	1.95	0.96	0.49	0.66
2010	plh0186R	18850	2.52	0.86	0.33	0.66
2010	plh0187	18852	2.38	1.02	0.58	0.66
2011	plh0184	20969	2.75	1.01	0.40	0.65
2011	plh0185	20936	1.97	0.98	0.50	0.65
2011	plh0186R	20954	2.49	0.88	0.28	0.65
2011	plh0187	20965	2.41	1.02	0.57	0.65
2012	plh0184	20753	2.72	0.99	0.40	0.67
2012	plh0185	20741	1.91	0.96	0.50	0.67
2012	plh0186R	20739	2.45	0.85	0.35	0.67
2012	plh0187	20745	2.33	1.01	0.56	0.67
2013	plh0184	25919	2.81	1.00	0.42	0.68
2013	plh0185	25895	1.91	0.96	0.50	0.68
2013	plh0186R	25909	2.43	0.84	0.35	0.68
2013	plh0187	25921	2.35	1.01	0.58	0.68
2014	plh0184	27400	2.80	1.02	0.42	0.68
2014	plh0185	27379	1.90	0.97	0.50	0.68
2014	plh0186R	27384	2.39	0.84	0.35	0.68
2014	plh0187	27384	2.34	1.03	0.57	0.68

continues on next page

Table 1 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2015	plh0184	25338	2.79	1.00	0.42	0.67
2015	plh0185	25307	1.92	0.97	0.49	0.67
2015	plh0186R	25332	2.40	0.84	0.36	0.67
2015	plh0187	25320	2.33	1.02	0.56	0.67
2016	plh0184	24474	2.77	1.02	0.42	0.67
2016	plh0185	24451	1.96	0.98	0.49	0.67
2016	plh0186R	24464	2.40	0.84	0.32	0.67
2016	plh0187	24459	2.32	1.01	0.56	0.67
2017	plh0184	26742	2.75	1.00	0.41	0.67
2017	plh0185	26734	1.92	0.97	0.51	0.67
2017	plh0186R	26738	2.36	0.82	0.34	0.67
2017	plh0187	26733	2.31	1.01	0.55	0.67
2018	plh0184	25850	2.75	1.01	0.42	0.67
2018	plh0185	25836	1.90	0.97	0.50	0.67
2018	plh0186R	25845	2.37	0.83	0.36	0.67
2018	plh0187	25847	2.30	1.01	0.56	0.67
2019	plh0184	26013	2.72	1.02	0.42	0.67
2019	plh0185	25992	1.87	0.96	0.51	0.67
2019	plh0186R	26004	2.35	0.82	0.34	0.67
2019	plh0187	26009	2.28	1.00	0.56	0.67
2020	plh0184	26064	2.72	1.01	0.42	0.67
2020	plh0185	26063	2.06	1.03	0.49	0.67
2020	plh0186R	26062	2.38	0.84	0.36	0.67
2020	plh0187	26061	2.32	1.01	0.56	0.67

3.2.2 Anomie

Summary

Anomie describes the individual's subjective response to a community and social environment that is perceived to be threatening and unregulated. Anomie is expressed in an individual tendency towards loss of motivation and feelings of despair and helplessness (Srole, 1956). The four-item scale has been used in the SOEP at irregular intervals in the 90s and at regular five-year intervals since 2008.

Theoretical Background

Anomie refers to a condition of normlessness, that is, a lack of social norms. Durkheim (1897, 1951) introduced the concept of anomie in sociology to describe the erosion of social norms and societal rules under conditions of far-reaching structural change—for example, the conditions that arise in times of rapid social and economic transformation. The result is a breakdown of bonds between the individual and the community or society. Merton (1938) applied and expanded the concept of anomie in his theory of deviant behavior. He extended Durkheim's understanding by observing the factors that lead to anomie. In Merton's view, anomie may occur when (1) cultural goals and desires are prescribed as normative for a society as a whole; (2) the legitimate means used to achieve these goals are strictly regulated; and (3) these legitimate means are unequally distributed in the society. Anomie is then the result of a state in which the individual adheres to the society's main cultural ideas and principles but does not possess the legitimate means to attain them. The result may be various forms of deviant behavior such as criminal acts. Durkheim (1897, 1951) treated anomie primarily as a social condition, whereas Merton (1938) shifted the focus to the individual. Srole (1956) followed on Merton's (1938) understanding, contributing the social psychological construct of anomia to the constellation of themes surrounding anomie. Anomia relates to the individual, psychological side of a social condition that is perceived to be anomic. In this regard, the concept of anomia shows certain similarities with external locus of control.

Scale Development

The anomie scale taken from the German Welfare Survey (Duttenhöfer & Schröder, 1996) was shortened for the SOEP survey and one positive item was added.

References

Durkheim, E. (1951). *Suicide, a study in sociology*. Glencoe, Ill.: Free Press.

Duttenhöfer, S. & Schröder, H. (1996). *Die Wohlfahrtssurveys 1978-1993 - Variablenübersicht*. Zuma-Technischer Bericht 94/11. Mannheim.

Merton, R. K. (1938). *Social structure and anomie*. *American Sociological Review*, 3, 672-682.

Srole, L. (1956). *Social integration and certain corollaries: An exploratory study*. *American Sociological Review*, 21, 709-716.

Items

To what extent do the following statements apply to you (Wie sehr stimmen die folgenden Aussagen für Sie persönlich):

1. When I think about the future, I'm actually quite optimistic. (Wenn ich an die Zukunft denke, bin ich eigentlich sehr zuversichtlich.)
2. I often feel lonely. (Ich fühle mich oft einsam.)
3. I don't really enjoy my work. (Meine Arbeit macht mir eigentlich keine Freude.)
4. Things have gotten so complicated that I almost can't manage anymore. (Die Verhältnisse sind so kompliziert geworden, dass ich mich fast nicht mehr zurecht finde.)

Scale: 1 (Completely / Stimmt ganz und gar) to 4 (Not at all / Stimmt ganz und gar nicht)

Test-Retest Correlations

In 2005, this scale was included in a retest taken by a subsample (N = 126 completed the scale) within 30 to 49 days after the initial test. Test-retest correlations for the items were (in scale order) .39, .52, .30, and .52; scale scores correlated .60.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
1990	plh0188R	4432	2.73	0.84	0.20	0.46
1990	plh0189	4413	3.36	0.97	0.28	0.46
1990	plh0190	3430	3.31	0.88	0.23	0.46
1990	plh0191	4421	2.99	0.94	0.35	0.46
1991	plh0188R	4179	2.62	0.86	0.27	0.48
1991	plh0189	4161	3.33	0.94	0.26	0.48
1991	plh0190	3765	3.30	0.89	0.23	0.48
1991	plh0191	4157	2.89	0.92	0.34	0.48
1992	plh0188R	11000	2.68	0.80	0.27	0.55
1992	plh0189	10991	3.28	0.93	0.36	0.55
1992	plh0190	10498	3.31	0.85	0.31	0.55
1992	plh0191	10989	3.23	0.89	0.41	0.55
1993	plh0188R	13106	2.55	0.82	0.21	0.56
1993	plh0189	13098	3.25	0.94	0.37	0.56
1993	plh0190	12198	3.20	0.87	0.35	0.56
1993	plh0191	13061	3.10	0.93	0.44	0.56
1995	plh0188R	13698	2.77	0.75	0.29	0.59
1995	plh0189	13685	3.22	0.92	0.38	0.59
1995	plh0190	12791	3.26	0.83	0.37	0.59
1995	plh0191	13653	3.21	0.87	0.45	0.59

continues on next page

Table 2 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
1996	plh0188R	13464	2.71	0.77	0.30	0.59
1996	plh0189	13468	3.23	0.90	0.38	0.59
1996	plh0190	12594	3.24	0.82	0.37	0.59
1996	plh0191	13429	3.20	0.84	0.45	0.59
1997	plh0188R	13227	2.50	0.79	0.25	0.57
1997	plh0189	13205	3.27	0.90	0.37	0.57
1997	plh0190	12101	3.23	0.82	0.37	0.57
1997	plh0191	13190	3.18	0.86	0.44	0.57
2008	plh0188R	19614	2.62	0.77	0.31	0.60
2008	plh0189	19608	3.16	0.88	0.40	0.60
2008	plh0190	17957	3.25	0.81	0.36	0.60
2008	plh0191	19549	3.22	0.84	0.46	0.60
2013	plh0188R	19068	2.87	0.71	0.34	0.62
2013	plh0189	19076	3.24	0.83	0.41	0.62
2013	plh0190	17163	3.29	0.83	0.39	0.62
2013	plh0191	19009	3.35	0.83	0.46	0.62
2018	plh0188R	25769	2.82	0.76	0.28	0.59
2018	plh0189	25814	3.21	0.86	0.38	0.59
2018	plh0190	22900	3.24	0.85	0.37	0.59
2018	plh0191	25709	3.30	0.83	0.44	0.59

3.2.3 Basic Social Justice Orientations Scale

Summary

The Basic Social Justice Orientations (BSJO) is a short scale to measure individuals' support for equality, need, equity, and entitlement, which are four basic distributive principles in justice research (Hülle/Liebig/May 2017). After validation and use in several population surveys (LINOS-1, SOEP-IS 2012, ALLBUS 2014, ESS round nine), the module became part of the SOEP questionnaire program from 2019 onwards.

Theoretical Background

Starting as a equity theory (Adams 1963) the principles of justice were extended by a “multi-principle approach” in the mid-1970s, including equality and need (Deutsch 1975), and by the emphasis on entitlement (Miller 1976). These four principles of distributive justice form the four dimensions of the Basic Social Justice Orientations (BSJO) scale, which is used to analyze normative attitudes towards the resolution of distributive problems in society (Hülle/Liebig/May 2017). The dimension of equality implies a conception of justice according to which everyone receives an equal share of benefits and burdens. Within the 2nd dimension—equity—the distribution of benefits and burdens is considered equitable if these are linked to current individual contributions and efforts. The needs principle represents the 3rd dimension and corresponds to the idea of sharing benefits according to people's individual needs. The 4th dimension is the principle of entitlement. According to this principle, ascriptive characteristics (such as social origin) or status characteristics acquired in the past (such as occupational status) form entitlements on the basis of which benefits and burdens are to be distributed.

Scale Development

Pretests in LINOS-1, SOEP-IS 2012 and for the main survey of ALLBUS 2014 not only verified the validity of the items, they also reduced the number of items from 12 to 8 by using the criterion of factorial validity, so that each dimension was then covered by 2 items (Hülle/Liebig/May 2017). In ALLBUS 2014 two items (“It is just if all people have the same living conditions” and “It is just if income and wealth are equally distributed among the members of our society”) measured equality, two items (“A society is just if it takes care of those who are poor and needy” and “It is just if people taking care of their children or their dependent relatives receive special support and benefits”) measured

need, two items (“It is just if hard working people earn more than others” and “It is just if every person receives only that which has been acquired through their own efforts”) measured equity, and two items (“It is just if members of respectable families have certain advantages in their lives” and “It is fair if people on a higher level of society have better living conditions than those on the lower level”) measured entitlement (Sauer et al. 2014). For the European Social Survey round nine (2018/2019) previous results of the BSJO scale were used to develop and include a four-item-version with one item per dimension (Adriaans et al. 2020). In 2019, the SOEP included the ultra-brief version of with the plan to replicate them every 5 years in the questionnaire program. However, a focus module on perceptions of inequality was conducted in the SOEP 2021 (Adriaans et al. 2021). The BSJO scale was repeated in 2021 as a complement to the content of this module.

References

Adams, J. S. (1963). *Towards an understanding of inequity*. *The Journal of Abnormal and Social Psychology*, 67(5), 422–436.

Adriaans, J., Bohmann, S., Targa, M., Liebig, S., Hinz, T., Jasso, G., ... & Sabbagh, C. (2020). *Justice and fairness in europe: Topline results from round 9 of the european social survey*.

Adriaans, J., Griese, F., Auspurg, K., Bledow, N., Bohmann, S., Busemeyer, M. R., ... & Verwiebe, R. (2021). *Dokumentation zum Entwicklungsprozess des Moduls "Einstellungen zu sozialer Ungleichheit" im SOEP (v38) (No. 1071)*. *SOEP Survey Papers*.

Deutsch, M. (1975). *Equity, equality, and need: What determines which value will be used as the basis of distributive justice?* *Journal of Social Issues*, 31(3), 137–149.

Hülle, S., Liebig, S., & May, M. J. (2018). *Measuring attitudes toward distributive justice: The basic social justice orientations scale*. *Social Indicators Research*, 136(2), 663–692.

Miller, D. (1976). *Social justice*. Oxford: Clarendon Press.

Valet, P., May, M., Sauer, C., & Liebig, S. (2014). *LINOS-1: Legitimation of inequality over the life-span*.

Items

People have different ideas about what makes a society just. What’s your opinion about the following statements? (Es gibt unterschiedliche Vorstellungen darüber, wann eine Gesellschaft gerecht ist. Wie ist Ihre persönliche Meinung zu den folgenden Aussagen?)

1. A society is just when people who work hard earn more than others / Es ist gerecht, wenn Personen, die hart arbeiten, mehr verdienen als andere (Equity)
2. A society is just when people from respected families have advantages in life / Es ist gerecht, wenn Personen, die aus angesehenen Familien stammen, dadurch Vorteile im Leben haben (Entitlement)
3. A society is just when it takes care of the weak and needy / Eine Gesellschaft ist gerecht, wenn sie sich um die Schwachen und Hilfsbedürftigen kümmert (Need)
4. A society is just when the income and wealth in society are equally distributed among all people. / Es ist gerecht, wenn Einkommen und Vermögen in unserer Gesellschaft an alle Personen gleich verteilt werden. (Equality)

Scale: 1 (Disagree completely / Stimme überhaupt nicht zu) to 7 (Agree completely / Stimme voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2019	plh0396i01	25841	6.28	1.08	0.21	0.21
2019	plh0396i02	25792	2.09	1.45	0.12	0.21
2019	plh0396i03	25856	6.19	1.11	0.05	0.21
2019	plh0396i04	25742	3.10	1.91	0.08	0.21

3.2.4 Cognitive Competencies

Symbol-Digit Test and Animal Naming Task

Summary

The two ultra-short cognitive performance tasks allow for reliable assessment of general intellectual ability and distinguish between two components of intellectual functioning: cognitive mechanics and pragmatics (e.g., Lindenberger & Baltes, 1997). Each test takes 90 seconds and is completed on a laptop. The respondent completes the Symbol-Digit Test him/herself, whereas the interviewer documents the answers to the Animal Naming Task. Both tasks therefore require the survey mode of Computer Assisted Personal Interviewing (CAPI).

Theoretical Background

Cognitive mechanisms are hard-wired, biologically based capacities for information processing and are measured with the Symbol-Digit Test (SDT). The study of cognitive mechanics deals with differences in cognitive performance, for example, in the speed, accuracy, processing capacity, coordination, and inhibition of basic cognitive processes. Prime examples include perceptual speed, working memory, and the capacity for deductive reasoning. The cognitive mechanisms usually develop in a process continuing up to early adulthood, then begin to decline gradually and may deteriorate more rapidly in some areas in old age. Cognitive pragmatics are education- and experience-related competencies, which are measured with the Animal Naming Task (ANT). The development of cognitive pragmatics is the result of investments in the development of cognitive mechanics in selected behavioral areas early in the life course (e.g., in educational trajectories, in training). The cognitive pragmatics develop continuously throughout life, reaching their peak late in the life course and declining only marginally in old age. Unsurprisingly, the cognitive abilities that build on pragmatic intellectual abilities (e.g., knowledge, vocabulary, wisdom) usually correlate much more strongly with socio-economic resources such as education, income, and occupational prestige, whereas the development of basic cognitive mechanisms over the life course is much more strongly affected by individual sensory and psychomotor resources (Baltes, Lindenberger & Staudinger, 1998).

Scale Development

The individual tests of perceptual speed (Symbol-Digit Test) and word fluency (Animal Naming Test) were modified for use with a computer-assisted survey mode. The decisive factor in this was the need to be able to use the tests without any special interviewer training and to reduce sources of error in the framework of Computer-Assisted Personal Interviewing (CAPI) as much as possible. Further information on the development of the tests can be found in Lang (2005), Lang, Weiss, Stocker and von Rosenblatt (2007), and in Schupp, Herrmann, Jaensch and Lang (2008).

References

- Baltes, P. B., Lindenberger, U. & Staudinger, U. M. (1998). *Life-span theory in developmental psychology*. In R. M. Lerner (Ed.), *Handbook of child psychology (5th edition, Vol. 1: Theoretical models of human development)*, pp. 1029 – 1143. New York: Wiley.
- Lang, F. R. (2005). *Erfassung des kognitiven Leistungspotenzials und der “Big Five” mit Computer-Assisted-Personal-Interviewing (CAPI): Zur Reliabilität und Validität zweier ultrakurzer Tests und des BFI-S (Assessment of cognitive capabilities and the Big Five with Computer-Assisted Personal Interviewing (CAPI): Reliability and validity)*. German Institute of Economic Research. Berlin: DIW Berlin.
- Lang, F. R., Weiss, D., Stocker, A., & von Rosenblatt, B. (2007). *Assessing cognitive capacities in Computer-Assisted Survey Research: Two ultra-short tests of intellectual ability in the German Socio-Economic Panel (SOEP)*. *Schmollers Jahrbuch*, 127, 183-192.
- Lindenberger, U., & Baltes, P. B. (1997). *Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study*. *Psychology and Aging*, 12, 410-432.
- Schupp, J., Herrmann, S., Jaensch, P., & Lang, F. R. (2008). *Erfassung kognitiver Leistungspotentiale Erwachsener im Sozio-oekonomischen Panel (SOEP)*. Berlin: DIW Berlin.

Items

Symbol-Digit Test & Animal Naming Task

1. In 30 seconds.
2. In 30-60 seconds.
3. In 60-90 seconds.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2006	f99z30r	5790	8.09	4.09	0.86	0.94
2006	f99z60rneu	5790	9.05	4.10	0.91	0.94
2006	f99z90rneu	5790	8.61	3.62	0.86	0.94
2012	f99z30r	7342	8.97	3.82	0.78	0.90
2012	f99z60rneu	7342	10.22	3.55	0.85	0.90
2012	f99z90rneu	7342	9.98	3.07	0.77	0.90
2016	f99z30r	18155	9.50	6.13	0.39	0.68
2016	f99z60rneu	18155	10.62	3.72	0.65	0.68
2016	f99z90rneu	18155	10.25	3.23	0.61	0.68
2006	f96t30g	5790	11.83	6.02	0.40	0.65
2006	f96t60gneu	5790	7.18	4.60	0.57	0.65
2006	f96t90gneu	5790	5.06	4.19	0.47	0.65
2012	f96t30g	1285	12.64	5.18	0.40	0.68
2012	f96t60gneu	1285	8.86	4.44	0.58	0.68
2012	f96t90gneu	1285	6.34	4.14	0.52	0.68
2016	f96t30g	809	11.99	4.83	0.46	0.70
2016	f96t60gneu	809	8.88	4.18	0.60	0.70
2016	f96t90gneu	809	6.53	3.93	0.52	0.70

Multiple-Choice Vocabulary Intelligence Test (MWT)

Summary

The Multiple-Choice Vocabulary Intelligence Test (Mehrfachwahl-Wortschatz-Intelligenztest; MWT; Lehrl, 2005) aims to measure the education- and experience-related cognitive pragmatics. The test asks for knowledge and is therefore only minimally influenced by currently available cognitive capacities. The test takes about 5 minutes. The respondents are asked to find the existing and commonly known word in 37 groups of five words each with four words in each group being fictive and newly constructed (multiple-choice). The 37 groups are ordered by difficulty and the test is finished after three incorrect classifications.

Scale Development

The Multiple-Choice Vocabulary Intelligence Test (Version A) was modified for use with a computer-assisted survey mode. After pretesting in 2011 the test was introduced into the SOEP in 2012 (N = 6,864, M = 28.14, SD = 6.98).

References

Lehrl, S. (1991). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-A; (Parallelform zum MWT-B)*. Erlangen: perimed-Fachbuch-Verl.-Ges.

Items

The MWT items are not allowed to be published.

I-S-T 2000 R

Summary

In the year 2006, the SOEP included its first cognition test for adolescents in the range of SOEP survey instruments. Since then, the test has been carried out annually with each new cohort of adolescents. A specially designed questionnaire is used to measure adolescents' cognitive mechanics based on the I-S-T 2000 test in three modules (word analogies, number sequences, and matrices). In strict adherence to the test criteria, interviewer-based surveying is compulsory. Also, the use and especially design and structure of the questionnaire require that the test be conducted by an interviewer and that the time be monitored.

Theoretical Background

The cognitive test measures adolescents' cognitive mechanics in three different modules. Cognitive mechanics refer to the hard-wired, biologically based capacities for information processing. The task set "word analogies" measures respondents' verbal cognitive potentials. Respondents are asked to assign words to a given sequence of words according to a specific rule. The test uses the individual's own vocabulary to measure combinatory ability. The task set "number sequences" measures the numerical cognitive potential of the respondents. These tasks focus primarily on the adolescents' abstract combinatory ability and logical thinking. The task set "matrices" measures the figural cognitive abilities of the respondents.

Scale Development

To study the cognitive performance potential of adolescents, we developed a questionnaire based on the I-S-T 2000-Test (Amthauer, Brocke, Liepmann & Beauducel, 2001) that is appropriate for an individual panel survey. The modifications have been described in detail by Solga et al. (2005). Further information on test development can be found in Schupp and Herrmann (2009).

References

- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) – Handanweisung*. Göttingen: Hogrefe.
- Solga, H., Stern, E., von Rosenbladt, B., Schupp, J., & Wagner, G. G. (2005). *The measurement and importance of general reasoning potentials in schools and labour markets*. Berlin: DIW Berlin.
- Schupp, J., & Herrmann, S. (2009). *Kognitionspotenziale Jugendlicher. Ergänzung zum Jugendfragebogen der Längsschnittstudie Sozio-oekonomisches Panel (SOEP)*. Berlin: DIW Berlin.

Items

I-S-T 2000 R

1. Word analogies.
2. Number sequences.
3. Matrices.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2006-2020	analog	3570	8.18	3.65	0.53	0.69
2006-2020	rechenz	3545	12.59	4.91	0.50	0.69
2006-2020	matrize	3562	9.78	3.43	0.54	0.69

3.2.5 Conspiracy Mentality

Summary

Conspiracy mentality has been measured in the SOEP in 2021 using the 5-item Conspiracy Mentality Questionnaire (CMQ) - designed to assess the general propensity to believe in conspiracy theories. It has been included in the SOEP in 2021.

Theoretical Background

Conspiracy theories provide alternative explanations for important world events and have rapidly gained ground in public discourse throughout the past century. In addition, there is increasing evidence for the existence of stable individual differences in people's propensity to believe in conspiracy theories and the effect of such differences on both attitudes and behavior. If a person believes in one conspiracy theory, he or she will also be more likely to believe in other conspiracy theories (Swami et al., 2010). Individual differences in conspiracy mentality have important consequences as they predict prejudice against powerful societal groups. Consequences may be political disengagement or critical health behaviors such as vaccination uptake (Kata, 2010) and adherence to medication regimens (Bogart, Wagner, Galvan, & Banks, 2010).

Scale Development

The CMQ was developed and validated by Bruder et al. (2012) for integration in comparable international panel studies or surveys. It has been used to test cross-cultural differences in a large international sample ($N = 7,766$) and both its convergent and discriminant validity have been tested in four studies. The CMQ has proven to be the strongest predictor for beliefs in specific conspiracy theories, even when controlling for a broad range of other individual differences measures (Bruder et al., 2012). The SOEP has included the CMQ in its original form in 2021.

References

- Bogart, L. M., Wagner, G., Galvan, F. H., & Banks, D. (2010). *Conspiracy beliefs about HIV are related to antiretroviral treatment nonadherence among African American men with HIV*. *J AIDS-Journal of Acquired Immune Deficiency Syndromes*, 53, 648-655.
- Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). *Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire*. *Frontiers in Psychology*, 4, 255.
- Kata, A. (2010). *A postmodern Pandora's box: Anti-vaccination misinformation on the Internet*. *Vaccine*, 28, 1709-1716.
- Swami, V., Chamorro Premuzic, T., & Furnham, A. (2010). *Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs*. *Applied Cognitive Psychology*, 24, 749-761.

Items

How would you rate the following statements? Please indicate the probability on a scale from 0% to 100%. 0% means: you certainly do not believe that to be true. 100% means: you definitely believe this to be true. You can use the values in between to grade the probability. (Wie beurteilen Sie die folgenden Aussagen? Bitte geben Sie die Wahrscheinlichkeit auf einer Skala von 0% bis 100% an. 0% bedeutet: Das halten Sie mit Sicherheit nicht für wahr. 100% bedeutet: Das halten Sie mit Sicherheit für wahr. Mit den Werten dazwischen können Sie die Wahrscheinlichkeit abstufen.)

1. Many very important things happen in the world, which the public is never informed about / Es geschehen viele sehr wichtige Dinge in der Welt, über die die Öffentlichkeit nie informiert wird
2. Politicians usually do not tell us the true motives for their decisions / Politiker geben uns normalerweise keine Auskunft über die wahren Motive ihrer Entscheidungen
3. Government agencies closely monitor all citizens/ Regierungsbehörden überwachen alle Bürger genau
4. Events which superficially seem to lack a connection are often the result of secret activities/ Ereignisse, die auf den ersten Blick nicht miteinander in Verbindung zu stehen scheinen, sind oft das Ergebnis geheimer Aktivitäten

5. There are secret organizations that greatly influence political decisions / Es gibt geheime Organisationen, die großen Einfluss auf politische Entscheidungen haben

Scale: 0% (certainly not / mit Sicherheit nicht) to 100% (certain / mit Sicherheit)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2021	plh0425i01	17673	6.17	2.90	0.56	0.84
2021	plh0425i02	17526	6.12	2.52	0.57	0.84
2021	plh0425i03	17473	3.88	2.90	0.65	0.84
2021	plh0425i04	16976	3.60	2.86	0.74	0.84
2021	plh0425i05	17246	4.06	3.24	0.69	0.84

3.2.6 Effort-Reward Imbalance Model

Summary

The Effort-Reward Imbalance Model is designed to measure occupational and job stress. In this model, stress is conceptualized as high work-related effort coupled with low control over job-related rewards. The model also considers the personality trait of excessive willingness to overexert oneself, which exacerbates the negative health consequences of occupational stress. The SOEP uses a shortened version of the ERI questionnaire comprising a total of 16 four- or five- level items. The ERI questionnaire has been used in the SOEP at regular five-year intervals since 2006.

Theoretical Background

The ERI model is a newly developed instrument used to measure job stress and its negative health consequences (Siegrist, 1996). Job stress is conceptualized as the relation between demands and efforts on the one hand and rewards on the other. This conceptualization is based on the assumption that job relations are subject to the general social norm of reciprocity. The general reciprocity norm states that social interaction (here the performance of job tasks) always takes place in expectation of appropriate compensation. If this expectation is not fulfilled over the long term—that is, if the reciprocity norm is violated—in many cases the interaction will be terminated. The termination of employment relationships is often not possible on short notice or without far-reaching consequences, which means that employment relationships often have to be maintained while violating the norm of reciprocity. The resulting imbalance between high perceived demands and low perceived rewards is referred to as a “gratification crisis.” Gratification crises lead to lower job satisfaction and high stress levels, and thus constitute a risk factor for numerous psycho-social and stress-related illnesses (Siegrist, 1996). Alongside these external factors in work stress, the ERI model also contains an intrinsic dimension, “overcommitment,” or the excessive willingness to overexert oneself. This is based on the hypothesis that people with an excessive willingness to overexert themselves suffer more severely from gratification crises than individuals with a lower willingness to overexert themselves.

Scale Development

In the previous version, the instrument only contained 23 items. With the goal of achieving a more broadly applicable instrument, it was shortened to 16 items and used for the first time in this form in the 2006 SOEP questionnaire. Overcommitment was not shortened in the SOEP version, since the scale is already the product of various validation studies.

References

Department of Medical Sociology, Duesseldorf University (2008). *Effort-reward imbalance at work: Theory, measurement and evidence*. Düsseldorf.

Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1, 27.

Items

Please indicate to what degree you agree with the following statements (Bitte geben Sie an, in welchem Maße Sie den folgenden Aussagen zustimmen):

Effort

1. I have constant time pressure due to a heavy workload. (Aufgrund des hohen Arbeitsaufkommens besteht häufig großer Zeitdruck.)
2. I have many interruptions and disturbances while performing my job. (Bei meiner Arbeit werde ich häufig unterbrochen und gestört.)
3. Over the past few years, my job has become more and more demanding. (Im Laufe der letzten beiden Jahre ist meine Arbeit immer mehr geworden.)

Scale: 1 (Yes / Ja) to 2 (No /Nein); 1 (Not at all / Gar nicht) to 4 (Very heavily / Sehr stark) is recoded to 1 to 5

Reward

1. I receive the respect I deserve from my superior. (Ich erhalte von meinen Vorgesetzten die Anerkennung, die ich verdiene.) (R)
2. My job promotion prospects are poor. (Die Aufstiegschancen in meinem Betrieb sind schlecht.) (R)
3. I have experienced or I expect to experience an undesirable change in my work situation. (Ich erfahre – oder erwarte – eine Verschlechterung meiner Arbeitssituation.) (R)
4. My job security is poor. (Mein eigener Arbeitsplatz ist gefährdet.) (R)
5. Considering all my efforts and achievements, I receive the respect and prestige I deserve at work. (Wenn ich an all die erbrachten Leistungen und Anstrengungen denke, halte ich die erfahrene Anerkennung für angemessen.) (R)
6. Considering all my efforts and achievements, my job promotion prospects are adequate. (Wenn ich an all die erbrachten Leistungen und Anstrengungen denke, halte ich meine persönlichen Chancen des beruflichen Fortkommens für angemessen.) (R)
7. Considering all my efforts and achievements, my salary / income is adequate. (Wenn ich an all die erbrachten Leistungen denke, halte ich mein Gehalt / meinen Lohn für angemessen.) (R)

Scale: 1 (Yes / Ja) to 2 (No /Nein); 1 (Not at all / Gar nicht) to 4 (Very heavily / Sehr stark) is recoded to 1 to 5

Overcommitment

1. At work, I easily get into time pressure. (Beim Arbeiten komme ich leicht in Zeitdruck.)
2. I often am already thinking about work-related problems when I wake up. (Es passiert mir oft, dass ich schon beim Aufwachen an Arbeitsprobleme denke.)
3. When I get home, it is easy to switch off from thinking about work. (Wenn ich nach Hause komme, fällt mir das Abschalten von der Arbeit sehr leicht.) (R)
4. Those closest to me say I sacrifice too much for my career. (Diejenigen, die mir am nächsten stehen sagen, dass ich mich für meinen Beruf zu sehr aufopfere.)
5. Work seldom lets go of me; it stays in my head all evening. (Die Arbeit lässt mich selten los, das geht mir abends im Kopf rum.)
6. If I put off something that needs to be done that day, I can't sleep at night. (Wenn ich etwas verschiebe, was ich eigentlich heute tun müsste, kann ich nachts nicht schlafen.)

Scale: 1 (Yes / Ja) to 2 (No /Nein); 1 (Not at all / Gar nicht) to 4 (Very heavily / Sehr stark) is recoded to 1 to 5

Coding

Please note that effort and reward are measured by two questions in the SOEP that have to be recoded to a five-point Likert scale. The first of these questions asks whether the content of the items applies to the respondent. Depending

on the answer to this first question, the level of emotional burden is then evaluated in a second four-step question. An affirmative answer to the reward item “I get the recognition I deserve from my superiors” is encoded with the highest value (5), while a negative answer is encoded depending on the burden: No reward but no burden is encoded with the value 4. No reward and a high burden is encoded with 1 on the five-point Likert scale. The same procedure applies for effort items. When recoding, attention to the five-point scale is required because of the inverse direction of some reward items. A negative answer to an inverse reward item—for example, “The opportunities for advancement in my company are bad”—is coded with the highest value (5) for reward. An affirmative answer to this inverse reward item must be encoded with 1, 2, 3, or 4 depending on how burdened the person feels. For example: An affirmative answer to “advancement in company is bad” and a high rating for burden equals 1 for reward on the five-point Likert scale. The Effort-Reward Ratio (ERR) is calculated by the formula $ERR = e/(r \cdot c)$, where “e” stands for total effort, “r” stands for total reward, and “c” is a correction factor for the ratio of the number of items used. In this case, “c” equals 3/7 because effort is measured with 3 items and reward is measured with 7 items. Thus, the total effort score ranges from 3 to 15, as three effort items with a five-point Likert scale are used in SOEP. High values correspond to high perceived effort, low values to low effort. The total score for reward is constructed analogously and ranges between 7 and 35 for seven reward items. Low values mean low reward. A ERR value close to 0 means low perceived effort at high perceived reward. A ERR value greater than 1 indicates the opposite. In this case, there is an effort-reward imbalance.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2006	eri1	12315	2.43	1.31	0.61	0.74
2006	eri2	12313	2.19	1.29	0.51	0.74
2006	eri3	12232	2.39	1.31	0.57	0.74
2011	eri1	11594	2.36	1.30	0.61	0.74
2011	eri2	11581	2.23	1.28	0.53	0.74
2011	eri3	11444	2.35	1.31	0.57	0.74
2012	eri1	5130	2.34	1.31	0.57	0.71
2012	eri2	5131	2.07	1.26	0.48	0.71
2012	eri3	4976	2.29	1.28	0.55	0.71
2016	eri1	592	2.21	1.29	0.63	0.77
2016	eri2	592	2.20	1.28	0.55	0.77
2016	eri3	586	2.36	1.29	0.62	0.77
2006	eri4	11594	4.31	1.07	0.58	0.79
2006	eri5	11809	3.83	1.19	0.51	0.79
2006	eri6	12195	4.25	1.29	0.50	0.79
2006	eri7	12179	4.48	1.15	0.40	0.79
2006	eri8	11720	4.30	1.07	0.61	0.79
2006	eri9	11592	4.24	1.10	0.61	0.79
2006	eri10	11872	3.77	1.31	0.46	0.79
2011	eri4	10868	4.30	1.09	0.63	0.80
2011	eri5	10900	3.92	1.16	0.53	0.80
2011	eri6	11524	4.47	1.12	0.46	0.80
2011	eri7	11499	4.70	0.90	0.33	0.80
2011	eri8	11156	4.28	1.09	0.66	0.80
2011	eri9	10910	4.28	1.09	0.64	0.80
2011	eri10	11273	3.90	1.31	0.50	0.80
2012	eri4	4713	4.38	1.04	0.62	0.78
2012	eri5	4775	4.04	1.11	0.50	0.78
2012	eri6	5042	4.52	1.06	0.45	0.78
2012	eri7	5033	4.73	0.89	0.33	0.78
2012	eri8	4757	4.35	1.04	0.64	0.78
2012	eri9	4623	4.30	1.06	0.62	0.78
2012	eri10	4884	3.80	1.32	0.42	0.78

continues on next page

Table 7 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2016	eri4	543	4.36	1.05	0.65	0.80
2016	eri5	555	4.01	1.05	0.52	0.80
2016	eri6	585	4.58	1.01	0.47	0.80
2016	eri7	589	4.72	0.87	0.36	0.80
2016	eri8	550	4.35	1.03	0.66	0.80
2016	eri9	533	4.28	1.06	0.68	0.80
2016	eri10	567	3.93	1.24	0.43	0.80
2006	plb0112	12348	2.40	0.88	0.45	0.79
2006	plb0113	12346	2.16	0.95	0.67	0.79
2006	plb0114R	12328	2.29	0.94	0.41	0.79
2006	plb0115	12299	2.26	0.94	0.50	0.79
2006	plb0116	12347	2.18	0.92	0.72	0.79
2006	plb0117	12340	1.97	0.88	0.47	0.79
2011	plb0112	11624	2.33	0.89	0.45	0.79
2011	plb0113	11621	2.12	0.96	0.68	0.79
2011	plb0114R	11601	2.27	0.96	0.43	0.79
2011	plb0115	11571	2.23	0.96	0.52	0.79
2011	plb0116	11609	2.13	0.92	0.74	0.79
2011	plb0117	11615	1.93	0.88	0.49	0.79
2012	plb0112	5149	2.47	0.95	0.42	0.77
2012	plb0113	5149	2.05	0.98	0.64	0.77
2012	plb0114R	5138	2.18	0.99	0.41	0.77
2012	plb0115	5117	2.14	0.98	0.51	0.77
2012	plb0116	5140	2.06	0.94	0.70	0.77
2012	plb0117	5134	1.84	0.90	0.43	0.77
2016	plb0112	15107	2.37	0.92	0.42	0.77
2016	plb0113	15106	2.00	0.95	0.65	0.77
2016	plb0114R	15087	2.16	0.99	0.42	0.77
2016	plb0115	14975	2.10	0.97	0.50	0.77
2016	plb0116	15105	2.00	0.92	0.71	0.77
2016	plb0117	15084	1.80	0.87	0.44	0.77

Note

It should be noted that effort and reward are measured by two questions that have to be recoded to a five-point Likert scale in the SOEP data. Consequently, ERI1 to ERI10 refers to the following variable labels: ERI1 (plb0118/19), ERI2 (plb0120/21), ERI3 (plb0122/23), ERI4 (plb0130/31), ERI5 (plb0124/25), ERI6 (plb0126/27), ERI7 (plb0128/29), ERI8 (plb0132/33), ERI9 (plb0134/35), ERI10 (plb0136/37).

3.2.7 Impulsiveness & Patience**Summary**

The measures of impulsiveness and patience are simple and ultra-short survey measures with the distinct advantage of gathering information on time preferences that are behaviorally relevant while at the same time being cheap to measure on a large and representative scale. Impulsiveness and patience has been measured at regular five-year intervals since 2008.

Theoretical Background

Essentially, all economic decisions involve a time dimension and thus a trade-off between different payoffs or costs that accrue at different points in time. Patience (or the rate of time preference) is the central preference parameter that

shapes such decisions over time.

Scale Development

So far, there is still no reliable measure of patience in large-scale representative surveys, since direct measures of patience are typically elicited in laboratory experiments among particular (student) subject pools only. To fill this gap, a survey measure of patience and impulsiveness has been included in the SOEP. The measures for impulsiveness and patience were evaluated with an incentive-compatible intertemporal choice experiment for impatience. Individuals who state that they are more impatient also exhibit a higher degree of impatience in the incentivized choice experiment (see Vischer et al., 2013).

References

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. G. (2013). *Validating an ultra-short survey measure of patience. Economics Letters, 120, 142-145.*

Items

How would you describe yourself (Wie schätzen Sie sich persönlich ein):

Impulsiveness

1. Do you generally think things over for a long time before acting – in other words, are you not impulsive at all? Or do you generally act without thinking things over a long time – in other words, are you very impulsive? (Sind Sie im Allgemeinen ein Mensch, der lange überlegt und nachdenkt, bevor er handelt, also gar nicht impulsiv ist? Oder sind Sie ein Mensch, der ohne lange zu überlegen handelt, also sehr impulsiv ist?)

Scale: 0 (Not at all impulsive / Gar nicht impulsiv) to 10 (Very impulsive / Sehr impulsiv)

Patience

2. Are you generally an impatient person, or someone who always shows great patience? (Sind Sie im Allgemeinen ein Mensch, der ungeduldig ist, oder der immer sehr viel Geduld aufbringt?)

Scale: 0 (Very impatient / Sehr ungeduldig) to 10 (Very patient / Sehr geduldig)

Items and Scale Statistics

year	variable	count	mean	sd
2008	plh0254	19635	5.09	2.19
2008	plh0253	19643	6.07	2.28
2013	plh0254	19107	5.18	2.21
2013	plh0253	19114	6.18	2.34
2018	plh0254	25910	5.03	2.30
2018	plh0253	25936	6.00	2.45

3.2.8 Life Goals

Summary

Respondents' life goals have been measured in the SOEP in 1990, 1992, 1995, since 2004 at four-year intervals and since 2016 at five-year intervals. They have been used with ten items that can be grouped into three scales (success, family life, and altruism).

Theoretical Background

Life goals can be thought of as “relatively long- term, value-laden life objectives” (Meier et al., 1959). They have been conceptualized as “organizers” of developmental self-regulation that individuals use to influence their own development as they adapt to the constraints of a given situation (Heckhausen, 1999). Individual life goals correspond to societal

values to the extent that they are aligned with cultural preferences, if not culturally prescribed norms. A classic example is the goal of material success as embodied in the achievement of money, power, and security, which in turn lead to prestige and social recognition. However, overarching non-monetary goals such as “the good life” or “personal development” are also taking on an increasingly important role in individuals’ lives (Meier et al., 1959). Research on life goals has investigated their impact on individuals’ future orientations and their occupational, educational, and family-related decision making (Chang et al., 2006). Life goals also play a prominent role in motivational theories of life-span development, which focus on goal commitment, planning, and eventual goal attainment (Heckhausen et al., 2010). Another strand of research is concerned with the prioritization of life goals and perceived control over goal attainment (Heckhausen, 1999). Recent evidence that life goals play a role in life satisfaction has contributed to the research on subjective well-being and life satisfaction (Headey 2008).

Scale Development

The items are based (albeit with some changes of wording) on a classification of goals and measures initially developed by Kluckhohn and Strodtbeck (1961) and translated into German by Bielenski and Strümpel (1988). More information on the development of the scale is given by Headey (2008).

References

- Bielenski, H. & Strümpel, B. (1988). *Eingeschränkte Erwerbsarbeit bei Frauen und Männern. Fakten - Wünsche - Realisierungschancen*. Berlin: Edition Sigma.
- Chang, E. S., Chen, C., Greenberger, E., Dooley, D., & Heckhausen, J. (2006). What do they want in life?: The life goals of a multi-ethnic, multi-generational sample of high school seniors. *Journal of Youth and Adolescence*, 35, 302-313.
- Headey, B. W. (2008). Life goals matter to happiness: A revision of set-point theory. *Social Indicators Research*, 86, 213-231.
- Heckhausen, J. (1999). *Developmental regulation in adulthood: Age-normative and sociostructural constraints as adaptive challenges*. New York: Cambridge University Press.
- Heckhausen, J., Wrosch, C., & Schulz, R. (2010). A motivational theory of life-span development. *Psychological Review*, 117, 32-60.
- Kluckhohn, F. R., & Strodtbeck, F. L. (1961). *Variations in value orientations*. Evanston, Illinois: Row, Peterson.
- Meier, D. L., & Bell, W. (1959). Anomia and differential access to the achievement of life goals. *American Sociological Review*, 24, 189-202.

Items

Are the following things currently ... for you (Sind für Sie persönlich die folgenden Dinge heute ...):

Success

1. Being able to afford to buy things for myself (Sich etwas leisten können).
2. Being fulfilled (Sich selbst verwirklichen).
3. Being successful in my career (Erfolg im Beruf haben).
4. Seeing the world and/or traveling extensively (Die Welt sehen, viele Reisen machen).

Scale: 1 (Very important / Sehr wichtig) to 4 (Not at all important / Ganz unwichtig)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
1990	plh0104	7136	2.07	0.62	0.36	0.60
1990	plh0106	7101	2.15	0.76	0.45	0.60
1990	plh0107	6845	2.18	0.92	0.43	0.60
1990	plh0112	7143	2.47	0.84	0.31	0.60

continues on next page

Table 9 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
1992	plh0104	11018	1.96	0.62	0.38	0.61
1992	plh0106	10927	2.13	0.79	0.47	0.61
1992	plh0107	10672	2.11	0.95	0.44	0.61
1992	plh0112	11008	2.47	0.83	0.30	0.61
1995	plh0104	11663	1.98	0.61	0.36	0.61
1995	plh0106	11585	2.09	0.76	0.49	0.61
1995	plh0107	11200	2.07	0.90	0.45	0.61
1995	plh0112	11639	2.54	0.81	0.30	0.61
2004	plh0104	21933	1.98	0.59	0.37	0.61
2004	plh0106	21809	2.15	0.75	0.48	0.61
2004	plh0107	20840	2.13	0.87	0.43	0.61
2004	plh0112	21917	2.56	0.82	0.30	0.61
2008	plh0104	19632	2.01	0.60	0.37	0.63
2008	plh0106	19547	2.23	0.77	0.50	0.63
2008	plh0107	18612	2.21	0.90	0.45	0.63
2008	plh0112	19596	2.64	0.82	0.33	0.63
2010	plh0104	7788	2.06	0.66	0.39	0.59
2010	plh0106	7768	2.05	0.71	0.45	0.59
2010	plh0107	7785	2.06	0.71	0.41	0.59
2010	plh0112	7780	2.58	0.81	0.27	0.59
2012	plh0104	27925	2.06	0.63	0.36	0.61
2012	plh0106	27807	2.13	0.74	0.49	0.61
2012	plh0107	26216	2.16	0.86	0.44	0.61
2012	plh0112	27903	2.56	0.83	0.29	0.61
2016	plh0104	24603	2.06	0.63	0.35	0.62
2016	plh0106	24483	2.14	0.74	0.48	0.62
2016	plh0107	23229	2.20	0.85	0.45	0.62
2016	plh0112	24580	2.49	0.84	0.33	0.62

Items

Are the following things currently ... for you (Sind für Sie persönlich die folgenden Dinge heute ...):

Family Life

1. Owning a house (Ein eigenes Haus haben).
2. Having a happy marriage/relationship (Eine glückliche Ehe / Partnerschaft haben).
3. Having children (Kinder haben).

Scale: 1 (Very important / Sehr wichtig) to 4 (Not at all important / Ganz unwichtig)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
1990	plh0108	7125	2.39	0.99	0.32	0.56
1990	plh0109	7077	1.44	0.74	0.40	0.56
1990	plh0110	7067	1.88	0.88	0.42	0.56
1992	plh0108	10947	2.41	1.01	0.31	0.56
1992	plh0109	10936	1.40	0.73	0.42	0.56
1992	plh0110	10888	1.79	0.90	0.41	0.56

continues on next page

Table 10 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
1995	plh0108	11592	2.33	0.98	0.28	0.54
1995	plh0109	11568	1.39	0.71	0.40	0.54
1995	plh0110	11529	1.77	0.89	0.40	0.54
2004	plh0108	21811	2.32	0.98	0.31	0.55
2004	plh0109	21730	1.43	0.69	0.41	0.55
2004	plh0110	21572	1.80	0.89	0.39	0.55
2008	plh0108	19534	2.41	0.98	0.30	0.54
2008	plh0109	19464	1.45	0.70	0.39	0.54
2008	plh0110	19391	1.76	0.88	0.38	0.54
2010	plh0108	7786	2.28	0.92	0.24	0.47
2010	plh0109	7786	1.27	0.53	0.37	0.47
2010	plh0110	7783	1.34	0.56	0.34	0.47
2012	plh0108	27801	2.41	0.98	0.29	0.52
2012	plh0109	27663	1.45	0.70	0.38	0.52
2012	plh0110	27612	1.67	0.84	0.35	0.52
2016	plh0108	24490	2.38	0.98	0.28	0.51
2016	plh0109	24380	1.46	0.71	0.38	0.51
2016	plh0110	24331	1.65	0.84	0.34	0.51

Items

Are the following things currently ... for you (Sind für Sie persönlich die folgenden Dinge heute ...):

Altruism

1. Being there for others (Für andere da sein).
2. Spending a lot of time with friends (Viel mit Freunden zusammen sein).
3. Being politically and/or socially involved (Sich politisch, gesellschaftlich einsetzen).

Scale: 1 (Very important / Sehr wichtig) to 4 (Not at all important / Ganz unwichtig)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
1990	plh0105	7142	1.90	0.59	0.19	0.40
1990	plh0090	7136	2.13	0.71	0.27	0.40
1990	plh0111	7125	2.99	0.76	0.25	0.40
1992	plh0105	11010	1.88	0.60	0.27	0.46
1992	plh0090	10993	2.10	0.71	0.33	0.46
1992	plh0111	10953	3.16	0.73	0.26	0.46
1995	plh0105	11644	1.82	0.59	0.24	0.43
1995	plh0090	11619	2.09	0.70	0.32	0.43
1995	plh0111	11597	3.16	0.71	0.24	0.43
2004	plh0105	21926	1.83	0.58		
2004	plh0111	21830	2.93	0.75		
2008	plh0105	19626	1.83	0.57		
2008	plh0111	19518	3.10	0.75		
2010	plh0105	7788	1.51	0.55		
2010	plh0111	7771	2.69	0.79		
2012	plh0105	27910	1.71	0.57		

continues on next page

Table 11 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2012	plh0111	27853	2.81	0.77		
2016	plh0105	24604	1.69	0.57		
2016	plh0111	24520	2.79	0.78		

3.2.9 Life Satisfaction

Summary

General life satisfaction has been measured with an individual item in the SOEP on an annual basis since 1984. Domain-specific life satisfaction was initially measured with seven items (1984-1990), and since 2008 with ten items. Items 1 to 5 can be combined into one scale in all survey years (Schimmack, Krause, Wagner, & Schupp, 2009). In addition, satisfaction in 11 further domains is surveyed at irregular intervals.

Theoretical Background

Since its inception, the SOEP has included cognitive measures of well-being. The first measure is the global 11-point rating of life satisfaction (Schimmack, Schupp, & Wagner, 2008). This item is used almost exclusively as a measure of well-being in the SOEP. The reasons for its popularity are its high face validity and the widespread use of life satisfaction ratings in the well-being literature. Single-item measures of life satisfaction are a reasonably valid and common way to measure general life satisfaction: moderate associations with other well-being measures, including written interviews, informant reports, and measures of daily affect are reported by Sandvik, Diener, and Seidlitz (1993). Research using the World Value Survey (<http://www.worldvaluessurvey.org>) found single-item measures of life satisfaction to be positively related to affect balance and positive affect, and inversely related to negative affect (Suh et al., 1998). The second measure is the average of various domain satisfactions that are routinely assessed in the SOEP (health, household income, dwelling, and leisure time). This measure has two drawbacks. First, it does not weigh domains by their subjective importance. Second, the measure fails to capture aspects of well-being that are not covered by the domains included in the survey (Schimmack, 2008). A key advantage of this measure is that it relies not solely on respondents' ability to summarize and weigh all relevant aspects of their lives in response to a single question about satisfaction with life in general.

Scale Development

Information on the further development, reliability, and validity of the items can be found in Kroh (2006), in Schimmack (2008) and in Schimmack, Krause, Wagner, & Schupp (2009).

References

- Kroh, M. (2006). *An experimental evaluation of popular well-being measures*. Berlin: DIW Berlin.
- Sandvik, E., Diener, E., & Seidlitz, L. (1993). *Subjective well-being: The convergence and stability of self-report and non-self-report measures*. *Journal of Personality*, 61, 317–342.
- Schimmack, U., Schupp, J., & Wagner, G. G. (2008). *The influence of environment and personality on the affective and cognitive component of subjective well-being*. *Social Indicators Research*, 89, 41–60.
- Schimmack, U., Krause, P., Wagner, G. G., & Schupp, J. (2009). *Stability and change of Well Being: An experimentally enhanced Latent State-Trait-Error Analysis*. *Social Indicators Research*, 95, 19–31.
- Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). *The shifting basis of life satisfaction judgments across cultures: Emotions versus norms*. *Journal of Personality and Social Psychology*, 74, 482–493.

Items

1. How satisfied are you with your life, all things considered? (Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?) How satisfied are you with (Wie zufrieden sind Sie):
2. your health (mit ihrer Gesundheit)?

3. your household income (mit dem Einkommen Ihres Haushalts)?
4. your dwelling (mit Ihrer Wohnung)?
5. your free time (mit Ihrer Freizeit / in den Jahren 1995, 1996: mit Ihrer Freizeittätigkeit)?
6. your job (mit ihrer Arbeit)?
7. your housework (mit ihrer Tätigkeit im Haushalt)?
8. the child care available (mit den vorhandenen Möglichkeiten der Kinderbetreuung)?
9. your personal income (mit Ihrem persönlichen Einkommen)?
10. your family life (mit Ihrem Familienleben)?
11. your sleep (mit Ihrem Schlaf)?

Scale: 0 (Completely dissatisfied / Ganz und gar unzufrieden) to 10 (Completely satisfied / Ganz und gar zufrieden)

Test-Retest Correlations

In 2005, 2006 and 2009, item 1 to 5 were included in retests taken by subsamples within 30 to 49 days after the initial test. Pooled across the years (minimal N = 603), test-retest correlations were (in scale order, item 1 to 5) .66, .64, .71, .67, and .56. The test-retest correlation of the scale based on these five items was .77.

Items and Scale Statistics

year	variable	count	mean	sd
1984	plh0182	12192	7.43	2.14
1984	plh0171	12224	7.00	2.67
1984	plh0175	12042	6.41	2.62
1984	plh0177	12167	7.60	2.54
1984	plh0178	12172	7.29	2.55
1984	plh0173	7024	7.65	2.28
1984	plh0174	7843	6.95	2.44
1985	plh0182	11050	7.24	2.05
1985	plh0171	11035	6.93	2.47
1985	plh0175	10872	6.46	2.46
1985	plh0177	10979	7.59	2.38
1985	plh0178	11006	7.14	2.44
1985	plh0173	6338	7.52	2.14
1985	plh0174	6433	6.95	2.26
1986	plh0182	10611	7.29	1.94
1986	plh0171	10618	6.89	2.44
1986	plh0175	10507	6.50	2.35
1986	plh0177	10587	7.56	2.29
1986	plh0178	10595	7.01	2.44
1986	plh0173	6169	7.41	2.12
1986	plh0174	6388	6.86	2.20
1987	plh0182	10482	7.14	1.96
1987	plh0171	10498	6.85	2.39
1987	plh0175	10400	6.52	2.28
1987	plh0177	7943	7.89	2.08
1987	plh0178	10459	7.07	2.37
1987	plh0173	6155	7.40	2.07
1987	plh0174	6363	6.91	2.11
1988	plh0182	9980	7.08	1.96

continues on next page

Table 12 – continued from previous page

year	variable	count	mean	sd
1988	plh0171	10001	6.78	2.39
1988	plh0175	9905	6.51	2.29
1988	plh0177	9978	7.52	2.24
1988	plh0178	9978	6.92	2.37
1988	plh0173	5837	7.25	2.09
1988	plh0174	6113	6.80	2.12
1989	plh0182	9679	7.10	1.94
1989	plh0171	9683	6.71	2.42
1989	plh0175	9600	6.53	2.25
1989	plh0177	9653	7.48	2.24
1989	plh0178	9651	6.85	2.37
1989	plh0173	5692	7.24	2.03
1989	plh0174	6065	6.69	2.13
1990	plh0182	13904	7.05	1.88
1990	plh0171	13940	6.75	2.43
1990	plh0175	13813	6.27	2.32
1990	plh0177	13890	7.29	2.43
1990	plh0178	9482	7.09	2.15
1990	plh0173	9079	7.22	2.15
1990	plh0174	6393	6.69	2.13
1990	plh0179	1483	7.59	2.65
1991	plh0182	13535	6.95	1.90
1991	plh0171	13643	6.75	2.34
1991	plh0175	13531	6.19	2.40
1991	plh0177	13622	7.29	2.32
1991	plh0178	13605	6.59	2.49
1991	plh0173	8534	6.96	2.28
1992	plh0182	13307	6.92	1.82
1992	plh0171	13349	6.81	2.32
1992	plh0175	13257	6.12	2.26
1992	plh0177	13307	7.14	2.31
1992	plh0178	13314	6.64	2.38
1992	plh0173	8050	7.23	1.98
1993	plh0182	13113	6.88	1.88
1993	plh0171	13146	6.66	2.31
1993	plh0175	12831	6.15	2.30
1993	plh0177	13095	7.22	2.32
1993	plh0178	13102	6.58	2.42
1993	plh0173	7786	7.05	2.06
1993	plh0174	8623	6.51	2.12
1994	plh0182	13344	6.86	1.85
1994	plh0171	13383	6.60	2.31
1994	plh0175	13096	6.05	2.33
1994	plh0177	13333	7.30	2.27
1994	plh0178	13323	6.55	2.43
1994	plh0173	7803	7.00	2.07
1994	plh0174	8986	6.51	2.12
1995	plh0182	13696	6.89	1.83
1995	plh0171	13709	6.67	2.25
1995	plh0175	13449	6.12	2.29

continues on next page

Table 12 – continued from previous page

year	variable	count	mean	sd
1995	plh0177	13626	7.29	2.24
1995	plh0178	13678	6.80	2.22
1995	plh0173	8537	6.89	2.20
1995	plh0174	9429	6.52	2.11
1996	plh0182	13489	6.90	1.78
1996	plh0171	13497	6.62	2.24
1996	plh0175	13242	6.18	2.27
1996	plh0177	13426	7.36	2.19
1996	plh0178	13455	6.79	2.33
1996	plh0173	8367	6.88	2.16
1996	plh0174	9294	6.54	2.06
1997	plh0182	13257	6.79	1.79
1997	plh0171	13250	6.59	2.21
1997	plh0175	12980	6.01	2.24
1997	plh0177	13170	7.35	2.13
1997	plh0178	13202	6.73	2.33
1997	plh0173	8073	6.85	2.12
1997	plh0174	9201	6.50	2.05
1997	plh0179	3612	6.53	2.40
1998	plh0182	14631	6.95	1.78
1998	plh0171	14615	6.65	2.23
1998	plh0175	14300	6.11	2.28
1998	plh0177	14527	7.48	2.08
1998	plh0178	14575	6.83	2.30
1998	plh0173	8762	6.89	2.15
1998	plh0174	10202	6.56	2.06
1998	plh0179	3811	6.67	2.43
1999	plh0182	14054	6.97	1.78
1999	plh0171	14044	6.61	2.24
1999	plh0175	13767	6.18	2.25
1999	plh0177	13964	7.53	2.04
1999	plh0178	13993	6.86	2.29
1999	plh0173	8555	6.90	2.12
1999	plh0174	9996	6.52	2.05
1999	plh0179	3746	6.63	2.42
2000	plh0182	24510	7.09	1.78
2000	plh0171	24534	6.76	2.27
2000	plh0175	24069	6.40	2.30
2000	plh0177	24405	7.72	2.02
2000	plh0178	24467	7.01	2.29
2000	plh0173	13485	7.05	2.20
2000	plh0174	17626	6.60	2.11
2000	plh0179	6064	6.61	2.62
2001	plh0182	22299	7.10	1.74
2001	plh0171	22270	6.77	2.24
2001	plh0175	21971	6.49	2.24
2001	plh0177	22194	7.74	1.95
2001	plh0178	22244	7.02	2.25
2001	plh0173	13262	7.06	2.16
2001	plh0174	16297	6.62	2.08

continues on next page

Table 12 – continued from previous page

year	variable	count	mean	sd
2001	plh0179	4482	6.54	2.70
2002	plh0182	23844	7.05	1.74
2002	plh0171	23838	6.75	2.21
2002	plh0175	23443	6.50	2.22
2002	plh0177	23723	7.78	1.91
2002	plh0178	23770	7.00	2.20
2002	plh0173	14486	7.04	2.14
2002	plh0174	17140	6.57	2.06
2002	plh0179	4257	6.53	2.54
2003	plh0182	22568	6.96	1.78
2003	plh0171	22571	6.73	2.19
2003	plh0175	22229	6.34	2.29
2003	plh0177	22494	7.76	1.92
2003	plh0178	22522	7.02	2.19
2003	plh0173	13465	6.96	2.17
2003	plh0174	16753	6.59	2.02
2003	plh0179	9420	6.33	2.47
2004	plh0182	21964	6.80	1.82
2004	plh0171	21940	6.61	2.24
2004	plh0175	21522	6.22	2.30
2004	plh0177	21858	7.85	1.86
2004	plh0178	21873	6.98	2.18
2004	plh0173	12964	6.87	2.21
2004	plh0174	16460	6.56	1.99
2004	plh0179	3986	6.35	2.59
2004	plh0176	21197	5.65	2.60
2005	plh0182	21040	6.95	1.83
2005	plh0171	21041	6.64	2.24
2005	plh0175	20710	6.25	2.34
2005	plh0177	20968	7.82	1.89
2005	plh0178	20993	6.98	2.24
2005	plh0173	12329	6.85	2.22
2005	plh0174	15904	6.58	2.03
2005	plh0179	3800	6.30	2.74
2005	plh0176	20340	5.65	2.66
2006	plh0182	22284	6.91	1.80
2006	plh0171	22306	6.64	2.23
2006	plh0175	21975	6.23	2.33
2006	plh0177	22230	7.81	1.89
2006	plh0178	22254	6.98	2.25
2006	plh0173	12865	6.89	2.19
2006	plh0174	17048	6.62	2.02
2006	plh0179	3650	6.37	2.71
2006	plh0176	21678	5.66	2.62
2006	plh0180	21971	7.71	1.98
2007	plh0182	20834	6.95	1.78
2007	plh0171	20843	6.58	2.20
2007	plh0175	20560	6.26	2.27
2007	plh0177	20761	7.82	1.82
2007	plh0178	20793	6.96	2.21

continues on next page

Table 12 – continued from previous page

year	variable	count	mean	sd
2007	plh0173	12260	6.85	2.16
2007	plh0174	16068	6.64	1.98
2007	plh0179	3485	6.40	2.70
2007	plh0176	20286	5.63	2.59
2007	plh0180	20523	7.68	1.97
2008	plh0182	19643	6.98	1.75
2008	plh0171	19655	6.56	2.18
2008	plh0175	19361	6.29	2.28
2008	plh0177	19598	7.96	1.77
2008	plh0178	19625	7.08	2.13
2008	plh0173	11644	6.90	2.10
2008	plh0174	15114	6.68	1.94
2008	plh0179	2924	6.78	2.53
2008	plh0176	19131	5.70	2.56
2008	plh0180	19526	7.80	1.99
2008	plh0172	19652	6.82	2.26
2009	plh0182	20733	6.98	1.78
2009	plh0171	20760	6.57	2.23
2009	plh0175	20484	6.40	2.30
2009	plh0177	20701	7.91	1.82
2009	plh0178	20728	7.13	2.15
2009	plh0173	12323	6.88	2.21
2009	plh0174	16271	6.78	1.96
2009	plh0179	3121	6.78	2.76
2009	plh0176	20360	5.79	2.61
2009	plh0180	20650	7.84	1.99
2009	plh0172	20759	6.83	2.27
2010	plh0182	26674	7.25	1.74
2010	plh0171	26678	6.80	2.24
2010	plh0175	26387	6.35	2.36
2010	plh0177	26623	7.79	1.98
2010	plh0178	26643	6.93	2.25
2010	plh0173	16095	7.01	2.24
2010	plh0174	21913	6.82	2.00
2010	plh0179	7397	7.05	2.68
2010	plh0176	26119	5.74	2.69
2010	plh0180	26545	8.00	1.95
2010	plh0172	26676	6.83	2.31
2011	plh0182	26131	7.18	1.74
2011	plh0171	26127	6.73	2.24
2011	plh0175	25821	6.45	2.34
2011	plh0177	26071	7.79	1.95
2011	plh0178	18427	7.17	2.16
2011	plh0173	16459	7.07	2.17
2011	plh0174	21705	6.79	1.96
2011	plh0179	7313	7.41	2.42
2011	plh0176	25225	5.91	2.62
2011	plh0180	25888	8.01	1.90
2011	plh0172	26134	6.81	2.30
2012	plh0182	26733	7.19	1.74

continues on next page

Table 12 – continued from previous page

year	variable	count	mean	sd
2012	plh0171	26750	6.73	2.21
2012	plh0175	26431	6.54	2.30
2012	plh0177	26695	7.81	1.92
2012	plh0178	19510	7.22	2.14
2012	plh0173	16624	7.13	2.07
2012	plh0174	22362	6.83	1.94
2012	plh0179	6770	7.49	2.28
2012	plh0176	25855	6.01	2.58
2012	plh0180	26416	8.02	1.88
2012	plh0172	26748	6.80	2.29
2013	plh0182	30884	7.31	1.75
2013	plh0171	30918	6.83	2.25
2013	plh0175	30613	6.59	2.30
2013	plh0177	30878	7.77	2.00
2013	plh0178	18896	7.26	2.13
2013	plh0173	19386	7.17	2.10
2013	plh0174	21663	6.84	1.96
2013	plh0179	6147	7.58	2.25
2013	plh0176	25018	6.14	2.56
2013	plh0180	25525	8.00	1.88
2013	plh0172	25959	6.79	2.27
2014	plh0182	27367	7.25	1.73
2014	plh0171	27439	6.75	2.21
2014	plh0175	27043	6.66	2.26
2014	plh0177	27362	7.81	1.95
2014	plh0178	27382	7.05	2.21
2014	plh0173	17322	7.14	2.10
2014	plh0174	22853	6.89	1.96
2014	plh0179	6417	7.42	2.26
2014	plh0176	26475	6.11	2.56
2014	plh0180	27136	8.00	1.86
2014	plh0172	27442	6.79	2.30
2015	plh0182	27120	7.38	1.73
2015	plh0171	25354	6.75	2.19
2015	plh0175	25008	6.79	2.20
2015	plh0177	25278	7.84	1.92
2015	plh0178	25295	7.08	2.18
2015	plh0173	15832	7.18	2.02
2015	plh0174	21234	6.89	1.92
2015	plh0179	5317	7.52	2.20
2015	plh0176	26212	6.24	2.53
2015	plh0180	25098	7.97	1.88
2015	plh0172	25357	6.81	2.23
2016	plh0182	28941	7.36	1.81
2016	plh0171	28928	6.91	2.27
2016	plh0175	24156	6.85	2.19
2016	plh0177	24417	7.85	1.92
2016	plh0178	24424	7.09	2.16
2016	plh0173	15362	7.22	1.98
2016	plh0174	20826	6.91	1.90

continues on next page

Table 12 – continued from previous page

year	variable	count	mean	sd
2016	plh0179	4883	7.48	2.15
2016	plh0176	24314	6.33	2.50
2016	plh0180	24230	8.00	1.84
2016	plh0172	24480	6.76	2.26
2017	plh0182	32420	7.34	1.77
2017	plh0171	32322	6.91	2.29
2017	plh0175	26422	6.89	2.15
2017	plh0177	26679	7.84	1.92
2017	plh0178	26702	7.11	2.12
2017	plh0173	17096	7.19	1.98
2017	plh0174	23368	6.90	1.91
2017	plh0179	4503	7.40	2.23
2017	plh0176	28895	6.36	2.46
2017	plh0180	26464	7.96	1.86
2017	plh0172	26761	6.69	2.24
2018	plh0182	30241	7.40	1.75
2018	plh0171	30201	6.85	2.29
2018	plh0175	25582	6.85	2.17
2018	plh0177	25791	7.82	1.92
2018	plh0178	25796	7.11	2.14
2018	plh0173	16616	7.16	1.99
2018	plh0174	22729	6.87	1.91
2018	plh0179	4259	7.26	2.30
2018	plh0176	29128	6.24	2.48
2018	plh0180	25586	7.95	1.87
2018	plh0172	25857	6.63	2.26
2019	plh0182	29823	7.54	1.71
2019	plh0171	29885	6.92	2.23
2019	plh0175	25735	7.18	2.14
2019	plh0177	25949	8.00	1.87
2019	plh0178	25959	7.22	2.13
2019	plh0173	16873	7.38	1.92
2019	plh0174	22859	7.00	1.91
2019	plh0179	4063	7.46	2.25
2019	plh0176	28998	6.64	2.40
2019	plh0180	25681	8.02	1.85
2019	plh0172	26028	6.76	2.22
2020	plh0182	30397	7.59	1.67
2020	plh0171	30423	7.04	2.22
2020	plh0175	25697	7.11	2.13
2020	plh0177	25969	7.92	1.92
2020	plh0178	25989	7.11	2.17
2020	plh0173	17006	7.26	1.99
2020	plh0174	22942	7.01	1.93
2020	plh0179	4480	7.02	2.58
2020	plh0176	28962	6.59	2.40
2020	plh0180	25614	7.90	1.95
2020	plh0172	26071	6.73	2.23

Items

How satisfied are you with (Wie zufrieden sind Sie):

1. Your social security (soziale Sicherung)?
2. School education (Ausbildung)?
3. Duration of leisure (Freizeitumfang)?
4. Living standard (Lebensstandard)?
5. Your area (Wohngegend)?
6. The supply of goods and services in your area (Angebot von Waren und Dienstleistungen)?
7. The environmental conditions in your area (Umweltzustand)?
8. Local access to public transport (mit dem Angebot an Waren und Dienstleistungen hier am Ort)?
9. Democracy as it exists in Germany (Demokratie in Deutschland)?
10. Social life (Freundes-/Bekanntenkreis)?
11. Your volunteer work in clubs, associations, or other social service organizations (ehrenamtliche Tätigkeit)?
12. The level of social justice in Germany (mit der Verwirklichung der sozialen Gerechtigkeit in Deutschland)?

Scale: 0 (Completely dissatisfied / Ganz und gar unzufrieden) to 10 (Completely satisfied / Ganz und gar zufrieden)

Items and Scale Statistics

year	variable	count	mean	sd
1987	plh0155	10281	6.75	2.15
1989	plh0164	8647	6.75	2.37
1990	plh0153	9482	6.72	2.41
1990	plh0162	4435	6.36	2.13
1990	plh0156	4427	6.70	3.00
1990	plh0157	4430	3.07	2.44
1990	plh0160	4436	3.02	2.61
1991	plh0162	13624	6.80	2.04
1991	plh0156	13606	7.32	2.35
1991	plh0157	13625	6.64	2.56
1991	plh0160	13635	5.63	2.51
1992	plh0155	13144	6.29	2.04
1992	plh0162	13309	6.66	1.98
1992	plh0157	13319	6.98	2.36
1992	plh0160	13321	5.64	2.29
1993	plh0164	11155	6.74	2.30
1993	plh0162	13118	6.86	1.98
1993	plh0160	13131	5.51	2.22
1994	plh0156	13350	7.39	2.19
1994	plh0157	13351	6.72	2.47
1994	plh0160	13354	6.29	2.16
1995	plh0153	13674	6.58	2.49
1995	plh0162	13694	6.88	1.97
1995	plh0160	13692	5.95	2.12
1996	plh0162	13477	7.01	1.90
1996	plh0157	13450	6.66	2.48
1996	plh0160	13477	6.15	2.04
1997	plh0155	12945	5.22	2.31

continues on next page

Table 13 – continued from previous page

year	variable	count	mean	sd
1997	plh0162	13227	6.65	1.90
1997	plh0160	13219	6.09	1.98
1998	plh0162	14604	6.88	1.88
1998	plh0157	14551	6.54	2.47
1998	plh0160	14582	6.34	1.97
1998	plh0159	14475	6.32	2.57
1999	plh0162	14027	7.00	1.87
1999	plh0156	14015	7.46	1.92
1999	plh0157	13972	6.55	2.48
1999	plh0160	14020	6.87	1.88
2000	plh0164	22649	7.02	2.13
2000	plh0162	24504	7.19	1.84
2000	plh0157	24388	6.44	2.70
2000	plh0160	24451	6.64	1.93
2001	plh0162	22252	7.19	1.79
2001	plh0157	22186	6.43	2.64
2001	plh0160	22221	6.75	1.87
2002	plh0155	23378	5.63	2.19
2002	plh0162	23817	7.12	1.81
2002	plh0160	23794	6.44	1.87
2003	plh0162	22561	7.07	1.84
2003	plh0157	22468	6.40	2.57
2003	plh0160	22522	6.73	1.81
2003	plh0159	22356	6.38	2.56
2004	plh0164	20447	7.01	2.09
2004	plh0162	21896	7.05	1.84
2004	plh0157	21829	6.31	2.61
2005	plh0162	21023	7.07	1.89
2005	plh0157	20962	6.41	2.63
2005	plh0152_v2	20871	5.07	2.42
2006	plh0162	22285	7.19	1.89
2006	plh0181	22233	7.53	1.87
2007	plh0155	20597	5.27	2.19
2007	plh0167	5071	6.75	2.55
2008	plh0164	18154	7.28	2.03
2009	plh0167	4721	6.89	2.45
2010	plh0152_v2	26443	5.41	2.33
2011	plh0181	26028	7.75	1.87
2012	plh0155	19249	6.27	2.10
2012	plh0181	7161	8.05	1.75
2013	plh0162	24032	7.52	1.92
2013	plh0181	6822	8.05	1.74
2014	plh0164	25414	7.53	2.08
2015	plh0355	24881	5.05	2.23
2016	plh0152_v2	24066	5.80	2.42
2016	plh0181	24375	7.68	1.84
2017	plh0155	26192	6.80	1.99
2017	plh0156	26676	7.97	1.81
2018	plh0162	25788	7.64	1.82
2019	plh0164	24325	7.76	1.97

continues on next page

Table 13 – continued from previous page

year	variable	count	mean	sd
2020	plh0152_v2	25631	6.34	2.35

3.2.10 Locus of Control

Summary

The concept of locus of control describes the extent to which individuals believe they can determine events in their own lives, or conversely, the extent to which they feel dependent on factors that are outside of their control (Rotter, 1966). Locus of Control has been measured in SOEP in the years 1999, 2005, 2010, 2015, and 2020.

Theoretical Background

Locus of control may be internally or externally oriented. Individuals with an internal locus of control believe that their lives are the result of their own decisions and their own behavior. Individuals with an external locus of control believe that the things that happen to them in their lives are outside their personal control. For example, people with an external locus of control tend to avoid situations in which they feel overwhelmed. People with an internal locus of control tend to set higher goals for themselves, to persevere in difficult situations, and to pursue their goals successfully (Strauser, Ketz, & Keim, 2002).

Scale Development

The Locus of Control Scale was developed by Helmut Nolte (Nolte et al., 1997). In the original conception, it covered four dimensions: 1. Internal locus of control (3 Items), 2. External locus of control (E, 5 Items), 3. Attitudes about fairness or justice (1 Item) and 4. Individual vs. collective orientation (1 Item). Seven of the 10 items can be combined into an overall scale with good reliability (Specht, Egloff, & Schmukle, 2013).

References

- Nolte, H., Weischer, C., Wilkesmann, U., Maetzel, J., & Tegethoff, H. G. (1997). *Kontrolleinstellungen zum Leben und zur Zukunft. Auswertung eines neuen, sozialpsychologischen Itemblocks im Sozioökonomischen Panel. Ruhr-Universität Bochum: Diskussionspapiere aus der Fakultät für Sozialwissenschaft.*
- Rotter, J. B. (1996). *Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs: General and Applied*, 80, 1-28.
- Specht, J., Egloff, B., & Schmukle, S. C. (2013). *Everything under control? The effects of age, gender, and education on trajectories of perceived control in a nationally representative German sample. Developmental Psychology*, 49, 353-364.
- Strauser, D. R., Ketz, K., & Keim, J. (2002). *The relationship between self-efficacy, locus of control and work personality. The Journal of Rehabilitation*, 68, 20-26.

Items

The following statements apply to different attitudes towards life and the future. To what degree to you personally agree with the following statements (Die folgenden Aussagen kennzeichnen verschiedene Einstellungen zum Leben und zur Zukunft. In welchem Maße stimmen Sie persönlich den einzelnen Aussagen zu):

1. How my life goes depends on me (Wie mein Leben verläuft, hängt von mir selbst ab).
2. Compared to other people, I have not achieved what I deserve (Im Vergleich mit anderen habe ich nicht das erreicht, was ich verdient habe).
3. What a person achieves in life is above all a question of fate or luck (Was man im Leben erreicht, ist in erster Linie eine Frage von Schicksal oder Glück).
4. If a person is socially or politically active, he/she can have an effect on social conditions (Wenn man sich sozial oder politisch engagiert, kann man die sozialen Verhältnisse beeinflussen).

5. I frequently have the experience that other people have a controlling influence over my life (Ich mache häufig die Erfahrung, dass andere über mein Leben bestimmen).
6. One has to work hard in order to succeed (Erfolg muss man sich hart erarbeiten).
7. If I run up against difficulties in life, I often doubt my own abilities (Wenn ich im Leben auf Schwierigkeiten stoße, zweifle ich oft an meinen Fähigkeiten).
8. The opportunities that I have in life are determined by the social conditions (Welche Möglichkeiten ich im Leben habe, wird von den sozialen Umständen bestimmt).
9. Innate abilities are more important than any efforts one can make (Wichtiger als alle Anstrengungen sind die Fähigkeiten, die man mitbringt).
10. I have little control over the things that happen in my life (Ich habe wenig Kontrolle über die Dinge, die in meinem Leben passieren).

Scale: 1 (Not at all / Stimme überhaupt nicht zu) to 7 (Absolutely / Stimme voll zu); in 1999: 1 (Absolutely / Stimme voll zu) to 4 (Not at all / Stimme überhaupt nicht zu)

Test-Retest Correlations

In 2005, this scale was included in a retest taken by a subsample (N = 160) within 30 to 49 days after the initial test. Test-retest correlations for the scale-relevant items 1, 2, 3, 5, 7, 8, and 10 were .45, .47, .45, .33, .43, .26, .21; scale scores correlated .56.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
1999	plh0379_v1R	14026	3.26	0.71	0.31	0.69
1999	plh0378_v1	13926	2.85	0.86	0.41	0.69
1999	plh0380_v1	13976	2.67	0.85	0.38	0.69
1999	plh0386_v1	13979	2.99	0.83	0.47	0.69
1999	plh0383_v1	13975	2.94	0.79	0.43	0.69
1999	plh0384_v1	13969	2.28	0.74	0.29	0.69
1999	plh0381_v1	13971	3.14	0.77	0.52	0.69
2005	plh0379_v2	20992	5.44	1.38	0.28	0.69
2005	plh0378_v2R	20813	4.70	1.80	0.41	0.69
2005	plh0380_v2R	20918	4.33	1.70	0.38	0.69
2005	plh0386_v2R	20939	4.86	1.73	0.47	0.69
2005	plh0383_v2R	20932	4.66	1.69	0.40	0.69
2005	plh0384_v2R	20844	3.46	1.51	0.33	0.69
2005	plh0381_v2R	20929	5.24	1.57	0.54	0.69
2010	plh0379_v2	18845	5.44	1.32	0.30	0.70
2010	plh0378_v2R	18686	4.72	1.78	0.42	0.70
2010	plh0380_v2R	18786	4.45	1.66	0.37	0.70
2010	plh0386_v2R	18799	4.90	1.70	0.49	0.70
2010	plh0383_v2R	18785	4.83	1.64	0.42	0.70
2010	plh0384_v2R	18755	3.51	1.46	0.33	0.70
2010	plh0381_v2R	18799	5.29	1.51	0.55	0.70
2015	plh0379_v2	26984	5.59	1.30	0.26	0.68
2015	plh0378_v2R	26539	4.76	1.77	0.40	0.68
2015	plh0380_v2R	26895	4.33	1.70	0.35	0.68
2015	plh0386_v2R	26924	4.96	1.70	0.46	0.68
2015	plh0383_v2R	26903	4.72	1.67	0.41	0.68
2015	plh0384_v2R	26757	3.56	1.50	0.31	0.68
2015	plh0381_v2R	26936	5.28	1.53	0.52	0.68

continues on next page

Table 14 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2016	plh0379_v2	3920	5.60	1.73	0.05	0.50
2016	plh0378_v2R	3737	3.75	2.13	0.21	0.50
2016	plh0380_v2R	3981	3.38	2.12	0.26	0.50
2016	plh0386_v2R	4011	5.30	2.10	0.30	0.50
2016	plh0383_v2R	3899	4.50	2.13	0.33	0.50
2016	plh0384_v2R	3763	2.87	1.78	0.19	0.50
2016	plh0381_v2R	3924	4.20	2.09	0.31	0.50
2020	plh0379_v2	25993	5.63	1.22	0.26	0.68
2020	plh0378_v2R	25558	4.94	1.74	0.42	0.68
2020	plh0380_v2R	25913	4.51	1.62	0.36	0.68
2020	plh0386_v2R	25932	5.03	1.65	0.46	0.68
2020	plh0383_v2R	25886	4.72	1.64	0.41	0.68
2020	plh0384_v2R	25797	3.62	1.47	0.30	0.68
2020	plh0381_v2R	25934	5.29	1.49	0.53	0.68

Note

In 1999 a different scale format was used. Therefore, means and standard deviations cannot be compared without transformations. Although the questionnaire contains 10 items, only a subset of 7 items can be aggregated into a scale with acceptable internal consistency (cf. Specht et al., 2013). Scale is scored so that higher values indicate an internal locus of control.

3.2.11 Loneliness**Summary**

Loneliness has been measured in the SOEP since 2013 in four-year intervals using the 3-item short version of the UCLA loneliness scale developed by Hughes et al. (2004). In addition, loneliness has been assessed in both waves of the soep-cov survey.

Theoretical Background

Loneliness describes the perceived discrepancy between one's desired and one's actual relationships, in quantity and especially in quality (Peplau & Perlman 1982). Thus, loneliness is subjective and different from objective measures of social isolation. Loneliness increases the risk for morbidity and all-cause mortality (Hawkley & Cacioppo, 2010; Luo, Hawkley, Waite, & Cacioppo, 2012). In fact, according to a meta-analysis, the mortality risk of lonely people increases by 26 percent (Holt-Lunstad, Smith, Baker, Harris, & Stephenson, 2015). For these reasons, loneliness has been termed a major health risk by the world health organization and research on loneliness is of major interest to many disciplines in the social sciences.

Scale Development

The SOEP assesses loneliness using the 3-item short version of the UCLA loneliness scale (Russel, 1996) developed by Hughes et al (2004). The items were translated by into German by an bilingual expert. Details on the scale development and validation can be found in Hughes et al. (2004).

References

- Hawkley, L. C., & Cacioppo, J. T. (2010). *Loneliness matters: A theoretical and empirical review of consequences and mechanisms. Annals of Behavioral Medicine*, 40(2), 218-227.
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). *Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspectives on Psychological Science*, 10(2), 227-237.

Hughes, M. E., Waite, L. J., Hawkey, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on Aging*, 26(6), 655-672.

Luo, Y., Hawkey, L. C., Waite, L. J., & Cacioppo, J. T. (2012). Loneliness, health, and mortality in old age: A national longitudinal study. *Social Science & Medicine*, 74(6), 907-914.

Peplau, L. A., & Perlman, D. (1982). Perspectives on loneliness. In L. A. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 1-20). New York: John Wiley & Sons.

Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, 66(1), 20-40.

Items

How often do you feel... (Wie oft haben Sie das Gefühl, ...)

1. that you lack companionship? (... dass Ihnen die Gesellschaft anderer fehlt?)
2. left out? (... außen vor zu sein?)
3. isolated from others? (... dass Sie sozial isoliert sind?)

Answers: 1 = very often (Sehr oft), 2 = often (Oft), 3 = sometimes (Manchmal), 4 = seldom (Selten), 5 = never (Nie); all items need to be reversed before forming the composite score.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2013	plj0587	25762	3.62	0.94	0.56	0.78
2013	plj0588	25684	3.96	0.89	0.68	0.78
2013	plj0589	25738	4.37	0.89	0.61	0.78
2016	plj0587	4281	3.07	1.34	0.54	0.78
2016	plj0588	4218	3.56	1.27	0.68	0.78
2016	plj0589	4260	3.67	1.31	0.62	0.78
2017	plj0587	29718	3.61	1.00	0.56	0.79
2017	plj0588	29632	3.92	0.96	0.68	0.79
2017	plj0589	29665	4.30	0.96	0.64	0.79
2018	plj0587	424	3.16	1.35	0.60	0.81
2018	plj0588	424	3.66	1.30	0.73	0.81
2018	plj0589	426	3.85	1.27	0.66	0.81
2019	plj0587	277	2.96	1.27	0.44	0.71
2019	plj0588	276	4.00	1.12	0.60	0.71
2019	plj0589	277	4.03	1.13	0.57	0.71

3.2.12 Narcissistic Admiration and Rivalry Questionnaire Short Scale (NARQ-S)

Summary

The Narcissistic Admiration and Rivalry Questionnaire Short Scale (NARQ-S) was developed to assess the agentic (admiration) and antagonistic (rivalry) aspects of narcissism (Back et al., 2013). It is included in the SOEP in 2018 and 2023 at five-year intervals.

Theoretical Background

Narcissism is best understood as a multidimensional construct, incorporating agentic aspects such as dominance, charm, self-assuredness, and humor as well as antagonistic aspects such as selfishness, hostility, entitlement, and arrogance. The narcissistic admiration and rivalry concept (NARC) is a self-regulatory process model of grandiose narcissism that differentiates these two interrelated dimensions of narcissism: narcissistic admiration (agentic aspects driven by

self-enhancement) and narcissistic rivalry (antagonistic aspects driven by self-defense). The Narcissistic Admiration and Rivalry Questionnaire (NARQ) assesses the agentic (admiration) and antagonistic (rivalry) aspects of narcissism according to the NARC and was developed and validated by Back et al. (2013). The NARQ-S consists of six items, three for each dimension, that can be answered on a six-point Likert-type scale ranging from 1 (not at all) to 6 (agree completely). Each dimension has three subscales that are measured by one item each and contain content addressing narcissists' affective-motivational, cognitive, and behavioral processes. For admiration, these subscales are grandiosity, strive for uniqueness, and charmingness. The rivalry dimension consists of the subscales devaluation, strive for supremacy, and aggressiveness. A total score of the NARQ-S can be calculated in cases where researchers are interested in a global assessment of grandiose narcissism. However, calculating and using the admiration and rivalry dimensions of NARQ-S is recommended because they provide a nuanced insight into the (at times paradoxical) effects of narcissism.

Scale Development

For each dimension of the NARQ-S, item inclusion was based on the highest factor loading of the respective subscale of the NARQ (Back et al., 2013). A validation study by Leckelt et al. (2018) used data from a large convenience sample (total N = 11,937) as well as data from a large representative sample (total N = 4,433) that included responses to other narcissism measures as well as related constructs, including the other Dark Triad traits, Big Five personality traits, and self-esteem. The study sought to validate the factor structure, provide representative descriptive data and reliability estimates, assess the reliability across the trait spectrum, and examine the nomological network of the NARQ-S. Results suggest that the NARQ-S shows a robust factor structure and is a reliable and valid short measure of the agentic and antagonistic aspects of grandiose narcissism. The SOEP has included the NARQ-S with both subscales in its original form. It has been included in 2018 and 2023 at five-year intervals.

References

- Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). *Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism*. *Journal of Personality and Social Psychology*, 105, 1013–1037.
- Leckelt, M., Wetzel, E., Gerlach, T. M., Ackerman, R.A., Miller, J.D., Chopik, W.J., Penke, L., Geukes, K., Küfner, A. C. P., Hutteman, R., Richter, D., Renner, K.-H., Allroggen, M., Brecheen, C., Campbell, W. K., Grossmann, I. & Back, M. D. (2018). *Validation of the Narcissistic Admiration and Rivalry Questionnaire Short Scale (NARQ-S) in convenience and representative samples*. *Psychological Assessment*, 30, 86–96.

Items

Please indicate how much the following statements apply to you using a response format ranging from “1 = not agree at all” to “6 = agree completely”. (Wie sehr treffen die folgenden Aussagen auf Sie zu? Antworten Sie bitte anhand der folgenden Skala. Der Wert 1 bedeutet: trifft überhaupt nicht zu, der Wert 6 bedeutet: trifft vollkommen zu. Mit den Werten dazwischen können Sie Ihre Einschätzung abstimmen.)

1. Being a very special person gives me a lot of strength / Ich ziehe viel Kraft daraus, eine ganz besondere Person zu sein (ADM-1)
2. I manage to be the center of attention with my outstanding contributions / Mit meinen besonderen Beiträgen schaffe ich es, im Mittelpunkt zu stehen (ADM-2)
3. I react annoyed if another person steals the show from me / Ich reagiere genervt, wenn eine andere Person mir die Schau stiehlt (RIV-1)
4. I deserve to be seen as a great personality / Ich habe es verdient, als große Persönlichkeit angesehen zu werden (ADM-3)
5. I want my rivals to fail / Ich will, dass meine Konkurrenten scheitern (RIV-2)
6. Most people are somehow losers / Die meisten Menschen sind ziemlich Versager (RIV-3)

Scale: 1 (not agree at all / trifft überhaupt nicht zu) to 6 (agree completely / trifft vollkommen zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2018	plh0361	25187	2.60	1.50	0.57	0.78
2018	plh0362	25192	2.40	1.35	0.59	0.78
2018	plh0363	25283	1.76	1.10	0.52	0.78
2018	plh0364	25128	2.07	1.30	0.62	0.78
2018	plh0365	25169	1.61	1.08	0.45	0.78
2018	plh0366	25124	1.86	1.11	0.39	0.78

3.2.13 Optimism/Pessimism – Attitudes toward the Future

Summary

The SOEP measured respondents' attitudes toward the future with an individual item in 1999, 2005, 2009, 2014, and 2019.

Theoretical Background

For pragmatic reasons and to save time, surveys usually measure attitudes about the future and future orientations by asking respondents whether they see the future positively or negatively, that is, optimistically or pessimistically. In the analysis of attitudes toward the future, questions on the nature and emergence of future orientations are of interest (e.g., whether attitudes about the future are less positive in adults than in young people, or whether young women are less optimistic about the future than young men; Trommsdorff, 1994). Furthermore, making assumptions about the future makes it possible to plan and anticipate one's future actions. Attitudes about the future should therefore provide the basis for decisions and planning behavior.

Scale Development

The question about attitudes toward the future was included in the SOEP on the recommendation of Gisela Trommsdorff. Additional information on the validity of this individual item can be found in Trommsdorff (1994).

References

Trommsdorff, G. (1994). *Zukunft als Teil individueller Handlungsorientierungen*. In E. Holst, J. P. Rinderspacher & J. Schupp (Hrsg.), *Erwartungen an die Zukunft. Zeithorizonte und Wertewandel in der sozialwissenschaftlichen Diskussion* (pp. 45-76). Frankfurt a. M.: Campus.

Items

When you think about the future (Wenn Sie an die Zukunft denken):

1. Are you ... (Sind Sie da ...)

Scale: 1 (optimistic / optimistisch) to 4 (pessimistic / pessimistisch)

Test-Retest Correlations

In 2009, this item was included in a retest taken by a subsample (N = 174) within 30 to 49 days after the initial test. Test-retest correlation was .60.

Items and Scale Statistics

year	variable	count	mean	sd
1999	plh0244	13989	1.99	0.80
2005	plh0244	20981	2.17	0.81
2009	plh0244	20699	2.23	0.81
2014	plh0244	27275	1.89	0.75

continues on next page

Table 17 – continued from previous page

year	variable	count	mean	sd
2019	plh0244	25816	1.94	0.75

3.2.14 Parenting Goals

Summary

The questions on parenting goals were taken from Kohn (1977). The 18 items can be combined into two scales: “autonomy” and “conformity.” Parenting goals have been measured in the SOEP since 2010 in the questionnaire for parents of children between the ages of 7 and 8.

Theoretical Background

Kohn (1977) found an association between parenting goals (conformity and autonomy) and the parents’ social class: Working-class parents emphasized discipline, manners, cleanliness, good behavior in school, honesty, and obedience, whereas middle-class parents emphasized consideration, interest in why and how things happen, responsibility, and self-control. This suggests that lower-class parents place more value on conformity and higher-class parents place more value on personal autonomy. The different values of parents from different social classes regarding childrearing are transferred to their children through their own behavior and shape their children’s social character.

Scale Development

The scale measuring parenting goals was taken from the German General Social Survey (ALLBUS). Further information on the scale can be found in Terwey (2000).

References

- Kohn, M. (1977). *Class and conformity: A study in values*. Chicago: University of Chicago Press.
- Terwey, M. (2000). ALLBUS: A German General Social Survey. *Schmollers Jahrbuch*, 120, 151-158.

Items

In the following, we will list a few traits and abilities that parents can foster in their children through their approach to parenting. How important do you consider the following parenting goals? That the child... (Im Folgenden werden einige Eigenschaften und Fähigkeiten genannt, die man durch Erziehung fördern kann. Für wie wichtig halten Sie persönlich die folgenden Erziehungsziele? Dass das Kind...)

Autonomy / Selbstständigkeit

1. Is interested in how and why things happen (sich dafür interessiert, wie und warum bestimmte Dinge passieren)
2. Is honest (ehrlich ist)
3. Is responsible (verantwortungsbewusst ist)
4. Has good judgment (ein gutes Urteilsvermögen besitzt)
5. Strives to achieve his/her goals (sich bemüht, seine Ziele zu erreichen)
6. Learns to overcome obstacles in life (lernt, sich im Leben auch gegen Widerstände durchzusetzen).

Scale: 1 (Not at all important / Überhaupt nicht wichtig) to 5 (Very important / Sehr wichtig)

Obedience / Conformity (Gehorsam / Konformität)

1. Is good in school (ein guter Schüler wird)
2. Gets along with other children (sich gut mit anderen Kindern versteht)
3. Behaves like normal girl/boy (sich wie ein normales Mädchen bzw. wie ein normaler Junge verhält)

4. Has good manners (gute Umgangsformen hat)
 5. Has good self-control (Selbstbeherrschung besitzt)
 6. Is considerate of others (auf andere Rücksicht nimmt)
 7. Obeys his/her parents (seinen Eltern gehorcht)
 8. Is neat and clean (ordentlich und sauber ist)
 9. Fits in well in groups (sich gut in Gruppen einfügen kann)
 10. Is satisfied with his/her own abilities (zufrieden mit dem ist, was es hat und kann)
 11. Learns to avoid risks in life (lernt, Risiken im Leben zu meiden)
 12. Is liked by others, friendly (von anderen gemocht wird, liebenswert ist)
- Scale: 1 (not at all important / überhaupt nicht wichtig) to 5 (very important / sehr wichtig)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2010-2020	edgoal3	14854	4.50	0.60	0.44	0.77
2010-2020	edgoal5	14879	4.82	0.42	0.43	0.77
2010-2020	edgoal8	14879	4.63	0.53	0.59	0.77
2010-2020	edgoal11	14842	4.44	0.61	0.57	0.77
2010-2020	edgoal13	14878	4.54	0.56	0.55	0.77
2010-2020	edgoal15	14855	4.53	0.59	0.53	0.77
2010-2020	edgoal1	14896	4.37	0.67	0.52	0.86
2010-2020	edgoal2	14903	4.56	0.55	0.45	0.86
2010-2020	edgoal4	14814	4.10	0.96	0.56	0.86
2010-2020	edgoal6	14890	4.57	0.57	0.57	0.86
2010-2020	edgoal7	14868	4.36	0.65	0.61	0.86
2010-2020	edgoal9	14869	4.58	0.56	0.48	0.86
2010-2020	edgoal10	14862	4.29	0.70	0.60	0.86
2010-2020	edgoal12	14878	4.24	0.70	0.64	0.86
2010-2020	edgoal14	14861	4.27	0.65	0.60	0.86
2010-2020	edgoal16	14854	4.39	0.72	0.43	0.86
2010-2020	edgoal17	14825	3.72	1.02	0.59	0.86
2010-2020	edgoal18	14859	3.97	0.80	0.60	0.86

3.2.15 Parenting Role

Summary

The questions on the parenting role were taken from the Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) project's parenting questionnaire (Wendt et. al., 2011). The ten items can be combined into three scales: "autonomy," "hostile attributions," and "willingness to make sacrifices." The questions on the parenting role have been asked in the SOEP since 2010 in the questionnaire for parents of children aged 7-8 years.

Theoretical Background

The four items assessing "autonomy in the parenting role" are based on an instrument developed by Skinner and Regan (1992). The negative items measure the parents' feelings of irksome dependence in their interaction with the child. Furthermore, it is assumed that the feeling of autonomy in the parenting role is also expressed as positive feelings towards the child. The three items of the scale "hostile attributions" are newly developed by pairfam and measure the parental disposition to interpret the child's behavior as intentionally hostile or egoistic. The scale "willingness to

make sacrifices” is an adapted version from the AGAPE scale developed by Bierhoff, Grau, & Ludwig (1993) to assess parents’ willingness to make sacrifices in their relationship with their child.

Scale Development

Information on the scale development can be found in the documentation provided by pairfam (Wendt et. al., 2011).

References

Bierhoff, H. W., Grau, I., & Ludwig, A. (1993). *Marburger Einstellungsinventar für Liebesstile (MEIL)*. Göttingen: Hogrefe.

Skinner, E. A., & Regan, C. (1992). *Parenting sense of autonomy. Technical Report*. University of Rochester, Rochester, NY.

Wendt, E.-V., Schmahl, F., Thönnissen, C., Schaer, M., & Walper, S. (2011). *Scales Manual of the German Family Panel. pairfam (Panel Analysis of Intimate Relationships and Family Dynamics)*. Bremen, Chemnitz, Munich.

Items

Autonomy

1. I have the feeling that taking care of my child/my children takes up all my strength and that my whole life revolves around it (R) (Ich habe das Gefühl, dass Betreuung und Erziehung meines Kindes mich völlig in Beschlag nehmen, mein ganzes Leben bestimmen).
2. I wish I didn’t feel so trapped by my parental duties (R) (Ich wünschte, ich würde mich durch meine Elternpflichten nicht so gefangen fühlen).
3. When I am with my child/children there is nothing else I’d rather be doing (Wenn ich mit meinem Kind zusammen bin, gibt es nichts anderes, was ich lieber täte).
4. I look forward to spending time with my child/children (Ich freue mich darauf, mit meinem Kind zusammen zu sein).

Scale: 1 (Not at all / stimme überhaupt nicht zu) to 5 (Absolutely / stimme voll und ganz zu)

Hostile Attributions

1. When my child disobeys and breaks rules, he/she just wants to annoy me (Wenn mein Kind nicht gehorcht und etwas Verbotenes tut, will es mich ärgern).
2. If there are any problems with the way I raise my child, then it’s my child’s fault (Wenn es Probleme in der Erziehung gibt, liegt das an meinem Kind).
3. It seems to me that when my child misbehaves, he/she does it intentionally (Ich denke, wenn mein Kind sich falsch verhält, macht es das mit Absicht).

Scale: 1 (Not at all / stimme überhaupt nicht zu) to 5 (Absolutely / stimme voll und ganz zu)

Readiness to Make Sacrifices

1. I am usually willing to sacrifice my own desires to satisfy those of my child (Gewöhnlich bin ich bereit, meine eigenen Wünsche denen meines Kindes zu opfern).
2. I would put up with anything for the good of my child (Ich würde alles aushalten für das Wohl meines Kindes).
3. I often stop what I am doing to offer help to my child (Ich lasse oft alles stehen und liegen, um mein Kind zu unterstützen).

Scale: 1 (Not at all / Stimme überhaupt nicht zu) to 5 (Absolutely / Stimme voll und ganz zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2010-2020	bepar3R	14814	3.57	1.14	0.37	0.52
2010-2020	bepar4R	14778	3.89	1.05	0.49	0.52
2010-2020	bepar6	14834	3.88	0.99	0.16	0.52
2010-2020	bepar10	14868	4.63	0.59	0.28	0.52
2010-2020	bepar2	14829	2.03	1.06	0.47	0.64
2010-2020	bepar5	14809	1.59	0.81	0.38	0.64
2010-2020	bepar8	14825	1.73	0.87	0.53	0.64
2010-2020	bepar1	14849	3.98	0.88	0.45	0.65
2010-2020	bepar7	14828	4.32	0.85	0.47	0.65
2010-2020	bepar9	14856	3.72	0.97	0.46	0.65

3.2.16 Parenting Style

Summary

The questions on the parenting style were taken from the Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) project's parenting questionnaire (Wendt et. al., 2011). The 18 items can be combined into six scales: "emotional warmth," "inconsistent parenting," "monitoring," "negative communication," "psychological control," and "strict control." The questions on parenting style have been asked in the SOEP since 2010 in the questionnaire for parents of children aged 7-8 years.

Theoretical Background

The scale "emotional warmth" comprises three items indicating the degree of affirmative attention and care in parenting. The items are based on mothers' and fathers' actual parenting behavior (see the corresponding scale of Jaursch, 2003, based on Perris et al., 1980, and Schuhmacher, Eisemann, & Brähler, 1999). The scale "inconsistent parenting" comprises three items indicating the degree of inconsistent behavior in parenting. The items are based on a questionnaire on parenting from Reichle and Franiek (2005). The scale "monitoring" comprises three items indicating the degree to which parents are informed about their child's activities and social contacts. The items are based on a questionnaire on parenting from Reichle and Franiek (2005). The scale "negative communication" comprises three items indicating the degree to which parents behave negatively toward their child. The items are based on the instrument developed by Schwarz, Walper, Gödde, and Jurasic (1997). The scale "psychological control" consists of three items. The items assess negative intrusive thoughts, feelings, and behavior of parents toward their child and are a shortened and adapted version of the scale "psychological pressure" from the "Zurich Brief Questionnaire for the Assessment of Parental Behaviors" (Reitzle et al. 2001). The scale "strict control" comprises three items indicating harsh control and authoritarian behavior of parents. The items are based on the instrument of Schwarz et al. (1997).

Scale Development

Information on the scale development can be found in the documentation of pairfam (Wendt et. al., 2011).

References

- Perris, C., Jacobsson, L., Lindström, H., von Knorring, L., & Perris, H. (1980). Development of a new inventory for assessing memories of parental rearing behaviour. *Acta Psychiatrica Scandinavica*, 61, 265-274.
- Jaursch, Stefanie, 2003. *Erinnertes und aktuelles Erziehungsverhalten von Müttern und Vätern: Intergenerationale Zusammenhänge und kontextuelle Faktoren. Dissertation. Friedrich- Alexander-Universität Erlangen-Nürnberg.*
- Reichle, B., & Franiek, S. (2005). *Erziehungsstil aus Elternsicht - Erweiterte deutsche Version des Alabama Parenting Questionnaire (EDAPQ) (Expanded German Version of the Alabama Parenting Questionnaire). Hochschule Ludwigsburg: Institut für Pädagogische Psychologie und Soziologie.*
- Reitzle, M., Winkler Metzke, C., & Steinhausen, H.-C. (2001). *Eltern und Kinder: Der Zürcher Kurzfragebogen zum Erziehungsverhalten (Zurich Brief Questionnaire for the Assessment of Parental Behaviors). Diagnostica*, 47, 196-207.

Schumacher, J., Eisemann, M., & Brähler, E. (1999). *Rückblick auf die Eltern: Der Fragebogen zum erinnerten elterlichen Erziehungsverhalten (FEE)*. *Diagnostika*, 45, 194-204.

Schwarz, B., Walper, S., Gödde, M., & Jurasic, S. (1997). *Dokumentation der Erhebungsinstrumente der 1. Haupterhebung (überarb. Version)*. *Berichte aus der Arbeitsgruppe "Familienentwicklung nach der Trennung,"* 14.

Wendt, E.-V., Schmahl, F., Thönnissen, C., Schaer, M., & Walper, S.). *Scales Manual of the German Family Panel. pairfam (Panel Analysis of Intimate Relationships and Family Dynamics)*. Bremen, Chemnitz, Munich.

Items

Emotional Warmth

1. I show my child with words and gestures that I care about him/her (Ich zeige meinem Kind mit Worten und Gesten, dass ich es gerne habe).
2. I console child up when he/she is sad (Ich tröste mein Kind, wenn es traurig ist).
3. I praise my child (Ich lobe mein Kind).

Scale: 1 (Never / nie) to 5 (Frequently / sehr häufig)

Inconsistent Parenting

1. I reduce punishments or end them early (Ich schwäche eine Bestrafung ab oder hebe sie vorzeitig auf).
2. I threaten my child with a punishment but don't actually follow through (Ich drohe meinem Kind eine Strafe an, bestrafe es aber dann doch nicht).
3. I find it hard to set and keep consistent rules for my child (Es fällt mir schwer in meiner Erziehung konsequent zu sein).

Scale: 1 (Never / nie) to 5 (Frequently / sehr häufig)

Monitoring

1. I try to actively influence my child's circle of friends (Ich versuche den Freundeskreis meines Kindes aktiv zu beeinflussen).
2. When my child goes out, I ask what he/she did and experienced (Wenn mein Kind unterwegs war, frage ich nach, was es getan und erlebt hat).
3. When my child goes out, I know exactly where he/she is (Wenn mein Kind außer Haus ist, weiß ich genau, wo es sich aufhält).

Scale: 1 (Never / nie) to 5 (Frequently / sehr häufig)

Negative Communication

1. I criticize my child (Ich kritisiere mein Kind).
2. I yell at my child when he/she does something wrong (Ich schreie mein Kind an, wenn es etwas falsch gemacht hat).
3. I scold my child when I am angry at him/her (Ich beschimpfe mein Kind, weil ich wütend auf es bin).

Scale: 1 (Never / nie) to 5 (Frequently / sehr häufig)

Psychological Control

1. I am disappointed and sad when my child misbehaves (Ich bin enttäuscht und traurig, wenn sich mein Kind schlecht benommen hat).
2. I think my child is ungrateful when he/she does not obey me (Ich halte mein Kind für undankbar, wenn es mir nicht gehorcht).

3. I don't talk to my child for a while when he/she does something wrong (Ich rede eine Zeit lang nicht mit meinem Kind, wenn es etwas angestellt hat).

Scale: 1 (Never / nie) to 5 (Frequently / sehr häufig)

Strict Control

1. I tend to be a strict parent (Ich bin eher streng zu meinem Kind).
2. If my child does something against my will, I punish him/her (Wenn mein Kind etwas gegen meinen Willen tut, bestrafe ich es).
3. I make it clear to my child that he/she is not to break my rules or question my decisions (Ich gebe meinem Kind zu verstehen, dass es sich meinen Anordnungen und Entscheidungen nicht widersetzen soll).

Scale: 1 (Never / nie) to 5 (Frequently / sehr häufig)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2010-2020	edbeh1	14673	4.47	0.64	0.55	0.73
2010-2020	edbeh8	14652	4.50	0.68	0.54	0.73
2010-2020	edbeh13	14634	4.36	0.66	0.56	0.73
2010-2020	edbeh5	14635	2.47	1.00	0.55	0.69
2010-2020	edbeh16	14518	2.66	0.98	0.50	0.69
2010-2020	edbeh18	14617	2.47	0.95	0.47	0.69
2010-2020	edbeh3	14653	4.34	0.71	0.29	0.36
2010-2020	edbeh6	14636	4.45	0.78	0.27	0.36
2010-2020	edbeh15	14604	2.18	1.05	0.11	0.36
2010-2020	edbeh2	14649	3.08	0.73	0.35	0.62
2010-2020	edbeh9	14654	2.26	0.83	0.51	0.62
2010-2020	edbeh14	14627	1.85	0.84	0.44	0.62
2010-2020	edbeh10	14600	1.54	0.76	0.38	0.51
2010-2020	edbeh11	14636	1.48	0.80	0.31	0.51
2010-2020	edbeh17	14610	2.69	0.91	0.29	0.51
2010-2020	edbeh4	14645	2.63	0.81	0.42	0.55
2010-2020	edbeh7	14627	2.95	0.90	0.33	0.55
2010-2020	edbeh12	14574	2.95	1.00	0.33	0.55

3.2.17 Patient Health Questionnaire – 4 (PHQ-4)

Summary

The Patient Health Questionnaire-4 (PHQ-4) is an ultra-brief screening instrument to assess the two leading symptoms of a depression and anxiety (Kroenke et al., 2009). It is included in the SOEP in its original form in 2016 and since 2019 at two-year intervals.

Theoretical Background

The Patient Health Questionnaire-4 (PHQ-4) is an ultra-brief screening instrument to assess the two leading symptoms of a depression and anxiety. It was developed and validated by Kroenke et al. (2009) to account for the fact that those two mental disorders are the most prevalent mental disorders in the general population, both are often comorbid, and patients that are affected by either or both of the disorders often lack the concentration to fill in long and detailed questionnaires. The PHQ-4 consists of four items that can be answered on a four-point Likert-type scale (0 = not at all, 1 = several days, 2 = more than half days, 3 = nearly every day). Two items measure the core symptoms of depression and form the PHQ-2 subscale of the PHQ-4. Two items measure the core symptoms of anxiety and form the

GAD-2 subscale of the PHQ-4. Both subscales have independently been shown to be good brief screening tools. The total PHQ-4 score complements the subscale scores as an overall measure of symptom burden, as well as functional impairment and disability. An elevated PHQ-4 score is not diagnostic, but is, instead, an indicator for further inquiry to establish the presence or absence of a clinical disorder warranting treatment.

Scale Development

The two-item PHQ-2 measures anhedonia (“little interest or pleasure in doing things”) and depressed mood (“Feeling down, depressed or hopeless”) over the past two weeks. The final sum score of both item ranges from 0 to 6. Scores of 3 and above indicate a depressive disorder with a sensitivity of 79 % and a specificity of 86 %. The two-item GAD-2 measures general irritability (“feeling nervous, anxious, or on edge”) and worries (“not being able to stop or control worrying”) over the past two weeks. The GAD-2 final sum scores also range from 0 to 6. A cutoff of 3 has reasonable sensitivity for generalized anxiety disorder (88%), panic disorder (76%), and social anxiety disorder (70%), and moderate sensitivity for posttraumatic stress disorder (59%), and good specificity (81%– 83%) for all four disorders. The SOEP has included the PHQ-4 with both subscales in its original form. It has been included since 2016 at two- or three-year intervals. So far it has been assessed in 2016, 2019, and 2021. In addition, it has been assessed in both soep-cov survey waves.

References

Arroll B, Goodyear-smith F, Crengle S, et al. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine*, 8, 4, 348-53.

Kroenke K, Spitzer RL, Williams JB. (2003). The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care*, 1284-1292.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Löwe, B. (2009) An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics*, 50, 613-21.

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., ... & Brähler, E. (2010). A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of affective disorders*, 122(1-2), 86-95.

Items

Over the last two weeks, how often have you been bothered by any if the following problems? (Jetzt geht es um die letzten zwei Wochen. Wie oft fühlten Sie sich da durch die folgenden Beschwerden beeinträchtigt?)

1. Little interest or pleasure in doing things / Wenig Interesse oder Freude an Ihren Tätigkeiten (PAQ-2)
2. Feeling down, depressed, or hopeless / Niedergeschlagenheit, Schwermut oder Hoffnungslosigkeit (PAQ-2)
3. Feeling nervous, anxious, or on edge / Nervosität, Ängstlichkeit oder Anspannung. (GAD-2)
4. Unable to stop or control worrying / Nicht in der Lage sein, Sorgen zu stoppen oder zu kontrollieren. (GAD-2)

Answers: 1 = not at all (überhaupt nicht), 2 = several days (an einzelnen Tagen), 3 = more than half days (an mehr als der Hälfte der Tage), 4 = nearly every day ((fast) jeden Tag); Items need to be recoded to 0 to 3 instead of 1 to 4.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2016	plh0339	28550	1.71	0.74	0.51	0.80
2016	plh0340	28678	1.53	0.74	0.70	0.80
2016	plh0341	28720	1.66	0.75	0.63	0.80
2016	plh0342	28585	1.45	0.73	0.64	0.80
2019	plh0339	29339	1.61	0.76	0.56	0.82
2019	plh0340	29450	1.43	0.71	0.72	0.82
2019	plh0341	29462	1.51	0.72	0.65	0.82

continues on next page

Table 21 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2019	plh0342	29412	1.32	0.65	0.65	0.82

3.2.18 Personality – Big Five

Big Five (Adults)

Summary

The Big Five concept makes it possible to describe human personality traits in terms of individually differing behaviors and experiences. At the heart of this approach lies the assumption that there exist five broad, non-overlapping dimensions of personality that can be used to describe human personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The scale consists of 15 items measuring the “Big Five” and was used in the SOEP in 2005, 2009, 2013, and 2017. From 2009 on the scale was extended by one additional item measuring openness. In 2019 the concept was used ahead of its regular interval because both, sample P (“Top Shareholder Sample”) and sample Q (Sample P “LGB*”) started in that year. The next replication is in 2023.

Theoretical Background

The Big Five Model of personality traits, or the “Five-Factor Model,” is a psychological approach aimed at comprehensively describing the human personality. According to this model, intra-individual personality differences can be measured in five broad dimensions – the so-called “Big Five.” Personality traits are coherent and constant structures of feelings, thoughts, and forms of behavior that influence psychological conditions and individual actions (Costa & McCrae, 1985). The Five-Factor Model assumes that values in these different dimensions vary from one individual to the next, but that the basic structure of the personality, at least for individuals from Western cultures, remains constant. The five personality dimensions are: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Individuals with high openness to experience describe themselves as having a lively imagination, being original and creative, and appreciating artistic and aesthetic experiences. Individuals with high conscientiousness describe themselves as highly motivated, success-oriented, thorough in their work, and as completing their tasks effectively and efficiently. Individuals with high extroversion see themselves as communicative, talkative, open, and outgoing. In contrast, introverted individuals describe themselves as reserved or shy. Individuals with high values for neuroticism describe themselves as nervous, often worried, and as having difficulties dealing with stress. The antithesis of neuroticism is emotional stability. Finally, individuals with high agreeableness describe themselves as considerate and friendly, willing to forgive, and as not being unkind to others. The Big Five approach has its origins in two distinct traditions in psychology: first, in the psycho-lexical tradition of Allport (1937) and Cattell (1946), and second, in the differential diagnosis or clinical tradition of personality research pioneered by Stern (1911) and Eysenck (1947). Both lines of research were ultimately brought together into a single model and formalized by Costa and McCrae (1985). Since then, the validity of this approach has been confirmed in a number of studies of different groups of individuals from different countries including the USA, England, Germany, Spain, Sweden, and China.

Scale Development

For reasons of space and time, the SOEP required a short and efficient scale that is nevertheless capable of reflecting the basic structure of the Five-Factor Model in a robust and reliable way. The choice of items was made based on five criteria: (1) the conditions and limitations that result from the SOEP questionnaire, (2) the aim of achieving balanced coverage of the basic personality dimensions, (3) the internal consistency of the scale, (4) the dimensionality of the items, and (5) the correspondence with other, established measurement instruments. The guideline in developing the short scale was to use at least three items per latent construct. The development and validation of the scale took place in the framework of a SOEP pretest, which represents an independent representative survey of the resident population of Germany. The items were taken from the Ten-Item Personality Inventory (TIPI; Gosling et al., 2003) and the BFI-25. The BFI-25 was developed based on a principal component analysis carried out with the entire inventory by John et al. (1991): the five items for each personality dimension with the highest factor loadings were included. Further information can be found in Gerlitz und Schupp (2005) and in Lang et al. (2011).

References

- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Henry Holt.
- Cattell, R. B. (1946). *Description and measurement of personality*. Oxford, England: World Book Company.
- Costa, P. T. & McCrae, R. R. (1985). *The NEO personality inventory: Manual, form S and form R*. Psychological Assessment Resources.
- Eysenck, H. J. (1947). *Dimensions of personality*. London: Routledge & Kegan Paul.
- Gerlitz, J.-Y., & Schupp, J. (2005). *Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP (The measurement of the Big Five personality traits in the SOEP)*. Berlin: DIW Berlin.
- Gosling, S. D., Rentfrow, P. J. & Swann, W. B. (2003). A very brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality*, 37, 504-528.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" Inventory – Versions 4a and 54*. Berkeley, University of California: Institute of Personality and Social Research.
- John, O. P., & Srivastava, S. (1999). *The Big Five Trait taxonomy: History, measurement, and theoretical perspectives*. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2. ed., pp. 102-138). New York: Guilford Press.
- Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the Big Five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, 43, 548-567.
- Stern, W. (1911). *Die differentielle Psychologie in ihren methodischen Grundlagen*. Leipzig: Barth.

Items

I see myself as someone who (Ich bin jemand, der):

Openness

1. is original, comes up with new ideas (originell ist, neue Ideen einbringt).
2. values artistic, aesthetic experiences (künstlerische, ästhetische Erfahrungen schätzt).
3. has an active imagination (eine lebhaft Phantasie, Vorstellungen hat).
4. is eager for knowledge (wissbegierig ist).

Scale: 1 (Not at all / Trifft überhaupt nicht zu) to 7 (Absolutely / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0215	20909	4.57	1.44	0.46	0.63
2005	plh0220	20892	4.11	1.82	0.39	0.63
2005	plh0225	20954	4.80	1.53	0.47	0.63
2009	plh0255	20675	5.34	1.35	0.45	0.67
2009	plh0215	20603	4.51	1.44	0.49	0.67
2009	plh0225	20642	4.70	1.55	0.49	0.67
2009	plh0220	20565	3.99	1.85	0.42	0.67
2012	plh0215	7140	4.89	1.34	0.44	0.62
2012	plh0220	7080	4.11	1.89	0.37	0.62
2012	plh0225	7133	4.98	1.52	0.44	0.62
2012	plh0255	7141	5.52	1.25	0.39	0.62
2013	plh0220	18939	4.25	1.83	0.41	0.66
2013	plh0225	19014	4.83	1.50	0.48	0.66
2013	plh0255	19045	5.48	1.26	0.44	0.66

continues on next page

Table 22 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2013	plh0215	18992	4.67	1.39	0.47	0.66
2017	plh0255	29457	5.64	1.26	0.44	0.66
2017	plh0215	29395	4.82	1.43	0.45	0.66
2017	plh0225	29383	4.98	1.55	0.47	0.66
2017	plh0220	29189	4.44	1.87	0.43	0.66
2019	plh0255	25934	5.56	1.24	0.43	0.65
2019	plh0220	25784	4.26	1.90	0.40	0.65
2019	plh0225	25901	4.81	1.56	0.47	0.65
2019	plh0215	25884	4.70	1.40	0.46	0.65

Conscientiousness

1. does a thorough job (gründlich arbeitet).
2. tends to be lazy (R) (eher faul ist).
3. does things effectively and efficiently (Aufgaben wirksam und effizient erledigt).

Scale: 1 (Not at all / Trifft überhaupt nicht zu) to 7 (Absolutely / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0212	20962	6.18	1.03	0.53	0.62
2005	plh0218R	20945	5.74	1.53	0.37	0.62
2005	plh0222	20916	5.78	1.13	0.45	0.62
2009	plh0212	20628	6.14	1.06	0.50	0.59
2009	plh0218R	20615	5.57	1.58	0.33	0.59
2009	plh0222	20609	5.75	1.15	0.43	0.59
2012	plh0212	7150	6.14	0.96	0.49	0.60
2012	plh0218R	7146	5.60	1.57	0.37	0.60
2012	plh0222	7134	5.80	1.03	0.45	0.60
2013	plh0212	19020	6.16	0.99	0.49	0.58
2013	plh0218R	18997	5.54	1.59	0.32	0.58
2013	plh0222	19004	5.79	1.08	0.43	0.58
2017	plh0212	29410	6.12	1.01	0.51	0.61
2017	plh0218R	29461	5.45	1.64	0.39	0.61
2017	plh0222	29416	5.81	1.07	0.46	0.61
2019	plh0212	25914	6.11	1.02	0.50	0.59
2019	plh0218R	25884	5.50	1.60	0.34	0.59
2019	plh0222	25911	5.75	1.09	0.45	0.59

Extraversion

1. is communicative, talkative (kommunikativ, gesprächig ist).
2. is outgoing, sociable (aus sich herausgehen kann, gesellig ist).
3. is reserved (R) (zurückhaltend ist).

Scale: 1 (Not at all / Trifft überhaupt nicht zu) to 7 (Absolutely / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0213	20991	5.50	1.34	0.54	0.66
2005	plh0219	20977	5.09	1.45	0.53	0.66
2005	plh0223R	20987	3.90	1.63	0.36	0.66
2009	plh0213	20661	5.43	1.37	0.54	0.66
2009	plh0219	20634	4.98	1.46	0.53	0.66
2009	plh0223R	20624	3.91	1.59	0.36	0.66
2012	plh0213	7149	5.74	1.24	0.54	0.68
2012	plh0219	7146	5.26	1.40	0.54	0.68
2012	plh0223R	7140	4.21	1.63	0.42	0.68
2013	plh0213	19039	5.55	1.30	0.54	0.66
2013	plh0219	19024	5.15	1.41	0.54	0.66
2013	plh0223R	19029	3.87	1.59	0.36	0.66
2017	plh0213	29502	5.63	1.33	0.55	0.66
2017	plh0219	29469	5.25	1.45	0.53	0.66
2017	plh0223R	29388	3.98	1.66	0.36	0.66
2019	plh0213	25944	5.56	1.33	0.57	0.69
2019	plh0219	25892	5.10	1.45	0.55	0.69
2019	plh0223R	25927	4.09	1.62	0.41	0.69

Agreeableness

1. is sometimes somewhat rude to others (R) (manchmal etwas grob zu anderen ist).
2. has a forgiving nature (verzeihen kann).
3. is considerate and kind to others (rücksichtsvoll und freundlich mit anderen umgeht).

Scale: 1 (Not at all / Trifft überhaupt nicht zu) to 7 (Absolutely / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0214R	20974	5.07	1.66	0.29	0.51
2005	plh0217	20974	5.51	1.32	0.27	0.51
2005	plh0224	21005	5.79	1.10	0.45	0.51
2009	plh0214R	20649	4.96	1.64	0.29	0.50
2009	plh0217	20658	5.38	1.35	0.27	0.50
2009	plh0224	20662	5.70	1.12	0.43	0.50
2012	plh0214R	7146	4.86	1.72	0.23	0.42
2012	plh0217	7148	5.54	1.29	0.19	0.42
2012	plh0224	7142	5.79	1.04	0.39	0.42
2013	plh0214R	19019	4.96	1.67	0.28	0.48
2013	plh0217	19042	5.43	1.32	0.26	0.48
2013	plh0224	19035	5.80	1.06	0.41	0.48
2017	plh0214R	29465	5.02	1.72	0.32	0.51
2017	plh0217	29484	5.55	1.36	0.28	0.51
2017	plh0224	29481	5.91	1.05	0.43	0.51
2019	plh0214R	25916	5.04	1.67	0.29	0.49
2019	plh0217	25912	5.39	1.38	0.25	0.49
2019	plh0224	25946	5.82	1.05	0.43	0.49

Neuroticism

1. worries a lot (sich oft Sorgen macht).
2. gets nervous easily (leicht nervös wird).
3. is relaxed, handles stress well (R) (entspannt ist, mit Stress gut umgehen kann).

Scale: 1 (Does not apply to me at all / Trifft überhaupt nicht zu) to 7 (Applies to me perfectly / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0216	20977	4.72	1.67	0.35	0.60
2005	plh0221	20964	3.71	1.73	0.49	0.60
2005	plh0226R	20975	3.45	1.51	0.39	0.60
2009	plh0216	20673	4.35	1.68	0.40	0.62
2009	plh0221	20650	3.61	1.68	0.49	0.62
2009	plh0226R	20650	3.52	1.50	0.40	0.62
2012	plh0216	7152	4.71	1.71	0.41	0.63
2012	plh0221	7143	3.58	1.70	0.50	0.63
2012	plh0226R	7139	3.42	1.51	0.42	0.63
2013	plh0216	19040	4.24	1.68	0.39	0.62
2013	plh0221	19019	3.65	1.69	0.50	0.62
2013	plh0226R	19036	3.41	1.48	0.41	0.62
2017	plh0216	29490	4.42	1.76	0.36	0.59
2017	plh0221	29481	3.60	1.72	0.46	0.59
2017	plh0226R	29465	3.33	1.55	0.38	0.59
2019	plh0216	25940	3.97	1.77	0.45	0.66
2019	plh0221	25932	3.39	1.67	0.53	0.66
2019	plh0226R	25933	3.38	1.51	0.43	0.66

Test-Retest Correlations

The Big Five inventory was included in retests taken by subsamples in 2005 and 2009 within 30 to 49 days after the initial test. Pooled across both occasions, test-retest correlations were .64 for the 3-item version of openness (N = 325) and .66 for the 4-item version (N = 167); .53 for conscientiousness; .64 for extraversion; .57 for agreeableness; and .62 for neuroticism.

Big Five (Youth)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2006-2020	jl0368	6964	4.86	1.35	0.41	0.60
2006-2020	jl0373	6896	4.02	1.89	0.38	0.60
2006-2020	jl0378	6962	5.23	1.52	0.41	0.60
2006-2020	jl0380	6941	5.11	1.42	0.35	0.60
2006-2020	jl0365	6972	5.28	1.34	0.58	0.67
2006-2020	jl0371R	6967	4.10	1.80	0.42	0.67
2006-2020	jl0375	6958	5.24	1.22	0.50	0.67
2006-2020	jl0366	6979	5.23	1.52	0.63	0.74
2006-2020	jl0372	6953	5.14	1.48	0.60	0.74
2006-2020	jl0376R	6963	4.15	1.75	0.48	0.74

continues on next page

Table 27 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2006-2020	jl0367R	6958	4.70	1.64	0.25	0.47
2006-2020	jl0370	6970	5.58	1.28	0.25	0.47
2006-2020	jl0377	6971	5.82	1.10	0.41	0.47
2006-2020	jl0369	6970	4.49	1.73	0.36	0.57
2006-2020	jl0374	6970	3.96	1.69	0.45	0.57
2006-2020	jl0379R	6966	3.49	1.53	0.33	0.57

Big Five (13-14 year olds, self-report)**Items and Scale Statistics**

year	variable	count	mean	sd	itemrestcorr	alpha
2016-2020	char14	2825	4.77	1.52	0.41	0.59
2016-2020	char19	2796	4.03	2.12	0.37	0.59
2016-2020	char24	2823	5.23	1.65	0.41	0.59
2016-2020	char26	2783	4.97	1.57	0.34	0.59
2016-2020	char11	2828	5.06	1.38	0.60	0.70
2016-2020	char17R	2827	4.15	1.89	0.43	0.70
2016-2020	char21	2827	4.99	1.37	0.57	0.70
2016-2020	char12	2829	5.28	1.52	0.52	0.64
2016-2020	char18	2809	4.96	1.57	0.50	0.64
2016-2020	char22R	2817	4.18	1.78	0.35	0.64
2016-2020	char13R	2824	5.16	1.56	0.30	0.52
2016-2020	char16	2830	5.49	1.46	0.29	0.52
2016-2020	char23	2829	5.81	1.21	0.44	0.52
2016-2020	char15	2827	3.57	1.88	0.36	0.55
2016-2020	char20	2828	3.47	1.74	0.45	0.55
2016-2020	char25R	2822	3.50	1.59	0.27	0.55

Big Five (11-12 year olds, self-report)**Items and Scale Statistics**

year	variable	count	mean	sd	itemrestcorr	alpha
2014-2020	char14	4290	4.85	1.57	0.42	0.58
2014-2020	char19	4269	4.35	2.07	0.35	0.58
2014-2020	char24	4306	5.41	1.61	0.39	0.58
2014-2020	char26	4255	5.02	1.63	0.31	0.58
2014-2020	char11	4319	4.94	1.41	0.56	0.68
2014-2020	char17R	4320	4.54	1.85	0.42	0.68
2014-2020	char21	4293	4.91	1.46	0.54	0.68
2014-2020	char12	4310	5.38	1.51	0.46	0.57
2014-2020	char18	4257	5.05	1.54	0.43	0.57
2014-2020	char22R	4294	4.29	1.81	0.29	0.57
2014-2020	char13R	4301	5.25	1.58	0.31	0.53
2014-2020	char16	4316	5.64	1.44	0.29	0.53

continues on next page

Table 29 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2014-2020	char23	4316	5.86	1.19	0.47	0.53
2014-2020	char15	4310	3.44	1.88	0.35	0.51
2014-2020	char20	4304	3.41	1.74	0.41	0.51
2014-2020	char25R	4286	3.74	1.61	0.23	0.51

Big Five (9-10 year olds & 5-6 year olds, parent reports)

References

Asendorpf, J. B., & van Aken, M. A. G. (2003). *Validity of Big Five personality judgments in childhood: A 9 year longitudinal study. European Journal of Personality, 17, 1-17.*

Weinert, S., Asendorpf, J. B., Beelmann, A., Doil, H., Frevert, S., Lohaus, A., & Hasselhorn, M. (2007). *Expertise zur Erfassung von psychologischen Personmerkmalen bei Kindern im Alter von fünf Jahren im Rahmen des SOEP. Berlin: DIW Berlin.*

Items

How would you rank your child in comparison to other children of the same age (Wie würden Sie Ihr Kind im Vergleich zu anderen Kindern gleichen Alters beurteilen):

Openness/Intellect

1. Child is not that interested – hungry for knowledge (ist wenig interessiert – ist wissensdurstig).
2. Child understands quickly – needs more time (begreift schnell – braucht mehr Zeit).

Scale: 0 to 10

Conscientiousness

1. Child is tidy – untidy (ist unordentlich – ist ordentlich).
2. Child is focused – easy to distract (ist konzentriert – ist leicht ablenkbar).

Scale: 0 to 10

Extraversion

1. Child is talkative – quiet (ist gesprächig – ist still).
2. Child is withdrawn – sociable (ist zurückgezogen – ist kontaktfreudig).

Scale: 0 to 10

Agreeableness

1. Child is good-natured – irritable (ist gutmütig – ist reizbar).
2. Child is obstinate – compliant (ist trotzig – ist fügsam).

Scale: 0 to 10

Neuroticism

1. Child is self-confident – insecure (hat Selbstvertrauen – ist unsicher).
2. Child is fearful – fearless (ist ängstlich – ist unängstlich).

Scale: 0 to 10

Items and Scale Statistics

year	variable	count	mean	sd
2012-2020	char7	7112	7.68	2.27
2012-2020	char4R	7112	6.95	2.86
2012-2020	char5	7122	5.03	2.90
2012-2020	char2R	7100	5.44	3.12
2012-2020	char1bR	7124	7.26	2.63
2012-2020	char9	7110	7.60	2.33
2012-2020	char6R	7092	7.00	2.67
2012-2020	char3	7097	5.84	2.57
2012-2020	char8	7109	3.46	2.71
2012-2020	char10R	7108	3.69	2.58

Items and Scale Statistics

year	variable	count	mean	sd
2008-2020	char7	7385	8.27	1.99
2008-2020	char4R	7388	7.54	2.48
2008-2020	char5	7403	5.62	2.60
2008-2020	char2R	7386	6.07	2.82
2008-2020	char1bR	7398	7.63	2.49
2008-2020	char9	7379	7.88	2.19
2008-2020	char6R	7376	7.15	2.44
2008-2020	char3	7376	5.71	2.46
2008-2020	char8	7383	3.16	2.59
2008-2020	char10R	7388	3.61	2.51

Big Five (2-3 year olds, parent reports)

References

Asendorpf, J. B., & van Aken, M. A. G. (2003). *Validity of Big Five personality judgments in childhood: A 9 year longitudinal study. European Journal of Personality, 17, 1-17.*

Weinert, S., Asendorpf, J. B., Beelmann, A., Doil, H., Frevert, S., Lohaus, A., & Hasselhorn, M. (2007). *Expertise zur Erfassung von psychologischen Personmerkmalen bei Kindern im Alter von fünf Jahren im Rahmen des SOEP. Berlin: DIW Berlin.*

Items

How would you rank your child in comparison to other children of the same age (Wie würden Sie Ihr Kind im Vergleich zu anderen Kindern gleichen Alters beurteilen):

Openness/Intellect

1. Quick at learning new things – needs more time (begreift eher schnell – braucht mehr Zeit).

Scale: 0 to 10

Conscientiousness

1. Focused – easily distracted (ist konzentriert – ist leicht ablenkbar).

Scale: 0 to 10

Extraversion

1. Shy – outgoing (ist eher schüchtern– ist eher kontaktfreudig).

Scale: 0 to 10

Agreeableness

1. Obstinate – obedient (ist eher trotzig – ist eher fügsam/folgsam).

Scale: 0 to 10

Items and Scale Statistics

year	variable	count	mean	sd
2005-2020	char4	7018	2.25	2.31
2005-2020	char2	7006	4.15	2.66
2005-2020	char1a	7020	7.49	2.46
2005-2020	char3	7001	4.89	2.50

3.2.19 Reciprocity

Summary

Reciprocity is understood as the tendency to respond in kind to the actions of other people, that is, to respond to actions directed at oneself by acting in the same or a similar way (Fehr & Schmidt, 2005; Perugini et al., 2003). A distinction is drawn here between negative and positive reciprocity. Negative reciprocity is the tendency to respond to bad treatment by another person in a similarly negative way. Positive reciprocity relates to the tendency to respond positively to a positive experience. Reciprocity was measured in the SOEP in 2005, 2010, 2015, and 2020.

Theoretical Background

The scale is designed to measure reciprocity as an internalized norm within individuals. Reciprocity is understood as the tendency to respond to the actions of other people in kind, that is, to respond to actions directed at oneself by acting in the same or a similar way (Fehr & Schmidt, 2005; Perugini et al., 2003). The action may either have already occurred or be expected in the future. The concept of reciprocity can be used on three levels: 1. To describe real patterns of social exchange in society, 2. To describe the belief of social actors themselves that (and to what extent), from their point of view, relationships are regulated by the norm of reciprocity, 3. To describe the extent to which reciprocity is a norm that governs social actors in their daily interactions and internalized by individuals (Gouldner, 1960). Gouldner also argues that reciprocity is a basic norm that can be found in all historically known societies. Negative reciprocity can be distinguished from positive reciprocity. Negative reciprocity describes the tendency to respond negatively to negative treatment by another person and thus to negatively sanction this behavior. Positive reciprocity describes the tendency to approve and positively sanction positive interpersonal experiences, that is, to return favors. This distinction is required because different individuals respond more or less sensitively to positive or negative experiences and tend toward more or less positive or negative sanctions. Thus, individuals may be careful to always return favors but may at the same time (possibly motivated by Christian norms of forgiveness) refrain from taking revenge or imposing penalties on others (Perugini et al. 2003). Empirically, positive and negative reciprocity are uncorrelated (Egloff, Richter & Schmukle, 2013). Both positive and negative reciprocity could be of interest in research on the motives for providing support of various kinds to friends and relatives or to engage in volunteer work, behavior in dilemma situations (game theory), and on forms of reciprocity as determinants of stability in social relationships.

Scale Development

The original scale (Perugini et al., 2003) consists of 27 items (nine per subscale) that were shortened for use in a survey to a total of nine (three per subscale), taking the three items with the highest factor loadings for each of the subscales. The three original dimensions could not, however, be extracted by means of principal component analysis in the 2004 SOEP pretest: the dimension “belief in reciprocity” is comprised of the dimensions “positive reciprocity” and “negative reciprocity.” That is, the theoretically postulated dimension “belief in reciprocity” is not empirically orthogonal to

behavior. Thus, only the two factors “negative reciprocity” and “positive reciprocity” can be found consistently in the sample. Thus, in the shortened version of the questionnaire, only the six items in these two dimensions were used in the main SOEP survey. Further information on scale development and validity can be found in Dohmen et al. (2008, 2009).

References

- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). *Representative trust and reciprocity: Prevalence and determinants*. *Economic Inquiry*, 46, 81-90.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2009). *Homo reciprocans: Survey evidence on behavioral outcomes*. *Economic Journal*, 119, 592 - 612.
- Egloff, B., Richter, D., & Schmukle, S. C. (2013). *Need for conclusive evidence that positive and negative reciprocity are unrelated. Comment on Yamagishi, T., et al. (2012) Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity*. *Proceedings of the National Academy of Sciences of the USA*, 110, 786.
- Fehr, E. & Schmidt, K. M. (2006). *The economics of fairness, reciprocity and altruism. Experimental evidence and new theories*. *Handbook of the Economics of Giving, Altruism and Reciprocity*, 615-691.
- Gouldner, A. W. (1960). *The Norm of Reciprocity: A Preliminary Statement*. *American Sociological Review*, 25, 161-178.
- Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). *The personal norm of reciprocity*. *European Journal of Personality*, 17, 251-283.

Items

To what degree do the following statements apply to you personally (In welchem Maße treffen die folgenden Aussagen auf Sie persönlich zu):

Positive

1. If someone does me a favor, I am prepared to return it (Wenn mir jemand einen Gefallen tut, bin ich bereit, dies zu erwidern)
2. I go out of my way to help somebody who has been kind to me in the past (Ich stenge mich besonders an, um jemandem zu helfen, der mir früher schon mal geholfen hat)
3. I am ready to assume personal costs to help somebody who helped me in the past (Ich bin bereit, Kosten auf mich zu nehmen, um jemanden zu helfen, der mir früher geholfen hat)

Scale: 1 (Does not apply to me at all / Trifft überhaupt nicht zu) to 7 (Applies to me perfectly / Trifft voll zu)

Test-Retest Correlations

Positive reciprocity was included in a retest taken by subsamples (N = 158) in 2005 within 30 to 49 days after the initial test. Test-retest correlations of the items were (in scale order) .25, .35, and .36; test-retest correlation of scale scores was .42.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0206i01	21015	6.45	0.88	0.41	0.64
2005	plh0206i04	20975	5.90	1.17	0.55	0.64
2005	plh0206i06	20949	5.28	1.48	0.45	0.64
2010	plh0206i01	18857	6.42	0.90	0.38	0.61
2010	plh0206i04	18822	5.88	1.18	0.50	0.61
2010	plh0206i06	18800	5.22	1.49	0.43	0.61
2015	plh0206i01	26987	6.40	0.94	0.35	0.61
2015	plh0206i04	26967	5.90	1.18	0.52	0.61

continues on next page

Table 33 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2015	plh0206i06	26877	5.31	1.49	0.43	0.61
2016	plh0206i01	4400	6.79	0.77	0.31	0.55
2016	plh0206i04	4393	6.78	0.71	0.46	0.55
2016	plh0206i06	4345	6.45	1.12	0.37	0.55
2017	plh0206i01	2883	6.82	0.62	0.31	0.56
2017	plh0206i04	2874	6.75	0.71	0.45	0.56
2017	plh0206i06	2854	6.42	1.16	0.45	0.56
2018	plh0206i01	426	6.81	0.62	0.28	0.53
2018	plh0206i04	422	6.72	0.84	0.44	0.53
2018	plh0206i06	422	6.34	1.29	0.40	0.53
2019	plh0206i01	280	6.71	0.76	0.52	0.71
2019	plh0206i04	277	6.48	1.16	0.61	0.71
2019	plh0206i06	279	6.14	1.38	0.54	0.71
2020	plh0206i01	30344	6.42	0.96	0.38	0.65
2020	plh0206i04	30309	5.97	1.19	0.55	0.65
2020	plh0206i06	30253	5.57	1.40	0.48	0.65

Items

To what degree do the following statements apply to you personally (In welchem Maße treffen die folgenden Aussagen auf Sie persönlich zu):

Negative

1. If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the cost (Wenn mir schweres Unrecht zuteilwird, werde ich mich um jeden Preis bei der nächsten Gelegenheit dafür rächen)
2. If somebody puts me in a difficult position, I will do the same to him/her (Wenn mich jemand in eine schwierige Lage bringt, werde ich das Gleiche mit ihm zu machen)
3. If somebody offends me, I will offend him/her back (Wenn mich jemand beleidigt, werde ich mich ihm gegenüber beleidigend verhalten)

Scale: 1 (Does not apply to me at all / Trifft überhaupt nicht zu) to 7 (Applies to me perfectly / Trifft voll zu)

Test-Retest Correlations

Negative reciprocity was included in a retest taken by a subsample (N = 158) in 2005 within 30 to 49 days after the initial test. Test-retest correlations of the items were (in scale order) .44, .42, and .58; test-retest correlation of scale scores was .64.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005	plh0206i02	20947	3.21	1.73	0.71	0.83
2005	plh0206i03	20925	2.88	1.63	0.73	0.83
2005	plh0206i05	20952	3.24	1.73	0.61	0.83
2010	plh0206i02	18822	3.18	1.69	0.70	0.82
2010	plh0206i03	18785	2.83	1.59	0.73	0.82
2010	plh0206i05	18805	3.12	1.70	0.59	0.82
2015	plh0206i02	26921	2.77	1.67	0.68	0.81
2015	plh0206i03	26936	2.47	1.53	0.73	0.81
2015	plh0206i05	26939	2.85	1.70	0.56	0.81

continues on next page

Table 34 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2016	plh0206i02	4298	1.86	1.67	0.58	0.75
2016	plh0206i03	4324	1.58	1.35	0.64	0.75
2016	plh0206i05	4341	1.89	1.65	0.55	0.75
2017	plh0206i02	2815	1.68	1.47	0.56	0.72
2017	plh0206i03	2842	1.62	1.43	0.57	0.72
2017	plh0206i05	2857	1.85	1.65	0.50	0.72
2018	plh0206i02	413	1.90	1.67	0.52	0.68
2018	plh0206i03	414	1.76	1.62	0.57	0.68
2018	plh0206i05	421	2.32	2.00	0.42	0.68
2019	plh0206i02	271	2.12	1.68	0.58	0.75
2019	plh0206i03	276	2.02	1.60	0.65	0.75
2019	plh0206i05	276	2.43	1.88	0.53	0.75
2020	plh0206i02	30253	2.57	1.60	0.67	0.79
2020	plh0206i03	30278	2.30	1.47	0.70	0.79
2020	plh0206i05	30285	2.72	1.67	0.55	0.79

3.2.20 Risk Aversion

Summary

The global and dimension-specific measurement of risk aversion consists of seven items and was used in the SOEP in 2004 and 2009. An individual item on general risk aversion has been included in the SOEP at two-year intervals since 2004 and annually since 2008.

Theoretical Background

Risks and uncertainties play a role in almost all important life decisions. In their pioneering work on the subject, Kahneman and Tversky (1979) developed a theoretical foundation for the analysis of risk attitudes in the form of “Prospect Theory.” This theory describes the process of decision making in situations of uncertainty, and postulates that the consequences of a decision are evaluated relative to a reference point as either gains or losses, also considering the probability of their occurrence. Here, individual risk attitudes affect not only how individuals evaluate the potential gains and losses but also how they evaluate the respective probabilities of the event’s occurrence. In the economic research, individual risk attitudes are used to explain decisions about financial investments, labor market behavior and success, and health behavior (e.g., Dohmen et al., 2011). More recent work in the field of educational sociology has studied the influence of individual risk aversion on educational decisions (Hartlaub & Schneider, 2013).

Scale Development

The scale was first piloted in the 2003 SOEP pretest and was then used in the 2004 main SOEP survey. Simultaneously to the SOEP pretest, a behaviorally oriented experiment on risk attitudes was carried out with respondents (see Dohmen et al., 2011). In this experiment, respondents were shown a table consisting of 20 text lines and were given the task of deciding in each line whether they would prefer to receive a fixed amount of money or take part in a lottery with a 50% chance of winning 300 euros. In the case of the “safe” option, the amount paid increased progressively from one line to the next (0 euros in the first, 10 euros in the second, and so on up to 190 euros in line 20). Since the expected lottery winnings would be 150 euros, more risk-averse respondents should take the safe option below this value, while risk-seeking participants can be expected to decide for the lottery even with safe options of 160, 170, 180, or 190 euros. The amount paid for the safe option at the time of switching to this option was regressed on the value of the questionnaire on risk aversion. Here, the behavior in the experiment could be predicted well based on the survey responses ($\beta = .61$, $p < .001$).

References

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). *Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences*. *Journal of the European Economic Association*, 9, 522-550.

Hartlaub, V. & Schneider, T. (2013). *Educational choice and risk preferences: How important is relative vs. individual risk preference?* Manuscript submitted for publication.).

Kahneman, D. & Tversky, A. (1979). *Prospect Theory: An analysis of decision under risk*. *Econometrica: Journal of the Econometric Society*, 47, 263-291.

Items

How do you see yourself (Wie schätzen Sie sich persönlich ein):

Risk Aversion in General

1. Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? (Sind Sie im Allgemeinen ein risikobereiter Mensch oder versuchen Sie, Risiken zu vermeiden?)

Scale: 0 (Risk averse / Gar nicht risikobereit) to 10 (Fully prepared to take risks / Sehr risikobereit)

Test-Retest Correlations

Risk aversion in general was included in retests taken by subsamples in 2005, 2006, and 2009 within 30 to 49 days after the respective initial tests. Test-retest correlation pooled across all three waves (N = 607) was .60.

Items and Scale Statistics

year	variable	count	mean	sd
2004	plh0204_v2	21881	4.42	2.38
2006	plh0204_v2	22210	4.77	2.29
2008	plh0204_v2	19639	4.46	2.31
2009	plh0204_v2	20707	3.74	2.21
2010	plh0204_v2	26628	4.42	2.35
2011	plh0204_v2	21011	4.54	2.27
2012	plh0204_v2	27903	4.86	2.26
2013	plh0204_v2	19117	4.51	2.40
2014	plh0204_v2	27275	4.77	2.42
2015	plh0204_v2	27116	4.87	2.44
2016	plh0204_v2	28765	4.86	2.61
2017	plh0204_v2	32247	4.66	2.60
2018	plh0204_v2	30132	4.29	2.62
2019	plh0204_v2	29792	4.99	2.57
2020	plh0204_v2	30334	4.84	2.61

year	variable	count	mean	sd
2006-2020	jl0349	6994	5.88	2.24

year	variable	count	mean	sd
2016-2020	char30	2855	4.89	2.48

year	variable	count	mean	sd
2014-2020	char30	4340	5.09	2.57

Items

How would you rate your willingness to take risks in the following areas? How is it ... (Wie würden Sie Ihre Risikobereitschaft in Bezug auf die folgenden Bereiche einschätzen? Wie ist das ...):

Risk Aversion in different Domains

1. While driving (beim Autofahren)?
2. In financial matters (bei Geldanlagen)?
3. During leisure and sport (bei Freizeit und Sport)?
4. In your occupation (bei Ihrer beruflichen Karriere)?
5. With your health (bei Ihrer Gesundheit)?
6. Your faith in other people (bei Vertrauen in fremde Menschen)?

Scale: 0 (Risk averse / Gar nicht risikobereit) to 10 (Fully prepared to take risks / Sehr risikobereit)

Test-Retest Correlations

Risk aversion in different domains was included in a retest of a subsample (N = 120) in 2009 within 30 to 49 days after the respective initial tests. Test-retest correlation of the items were (in scale order) .67, .48, .58, .69, .51, and .37; scale scores had a test-retest correlation of .69.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2004	plh0197	20604	2.93	2.53	0.63	0.85
2004	plh0198	21691	2.41	2.23	0.62	0.85
2004	plh0199	21574	3.49	2.61	0.68	0.85
2004	plh0200	19902	3.60	2.71	0.68	0.85
2004	plh0201	21868	2.93	2.47	0.65	0.85
2004	plh0202	21890	3.35	2.40	0.50	0.85
2009	plh0197	19426	2.98	2.57	0.62	0.82
2009	plh0198	20472	1.90	2.14	0.56	0.82
2009	plh0199	20274	3.20	2.63	0.67	0.82
2009	plh0200	18077	3.21	2.72	0.64	0.82
2009	plh0201	20696	2.71	2.44	0.63	0.82
2009	plh0202	20712	3.23	2.39	0.43	0.82
2014	plh0197	25254	3.51	2.63	0.60	0.82
2014	plh0198	26442	2.37	2.25	0.59	0.82
2014	plh0199	26791	3.82	2.60	0.65	0.82
2014	plh0200	23532	3.87	2.65	0.64	0.82
2014	plh0201	27254	3.21	2.46	0.63	0.82
2014	plh0202	27290	3.57	2.41	0.44	0.82

3.2.21 Self Esteem

Summary

Global self-esteem—a person’s overall evaluation or appraisal of his or her worth— was measured in the SOEP in 2010, 2015, and 2020 with an individual item.

Theoretical Background

Self-esteem is a central construct in psychology and has value both as a predictor variable (e.g., for the occurrence of various life events) and as a target variable with great potential significance for numerous research questions.

Scale Development

In the SOEP, following on the paper by Robins, Hendin, & Trzesniewski (2001), an individual item was used to measure self-esteem.

References

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). *Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. Personality and Social Psychology Bulletin*, 27, 151-161.

Items

To what degree do the following statements apply to you personally (In welchem Maße treffen die folgenden Aussagen auf Sie persönlich zu):

1. I have a positive attitude toward myself (Ich habe eine positive Einstellung zu mir selbst).

Scale: 1 (Does not apply to me at all / Trifft überhaupt nicht zu) to 7 (Applies to me perfectly / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd
2010	plh0206i11	18817	5.58	1.28
2015	plh0206i11	26964	5.65	1.29
2016	plh0206i11	4195	6.27	1.18
2017	plh0206i11	2743	6.34	1.17
2018	plh0206i11	418	6.39	1.11
2019	plh0206i11	279	6.35	1.02
2020	plh0206i11	30301	5.74	1.29

year	variable	count	mean	sd
2010-2020	jl1380	5822	5.28	1.44

year	variable	count	mean	sd
2016	char27	2814	5.21	1.50

year	variable	count	mean	sd
2014-2020	char27	4264	5.24	1.44

3.2.22 Sources of Social Inequality

Summary

The scale measures respondents' attributions of upward mobility and success in society to the use of legitimate or illegitimate means, whereby the categories of "legitimate" and "illegitimate" are evaluated from the perspective of the prevailing universalistic ideology of meritocratic achievement. The scale has been used in the SOEP youth questionnaire since 2001.

Theoretical Background

The scale measures a non-normative aspect: namely, beliefs about the extent to which success can be achieved by specific means. Theoretically, the scale is implicitly oriented toward measuring respondents' perceptions of legitimacy or their critique of the social status assignment process. The legitimization of existing social inequality is necessary for the functioning of a society, independent of the form that this legitimization takes. In German society, as in most Western societies, social inequalities are legitimized meritocratically, that is, inequalities are considered justified if they are based on non-ascriptive characteristics such as individual abilities and achievements, as expressed, for example, in educational degrees. To this end, it must be determined, first, whether the individual supports the normative basis for this meritocratic ideology, and second, to what extent the respondent views these normative standards as being realized in the society. The second, cognitive aspect is measured with the items described below. Both aspects together allow for an estimation of whether the inequality that exists in a society is perceived as legitimate or as worthy of criticism.

Scale Development

For the SOEP surveys, 11 items from the original scale by Sandberger (1983) were selected. Item 11 (having the right gender) was added later.

References

Sandberger, J.-U. (1983). *Zwischen Legitimation und Kritik. Vorstellungen von Akademikern, Studenten und Bevölkerung zur sozialen Ungleichheit. Zeitschrift für Soziologie*, 12, 181-202.

Items

In your opinion, what are the most important factors for achieving success and improving one's social situation in Germany (Wovon hängt es Ihrer Meinung nach in Deutschland tatsächlich ab, ob jemand Erfolg hat und sozial aufsteigt):

Legitimate instruments (Legitime Mittel)

1. Working hard (Man muss sich anstrengen und fleißig sein)
2. Being gifted and intelligent (Man muss begabt und intelligent sein)
3. Having good training in a specialized field (Man muss gute Fachkenntnisse auf seinem Spezialgebiet haben)
4. Having good grades at school (Man muss einen möglichst guten Schulabschluss haben)
5. Being dynamic and exhibiting initiative (Man muss dynamisch sein und Initiative haben)

Scale: 1 (I agree completely/Stimme voll zu) to 4 (I don't agree at all/Stimme überhaupt nicht zu)

Illegitimate instruments (Illegitime Mittel)

1. Exploiting others (Man muss andere ausnutzen)
2. Being from the right family (Man muss aus der richtigen Familie stammen)
3. Having a lot of money and assets (Man muss Geld und Vermögen haben)

4. Being tough and ruthless (Man muss rücksichtslos und hart sein)
5. Having connections to the right people (Man muss Beziehungen zu den richtigen Leuten haben)
6. Being politically active in the right party (Man muss sich auf der richtigen Seite politische engagieren)
7. Having the right gender; men have better career opportunities (Man muss das richtige Geschlecht haben; Männer haben bessere Aufstiegschancen)

Scale: 1 (I agree completely / Stimme voll zu) to 4 (I don't agree at all / Stimme überhaupt nicht zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2001-2020	jl0337	8965	1.44	0.58	0.31	0.62
2001-2020	jl0339	8940	1.97	0.71	0.40	0.62
2001-2020	jl0341	8936	1.64	0.65	0.40	0.62
2001-2020	jl0343	8963	1.60	0.70	0.39	0.62
2001-2020	jl0348	8901	1.71	0.64	0.37	0.62
2001-2020	jl0338	8928	3.27	0.81	0.53	0.78
2001-2020	jl0340	8914	2.69	0.92	0.55	0.78
2001-2020	jl0342	8937	2.81	0.85	0.58	0.78
2001-2020	jl0344	8932	3.08	0.85	0.56	0.78
2001-2020	jl0345	8932	2.05	0.79	0.53	0.78
2001-2020	jl0346	8773	2.93	0.79	0.39	0.78
2001-2020	jl0347	8894	3.01	0.89	0.41	0.78

3.2.23 Strengths and Difficulties Questionnaire (SDQ)

Summary

The Strengths and Difficulties Questionnaire (SDQ; Goodman 1997; German version by Woerner et al. 2002) captures the child's behavior on five scales "hyperactivity," "emotional problems," "prosocial behavior," "conduct problems," and "peer problems." The Strengths and Difficulties Questionnaire has been given to parents of children aged 5-6 years since 2008 and to parents of children aged 9-10 years since 2012. Since 2014, children aged 11-12 years fill in the questionnaire themselves.

Theoretical Background

The SDQ is currently one of the most frequently used screening instruments for child and adolescent mental health throughout the world. The instrument includes four problem subscales of five items each that also yield a total difficulties score, in addition to a prosocial dimension.

Scale Development

The scale was shortened for the SOEP to 17 items for the bioage06 questionnaire and to 18 items for the bioage10 questionnaire. The answer format was modified to seven levels. For the 11-12 year olds, the full version of the SDQ with an answer format of three levels was used.

References

Goodman, R. (1997). *The Strengths and Difficulties Questionnaire: A research note. Journal of Child Psychology and Psychiatry*, 38, 581-586.

Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R. & Rothenberger, A. (2002). *Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 30, 105-112.

Strengths and Difficulties Questionnaire (5-6 year olds & 9-10 year olds, parent reports)

Items

To what extent do the following statements apply to your child (Inwieweit treffen folgende Aussagen auf ihr Kind zu):

Hyperactivity

1. Child is agitated, hyperactive, cannot sit still (ist unruhig, überaktiv, kann nicht lange stillsitzen)
2. Child is fidgety (ist ständig zappelig)
3. Child is easily distracted and lacks concentration (ist leicht ablenkbar, unkonzentriert)
4. Child finishes tasks, is able to concentrate (führt Aufgaben zu Ende; kann sich lange konzentrieren)
5. Child thinks before acting (Denkt nach bevor es handelt)

Scale: 1 (Does not apply at all / Trifft überhaupt nicht zu) to 7 (Applies completely / Trifft voll zu)

Emotional Problems

1. Child is often unhappy or dejected (unglücklich, niedergeschlagen)
2. Child is nervous or clingy in new situations, loses self-confidence easily (ist nervös oder anklammernd in neuen Situationen; verliert leicht das Selbstvertrauen)
3. Child has many fears, becomes frightened easily (hat viele Ängste; fürchtet sich leicht)

Scale: 1 (Does not apply at all/Trifft überhaupt nicht zu) to 7 (Applies completely /Trifft voll zu)

Prosocial Behavior

1. Child is considerate (ist rücksichtsvoll)
2. Child likes to share with others (sweets, toys, crayons) (teilt gerne mit anderen Kindern (Süßigkeiten, Spielzeug, Buntstifte usw.))
3. Child is helpful if others are hurt, sick, or sad (ist hilfsbereit, wenn andere verletzt, krank oder betrübt sind)
4. Child helps others of his/her own accord (parents, teachers, other children) (hilft anderen oft freiwillig (Eltern, Erziehern, anderen Kindern))

Scale: 1 (Does not apply at all / Trifft überhaupt nicht zu) to 7 (Applies completely / Trifft voll zu)

Conduct Problems

1. Child often has tantrums, has a temper (hat oft Wutanfälle, ist aufbrausend)
2. Child quarrels a lot with other children, picks on them (streitet oft, schikaniert Kinder)

Scale: 1 (Does not apply at all / Trifft überhaupt nicht zu) to 7 (Applies completely / Trifft voll zu)

Peer Problems

1. Child is a loner, usually plays by him/herself (ist ein Einzelgänger, spielt meist alleine)
2. Child is popular with other children (ist im Allgemeinen bei anderen Kindern beliebt)
3. Child is often made fun of or picked on by other children (wird von anderen gehänselt oder schikaniert)
4. Child gets along better with adults than with other children (kommt besser mit Erwachsenen aus)

Scale: 1 (Does not apply at all / Trifft überhaupt nicht zu) to 7 (Applies completely / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2008-2020	behav2	5877	3.25	1.89	0.59	0.78
2008-2020	behav7	5842	2.80	1.78	0.63	0.78
2008-2020	behav11	5859	3.30	1.75	0.60	0.78
2008-2020	behav17R	5871	3.20	1.65	0.51	0.78
2008-2020	behav9	5887	1.99	1.34	0.39	0.64
2008-2020	behav12	5862	2.91	1.73	0.48	0.64
2008-2020	behav16	5876	2.65	1.66	0.49	0.64
2008-2020	behav1	5864	5.52	1.28	0.45	0.68
2008-2020	behav3	5882	5.57	1.38	0.45	0.68
2008-2020	behav6	5871	5.96	1.28	0.48	0.68
2008-2020	behav14	5869	5.53	1.41	0.47	0.68
2008-2020	behav4	5876	2.90	1.72		
2008-2020	behav8	5877	2.11	1.40		
2008-2020	behav5	5880	2.29	1.63	0.37	0.57
2008-2020	behav10R	5869	2.10	1.17	0.35	0.57
2008-2020	behav13	5866	1.73	1.13	0.40	0.57
2008-2020	behav15	5846	2.82	1.65	0.33	0.57

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2012-2020	behav2	5821	3.07	1.96	0.57	0.81
2012-2020	behav7	5802	2.64	1.81	0.61	0.81
2012-2020	behav11	5799	3.43	1.92	0.67	0.81
2012-2020	behav17R	5804	3.22	1.78	0.61	0.81
2012-2020	behav18R	5796	3.29	1.61	0.54	0.81
2012-2020	behav9	5821	2.10	1.45	0.45	0.69
2012-2020	behav12	5794	2.74	1.73	0.52	0.69
2012-2020	behav16	5811	2.46	1.62	0.54	0.69
2012-2020	behav1	5807	5.68	1.36	0.49	0.71
2012-2020	behav3	5822	5.68	1.39	0.49	0.71
2012-2020	behav6	5815	6.14	1.27	0.52	0.71
2012-2020	behav14	5803	5.60	1.45	0.50	0.71
2012-2020	behav4	5820	2.83	1.78		
2012-2020	behav8	5811	2.06	1.46		
2012-2020	behav5	5824	2.23	1.62	0.42	0.64
2012-2020	behav10R	5795	2.14	1.25	0.45	0.64
2012-2020	behav13	5798	2.01	1.42	0.47	0.64
2012-2020	behav15	5787	2.82	1.73	0.38	0.64

Strengths and Difficulties Questionnaire (11-12 year olds & 13-14 year olds, self-report)

Items

Hyperactivity

1. I am restless, I cannot stay still for long (Ich bin oft unruhig, ich kann nicht lange stillsitzen)
2. I am constantly fidgeting or squirming (Ich bin dauernd in Bewegung und zappelig)
3. I am easily distracted, I find it difficult to concentrate (Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren)

4. I think before I do things (Ich denke nach, bevor ich etwas tue)
5. I finish the work I'm doing. My attention is good (Was ich angefangen habe, mache ich zu Ende; ich kann mich lange genug konzentrieren)

Scale: 1 = Not True (Trifft gar nicht zu), 2 = Somewhat True (Trifft teilweise zu), 3 = Certainly True (Trifft voll zu)

Emotional Problems

1. I get a lot of headaches, stomach-aches or sickness (Ich habe häufig Kopfschmerzen oder Bauchschmerzen oder mir wird oft schlecht)
2. I worry a lot (Ich mache mir häufig Sorgen)
3. I am often unhappy, depressed or tearful (Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen)
4. I am nervous in new situations. I easily lose confidence (Neue Situationen machen mich nervös; ich verliere leicht das Selbstvertrauen)
5. I have many fears, I am easily scared (Ich habe viele Ängste; ich fürchte mich leicht)

Scale: 1 = Not True (Trifft gar nicht zu), 2 = Somewhat True (Trifft teilweise zu), 3 = Certainly True (Trifft voll zu)

Prosocial Behavior

1. I try to be nice to other people. I care about their feelings (Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig)
2. I usually share with others, for example CD's, games, food (Ich teile gerne mit anderen)
3. I am helpful if someone is hurt, upset or feeling ill (Ich bin hilfsbereit, wenn andere verletzt, krank oder traurig sind)
4. I am kind to younger children (Ich bin nett zu jüngeren Kindern)
5. I often offer to help others (parents, teachers, children) (Ich helfe anderen oft freiwillig (Eltern, Lehrern oder Gleichaltrigen))

Scale: 1 = Not True (Trifft gar nicht zu), 2 = Somewhat True (Trifft teilweise zu), 3 = Certainly True (Trifft voll zu)

Conduct Problems

1. I get very angry and often lose my temper (Ich werde leicht wütend; ich verliere oft meine Beherrschung)
2. I usually do as I am told (Normalerweise tue ich, was man mir sagt)
3. I fight a lot (Ich schlage mich häufig mit anderen)
4. I can make other people do what I want (Ich kann andere zwingen zu tun, was ich will)
5. I am often accused of lying or cheating (Andere behaupten oft, dass ich lüge oder mogele)
6. I take things that are not mine from home, school or elsewhere (Ich nehme Dinge, die mir nicht gehören (von zu Hause, in der Schule oder anderswo))

Scale: 1 = Not True (Trifft gar nicht zu), 2 = Somewhat True (Trifft teilweise zu), 3 = Certainly True (Trifft voll zu)

Peer Problems

1. I would rather be alone than with people of my age (Ich bin meistens für mich allein; ich beschäftige mich lieber mit mir selbst)
2. I have one good friend or more (Ich habe einen oder mehrere gute Freunde oder Freundinnen)
3. Other people my age generally like me (Im Allgemeinen bin ich bei Gleichaltrigen beliebt)
4. Other children or young people pick on me or bully me (Ich werde von anderen gehänselt oder schikaniert)

5. I get along better with adults than with people my own age (Ich komme besser mit Erwachsenen aus als mit Gleichaltrigen)

Scale: 1 = Not True (Trifft gar nicht zu), 2 = Somewhat True (Trifft teilweise zu), 3 = Certainly True (Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2014-2020	behav2_full	4342	1.78	0.69	0.55	0.70
2014-2020	behav10_full	4342	1.82	0.75	0.44	0.70
2014-2020	behav16_full	4341	1.82	0.70	0.51	0.70
2014-2020	behav22_fullR	4342	1.72	0.60	0.36	0.70
2014-2020	behav26_fullR	4340	1.74	0.60	0.41	0.70
2014-2020	behav3_full	4344	1.47	0.63	0.27	0.63
2014-2020	behav8_full	4337	1.76	0.70	0.40	0.63
2014-2020	behav14_full	4339	1.31	0.54	0.43	0.63
2014-2020	behav17_full	4318	1.63	0.68	0.37	0.63
2014-2020	behav25_full	4335	1.45	0.62	0.46	0.63
2014-2020	behav1_full	4344	2.67	0.50	0.39	0.64
2014-2020	behav4_full	4341	2.43	0.58	0.37	0.64
2014-2020	behav9_full	4345	2.76	0.47	0.43	0.64
2014-2020	behav18_full	4348	2.78	0.47	0.37	0.64
2014-2020	behav21_full	4341	2.44	0.57	0.40	0.64
2014-2020	behav5_full	4328	1.66	0.69	0.36	0.59
2014-2020	behav7_fullR	4335	1.69	0.56	0.24	0.59
2014-2020	behav12_full	4343	1.16	0.42	0.40	0.59
2014-2020	behav13_full	4325	1.16	0.42	0.26	0.59
2014-2020	behav19_full	4330	1.34	0.58	0.39	0.59
2014-2020	behav23_full	4338	1.13	0.39	0.34	0.59
2014-2020	behav6_full	4343	1.67	0.69	0.24	0.54
2014-2020	behav11_fullR	4349	1.14	0.41	0.33	0.54
2014-2020	behav15_full	4255	2.41	0.60	0.30	0.54
2014-2020	behav20_full	4339	1.24	0.51	0.37	0.54
2014-2020	behav24_full	4312	1.49	0.62	0.31	0.54

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2016-2020	behav2_full	2805	1.71	0.68	0.52	0.67
2016-2020	behav10_full	2793	1.69	0.72	0.39	0.67
2016-2020	behav16_full	2805	1.77	0.68	0.50	0.67
2016-2020	behav22_fullR	2799	1.68	0.58	0.34	0.67
2016-2020	behav26_fullR	2797	1.72	0.60	0.36	0.67
2016-2020	behav3_full	2803	1.46	0.63	0.32	0.67
2016-2020	behav8_full	2797	1.82	0.71	0.44	0.67
2016-2020	behav14_full	2795	1.27	0.52	0.49	0.67
2016-2020	behav17_full	2788	1.65	0.68	0.41	0.67
2016-2020	behav25_full	2803	1.40	0.60	0.50	0.67
2016-2020	behav1_full	2802	2.68	0.51	0.44	0.66
2016-2020	behav4_full	2806	2.40	0.59	0.40	0.66
2016-2020	behav9_full	2800	2.77	0.47	0.46	0.66

continues on next page

Table 48 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2016-2020	behav18_full	2807	2.75	0.48	0.37	0.66
2016-2020	behav21_full	2800	2.41	0.58	0.42	0.66
2016-2020	behav5_full	2797	1.60	0.66	0.33	0.58
2016-2020	behav7_fullR	2802	1.67	0.56	0.21	0.58
2016-2020	behav12_full	2803	1.11	0.38	0.41	0.58
2016-2020	behav13_full	2779	1.23	0.48	0.29	0.58
2016-2020	behav19_full	2797	1.28	0.54	0.35	0.58
2016-2020	behav23_full	2802	1.14	0.40	0.39	0.58
2016-2020	behav6_full	2800	1.80	0.72	0.27	0.53
2016-2020	behav11_fullR	2806	1.14	0.42	0.33	0.53
2016-2020	behav15_full	2777	2.43	0.59	0.29	0.53
2016-2020	behav20_full	2797	1.20	0.46	0.37	0.53
2016-2020	behav24_full	2787	1.54	0.64	0.27	0.53

3.2.24 Supportive Parenting

Summary

Supportive parenting describes a style of childrearing based on warmth, love, constructive communication, and a generally positive way of interacting in the parent-child relationship. The scale has been used in the SOEP youth questionnaires since 2001.

Theoretical Background

Parenting is not understood here as a series of techniques and abilities, but as a special form of relationship between parents and children. The concept is based on the idea that efforts by parents to create a supportive, wholesome environment have a positive impact on the cognitive and social development of the children. A supportive parenting style is part of such a supportive environment (Simons et al., 1992) and can be described as follows: Supportive parents express interest in children's activities, talk to or play games with them, provide help with everyday problems and schoolwork, express enthusiasm and praise over accomplishments, and show affection and love (Amato, 1990).

Scale Development

The scale is based on the Supportive Parenting Scale (SPS) of Simons et al. (1992), and this version was translated into German by Schwarz et al. (1997).

References

Amato, Paul R. 1990. *Dimensions of the family environment as perceived by children: A multidimensional scaling analysis. Journal of marriage and family*, 52, 613-620.

Schwarz, B., Walper, S., Gödde, M., & Jurasic, S. (1997). *Dokumentation der Erhebungsinstrumente der 1. Hauptbefragung. Berichte aus der Arbeitsgruppe "Familienentwicklung nach der Trennung."* University of Munich: Unpublished manuscript.

Simons, R. L., Lorenz, F. O., Conger, R. D., & Wu, C.-I. (1992). *Support from spouse as mediator and moderator of the disruptive influence of economic strain on parenting. Child Development*, 63, 1282-1301.

Items

How often do the following situations occur with your mother and father? How often... (Wie häufig treten in Ihrer Beziehung zu Ihren Eltern die folgenden Situationen auf? Wie häufig kommt es vor...):

1. Do your parents talk to you about things you do or experience (dass Ihre Eltern mit Ihnen über Dinge sprechen, die Sie tun oder erlebt haben)?

2. Do your parents bring up things that bother or worry you (dass Ihre Eltern Dinge ansprechen, die Sie ärgern oder belasten)?
3. Do your parents ask you for your opinion before they decide something that affects you (dass Ihre Eltern nach Ihrer Meinung fragen, bevor sie etwas entscheiden, das Sie betrifft)?
4. Do your parents express their opinion when you do something that they like or approve of (dass, wenn Sie etwas tun, was Ihre Eltern gut finden, Ihre Eltern Ihnen auch zeigen, dass sie sich darüber freuen)?
5. Are you and your parents able to find a solution together to problems you have with each other (dass, wenn Sie und Ihre Eltern ein Problem miteinander haben, Sie dann gemeinsam eine Lösung finden können)?
6. Do your parents give you the impression that they really trust you (dass ihre Eltern Ihnen das Gefühl geben, dass sie Ihnen wirklich vertrauen)?
7. Do your parents ask for your opinion before they make decisions on family matters or issues (dass Ihre Eltern nach Ihrer Meinung fragen, bevor sie über Familienangelegenheiten entscheiden)?
8. Do your parents give you an explanation for their decisions (dass Ihre Eltern Ihnen gegenüber Entscheidungen begründen)?
9. Do your parents show you that they really love you (dass Ihre Eltern Ihnen zeigen, dass sie Sie wirklich lieben)?

Scale: 1 (Very often / Sehr häufig) to 5 (Never / Nie)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2001-2018	jl0040	7831	2.25	0.93	0.51	0.83
2001-2018	jl0043	7821	2.74	1.00	0.17	0.83
2001-2018	jl0044	7812	2.31	1.08	0.53	0.83
2001-2018	jl0046	7820	1.96	0.86	0.64	0.83
2001-2018	jl0048	7775	2.33	1.00	0.60	0.83
2001-2018	jl0050	7815	1.82	0.91	0.62	0.83
2001-2018	jl0052	7778	2.48	1.05	0.61	0.83
2001-2018	jl0054	7796	2.39	1.03	0.60	0.83
2001-2018	jl0056	7808	1.71	0.87	0.61	0.83
2001-2018	jl0041	7280	2.78	1.11	0.69	0.91
2001-2018	jl1043	7270	3.14	1.08	0.40	0.91
2001-2018	jl0045	7252	2.63	1.26	0.69	0.91
2001-2018	jl0047	7271	2.30	1.10	0.78	0.91
2001-2018	jl0049	7216	2.67	1.17	0.73	0.91
2001-2018	jl0051	7248	2.16	1.15	0.77	0.91
2001-2018	jl0053	7194	2.87	1.20	0.71	0.91
2001-2018	jl0055	7232	2.71	1.19	0.73	0.91
2001-2018	jl0057	7258	2.13	1.17	0.74	0.91

3.2.25 Temperament

Summary

In the SOEP since 2003, two dimensions of temperament in early childhood, “affect” and “activity,” have been measured with five items assessed by one of the parents for children below the age of 2 and for those aged 2-3.

Theoretical Background

Temperament describes the way a person acts and reacts and is comprised of emotions, motor, attention-related reactions, and self-regulation. The concept of the “early childhood temperament” is used in developmental psychopathology and in the research and theory on child and youth psychiatry. Temperament characteristics are treated here as risk factors for the emergence of behavioral disorders, as characteristics influencing parental interactive behavior, and as characteristics that show heightened vulnerability to unfavorable environmental conditions.

Scale Development

Based on an expert report by Pauen & Vonderlin (2007), the Infant Behavior Questionnaire (IBQ) of Rothbart (1981; German version by Pauli-Pott, Mertesacker, & Beckmann, 2003) was used as a point of orientation for the measurement of early child temperament, but shortened significantly.

References

Pauen, S., & Vonderlin, E. (2007). *Entwicklungsdiagnostik in den ersten drei Lebensjahren Empfehlungen zum Ausbau des Erhebungsinstrumentariums über Kinder im Sozio- oekonomischen Panel (SOEP)*. Berlin: DIW Berlin.

Pauli-Pott, U., Mertesacker, B., & Beckmann, D. (2003). *Ein Fragebogen zur Erfassung des frühkindlichen Temperaments im Elternurteil*. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 31, 99-110.

Rothbart, M. K. (1981). *Measurement of temperament in infancy*. *Child Development*, 52, 569-578.

Items

How do you see your child today (Wie sehen Sie Ihr Kind heute):

Affect

1. My child is generally happy and satisfied (Mein Kind ist meist fröhlich und zufrieden)
2. My child is easily irritated and cries frequently (Mein Kind ist leicht erregbar und weint häufig)
3. My child is difficult to console (Mein Kind ist schwer zu trösten)

Scale: 1 (Applies fully / Trifft voll zu) to 4 (Does not apply at all / Trifft gar nicht zu)

Activity

1. My child is curious and active (Mein Kind ist neugierig und aktiv)
2. (since 2007) My child is more reserved (Mein Kind ist eher zurückhaltend)
3. My child is communicative and likes to talk (Mein Kind ist gesprächig und redet gern)

Scale: 1 (Applies fully / Trifft voll zu) to 4 (Does not apply at all / Trifft gar nicht zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2003-2020	temp1	7155	1.23	0.47	0.36	0.62
2003-2020	temp2R	7130	1.67	0.72	0.52	0.62
2003-2020	temp3R	7134	1.46	0.66	0.45	0.62
2003-2020	temp4	7080	1.25	0.52		
2007-2020	temp5R	5866	1.69	0.70		

continues on next page

Table 50 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2005-2020	temp1	7576	1.26	0.48	0.31	0.54
2005-2020	temp2R	7549	1.94	0.76	0.41	0.54
2005-2020	temp3R	7543	1.60	0.73	0.38	0.54
2005-2020	temp4	7563	1.19	0.45		
2005-2020	temp6	7541	1.37	0.67		

3.2.26 Tendency to Forgive

Summary

The tendency to forgive is an individual disposition to forgive other people after being hurt or treated unjustly by them and was measured in SOEP in 2010, 2015, and 2020.

Theoretical Background

The individual tendency to forgive is the tendency to either maintain or overcome one's own anger or resentment after being personally injured: "the TTF is a simple, four-item scale designed to assess the extent to which individuals typically experience or engage in forgiveness when they have been wronged by others" (Brown, 2003). The scale is not designed to measure the process of forgiving (how or when one forgives), but to measure individual differences in the outcome of this process (to what extent an individual typically forgives). The tendency to forgive is understood as an individual disposition and not as a process linked to specific situations or actions. As a construct, the tendency to forgive shows similarities to other constructs such as revenge, feelings of aggression, desire for reconciliation, and negative reciprocity. For the research, it is of interest to study the consequences the individual tendency to forgive on mental health and life satisfaction. Furthermore, the individual tendency to forgive may play a role as a strategy in decision conflicts and dilemma situations.

Scale Development

The scale is the German translation of the already validated Tendency to Forgive Scale. Information on the validity of the original scale can be found in Brown and Phillips (2005). The scale was used for the first time in the SOEP in the 2009 SOEP pretest (representative random sample of the German population, N = 1.007) and displays, given the typical characteristics of population surveys (low number of items, high heterogeneity of survey respondents), satisfactory reliability and factorial validity (Weinhardt & Schupp, 2011). Based on the satisfactory results of the 2009 SOEP pretest, the Tendency to Forgive scale was included in the main SOEP survey in 2010.

References

Brown, R., & Phillips, A. (2005). *Letting bygones be bygones: Further evidence for the validity of the tendency to forgive scale. Personality and Individual Differences*, 38, 627-638.

Brown, R. P. (2003). *Measuring individual differences in the tendency to forgive: Construct validity and links with depression. Personality and Social Psychology Bulletin*, 29, 759-771.

Weinhardt, M., & Schupp, J. (2011). *Die Messung individueller Vergebungstendenz im SOEP: Skaleneigenschaften der deutschen Version der Tendency-to-Forgive-Scale. Berlin: DIW Berlin.*

Items To what degree do the following statements apply to you personally (In welchem Maße treffen die folgenden Aussagen auf Sie persönlich zu):

1. I get over it relatively quickly when someone hurts my feelings (Ich komme relativ leicht darüber hinweg, wenn jemand mich emotional verletzt).

2. When somebody has wronged me, I often think about it for quite a while (R) (Wenn jemand mir unrecht getan hat, denke ich oft lange darüber nach).
3. I tend to bear grudges (R) (Ich bin tendenziell nachtragend).
4. When other people wrong me I try to just forgive and forget (Wenn andere mir unrecht tun, versuche ich einfach zu vergeben und zu vergessen).

Scale: 1 (Does not apply to me at all / Trifft überhaupt nicht zu) to 7 (Applies to me perfectly / Trifft voll zu)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2010	plh0206i07	18799	3.62	1.63	0.34	0.57
2010	plh0206i08R	18805	3.26	1.68	0.33	0.57
2010	plh0206i09R	18820	4.79	1.67	0.39	0.57
2010	plh0206i10	18813	4.16	1.54	0.35	0.57
2015	plh0206i07	26916	3.59	1.68	0.33	0.57
2015	plh0206i08R	26951	3.29	1.75	0.34	0.57
2015	plh0206i09R	26897	4.81	1.74	0.40	0.57
2015	plh0206i10	26928	4.27	1.58	0.33	0.57
2016	plh0206i07	118	3.49	1.96	0.34	0.53
2016	plh0206i08R	118	3.36	1.90	0.41	0.53
2016	plh0206i09R	118	4.56	1.91	0.39	0.53
2016	plh0206i10	118	4.60	1.70	0.15	0.53
2020	plh0206i07	28997	3.67	1.68	0.34	0.58
2020	plh0206i08R	29041	3.49	1.77	0.35	0.58
2020	plh0206i09R	28888	5.02	1.72	0.41	0.58
2020	plh0206i10	29014	4.39	1.61	0.34	0.58

3.2.27 Trust, Trustworthiness, Fairness

Summary

Trust, trustworthiness, and fairness were measured in the SOEP at five-year intervals since 2003, using five items. Three of the items can be combined into a single scale for measuring generalized trust.

Theoretical Background

In sociology and political science, generalized trust describes the average expectation of trustworthiness in strangers (Stolle, 2002). Trust is interpreted as a measure of cooperative relationships among citizens that allows people to interact in a complex social environment and to reduce their transaction costs (Luhmann, 2000). For sociologist James Coleman, two points are characteristic for the concept of trust: on the one hand, trust means that the person trusting entrusts specific goods to another person without being able to directly control or sanction the actions of the other person. On the other hand, there must be a potential benefit as an incentive: the person must gain some advantage from having trusted another person if the latter proves to be trustworthy. In this conception, trust is less a personality characteristic than a specific behavior. Generalized trust (trust in strangers) also differs from trust in institutions and from trust in acquaintances. In social psychology, the concept is also important: “Interpersonal trust is defined here as an expectancy held by an individual or a group that the word, promise, verbal or written statement of another individual or group can be relied upon” (Rotter, 1967).

Scale Development

The scale used to measure trust was based on the scales used in the General Social Survey (GSS) and in the World Values Survey (WVS), and an additional (Item 2) was added. The scale was first piloted in the 2002 SOEP pretest and used in 2003 in the main SOEP survey. The scale was further validated in the 2003, 2004, 2005, and 2006 SOEP

pretests (through parallel surveying of the personality dimension “agreeableness: trust” from the NEO-PI-R and testing of sequence effects; Naef & Schupp, 2009). Further information can be found in Dohmen et al. (2008).

References

- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). *Representative trust and reciprocity: Prevalence and determinants*. *Economic Inquiry*, 46, 81-90.
- Naef, M., & Schupp, J. (2009). *Measuring Trust: Experiments and Surveys in Contrast and Combination*. Berlin: DIW Berlin.
- Luhmann, N. (2000). *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*. Stuttgart: Lucius & Lucius.
- Rotter, J. B. (1967). *A new scale for the measurement of interpersonal trust*. *Journal of Personality*, 35, 651-665.
- Stolle, D. (2002). *Trusting Strangers - Generalized trust in perspective*. *Österreichische Zeitschrift für Politikwissenschaft*, 31, 397-412.

Items

What is your opinion on the following three statements (Was ist Ihre Meinung zu den folgenden drei Aussagen):

1. On the whole, one can trust people (Im Allgemeinen kann man den Menschen vertrauen)
2. Nowadays one can't depend on anyone (Heutzutage kann man sich auf niemanden mehr verlassen)
3. When dealing with strangers, it is better to be cautious before trusting them (Wenn man mit fremden zu tun hat, ist es besser, vorsichtig zu sein, bevor man ihnen vertraut)

Scale: 1 (Agree completely/ Stimme voll zu) to 4 (Disagree completely/ Lehne voll ab)

Do you believe that most people ... (Glauben Sie, dass die meisten Leute...):

4. Would exploit you if they had the opportunity (Sie ausnützen würden, falls sie eine Möglichkeit dazu hätten) or would try to treat you fairly (oder versuchen würde, Ihnen gegenüber fair zu sein)?

Scale: 1 (Would exploit you if they had the opportunity / Sie ausnützen würden, falls sie eine Möglichkeit dazu hätten) or 2 (Or would try to treat you fairly / oder versuchen würde, Ihnen gegenüber fair zu sein?)

Would you say that the most of the time, people ... (Würden Sie sagen, dass die Leute die meiste Zeit...)

5. Attempt to be helpful? (versuchen, hilfsbereit zu sein?) Or only act in their own interests? (oder nur ihre eigenen Interessen verfolgen?)

Scale: 1 (Attempt to be helpful / versuchen, hilfsbereit zu sein?) or 2 (Or only act in their own interests / oder nur ihre eigenen Interessen verfolgen?)

Test-Retest Correlations

The three trust items (1-3) which form a scale were included in a retest of a subsample (N = 161) in 2005 within 30 to 49 days after the respective initial tests. Test-retest correlation of the items were (in scale order) .41, .42, and .28; scale scores had a test-retest correlation of .46.

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2003	plh0192R	22489	2.63	0.68	0.45	0.61
2003	plh0193	22463	2.60	0.77	0.50	0.61
2003	plh0194	22501	1.69	0.72	0.33	0.61
2003	plh0195	22218	1.53	0.50		
2003	plh0196	22269	1.64	0.48		
2008	plh0192R	19623	2.62	0.67	0.44	0.60
2008	plh0193	19597	2.63	0.77	0.49	0.60

continues on next page

Table 53 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2008	plh0194	19611	1.72	0.72	0.31	0.60
2008	plh0195	19347	1.54	0.50		
2008	plh0196	19391	1.62	0.49		
2013	plh0192R	25723	2.71	0.65	0.47	0.62
2013	plh0193	25698	2.68	0.76	0.51	0.62
2013	plh0194	25701	1.72	0.71	0.33	0.62
2013	plh0195	25210	1.59	0.49		
2013	plh0196	25316	1.56	0.50		
2018	plh0192R	29678	2.73	0.70	0.41	0.60
2018	plh0193	29610	2.66	0.81	0.47	0.60
2018	plh0194	29668	1.72	0.75	0.36	0.60
2018	plh0195	28820	1.61	0.49		
2018	plh0196	29066	1.53	0.50		

3.2.28 Vineland Adaptive Behavior Scales

Summary

The 20-item Vineland Adaptive Behavior Scale has been used in the SOEP since 2005 to assess parents' ratings of their 2-3-year-old children in the areas of language, everyday skills, motor skills, and social relationships.

Theoretical Background

Especially at an age when individual functions such as linguistic, mathematic, and social competencies are still relatively undifferentiated, adaptive behavior offers a true-to-life measure of child development in an everyday life context that is significant for the child. Measuring this adaptive behavior also does not require testing the child but rather takes place in the framework of a reporting procedure in which a person who is familiar with the child's adaptive behavior in everyday life provides specific descriptions of the child's behavior.

Scale Development

The scale is based on a German adaptation (Tietze et al. 1998) of the Vineland Adaptive Behavior Scale by Sparrow, Balla und Cicchetti (1984). For use in the SOEP survey framework, the scale was reduced to 20 items. Further information can be found in Schmiade, Spieß, & Tietze (2008).

References

- Schmiade, N., Spieß, C. K., & Tietze, W. (2008). *Zur Erhebung des adaptiven Verhaltens von zwei- und dreijährigen Kindern im Sozio-oekonomischen Panel (SOEP)*. Berlin: DIW Berlin.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scale. A revision of the Vineland Social Maturity Scale by Edbar A. Doll. Survey Form Manual*. Circle Pines, Minnesota: American Guidance Service.
- Tietze, W., Meischner, T., Gänsfuß, R., Grenner, K., Schuster, K.-M., Völkel, P., & Roßbach, H.-G. (1998). *Wie gut sind unsere Kindergärten? Eine Untersuchung zur pädagogischen Qualität in deutschen Kindergärten*. Neuwied: Luchterhand.

Items

For parents, it is always a big event when their child learns something new. Please tell us what new things your child can do (Für Eltern ist es immer ein großes Erlebnis, wenn das Kind schon wieder etwas Neues kann. Bitte geben Sie an, welche Dinge das bei Ihrem Kind sind):

Talking (Sprechen)

1. Understands brief instructions such as “go get your shoes” (Versteht kurze Anweisungen, wie z.B. „hole deine Schuhe”)
2. Forms sentences with at least two words (Bildet Sätze mit mindestens zwei Wörtern)
3. Speaks in full sentences (with four or more words) (Spricht in ganzen Sätzen (mit vier und mehr Wörtern))
4. Listens attentively to a story for five minutes or longer (Hört einer Geschichte 5 Minuten oder länger aufmerksam zu sein)
5. Passes on simple messages such as “dinner is ready” (Überbringt einfache Nachrichten wie z.B. „Essen ist fertig”)
6. (since 2010) Follows instructions (heard five minutes before) (Folgt Anweisungen, die es fünf Minuten zuvor gehört hat)
7. (since 2010) Tells his/her first and lastname when asked for it (Nennt seinen Vor- und Nachnamen, wenn es danach gefragt wird)
8. (since 2010) Listens attentively to stories for at least 15 Minutes (Hört in der Regel Geschichten mindestens 15 Min. aufmerksam zu)

Scale: 1 = Yes (ja), 2 = To some extent (teilweise), 3 = No (nein)

Everyday skills (Alltagsfertigkeiten)

1. Uses a spoon to eat without assistance and without dripping (Isst selbstständig mit dem Löffel, ohne zu kleckern)
2. Blows his/her nose without assistance (Putzt sich selbst die Nase)
3. Uses the toilet to do “number two” (Benutzt für „großes Geschäft” die Toilette)
4. Puts on pants and underpants the right way around (Zieht sich Hosen und Unterhosen selbst richtig herum an)
5. Brushes his/her teeth without assistance (Putzt sich selbst die Zähne)

Scale: 1 = Yes (ja), 2 = To some extent (teilweise), 3 = No (nein)

Movement (Bewegung):

1. Walks forward down the stairs (Läuft Treppen vorwärts herunter)
2. Opens doors with the door handle (Öffnet Türen mit Türklinke)
3. Climbs up playground climbing equipment and other high playground structures (Klettert auf Klettergerüste und andere hohe Spielgeräte)
4. Cuts paper with scissors (Schneidet mit einer Schere Papier durch)
5. Paints/draws recognizable shapes on paper (Malt auf Papier erkennbare Formen)
6. (since 2010) Holds pens in the right way (not fist grip) when drawing (Hält Stifte richtig (nicht mit der Faust), um zu malen)

Scale: 1 = Yes (ja), 2 = To some extent (teilweise), 3 = No (nein)

Social relationships (Soziale Beziehungen):

1. Calls familiar people by name; for example, says “mommy” and “daddy” or uses the father’s first name (Nennt vertraute Personen beim Namen; sagt z.B. „Mama.” „Papa” oder verwendet den Vornamen des Vaters)
2. Participates in games with other children (Nimmt an Spielen mit anderen Kindern teil)
3. Gets involved in role-playing games (“playing pretend”) (Beschäftigt sich mit Rollenspielen („tun als ob”))
4. Shows a special liking for particular playmates or friends (Zeigt eine Vorliebe für bestimmte Spielgefährten oder Freunde)

5. Calls his/her own feelings by name, for example, “sad,” “happy,” “scared” (Benennt eigene Gefühle, z. B. „traurig,” „freuen,” „Angst”)
6. (since 2010) Takes turns with other children when playing without being asked to do so (Wechselt sich mit anderen beim Spielen ab, ohne darum gebeten zu werden)

Scale: 1 = Yes (ja), 2 = To some extent (teilweise), 3 = No (nein)

Items and Scale Statistics

year	variable	count	mean	sd	itemrestcorr	alpha
2005-2013	spch1	6337	1.15	0.40	0.26	0.67
2005-2013	spch2	4162	1.08	0.35	0.54	0.67
2005-2013	spch3	4535	1.33	0.61	0.60	0.67
2005-2013	spch4	3672	1.56	0.68	0.31	0.67
2005-2013	spch5	4526	1.15	0.45	0.57	0.67
2010-2020	spch3	6947	1.38	0.64	0.57	0.70
2010-2020	spch5	6931	1.18	0.49	0.50	0.70
2010-2020	spch6	6553	1.56	0.62	0.37	0.70
2010-2020	spch7	6575	1.67	0.78	0.49	0.70
2010-2020	spch8	6565	1.70	0.72	0.38	0.70

year	variable	count	mean	sd	itemrestcorr	alpha
2005-2020	skll1	10268	1.46	0.60	0.35	0.67
2005-2020	skll2	10261	1.71	0.74	0.42	0.67
2005-2020	skll3	10257	1.99	0.90	0.42	0.67
2005-2020	skll4	8097	1.82	0.76	0.55	0.67
2005-2020	skll5	8099	1.62	0.69	0.37	0.67

year	variable	count	mean	sd	itemrestcorr	alpha
2005-2013	mvmn1	4539	1.07	0.31	0.30	0.52
2005-2013	mvmn2	6336	1.17	0.52	0.28	0.52
2005-2013	mvmn3	4536	1.20	0.48	0.33	0.52
2005-2013	mvmn4	4523	1.50	0.75	0.41	0.52
2005-2013	mvmn5	4533	1.81	0.75	0.29	0.52
2010-2020	mvmn1	6953	1.09	0.35	0.28	0.57
2010-2020	mvmn3	6944	1.24	0.53	0.31	0.57
2010-2020	mvmn4	6918	1.59	0.80	0.40	0.57
2010-2020	mvmn5	6945	1.79	0.76	0.36	0.57
2010-2020	mvmn6	8736	1.61	0.76	0.35	0.57

year	variable	count	mean	sd	itemrestcorr	alpha
2005-2013	sclr1	4173	1.02	0.16	0.24	0.51
2005-2013	sclr2	6701	1.40	0.59	0.20	0.51
2005-2013	sclr3	4538	1.24	0.52	0.25	0.51
2005-2013	sclr4	4528	1.33	0.60	0.40	0.51
2005-2013	sclr5	4532	1.28	0.55	0.40	0.51

continues on next page

Table 57 – continued from previous page

year	variable	count	mean	sd	itemrestcorr	alpha
2010-2020	sclr2	9116	1.34	0.55	0.17	0.57
2010-2020	sclr3	6927	1.35	0.64	0.30	0.57
2010-2020	sclr4	6909	1.36	0.63	0.42	0.57
2010-2020	sclr5	6937	1.31	0.59	0.40	0.57
2010-2020	sclr6	6527	1.58	0.69	0.37	0.57

Last change: Jul 24, 2023

3.3 Survey Concepts and Modes

Measuring stability and detecting changes means repeating (almost) identical measures over time. Furthermore, the SOEP questions capture stability and change by varying with regard to the time dimension, that is, asking about events in the past, the present, and the future. Conceptually, different measurements of time are used:

- Questions about a point in time (present), e.g., current employment status or current levels of satisfaction
- Retrospective questions about certain events in the past, e.g., how often have you changed jobs in the last ten years?
- Retrospective life event history since the age of 15 (in the past), e.g., employment or marital history
- Monthly calendar information on income and labor market participation (in the past), e.g., employment status January through December of last year
- Questions about a period of time (in the past), e.g., demographic changes since the last interview such as marriage or death of spouse
- Questions about the future, e.g., expected satisfaction with life five years from now, or job expectations

Survey Modes

The SOEP uses several different modes to collect the data. Originally, the respondent's answers were always recorded by an interviewer who filled in the answers in a paper questionnaire, the "pen-and-paper interview" or PAPI. The personal contact between interviewer and respondent is important for the success of the survey; however, before losing a respondent due to a scheduling conflict between interviewer and respondent, the SOEP has allowed respondents to mail in the questionnaire since the second wave of subsamples A-I. This is not the same as the concept of a regular mail survey, because the interviewer still maintains personal contact with the household and schedules appointments with respondents if possible. Starting with subsample J, only "computer-assisted personal interviews" (CAPI) are allowed, and thus it is no longer possible to mail in the questionnaires.

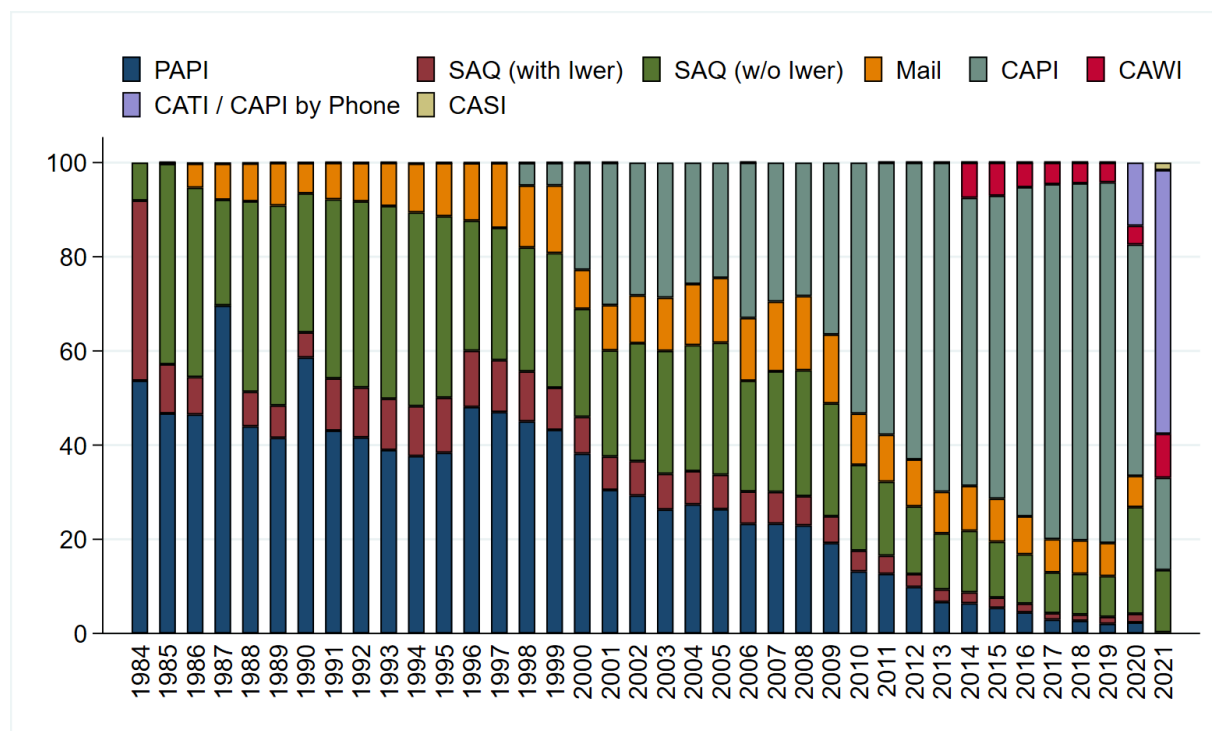
When visiting a household, the interviewer interviews household members one at a time and can also give questionnaires to other household members to complete without the interviewer's assistance (self-administered questionnaires, SAQ). This is a time-efficient approach because it allows different household members to complete their questionnaires at the same time.

In 1998, computers were used for the first time in the SOEP for computer-assisted personal interviews (CAPI). Compared to PAPI, the CAPI mode is much more efficient in converting the data into an electronic format, which was an important asset especially with the extensions to the panel starting in the year 2000. The CAPI mode was first used parallel with PAPI, meaning that interviewers and respondents were free to choose how they wanted to do the interview. This was important for the "older" sample members (respondents as well as interviewers), who were used to the PAPI concept. Only in the most recent samples (starting in subsample J) is CAPI the sole interview mode. The figure depicts the development of modes up to 2011, showing that the CAPI mode has gained importance since its implementation.

Since the questionnaires have to be identical in both modes, CAPI is implemented in a relatively simple way in the SOEP and does not utilize all the technical possibilities of this interview mode. For example, the SOEP basically does

not use any form of dependent interviewing (i.e., referring to respondent data from previous waves), because this cannot be easily implemented in the PAPI mode. Also, the filtering structure is very simple in the SOEP, because a respondent must be able to follow the interview path on paper on her/his own. Still, some technical features like the control of value ranges (e.g. month of birth, year of first marriage) or the randomization of scale items are implemented in the CAPI version of the questionnaire.

In the future, new modes will be introduced into the SOEP as they develop. The computer-assisted web interview (CAWI) is close to implementation, but will not be used as a replacement of the current CAPI and PAPI modes, but rather as an extension the respondents may use, similar to the mail-in or self-administered questionnaires. The core interview concept of the SOEP survey, the personal contact between respondent and interviewer, will not change.



Download STATA Code to create figure.

Last change: Jun 30, 2023

3.4 Panel Care

To cope with panel attrition and to keep longitudinal response rates high, the SOEP has implemented “panel care” efforts to maintain personal contact between respondents and the survey. Panel care can be divided into incentives given directly to the respondent and other measures undertaken to keep the respondent in the study.

Respondents have been given gifts as tokens of appreciation since the very beginning of the study. Most of these gifts are small in-kind incentives like flowers, for which the interviewers have their own budget. In addition, the interviewers are asked to hand out a brochure with recent results from the study. Up to 2007, respondents also received a lottery ticket as a thank-you upon completion of their interview. Proceeds from the lottery benefit social projects in Germany. Since 2008, the lottery ticket has been included with the contact letter that is sent out about two weeks prior to the interview. It is thus given unconditionally, as long as the person participated in the previous wave. After a successful interview, the respondent receives a thank-you letter from survey institute along with one postage stamp as a small additional gift.

In 2009, different incentive schemes were tested in the new subsample I to increase the first-wave response rates. The basic experiment included four randomized groups of households: (1) those with the default setup of the conditional lottery ticket; (2) those with a “low” cash incentive of 5 euros per household and 5 euros per adult respondent; (3) those with a “high” cash incentive of 5 euros per household and 10 euros per adult respondent; and (4) those with a choice between a “low” cash incentive and a lottery ticket. The results showed slightly higher response rates in the cash groups, although the extra money in group (3) did not pay off. The current incentive strategy for the different SOEP samples is shown here:

Sam- ples	A-H	J,K, L1, N, O, Q	L2, L3	P	M1/M2	M3-M6	M7/8
In- cen- tives for adult:	Lottery ticket (5,427 households and cash (636 households)	5 euros (households) and 10 euros (adults)	5 euros (households), 5 euros (adults), 10 euros (bonus payment)	Lot- tery Ticke	5 euros (households) and 10 euros (adults)	15 euros cash or voucher, if one household and one individual interview was given	5 eu- ros (house- holds) and 10 eu- ros (adults)
In- cen- tives for other re- spon- dents	Power bank (<i>Youth Questionnaire</i>), Bicycle repair kit (<i>Early Youth Questionnaire</i>), Small clock (<i>Pupils Questionnaire</i> CAPI/PAPI), Puzzle (<i>Pupils Questionnaire</i> MAIL)	Power bank (<i>Youth Questionnaire</i>), Bicycle repair kit (<i>Early Youth Questionnaire</i>), Small clock (<i>Pupils Questionnaire</i> CAPI/PAPI), Puzzle (<i>Pupils Questionnaire</i> MAIL)	5 euros (<i>Youth Questionnaire</i>), (<i>Youth Questionnaire</i> , <i>Early Youth Questionnaire</i> , <i>Pupils Questionnaire</i>) and 5 euros (<i>Mother-Child Instruments</i>)		Power bank (<i>Youth Questionnaire</i>), USB-Stick (<i>Early Youth Questionnaire</i>), Bicycle repair kit (<i>Pupils Questionnaire</i>)		

The survey institute also does additional work to keep response rates high. Addresses are checked throughout the year to ensure that current addresses are on file. This is done, for instance, by sending out brochures about recent research based on the SOEP data and seasonal greeting cards.

Face-to-face interviews also ensure a personal relationship between interviewer and respondent, which increases the likelihood that respondents will stay in the survey. Keeping the same interviewer over time is therefore an important goal of the survey. Some SOEP respondents have in fact had the same interviewer since the beginning in 1984.

Last change: Jul 24, 2023

TARGET POPULATION AND SAMPLES

The target population covered in the SOEP is defined as the population of private households residing within the current boundaries of the Federal Republic of Germany (FRG). Because of changes in these boundaries (in 1990) and changes in the population due to migration, various adaptations have been made to the initial sampling structure to maintain the sample's representativity. In addition, certain groups have been oversampled to increase the statistical power.

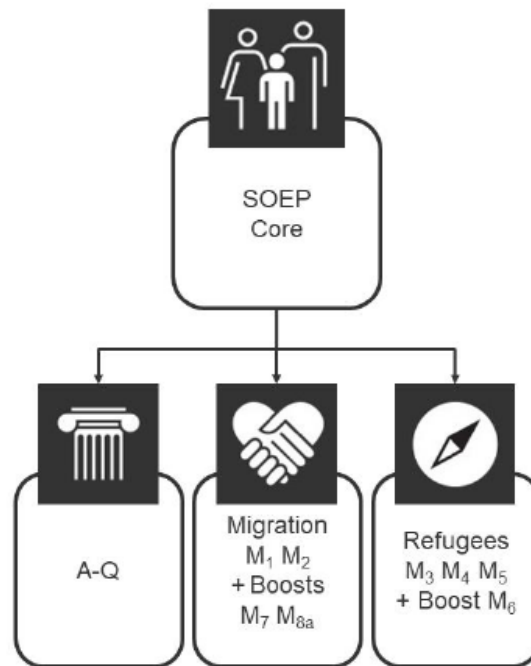
The different SOEP-Core subsamples constitute the centerpiece of the SOEP.

1. Within SOEP-Core, **samples A-Q** form the heart of the SOEP. They contain the oldest samples, beginning with the founding sample in A from 1984 and the highest number of participating households. Fieldwork traditionally starts at the beginning of February, and its questionnaires serve as a master for the other SOEP-Core subsamples.
2. The SOEP migration sample with its **samples M1 and M2** was established in 2013 and is designed to improve the representation of migrants living in Germany. Fieldwork started in April, using the questionnaires from samples A-Q, supplemented by translated questionnaires for five different languages.
3. In order to map recent migration and integration dynamics, SOEP refugee **samples M3 to M5** were installed beginning in the year 2016. In 2020, fieldwork began in August with a questionnaire that was tailored to issues of recent refugees while containing many questions from the SOEP samples A-Q as well.
4. **Sample M6** – a boost sample of refugees targeted the same population as the older refugee sample M5 - adult refugees who have applied for asylum in Germany since 1 January 2013 and are currently living in Germany – and the same sample design and sample frame were used.
5. The two boost samples, **samples M7 and M8a**, were added the SOEP migration sample system. Like the older migration samples M1 and M2, the Integrated Employment Biographies Sample (IEBS) of the Federal Employment Agency (BA) served as the sampling frame for both boost samples. Boost sample M7's goal was to capture migration dynamics and processes from 2016 to 2018 with a focus on EU migration. To ensure that statistically significant group comparisons can be made, sampling was restricted to the three most significant countries of origin in that time period: Romania, Bulgaria, and Poland. M8a, on the other hand, was designed to help evaluate the skilled worker immigration law (Fachkräfteeinwanderungsgesetz), which came into effect March 1, 2020, and targeted migrants from third countries that came to Germany between 2017 and 2018, sampling them as a control group for a treatment group that will be sampled at a later date.

In 1984, the survey started with a sample covering the entire population of then West Germany (FRG), where the five biggest groups of foreigners ("guest workers") were oversampled.

The SOEP was expanded to the territory of the German Democratic Republic in June 1990, only six months after the fall of the Berlin Wall. In 1994/95, a boost sample of migrants who came to Germany after 1984 was added to take the influx of ethnic Germans from former Soviet countries into account. In 2013 another sample of migrants which includes individuals who immigrated to Germany after 1995 or second-generation immigrants was added. Since then, multiple migration or refugees samples were added in cooperation with the IAB (Institut für Arbeitsmarkt- und Berufsforschung) or the BAMF (Bundesamt für Migration und Flüchtlinge)

Now and then samples that were representative of the entire population in Germany were added to counter effects of panel attrition and to increase the overall sample size.



The different samples in the SOEP are identified by letters: sample “A” refers to the German sample drawn in 1984, “C” to the East Germans from 1990, and so on. Even though these samples are kept separate, the respondents have received identical questionnaires for the most part, and distinctions by sample are usually not necessary in an analysis.

However, one of the ideas of the SOEP is that the users have full information available about survey methodological issues and survey design, which in this case means that you can identify the corresponding sample for each observation. In the following section, we present details on each of the samples, which unless stated otherwise are multi-stage random samples with regional clusters. The households are selected by random-walk routines.

For an extensive discussion on sampling (and weighting), see: [Survey methods](#).

4.1 The SOEP Samples in Detail

Sample A “Residents of the Federal Republic of Germany” covers individuals in private households with a household head who does not belong to one of the main groups of “guest workers” (i.e., Turkish, Greek, Yugoslavian, Spanish, or Italian households). Because only a few foreigners are in Sample A, it is often called the “West German Sample” of the SOEP. In 1984 it covered 4,528 households with a sampling probability of about 0.0002.

Sample B “Foreigners in the Federal Republic of Germany” adds individuals in private households with a Turkish, Greek, Yugoslavian, Spanish, or Italian household head, who in 1984 constituted the main groups of foreigners in the FRG. Compared to Sample A, the population of Sample B is oversampled with a sampling probability of about 0.002. In the first wave, Sample B included 1,393 households.

Sample C “German Residents of the German Democratic Republic (GDR)” consists of individuals in private households in which the household head was a citizen of the German Democratic Republic (GDR). This meant that approximately 1.7% of the residential population of the GDR in June 1990 was excluded from the sample as foreigners (most of whom were living in “institutionalized” housing). In total, the sample started with 2,179 households with a sampling probability of about 0.0005.

Sample D “Immigrants” started in 1994/95 with two different samples. In 1994, the first sample, D1, had 236 households and in 1995, the second sample, D2, had 295 households, leading to a total of 531 households (D1 and D2) in

1995. This sample consisted of households in which at least one household member had moved from abroad to West Germany after 1984. The sampling probability is about 0.0002.

Sample E “Refresher” was added in 1998, selected from the entire population of private households in Germany. The households were chosen independently of the ongoing panel and its subsamples A through D. The aim was to increase the number of observations of the general population and to preserve its representativity. The selection scheme used for sample E essentially resembles the one used in subsample A. The number of households in the first wave of subsample E was \$1,060\$, with a sampling probability of about 0.00005. With the 2012 data release, parts of subsample E were extracted into the SOEP Innovation Sample. It is also the first sample in which Computer-Assisted Personal Interview (CAPI) was used. At that time, interviews in Samples A-D were being conducted entirely using Paper-and-Pencil-Interviews (PAPI). To study mode effects, households from sample E were randomly allocated to either CAPI or PAPI.

Sample F “Refresher” was selected independently of all other subsamples from the population of private households in 2000. The selection scheme was slightly altered compared to the previous addition in Sample E: while the “German” households (all adults aged 16 or older in the household have German nationality) were selected with a sampling probability of \$0.00028\$, the ‘non-German’ households (at least one adult does not have German nationality) were oversampled with a probability of 0.0005. Overall, the number of added households in subsample F’s first wave amounts to 6,043.

Sample G “High-Income” entered the SOEP in 2002 independently from all other subsamples. The original selection scheme required that the responding households had a monthly income of at least DM 7,500 (EUR 3,835), which - due to the lack of an adequate sampling frame - were identified using a screening procedure. This sample of a total of 1,224 households increased the potential for analysis in the high-income bracket, which was previously difficult to study because of the low case numbers. The derived sampling probability is about 0.0014. Starting with Wave 2 in 2003, the selection scheme for this subsample was changed such that only households with a net monthly income of at least EUR 4,500 were followed.

Sample H “Refresher” started in 2006 as a random sample, again independently of all previous subsamples, covering all residential households in Germany. The added 1,506 households were sampled with a probability of 0.0001.

Sample I “Incentive Sample” started in 2009, where in the first wave, a new incentive scheme was tested to increase participation rates (see also [sec:PanelCare]). The sampling was independent of all other SOEP samples, adding a total number of 1,531 households to the SOEP. The sampling probability was 0.00013. This sample remained in the main data release for its first two waves (2010 and 2011, or waves Z and BA). With the 2012 data release, subsample I was extracted into the SOEP Innovation Sample.

Sample J “Refresher Sample” started in 2011 as a random sample, independently of all previous subsamples, covering residential households in Germany. The added 3,136 households were sampled with a probability of 0.0002.

Sample K “Refresher Sample” started in 2012 as a random sample, drawn independently of all previous subsamples, covering the residential households in Germany. The added 1,526 households were sampled with a probability of 0.0001.

Sample L1 “Cohort Sample” covers private households in Germany in which at least one household member was born between January 2007 and March 2010 and was therefore a child at that time. Again, migrants identified were oversampled using an onomastic procedure. Sample L1 (as well as L2 and L3) was part of the SOEP-related study “Families in Germany” (FiD), which was integrated into the SOEP in 2014. As part of an evaluation project by the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ) and the Federal Ministry of Finance (BMF), the study focused on public benefits in Germany for married people and families. Therefore, the survey instruments used in waves BA to BD differ in some respects from those used in the other samples.

Sample L2 “Family Types I” covers private households in Germany that meet at least one of the following criteria for household composition: single parents, low-income families, and large families with three or more children. Similar to Sample G, we face the problem that the eligible sub-population is relatively small and an adequate sampling frame is lacking. So again, a preceding telephone screening procedure identifies eligible households.

Sample L3 “Family Types II” covers private households in Germany that meet at least one of the following criteria for household composition: single parents or large families with three or more children. It is conducted analogously to Sample L2 to increase the number of cases in these sub-populations.

Sample M1 “Migration Sample” is a new migration sample added in 2013 with around 2,700 households drawn using register information from the German Federal Employment Agency. It includes individuals who immigrated to Germany after 1995 or second-generation immigrants.

Sample M2 “Migration Sample” was another migration sample added in 2015 with around 1,100 households drawn using register information from the German Federal Employment Agency. It includes individuals who immigrated to Germany between 2010 and 2013.

Sample M3 “Refugee Sample” was a new refugee sample added in 2016 for the IAB-BAMF-SOEP Refugee Survey in which roughly 1,769 refugee households were interviewed repeatedly. Respondents aged 18 and older who entered Germany between January 2013 and December 2016 and who had filed an asylum application by April 2016 (regardless of their current legal status) were interviewed along with the other members of their households.

Sample M4 “Refugee Family Sample”: the 2016 “IAB-BAMF-SOEP Survey of Refugees” (Samples M3 and M4) is a joint project of the Institute for Employment Research (IAB), the Research Center of the Federal Office for Migration and Refugees (BAMF-FZ) and the Socio-Economic Panel (SOEP). The target population of the samples consists of 1,769 households with individuals who arrived in Germany between January 2013 and January 2016 and had applied for asylum by June 2016 or were hosted as part of specific programs of the federal states (irrespective of their asylum procedure and their current legal status). The first part of the sample (M3) was financed with funds allocated to the IAB from the research budget of the Federal Employment Agency (BA). Sample M4 was funded by the Federal Ministry of Education and Research (BMBF) and has a focus on refugee families.

Sample M5 “Refugee Sample” M5 is the third boost sample of refugee households. The population of M5 covers adult refugees who applied for asylum in Germany between January 1, 2013, and December 31, 2016, and are currently living in Germany. The first wave of M5 was conducted in 2017. M5 added another 1,519 households of refugees who have migrated to Germany since 2013 to the SOEP framework.

Sample N “Refresher Sample (PIAAC-L)”: Sample N integrated 2,314 households of former participants in the Program for the International Assessment of Adult Competencies (PIAAC and PIAAC-L) in 2017. This is the most recent addition to the SOEP-Core samples. Fieldwork in sample N was conducted between mid-March and mid-August and thus slightly later than the majority of samples A–L1.

Sample O “Social City Sample”: Sample O includes 935 households located primarily in bigger cities. It was designed to enhance the potential of the data for analysis by incorporating more city-specific environments. The sample was selected in cooperation with BBSR using a new sampling design based on regional data in areas where the “Soziale Stadt” (social city) urban development project is being carried out. Based on the digital data available on the boundaries of the “Soziale Stadt” areas, it was possible to create a new variable going back to the year 2000 that shows whether or not a household’s address is within an area covered by the project.

Sample P “Top Shareholder Sample”: Sample P was conceptualized as a sample of highly affluent households in Germany. Against the backdrop of increasing income and wealth inequality in Germany, despite economic growth in recent decades, a lack of data on wealthy populations has become increasingly evident in the social sciences. Goals to be accomplished with sample P were to improve the empirical basis of the poverty and wealth report of the German government as well as laying the foundation for medium and long-term cross-sectional and longitudinal analyses. The gross sample of sample P consisted of 23,259 households.

Sample Q “LGB*”: Sample Q is a boost sample of a hard-to-survey population: lesbians, gays, bisexuals, transgender people, and those who identify as non-binary. While the actual percentage of LGBTQ+ people in the general population is unknown, this population was too scarcely represented in the SOEP to meaningfully analyze this group. 835 households were recruited via an approximately 9-month long telephone screening process. Of these households 477 participated between April and November.

Sample M6 “Refugee Sample”: M6 is the acronym for the fourth top-up sample for households that represents refugees. The population of M6 covers two groups: firstly, adult refugees who arrived in Germany between January 1, 2013 and December 2016 (“Refreshment”) and secondly adult refugees who came to Germany between January 1, 2017 and June 2019 (“Enlargement”) with a strongly disproportionate oversampling of refugees from East- and West-Africa.

Sample M7 “Migration Sample”: Like the older migration samples M1 and M2, the Integrated Employment Biographies Sample (IEBS) of the Federal Employment Agency (BA) served as the sampling frame for both boost samples.

Boost sample M7's goal was to capture migration dynamics and processes from 2016 to 2018 with a focus on EU migration. To ensure that statistically significant group comparisons can be made, sampling was restricted to the three most significant countries of origin in that time period: Romania, Bulgaria, and Poland.

Sample M8a “Migration Sample”: Like the older migration samples M1 and M2, the Integrated Employment Biographies Sample (IEBS) of the Federal Employment Agency (BA) served as the sampling frame for both boost samples. Boost sample M8a was designed to help evaluate the skilled worker immigration law (Fachkräfteeinwanderungsgesetz), which came into effect March 1, 2020, and targeted migrants from third countries that came to Germany between 2017 and 2018, sampling them as a control group for a treatment group that will be sampled at a later date.

More information about “Sample Sizes and Panel Attrition” can be found [here](#)

4.1.1 Sample-Specific Questionnaires

In SOEP it is common for special samples to receive extended, adapted, and/or integrated questionnaires in the first few years. This ensures that sample-specific questions that do not play a role in the main SOEP can also be included. In the following tables you can see which questionnaires the respective samples received, which years they ran, which raw data set they were included in, and which “long” data set they went into.

Sample Specific Instruments:

Sample A

Sample A							
Year	1984	1985	1986	1987	1988	1989	
Version	1	2	3	4	5	6	
Questionnaire / Wave	a	b	c	d	e	f	long Dataset
Household	Sh, Skind						hl, kidlong
Household (New Respondents)	Sh, Skind						hl, kidlong
Individual	\$p, Spkal, biolela*	\$p, Spkal					pl, pkal, biol
Individual (New Respondents)	\$p, Spkal, biolela*			\$p, Spkal	\$p, Spkal, biolela*	\$p, Spkal	pl, pkal, biol
Biography				biolela*		biolela*	biol
Catch-Up Individual (Re-Questioning employed)				Spluecke			plueckl
Catch-Up Individual (Re-Questioning unemployed)				Spluecke			plueckl
Financial Statement (Vermögensbilanz)					ev		
* Not part of the data distribution file, only available as long-file							
S: Wave abbreviation							
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.							

Sample B

Year	1984	1985	1986	1987	1988	1989	
Version	1	2	3	4	5	6	
Questionnaire / Wave	a	b	c	d	e	f	long Dataset
Household (Foreigners Version)				Sh, Skind			hl, kidlong
Household (Foreigners Version New Respondents)				Sh, Skind			hl, kidlong
Individual (Foreigners Version)	\$pausl, \$pkal, biolela*			\$pausl, \$pkal			pl, pkal, biol
Individual (Foreigners Version New Respondents)				\$pausl, \$pkal, biolela*	\$pausl, \$pkal, biolela*	\$pausl, \$pkal	pl, pkal, biol
Biography (Foreigners Version)				biolela*		biolela*	biol
Catch-Up Individual (Re-Questioning employed)				Spluecke			plueckl
Catch-Up Individual (Re-Questioning unemployed)				Spluecke			plueckl
Financial Statement (Foreigners)				ev			
* Not part of the data distribution file, only available as long-file							
S: Wave abbreviation							
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.							

From the start of Sample B (foreigners), respondents could complete the individual questionnaire in German or in the respective foreign language. Starting with wave 2 of the panel, there were “old” and “new” survey units (households, persons), and there were survey units with or without certain changes (e.g., households that had or had not moved; individuals who had or had not changed careers). The questionnaires took these changes into account for all subgroups. Survey procedures and tools were designed to ensure that each subgroup received the right questionnaire for them. This technique as well as the bilingual design of the foreigner questionnaires was retained for waves 3-6. In addition, retrospective information and missing information on temporary drop outs was collected. The “financial statement”, which is now a survey module, was a separate questionnaire in the year 1988.

Sample Specific Instruments: Sample C

Year	1990	1991	1992	1993	1994	1995	
Version	7	8	9	10	11	12	
Questionnaire / Wave	g	h	i	j	k	l	long Dataset
Household East	ghost		\$h, \$kind				hl, kidlong
Individual East	\$post, \$pkalost, biolela*	\$post, \$pkalost		\$p, \$pkal			pl, biol, pkal
Individual East (New Respondents)		\$post, \$pkalost, biolela*		\$p, \$pkal			pl, biol, pkal
Biography East					biolela*		biol

* Not part of the data distribution file, only available as long-file
\$: Wave abbreviation
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.

SOEP researchers were determined to seize the historic opportunity of German reunification to obtain a first baseline measurement of incomes in the “old” GDR currency. The questionnaire was prepared by an East-West working group including DIW Berlin, WZB, Collaborative Research Centre 3, and the ISS at the Academy of Sciences in the GDR, with the participation of Infratest and its partner organization in the GDR. The result was a questionnaire that covered many of the same themes and questions and was structured similarly to the West SOEP questionnaire, but which focused more on the specific situation in the GDR (e.g., the housing situation).

Sample Specific Instruments:**Sample J**

Year	2011	2012	2013	
Version	28	29	30	
Questionnaire / Wave	bb	bc	bd	long Dataset
Individual with Biography		\$p, \$lela*, \$pkal		pl, biol, pkal
Household		\$h, \$kind		hl, kidlong
Youth		\$jugend		jugendl
Individual		\$p, \$pkal		pl, pkal
Mother-Child (Newborns)		\$muki*		bioagel
Mother-Child (2-3-year-olds)		\$muki2*		bioagel
Mother-Child (5-6-year-olds)		\$muki3*		bioagel
Parents (7-8-year-olds)		\$elt*		bioagel
Mother-Child (9-10-year-olds)		\$muki5*		bioagel
Deceased Individual		\$vp		vpl
Lust auf DJ (cognitive Test)				cogdj
Grip Strength				gripstr

* Not part of the data distribution file, only available as long-file
\$: Wave abbreviation
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.

Sample K

Year	2012	2013	
Version	29	30	
Questionnaire / Wave	bc	bd	long Dataset
Individual with Biography	\$p, \$lela*, \$pkal		pl, biol
Household	\$h, \$kind		hl, kidlong
Youth	\$jugend		jugendl
Individual		\$p, \$pkal	pl, pkal
Youth			jugendl
Mother-Child (Newborns)		\$muki*	bioagel
Mother-Child (2-3-year-olds)		\$muki2*	bioagel
Mother-Child (5-6-year-olds)		\$muki3*	bioagel
Parents (7-8-year-olds)		\$elt*	bioagel
Mother-Child (9-10-year-olds)		\$muki5*	bioagel
Deceased Individual		\$vp	vpl
Lust auf DJ (cognitive Test)			cogdj
Grip Strength			gripstr
* Not part of the data distribution file, only available as long-file			
\$: Wave abbreviation			
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.			

A major shift in the design of SOEP questionnaires took place with Sample J. Due to the increased panel mortality from wave 1 to wave 2 that was observed for the refresher samples F (2000- 2001), H (2006-2007), and I (2009-2010), the biographical module, with an average interview length of 17 minutes, was integrated into wave 1. If this had not been done, no biographical data would have been collected for approximately 20% of all SOEP respondents who would probably not have participated in wave 2. In comparison to the longitudinal samples, data collection in the first wave was focused on the main three questionnaires: the household, the individual, and the youth questionnaire. As the fieldwork in these refresher samples was conducted exclusively by CAPI, it was feasible to include complex modules with event-triggered question loops.

Sample Specific Instruments: Samples L1-L3

Year	2010	2011	2012	2013		
Version	27	28	29	30		
Questionnaire / Wave	ba	bb	bc	bd	Samples	long Dataset
Cohort Sample						
Household	\$h, \$kind				L1	hl, kidlong
Household New Respondents		\$h, \$kind			L1	hl, kidlong
Individual with Biography	\$p, \$lela*, \$pkal				L1	pl, biol, pkal
Individual		\$p, \$pkal			L1	pl, pkal
Youth	\$jugend				L1	jugendl
Catch-Up Individual (Re-Questioning)		\$pluecke			L1	plueckel
Parents 1	\$muki1*				L1	bioagel
Parents 2	\$muki2*				L1	bioagel
Parents 3	\$muki3*				L1	bioagel
Parents 4	\$muki4*				L1	bioagel
Parents 5	\$elt*				L1	bioagel
Parents 6	\$muki5*				L1	bioagel
Screening Sample						
Household	\$h, \$kind				L2+L3	hl, kidlong
Household New Respondents		\$h, \$kind			L2+L3	hl, kidlong
Individual with Biography	\$p, \$lela*, \$pkal				L2+L3	pl, biol, pkal
Individual		\$p, \$pkal			L2+L3	pl, pkal
Youth	\$jugend				L2+L3	jugendl
Catch-Up Individual (Re-Questioning)		\$pluecke			L2+L3	plueckel
Parents 1	\$muki1*				L2+L3	bioagel
Parents 2	\$muki2*				L2+L3	bioagel
Parents 3	\$muki3*				L2+L3	bioagel
Parents 4	\$muki4*				L2+L3	bioagel
Parents 5	\$elt*				L2+L3	bioagel
Parents 6	\$muki5*				L2+L3	bioagel

* Not part of the data distribution file, only available as long-file

\$: Wave abbreviation

Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets

The main focus of Families in Germany (FiD) was on the families and children – the parental questionnaires (filled out by parents about their children) were about twice as long as the comparable questionnaires in SOEP-Core, and questionnaires for the 1-2-year-olds and the 9-10-year-olds were added (as of 2012, SOEP-Core had added a questionnaire for 9-10-year-olds that is partly comparable to the FiD version). In large part, FiD resembled the SOEP. Each adult was asked to answer an individual questionnaire, which, in the first two years, included retrospective questions on childhood, education, and early work experience. In addition, there were several questions designed to capture the challenges families face with regard to the return of mothers into the labor market – with respect to workplace, work schedule, overtime, daycare options, etc.

Sample Specific Instruments:

Sample M1 (IAB-SOEP-Migration Sample)

Year	2013	2014	2015	2016	
Version	30	31	32	33	
Questionnaire / Wave	bd	be	bf	bg	long Dataset
Household	\$h, \$kind				hl, kidlong
Individual with Biography	\$p, \$lela*, \$pkal				pl, biol, pkal
Individual	\$p, \$pkal				pl, pkal
Youth	\$jugendl				jugendl
Pre-Teen	\$school				biopupil
Mother-Child (Newborns)	\$muki*				bioagel
Mother-Child (2-3-year-olds)	\$muki2*				bioagel
Mother-Child (5-6-year-olds)	\$muki3*				bioagel
Parents (7-8-year-olds)	\$elt*				bioagel
Mother-Child (9-10-year-olds)	\$muki5*				bioagel
Deceased Individual	\$vp				vpl
Lust auf DJ (Cognitive Test)					cogdj
Catch-Up Individual	\$pluecke				plueckel
* Not part of the data distribution file, only available as long-file					
\$: Wave abbreviation					
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.					

Sample M2 (IAB-SOEP-Migration Sample)

Year	2015	2016	
Version	32	33	
Questionnaire / Wave	bf	bg	long Dataset
Household	\$h, \$kind		hl, kidlong
Individual with Biography	\$p, \$lela*, \$pkal		pl, biol, pkal
Youth	\$jugend		jugendl
Individual		\$p, \$pkal	pl, pkal
Pre-Teen		\$school	biopupil
Early Youth		\$school2	biopupil
Mother-Child (Newborns)		\$muki*	bioagel
Mother-Child (2-3-year-olds)		\$muki2*	bioagel
Mother-Child (5-6-year-olds)		\$muki3*	bioagel
Parents (7-8-year-olds)		\$elt*	bioagel
Mother-Child (9-10-year-olds)		\$muki5*	bioagel
Deceased Individual		\$vp	vpl
Lust auf DJ (Cognitive Test)			cogdj
Catch-Up Individual		\$pluecke	plueckel
* Not part of the data distribution file, only available as long-file			
\$: Wave abbreviation			
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.			

Following the design shift for refresher samples since Sample J in 2011, respondents have been surveyed on their life history using the “biography questionnaire”, which was integrated into the individual questionnaire from wave 1. This ensures that biographical information will be available for all target persons who provided an individual interview in participating households. Other supplementary questionnaires were not included in the survey instruments given to first-wave respondents to avoid “overburdening” respondents with an extremely lengthy first-wave interview. Questionnaires for the migration boost samples include questions that have been part of SOEP-Core for the last three decades. In addition, the survey covers each respondent’s complete migration history, education, training, and employment history in Germany and abroad, and numerous aspects of cultural and living environments relevant to the social integration of migrants. The household questionnaire is identical to the questionnaire used in the SOEP-Core sample.

Sample Specific Instruments: Samples M3-M5 (IAB-BAMF-SOEP Refugee Sample)

Year	2016	2017	2018	2019	2020	2021		
Version	33	34	35	36	37	38		
Questionnaire / Wave	bg	bh	bi	bj	bk	bl	Samples	long Dataset
Household	Sh						M3-M4	hl
Individual with Biography	Sp, S1e1a*, Spkal						M3-M4	pl, biol, pkal
Individual	Sp						M3-M4	pl, pkal
Individual with Biography non-fugitive	Sp, S1e1a*, bhpkal						M3-M4	pl, biol
Youth (12-17-year-olds)	Sschool, Sschool2, Sjugend						M3-M4	biopupil, jugendl
Children in household	Smuki1-Smuki5*, Skind						M3-M4	bioagel, kidlong
Assessment of declarative knowledge and general cognitive ability in refugee children and adolescents	cog_refu						M3-M4	
Household	Sh, Skind						M5	hl, kidlong
Individual with Biography	Sp, S1e1a*						M5	pl, biol
Living Area (Wohnumfeld)	Sbrutt17						M5	hbrutto
Household	Sh						M3-M5	hl
Individual	Sp, Spkal						M3-M5	pl, pkal
Individual non-fugitive	Sp, Spkal						M3-M5	pl, pkal
Individual with Biography	Sp, S1e1a*, Spkal						M3-M5	pl, biol
Individual with Biography non-fugitive	Sp, S1e1a*, Spkal						M3-M5	pl, biol
Youth (12-17-year-olds)	Sschool, Sschool2, Sjugend						M3-M5	biopupil, jugendl
Children in household	Smuki1-Smuki5*, Skind						M3-M5	bioagel, kidlong
Assessment of declarative knowledge and general cognitive ability in refugee children and adolescents	cog_refu						M3-M5	
Household	Sh, Skind						M3-M6	hl, kidlong
Individual	Sp, Spkal						M3-M6	pl, pkal
Individual with Biography (first time respondents)	Sp, S1e1a*, Spkal						M3-M6	pl, biol
Youth (12-17-year-olds)	Sschool, Sschool2, Sjugend						M3-M6	biopupil, jugendl
Children in household	Smuki1-Smuki5*						M3-M6	bioagel, kidlong
Living Environment Questions	Sp, S1e1a*, Spkal						M6	pl, biol, pkal

* Not part of the data distribution file, only available as long-file

As with every other previously established subsample of migrants in the SOEP (M1 and M2), there was a clear need for several deviations from standard SOEP-Core questionnaires to reflect the special characteristics of the target group. Several additional questions concerning migration and integration were incorporated into the individual questionnaire to better field the range of research questions and research goals of the project partners. These included topics such as ethnic background, experiences en route to Germany, language skills, integration courses in Germany, job experience, current occupation, educational background, health, attitudes, and values. The household questionnaire was much more SOEP-related than the individual questionnaire in order to establish longitudinal information on the households.

In the first wave of M6 three questionnaires were fielded: the individual questionnaire for first time respondents (including additional biographical questions) for all adult household members, which was administered in separate versions for refugees and for Germans or migrants respectively, and the household questionnaire for the anchor respondent. Like for the other refugee samples M3-5, a special SOEP individual and life-history questionnaire was developed that includes issues specific to refugees. The version for Germans and migrants was identical to the individual and life-history questionnaire in samples A-Q and M1/2. As is the usual approach for boost samples, no youth or child questionnaires were fielded in sample M6. All questionnaires were solely available in CAPI mode and provided in seven different language versions, although a small percentage of interviews was conducted via telephone in the CAPI environment.

Sample Specific Instruments: Sample P (Top Shareholders)

Year	2019	
Version	36	
Questionnaire / Wave	bj	long Dataset
Household	\$h	hl, kidlong
Biography	\$lela*	biol
Individual	\$p, \$pkal	pl, pkl
* Not part of the data distribution file, only available as long-file		
\$: Wave abbreviation		
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.		

Three different questionnaires were used to collect data in sample P. Apart from the regular household and individual questionnaires, a life-history questionnaire module was used to collect background information of all respondents. Computer-assisted personal interviewing (CAPI) was applied alongside paper questionnaires (PAPI or SELF) for all questionnaires. While the life history questionnaire was integrated into the individual questionnaire in the CAPI, it was administered as a separate questionnaire in the PAPI and SELF modes

Sample Specific Instruments: Sample Q (LGB*)

Year	2019	
Version	36	
Questionnaire / Wave	bj	long Dataset
Household	\$h	hl, kidlong
Biography	\$lela*	biol
Individual	\$p, \$pkal	pl, pkl
Youth	\$jugend	jugendl
Pre-Teen	\$school	biopupil
Early Youth	\$school2	biopupil
Mother-Child (Newborns)	\$muki1*	bioagel
Mother-Child (2-3-year-olds)	\$muki2*	bioagel
Mother-Child (5-6-year-olds)	\$muki3*	bioagel
Parents (7-8-year-olds)	\$elt*	bioagel
Mother-Child (9-10-year-olds)	\$muki5*	bioagel
* Not part of the data distribution file, only available as long-file		
\$: Wave abbreviation		
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.		

Eleven different questionnaires were used to collect data in sample Q. Apart from the regular household and individual questionnaires, a life-history questionnaire module was used to collect background information of all respondents. A special module regarding their sexual orientation was added in the individual questionnaire. Adolescents of the age 16 or 17, 13 or 14 and 11 or 12 were interviewed using specific youth questionnaires. Additionally, all mother and child /parent questionnaires were administered in this boost sample. Computer-assisted personal interviewing (CAPI) was applied exclusively for all questionnaires.

Sample Specific Instruments: Samples M7-M8 (IAB-SOEP-Migration Sample)

Year	2020	2021	
Version	37	38	
Questionnaire / Wave	bk	bl	long Dataset
Household	\$h		hl, kidlong
Biography	\$lela*	\$lela*	biol
Individual	\$p, \$pkal	\$p, \$pkal	pl, pkl
* Not part of the data distribution file, only available as long-file			
\$: Wave abbreviation			
Note that the samples are continued up to the current wave. Here only the sample specific instruments are shown with reference to the data sets.			

In the first waves of M7 and M8 three questionnaires were fielded: the individual questionnaire for first time respondents (including additional biographical questions) for all adult household members, which had the life-history module integrated in the CAPI-instrument and the household questionnaire for the anchor respondent. In addition to these instruments, anchor-respondents had to answer a short screening questionnaire in order to clarify their membership in the target populations of M7 and M8a respectively. Respondents had to have been born outside of Germany, their stay should not be temporary, and they were to have moved to Germany no earlier than 2016 (M7) or 2017 (M8a) respectively. All questionnaires were solely available in CAPI mode. Translation aides were provided only in paper form in four additional languages. With regards to questionnaire content, the household and individual questionnaires were almost identical to the ones used in samples M1/2.

Last change: Jul 24, 2023

4.2 Eligibility and Follow-up

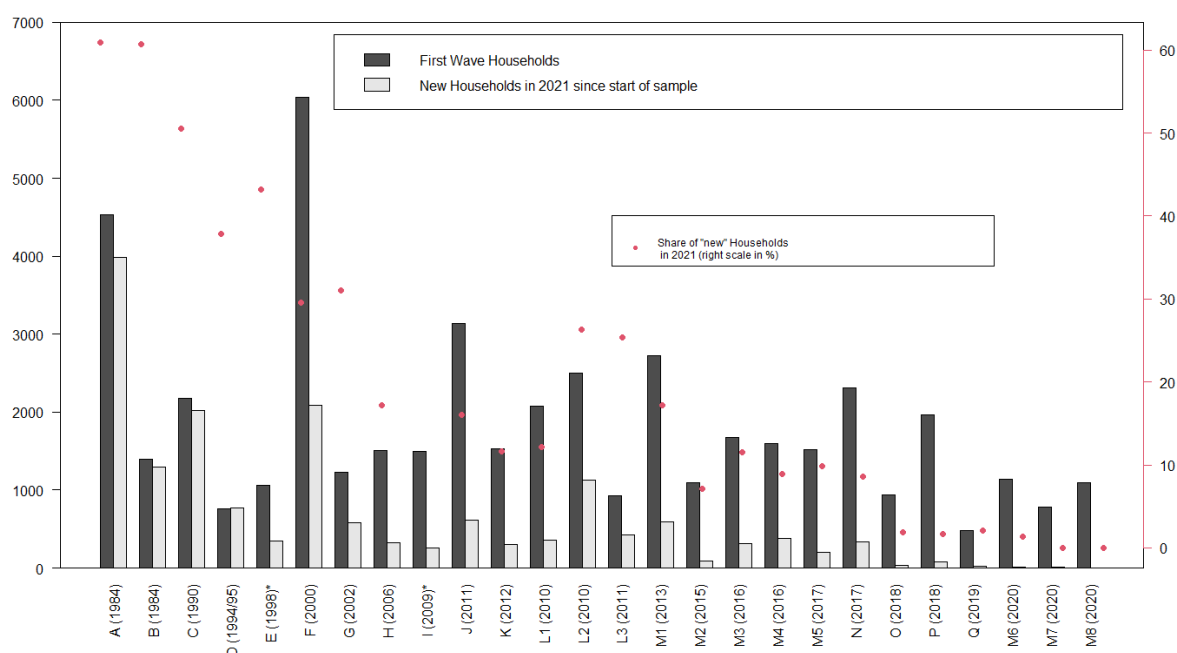
As mentioned, the SOEP's goal is to be representative of the residential population of Germany. All household members 16 and older are eligible for a personal interview, starting with the youth questionnaire for their age group, followed by "regular" individual questionnaires thereafter. As years go by, the children from the first wave reach age eligibility and become panel members. If they move out and start their own families, they and their new family members are also part of the survey. "New" individuals become part of the SOEP population by being born into SOEP households or as a result of residential mobility. If a person enters a SOEP household after the initial wave in which that household was surveyed, this person is asked to fill out the regular individual questionnaire if age-eligible or will be asked to participate once old enough. In the absence of panel attrition, this would make the SOEP a self-sustaining survey.

The concept of how to follow respondents and sample members over time is important for the representativeness of the study. The basic principle for follow-up in the SOEP is that all persons participating in a wave of any subsample will be surveyed in the following years as long as they stay within the boundaries of Germany. This rule also extends to respondents who entered a SOEP household after the first wave in which it was surveyed due to residential mobility or birth. If there is a "split-off", that is, if someone moves out of the household in which they were last interviewed, the members of the new household receive a new household identifier. The table conceptualizes how new sample members and households are surveyed in the SOEP. The figure shows that as a result of the follow-up concept, several thousand "new" households had become part of the SOEP population.

Individuals or households that could not be interviewed in a given year are termed "temporary drop-outs". They are followed until there are two consecutive waves of missing interviews for all household members or until the entire household refuses to participate further. In the case of a temporary drop-out, in which a respondent participates in the survey again after not participating in the previous wave, the respondent is asked to fill out an additional short questionnaire covering key information about employment and demographics in the year of their absence.

	Existing Households	New Households
Existing Individuals	Classic case: without change of address entire household moves	Respondent moves out of existing household and forms a new household
New Individuals	Born into household, move into household	Moved into or born into HH formed after a respondent moved out of existing HH

Changes to the Sample: Old and new household in the SOEP



[Download R Code to create figure](#)

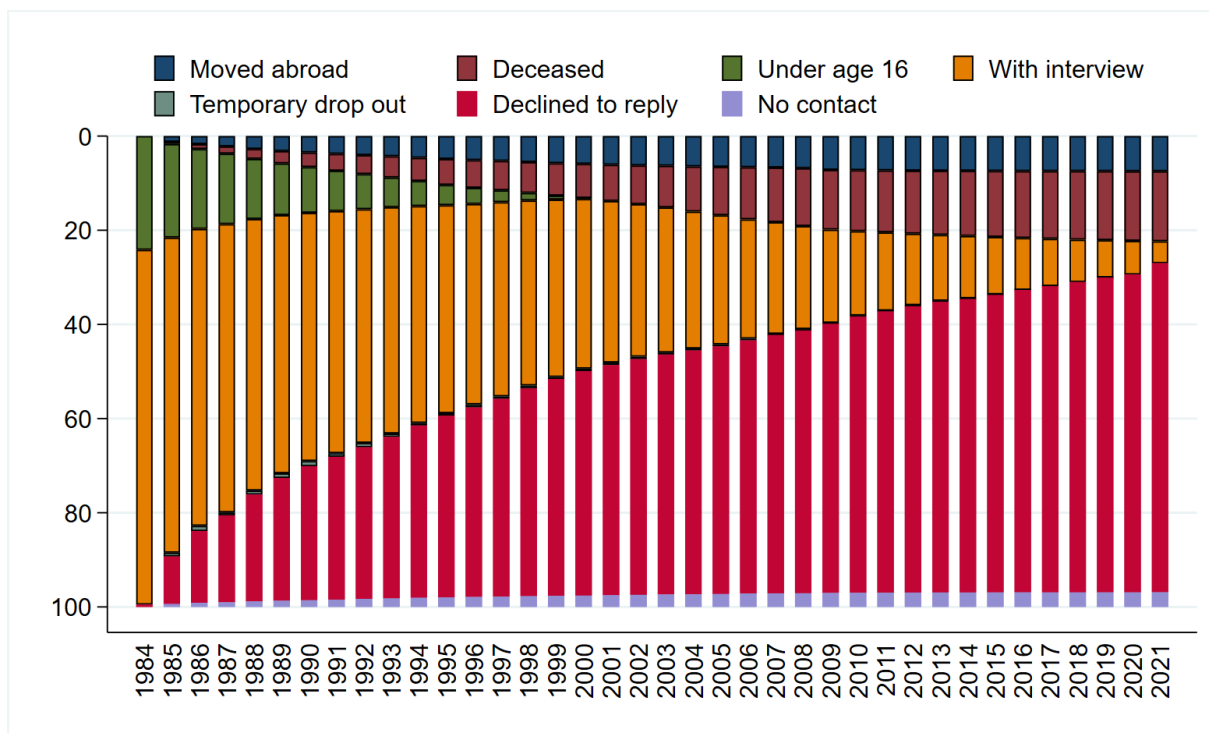
Last change: Jul 24, 2023

4.3 Development of Sample Sizes

Individuals who decline to take part in the survey or are not available for an interview are kept in the so-called “gross” sample of the study as long as they continue to live in households with at least one participating respondent. If the entire household declines to participate in two consecutive waves, all individuals in the household are removed from the SOEP. The table shows the starting sample sizes of samples A through M4, the years when the samples were first collected, as well as the percentage of those persons who were eligible for an interview but declined participation (“partial unit non-response”, PUNR) in the first wave. The figure illustrates the development of the number of successful person interviews since 1984. The reduction in the population size for all individual samples is mainly the result of individual-level drop-outs, refusals, moving abroad, etc. However, due to new persons moving into already existing households and children reaching the age of 16 and thereby increasing the sample size, this negative development is offset somewhat.

Starting Sample Size of the SOEP Samples

Longitudinal Development of the 1984 Population



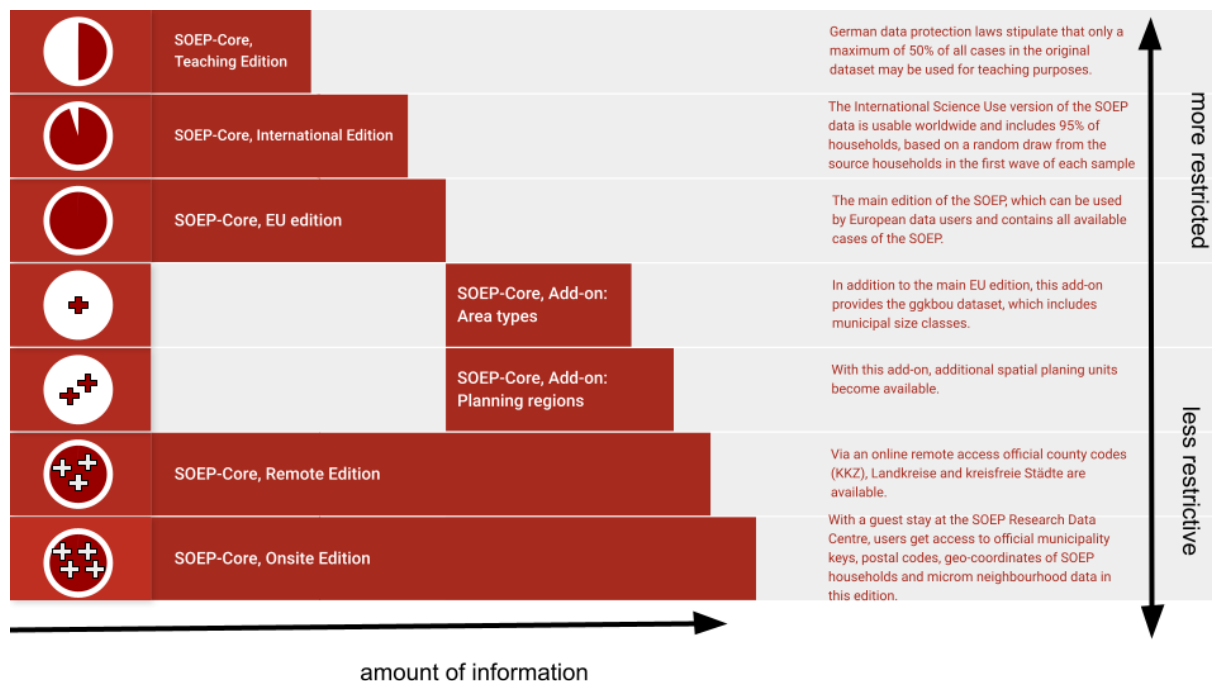
[Download Stata Code to create figure](#)

Last change: Jul 24, 2023

DATA STRUCTURE OF SOEP-CORE

5.1 Data Editions of SOEP-Core

Access to SOEP data is provided in compliance with the highest security standards to protect respondents' confidentiality and maintain their trust in the survey. The data are also provided solely for scientific research purposes, that is, they are only made available to members of the scientific community. This means that researchers are only given access to SOEP data after they have signed a data distribution contract with DIW Berlin. Different data packages, called "editions", reflect these requirements and can be differentiated by the amount of information contained in them, the level of data protection, and the mode of data access. The EU Edition is considered the standard edition. More restricted editions provide less information; less restricted editions provide more information but are only available under more restrictive conditions. The Teaching, International, and EU Editions *Teaching*, *International*, and *EU Edition* are made available as downloads under the standard data distribution contract, while the two add-ons Area Types and Planning Regions *Add-ons: Area Types and Planning Regions* require additional contracts. The Remote Edition *Remote Edition* can only be accessed through remote execution, and the Onsite Edition *Onsite Edition* can only be accessed on site at the SOEP Research Data Center at DIW Berlin.



In this figure, "more restrictive" means that existing variables from the EU Edition are left blank for reasons of data protection or not all cases are included. For example, variables that provide information at the federal state level are

not available in the International or Teaching Editions (which only distinguish between East and West Germany). A higher level of data protection makes it possible to provide more information with fewer restrictions. This makes the editions less restrictive in terms of the information available. In most cases, as more sensitive information is added to an edition, access to the data edition changes and the requirements for its use also change.

5.1.1 Teaching, International, and EU Edition

Only the standard data distribution contract is required for the EU Edition and the International Edition. The EU Edition includes 100% of all observations, the German federal states, and the urban/rural variable. This edition is only available to users from research institutions in the EU and countries with an “adequacy decision” (Angemessenheitsbeschluss)—Switzerland, Japan, Canada, Israel, and a few others.

The International Edition is available to users from research institutions in all other countries than those listed above. This edition contains 95% of all households from the first wave of each SOEP subsample based on a random sampling of the original households in each subsample and only the East/West versions of variables normally containing the federal states. The original variables (with information on the federal states) remain in the edition but they are assigned the missing code -7 “Only available in less restricted edition” if a variable cannot be made accessible in a specific edition. For more information on the missing codes in SOEP-Core, see the chapter *Missing Conventions*.

The least restrictive edition of the data but the one containing the least information is the Teaching Edition. Here, a data distribution contract is required for teaching staff; students only need to sign the data protection declaration, which the contract holder must keep on file. The contract holder is responsible for ensuring strict adherence to data protection. German data protection laws stipulate that a maximum of 50% of all cases in the original dataset may be used for teaching purposes. The Teaching Edition has the same data structure as the International Edition (with the exception of the EU-SILC Clone) but contains half the number of cases in the EU Edition. The Teaching Edition provided to students must be stored in a separate hard drive area to which the user guarantees controlled access. Students may under no circumstances take data home with them or transfer the data to any other device at the university.

5.1.2 Add-ons: Area Types and Planning Regions

In addition to the EU Edition, the SOEP offers additional datasets that can extend the standard file to include municipality size classes (add-on: Area Types) or even spatial planning units (add-on: Planning Regions). Access to these files is more restricted because they provide users with more sensitive information about the respondents.

For the add-on Area Types, a regional data contract is required in addition to the data distribution contract. This requires that the user submit a data protection concept to the SOEP. There is no template for this; users must develop this concept specifically for the workplace in which they want to use the data.

For the add-on Planning Regions, a regional data contract is also necessary, and the SOEP requires that users submit a data protection concept that they have developed themselves. For this add-on, however, the requirements for the data protection concept are significantly higher.

5.1.3 Remote Edition

Further information such as official county codes (KKZ), identifying administrative districts (Landkreise) and urban districts (kreisfreie Städte) can be accessed through remote execution using the Remote Edition (or on site). For this edition, users are required to submit an application to use SOEPrmote in addition to the data distribution contract. For the remote execution contract, no separate data protection concept is required, as users will only access the information remotely and no files are transmitted to computers outside of the Research Data Center of the SOEP (RDC SOEP).

To access the Remote Edition, there are two options available:

- SOEPrmote execution (e-mail processing)
- SOEPrmote access (on-site processing at special workstations)

With SOEPremote execution, users can email their Stata syntax to a remote server, which processes the syntax and returns the results to users by email. With SOEPremote access, users can use IGEL clients at RDC SOEP in Berlin. By using the IGEL clients, onsite users have the advantage of working directly with the Remote Edition instead of having to go through the email procedure. The disadvantage is having to plan and book a visit to the RDC SOEP in Berlin.

5.1.4 Onsite Edition

The Onsite Edition is the edition with all available information. Guests using RDC SOEP IGEL clients (*How to Use SOEP IGEL*) can access the additional information about the municipalities or postal codes of the SOEP households or data from microm GmbH on households' neighborhoods. Users can even analyze geocoded data. To access these data, researchers are first required to sign a data protection agreement, and a complete record is kept of all data access. The concept for providing the geo-coordinates of SOEP households is that the point coordinates are kept separate from the actual survey information throughout the entire process of analysis by data users due to privacy concerns. Researchers therefore never have simultaneous access to the SOEP survey data and the geo-coordinates of SOEP households. The results may only be published in completely anonymous form and are checked before they are transmitted from the secure server to the user.

- To apply to use a guest work station, click [here](#):
- For more information about your workplace at the SOEP Research Data Center see the section *How to Use SOEP IGEL*
- For more information about how to work with SOEP's spatial data see the section *Working with spatial data in R*

Last change: Mar 28, 2023

5.2 Principles of Data Analysis

All SOEPtutorials can be found on our [YouTube Channel](#)

The structure of panel data has three dimensions. First, the respective examination units (n) and a matrix of dependent and independent variables (y, x) are completely analogous to a cross-sectional design. Second, the dimension of time (t), whereby a distinction is made between two data formats for panel data structures - "wide" or "long" (with wide format the variable matrix is indexed with the dimension of time and with long format the respective examination units). Regardless of the selected data format, when using panel data with several survey waves, the data matrices often do not contain complete information due to the panel mortality of individual survey units or because data from new panel members are only collected at a later point in time. In both cases, the term "unbalanced panel data" is used. In contrast, the classical panel data structure, on the other hand, is "balanced", i.e., as many observations of dependent and independent variables are available for all study units as there are waves of data collection. Social science panel data often show a data structure characterized by many investigation units (large n) as well as, in relation to it, few waves and therefore measuring time (small t). When data from a panel study are available, even descriptive forms of data analysis are often of particular interest, since the identification of changes in a variable over time and the corresponding separation of interindividual and intraindividual changes can represent important social facts, particularly in the case of generalizable samples. It is of social scientific interest whether a constant 15% proportion of people whose income is below the poverty risk level is repeatedly found in the same person over time, or whether there was an even balance of increases and decreases in poverty risks and only half of the population was permanently exposed to the risk. The choice of complex analysis methods for panel data depends first and foremost on the respective measurement level of the dependent and independent variables, but also on whether they are time-constant variables (such as gender or migration background) or time-invariant variables. The statistical analysis models of panel data range from structural equation models, various regression models, event analysis, sequence data analysis, latent growth models to causal analyses using matching methods. A particular advantage of panel data is that the chronological sequence of changes can be modelled and calculated and the problem of unobserved heterogeneity, which is often encountered in the social sciences, can be significantly reduced, at least in comparison with cross-sectional data.

5.2.1 Cross-Sectional Data Structure (CS)

Cross-sectional data is a type of data that observes many subjects at the same point in time. Each person is assigned a row in the dataset and is only included once in such a dataset. By merging cross-sectional SOEP data across waves, you obtain a dataset in wide-format.

Row	ID	wave	sex	income
1	1	2015	m	1500
2	2	2015	m	1000
3	6	2015	f	2000
4	8	2015	m	5500

5.2.2 Data Structure in “Wide” Format (wide)

The SOEP data are available with different data structures. In the wide format, a respondent’s repeated responses are displayed in a single row and each response in a separate column. Each column represents a variable. We provide four datasets in the wide format: ppath, phrf, hpath, hhrf.

Row	ID	sex	income2015	income2016	income2017
1	1	m	1500	1500	2000
2	2	m	1000	1200	1200
3	6	f	2000	2000	2000
4	8	m	5500	6000	6500

5.2.3 Data Structure in “Long” Format (long)

The long format is a condensed and user-friendly dataset structure for longitudinal section analysis. Here, each person has one line per survey year. This means that you do not have several datasets for the different waves, but one dataset in which all survey waves are represented. A person can appear more than once in such a dataset. In the long format, one line describes a person-year combination.

Row	ID	syear	sex	income
1	1	2015	m	1500
2	1	2016	m	1500
3	1	2017	m	2000
4	2	2015	m	1000
5	2	2016	m	1200
6	2	2017	m	1200
7	6	2015	f	2000
8	6	2016	f	2000
9	6	2017	f	2000

In addition to the classic long format where one row in the dataset describes a person-year combination, there are also datasets that describe a longer period or a whole life, but only appear uniquely in the dataset without a survey year. These data sets can contain longitudinal information, but are constant over time. These time-constant data sets may include, for example, information on biological parents or employment history up to a certain age.

Row	ID	father_id	mother_id
1	20	18	19
2	21	18	19
3	35	34	35
4	36	34	35
5	37	34	35

5.2.4 Data Structure in Spell Format (spell)


In the strict sense of the word, spell data are about time periods with a defined start and end. When handling spell data it is necessary to take potential censoring into account. Censoring denotes that the beginning (left censored) or ending (right censored) of a spell is imprecise because of missing information or the beginning or ending of a spell is outside of the period of observation. It is quite conceivable that a person has only one spell over a given period, such as a male who is full-time employed. For a ten year period, there may be just the one spell “full-time employed”. In panel data, the same person would have 10 observations, one per year. A person may have many spells over a time period, and even have overlapping spells, like working part-time and receiving a disability pension. Spell data are useful for looking at stays in a certain state, and transitions in and out of that state.

Row	ID	spellnr	spelltype	begin	end	censored
1	1	1	Retired	1983	2007	left and right censored
2	1	2	Housewife/husband	1983	1984	left censored
3	1	3	Housewife/husband	1994	1994	uncensored
4	1	4	Housewife/husband	1998	1998	uncensored
5	2	1	Full-Time Employment	1984	1984	left censored
6	2	2	Full-Time Employment	1985	1985	uncensored

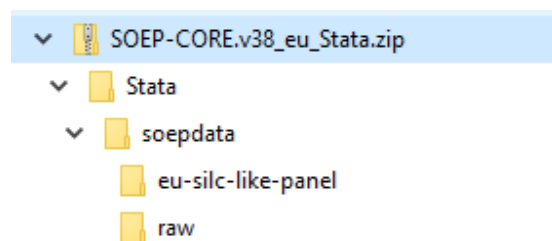
Last change: Mar 28, 2023

5.3 Data Distribution File

In the SOEP, each survey year is allocated to a data wave, which is abbreviated using the letters of the alphabet. One data wave may be released in several versions, which are displayed in SOEP with a “v” for version and the respective version number. The version number represents the survey years since the beginning of the survey. The SOEP has recently published the 38th version since the survey began in 1984. Within a data wave, updates may be made over time, such as v34.1. If updates have been made, users will be informed through various channels and be asked to order the data again. After ordering the data, the data will be sent to you in a zip file.




































>  SOEP-CORE.v38_eu_Stata.zip

Within this zip file you will find various datasets, a “raw” subdirectory and the “eu-silc-like-panel” subdirectory.



The datasets in the soepdata folder are a highly compressed and easy-to-analyze version of the SOEP data.

Note: SOEP strongly recommends that users use the soepdata top-level folder.

Name	Type
 eu-silc-like-panel	File folder
 raw	File folder
 abroad.dta	DTA File
 artkalen.dta	DTA File
 bioagel.dta	DTA File
 biobirth.dta	DTA File
 biocouplm.dta	DTA File
 biocouply.dta	DTA File
 bioedu.dta	DTA File
 bioimmig.dta	DTA File
 biojob.dta	DTA File
 biol.dta	DTA File
 biomarsm.dta	DTA File
 biomarsy.dta	DTA File
 bioparen.dta	DTA File
 biopupil.dta	DTA File
 biosib.dta	DTA File
 biotwin.dta	DTA File
 camces.dta	DTA File
 cog_refu.dta	DTA File
 cogdj.dta	DTA File
 cognit.dta	DTA File
 design.dta	DTA File
 gripstr.dta	DTA File
 hbrutt.dta	DTA File
 hbrutto.dta	DTA File
 hconsum.dta	DTA File
 health.dta	DTA File
 hgen.dta	DTA File
 hl.dta	DTA File
 hpath.dta	DTA File
 hpathl.dta	DTA File
 hwealth.dta	DTA File
 instrumentation.dta	DTA File
 interviewer.dta	DTA File

The data in SOEP-Core are no longer provided only as wave-specific individual files but are now pooled across all available years (in “long” format). In some cases, variables are harmonized to ensure that they are defined consistently over time. For example, the income information provided up to 2001 is given in euros, and categories are modified over time when versions of the questionnaire have been changed. The longitudinal nature of the data is one of the biggest

assets of the SOEP. This is why we provide longitudinal datasets such as PL or HL. The advantage of such a dataset is that longitudinal analyses can be carried out without great effort.

If you need more information about the “long” data structure, see chapter *Data Structure in “Long” Format (long)*.

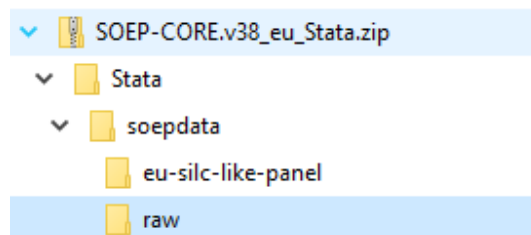
5.3.1 Core Datasets

The datasets in the soepdata folder:




































Tracking Data	Original Data	Survey Data	Generated Data	Spell Data
hbrutt	abroad	design	bioagel	artkalen
hbrutto	biol	pbr_hhch	biobirth	biocouplm
hpath	hl		bioedu	biocouply
hpathl	jugendl		bioimmig	biomarsm
pbr_exit	kidlong		biojob	biomarsy
pbrutto	more_docu		bioparen	lifespell
ppath	more_local		biopupil	migspell
ppathl	pl		biosib	pbiospe
instrumentation	plueckel		biotwin	refugspell
	vpl		camces	sozkalen
			cog_refu	
			cogdj	
			cognit	
			gripstr	
			hconsum	
			health	
			hgen	
			hwealth	
			interviewer	
			mihinc	
			pequiv	
			pflege	
			pgen	
			pkal	
			pwealth	
			timepref	
			trust	

5.3.2 Raw Datasets

In the “raw” directory, you will find all wave-specific datasets that were used to generate the long datasets on the previously presented level.



Attention: Please note that the datasets in the soepdata folder are completely sufficient for your data analysis. The datasets used to generate the SOEP-Core data can be found in the raw subdirectory. Detailed information about the raw datasets can be found here [Raw "raw"](#)

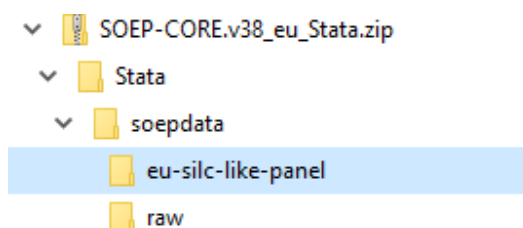
Name	Type
 ah.dta	DTA File
 ahbrutto.dta	DTA File
 ahgen.dta	DTA File
 akind.dta	DTA File
 ap.dta	DTA File
 apausl.dta	DTA File
 apbrutto.dta	DTA File
 apequiv.dta	DTA File
 apgen.dta	DTA File
 apkal.dta	DTA File
 bah.dta	DTA File
 bahbrutto.dta	DTA File
 bahgen.dta	DTA File
 bajugend.dta	DTA File
 bakind.dta	DTA File
 bap.dta	DTA File
 bapbrutto.dta	DTA File
 bapequiv.dta	DTA File
 bapgen.dta	DTA File
 bapkal.dta	DTA File
 bapluecke.dta	DTA File
 bavp.dta	DTA File
 bbh.dta	DTA File
 bbhbrutto.dta	DTA File
 bbhgen.dta	DTA File
 bbjugend.dta	DTA File
 bbkind.dta	DTA File
 bbp.dta	DTA File
 bbpbrutto.dta	DTA File
 bbpequiv.dta	DTA File
 bbpgen.dta	DTA File
 bbpkal.dta	DTA File
 bbpluecke.dta	DTA File
 bbvp.dta	DTA File
 bch.dta	DTA File

Within this “raw” directory, each wave is identified by letters of the alphabet: the first wave in 1984 is wave “A”, 1985 is wave “B”, and so on. To simplify the notation, the “\$” sign is used when referring to all waves of one group of datasets. For example, \$H refers to all household-level datasets from AH to now. For each year of SOEP data, there are single data files for households (e.g., \$H) as well as for individual respondents (e.g., \$P) and children (e.g., \$KIND) based on interview information. These observations make up the “net” population, with each of these files containing as many records as interviews could be conducted. Additional data files with a limited number of variables based on the “address log” constitute the “gross” number of households and persons, i.e., all households and their members that were eligible for an interview in any given year. Within the “raw” directory, the datasets are stored on a wave-specific basis and are the basis for generating the majority of the long datasets described above. In addition to these wave-specific datasets, the “RAW” directory also contains additional datasets in cross-sectional format that have not yet been distributed in long format (\$SCHOOL, \$SCHOOL2, EV, EXIT, \$PKALOST and PBR_HHCH).

Tracking Data	Original Data	Survey Data	Generated Data	Spell Data
ppfad	\$p	phrf	\$pgen	einkalen
hpfad	\$pausl	hhrf	\$hgen	sozkalen
\$pbrutto	\$pluecke	\$biorki	\$pequiv	
\$hbrutto	\$h	cirdef	\$pkal	
hbrutt\$\$	\$post	exit	\$pkalost	
cov_brutto	\$jugend			
cov_contact	\$school			
	\$school2			
	ev			
	\$vp			
	\$kind			

5.3.3 eu-silc-like-panel

The European Union Statistics on Income and Living Conditions (EU-SILC) contains data from across Europe on individual and household income, household living conditions, individual health, aspects of child care, employment, and self-assessed financial situation. EU-SILC offers both cross-sectional and longitudinal data. The German EU-SILC dataset currently contains only cross-sectional data. The eu-silc-like-panel dataset provided at DIW Berlin offers additional longitudinal information on private households in Germany based on data from the Socio-Economic Panel (SOEP) study since 2005. The eu-silc-like-panel is included in the annual SOEP data release since 2018 and requires a data distribution contract with DIW Berlin. The SOEP data are provided free of charge for scientific research. Researchers can compare all of the information in the dataset with longitudinal data on other European countries that can be obtained from Eurostat upon request.



Name	Type
 D-File.dta	DTA File
 H-File.dta	DTA File
 P-File.dta	DTA File
 R-File.dta	DTA File

The eu-silc-like-panel includes all of the four EU-SILC sub-datasets: The household register (D-File), the personal register (R-File), personal data (P-File), and household data (H-File). The clone datasets can be combined using the R-File, which includes both the current household and individual identifier. The identifiers in the eu-silc-like-panel are unique and do not vary among the four datasets. Complete documentation on the datasets can be found here: [Documentation EU-SILC](#).

Last change: Jul 24, 2023

5.4 Datasets SOEP-Core

SOEP-Core contains a multitude of different datasets. An overview of the documentation for the different datasets can be found on our website, under [Documentation of SOEP-Core](#).

To get an overview of the data types, a somewhat simplified categorization helps: There are *Tracking Data* and *Survey Data* files which describe the development of the sample, such that the user knows which individual or household was part of the interviewed sample in any given year. Then there are *Original Data* files, which contain the data from each year's questionnaires without any changes except for very basic consistency checks. To help the user with the data, there also are *Generated Data*. These contain consistently coded variables across all waves with common names, such that the users can easily use this information when combining datasets across waves. The SOEP also provides various data on the respondent's background, called biographical data. Biography data in general can conceptually be separated into biographical data which are unchanging (such as information on parent's education, or data from the Mother-Child Questionnaires) and data which may be updated through changes in a respondent's life (such as new children in the birth biography, or a job change in the job history). Some of the changing data are stored as *Spell Data*. For each spell there is a definition of the spell type, begin point, end point and the censoring status, indicating if a given employment or income spell is censored (left and/or right) or uncensored. One of the biggest assets of the SOEP data is their longitudinal nature, i.e., repeated observations of the same unit (individual or household) over time. That's why we provide longitudinal datasets, such as PL or HL. Finally, there are some files which cannot be easily categorized - some are one-time datasets, some provide information about the interviewers, some about respondents outside of Germany.

There are two datasets which should be the building block of any analysis, as they allow users to define longitudinal populations very easily: PPATHL and HPATHL. HPATHL includes all households which have been interviewed successfully at least once. Similarly, PPATHL contains all individuals who have ever lived in a household that has participated in the SOEP, i.e., that has been captured in HPATHL, including non-respondents and children. Both data files contain one record per household or individual, respectively, with wave-specific variables for each year's survey status. In addition to some time-invariant information (like gender, year of birth, migrant status), these files contain all necessary identifiers to combine other files with PPATHL and HPATHL. Although they provide essential information, PPATHL and HPATHL alone are of little use for actual analyses. The most often used sources for additional information in SOEP-Core are the cross-sectional data files provided in each survey year (or "wave") or the datasets in the "long" Format.

The SOEP datasets can be viewed based on their content classification (Tracking Data, Original Data, Survey Data, Generated Data and Spell Data), the data structure (cross-sectional (cs), wide, long, spell) and also from the respondent's perspective. From the respondent's perspective, datasets can contain gross or net information. In addition, some datasets provide information only at the household level and others provide information at the individual level.

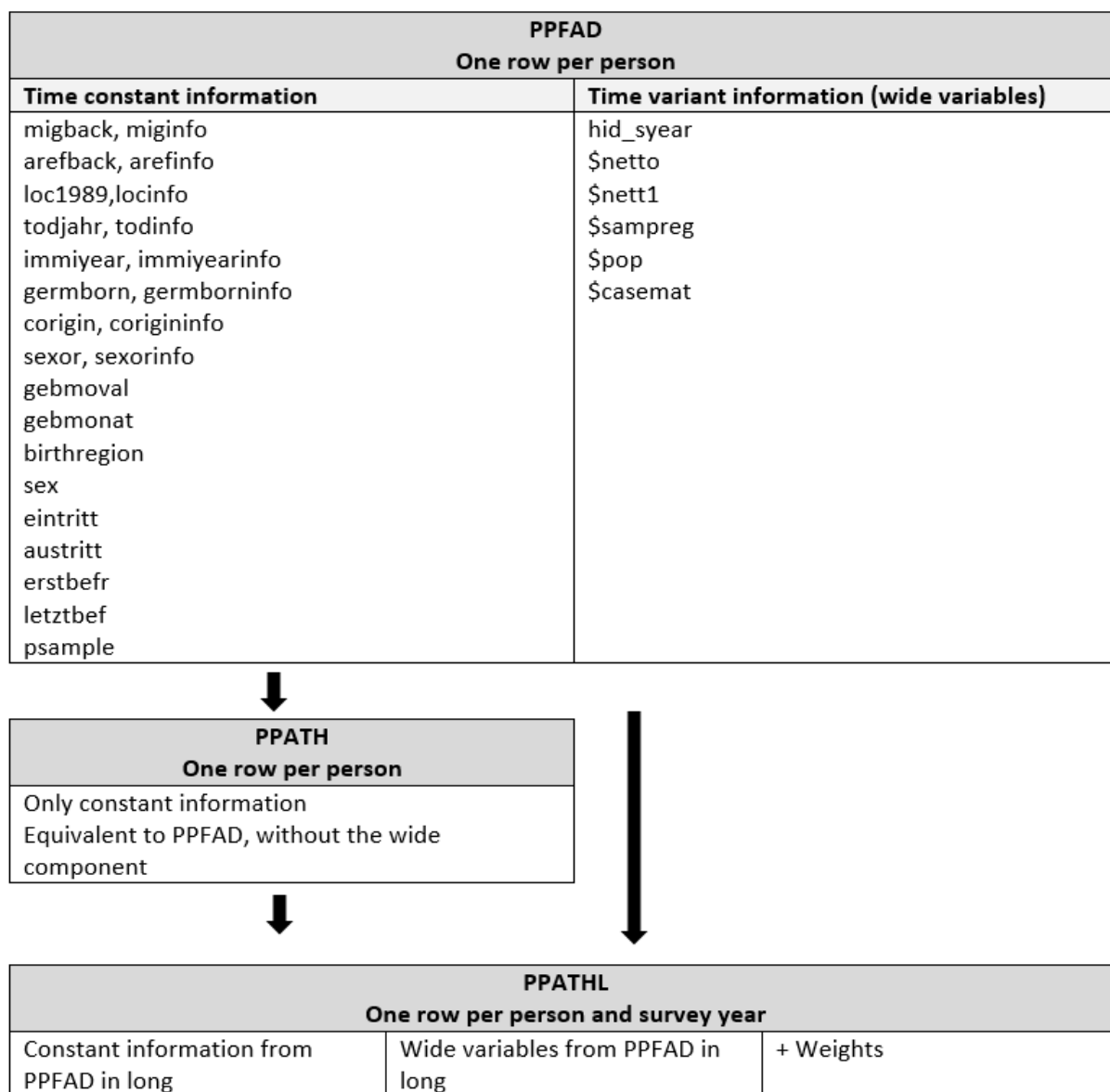
Individual Level Data		
Gross Sample	Net Sample	
pbrutto	Original Data	Generated Data
	pl	pgen
		pequiv
Household Level Data		
Gross Sample	Net Sample	
hbrutto	Original Data	Generated Data
	hl	hgen

Gross information at household or individual level is provided to users in the datasets hbrutto, hbrutt and pbrutto. Content information collected from household or individual questionnaires, for example, is original data and is stored in HL and PL. The SOEP team generates data from these original data, which are generated from the many SOEP questionnaires. New generated and user-friendly datasets such as pgen are created from the components of PL.

5.4.1 Tracking Data

Tracking data are the basis for linking your research-relevant variables. In addition to various demographic information, tracking data also provide information on how the interview was conducted. These datasets should be understood as initial data that you can use to merge your research-relevant variables via the individual and household numbers.

Dataset	Label	Format	Identifier (ID)	Additional Identifier
ppathl	Individual Tracking File	<i>long</i>	pid syear	hid cid parid
hpathl	Household Tracking File	<i>long</i>	hid syear	cid
pbrutto	Gross Individual Data	<i>long</i>	pid syear	hid cid intid hhnrold
hbrutto	Gross Household Data	<i>long</i>	hid syear	cid intid1 intid
hbrutt	Original gross population of a first wave sample	<i>long</i>	hid syear	cid
instrumentation	Data about the usage of the instruments	<i>long</i>	hid pid pid2 instrument syear	
pbr_exit	Cumulated Exit	<i>long</i>	pid syear	hid cid hnrold



hpathl “Household Tracking File” (long): HPATHL consists of all waves of the raw datasets HPATH and HHRF. For all years since 1984, the HPATHL dataset contains information on all households that have ever participated in the SOEP survey at any point in time. HPATHL is important for the delimitation of the unit of investigation (household), especially in longitudinal analysis. HPATHL is useful particularly for household analysis and can be used for pre-selection of specific households.

ppathl “Individual Tracking File” (long): PPATHL consists of all waves of the raw datasets PPATH and PHRF. For all years since 1984, the PPATHL dataset contains information on all individuals who have ever lived in a SOEP household at the point in time of a survey (i.e., all respondents, but also children under 17 years of age and individuals who have never given an interview). PPATHL is important for the delimitation of the units of investigation (individuals), especially for longitudinal analysis. It contains one record for each individual and year a individual has been a member of a respondent household. It is keyed on pid and syear, the survey year identifier. It contains the Household ID, the unvarying individual characteristics, individual weights, as well as the response status for that individual in each wave.

pbrutto “Gross Individual Data” (long): PBRUTTO consists of all waves of the raw datasets \$PBRUTTO. PBRUTTO covers all respondents who were either interviewed for the first time or contacted for the purpose of being interviewed again in a given wave. The dataset provides gross information on all SOEP respondents’ interviews as well as their positions in the panel framework.

hbrutto “Gross Household Data” (long): HBRUTTO consists of all waves of the raw datasets \$HBRUTTO. HBRUTTO covers all households that were successfully interviewed for the first time in a wave or were contacted for the purpose of being interviewed again. The datasets provide gross information on all SOEP households’ interviews as well as their positions in the panel framework.

hbrutt “Original gross population of a first wave sample” (long): The dataset HBRUTT contains demographic information and data on the interviews of all newly surveyed samples in the respective survey year that were successfully interviewed or contacted for the first time. All cross-sectional variables from HBRUTT\$\$ are used for hbrutt.

instrumentation “Data about the usage of the instruments” (long): The instrumentation data set records at the individual level which survey instruments are intended for each person and what the respective realization status is. If the instrument contains information about another person (e.g. for children’s instruments), the number of the third person is also stored in the data record. This way, each line is uniquely identified by the household ID, person ID and instrument ID (hid, pid, instrument and (pid2)). Basis for the decision which instruments are provided is the realized household matrix. If this is not available for a household, the information from the general sample is used instead. Missing information (e.g., which person in the household is supposed to provide information about a child) are stored in the data set with the code -1.

pbr_exit “Cumulated Exit” (long): The dataset pbr_exit is a supplement of pbrutto for individual dropouts. Individual dropouts are removed from the original pbrutto population, so that pbrutto covers all current household members. Pbr_exit contains the corresponding register information on individual dropouts from households.

5.4.2 Original Data

These datasets contain respondents’ direct information. The contents of these variables mirror the contents of the survey instruments. By searching the questionnaires, you can determine the exact wording of the question and obtain possible filter guidance.

Dataset	Label	Format	Identifier (ID)	Additional Identifier
pl	Individual questionnaire	<i>long</i>	pid, syear	hid, cid, intid
hl	Household questionnaire	<i>long</i>	hid, syear	cid, intid
biol	Biographical data	<i>long</i>	pid, syear	hid, cid, intid
jugendl	Youth questionnaire for first-time respondents at age 17	<i>long</i>	pid, syear	hid, cid, intid
kidlong	Data on children	<i>long</i>	pid, syear	hid, cid
more_docu	Dataset on the Mentoring of Refugees (MORE) Project	<i>long</i>	pid, syear	hid, cid
more_local	Dataset on the Mentoring of Refugees (MORE) Project	<i>long</i>	pid, syear	hid, cid
plueckel	Follow-up questionnaire	<i>long</i>	pid, syear	hid, cid, intid

continues on next page

Table 9 – continued from previous page

Dataset	Label	Format	Identifier (ID)	Additional Identifier
abroad	Questionnaire for respondents who have moved abroad	<i>long</i>	pid, year	hid, cid
vpl	Deceased individual	<i>long</i>	vpid, year	hid, cid, intid

pl “Individual questionnaire” (*long*): The PL dataset contains all waves of the \$P datasets from SOEP-Core. In addition, the PL file includes all variables of all waves of the datasets \$POST and \$PAUSL. This means that the PL dataset contains all variables from the individual questionnaire for all waves. In addition, the individual-specific data from the IAB-SOEP Migration Survey and IAB-BAMF-SOEP Refugee Survey are integrated into the PL dataset.

Attention: For large datasets we recommend the use of Stata/MP or Stata/SE on a computer with an internal memory of 16GB. Users can still work with the data in Stata/IC or on less powerful computers, but some modifications allow users to work effectively with even the largest datasets while placing low demands on their hard- and software.

```
clear
global data = "\\hume\rdc-prod\complete\soep-core\soep.v35"

* Search all available pl variables on paneldata.org: https://paneldata.
org/soep-core/data/pl
describe using "$data/pl.dta"

use pid hid cid syear plh0149-plh0151 using "$data/pl.dta"
```

hl “Household questionnaire” (*long*): HL contains all waves of the datasets \$H from SOEP-Core. This means that the HL dataset includes all questions of the household questionnaire. In addition, the household-specific data from the IAB-SOEP Migration Survey and IAB-BAMF-SOEP Refugee Survey are integrated into the original HL dataset.

biol “Biographical data” (*long*): BIOL contains cumulated individual-level raw data from the biographical questionnaire and from wave-specific biographical modules of the individual questionnaire. BIOL is intended to be used in addition to the generated biographical files (by advanced users) to complete (or modify) generated biographical variables.

jugendl “Youth questionnaire for first-time respondents at age 17” (*long*): JUGENDL contains the waves q (2000) up to the current wave of \$JUGEND in SOEP-Core. Since 2000 (wave Q), first-time respondents between the age of 16 and 17 have received a separate biographical questionnaire with additional age-group-specific questions, for instance, about their relationship to their parents or about what they do in their free time.

MORE_Docu “Dataset on the Mentoring of Refugees Projekt”: A dataset on the Mentoring of Refugees (MORE) project. Carried out in partnership with Start with a Friend (SWAF), this project aimed at bringing refugees and locals together to form friendships. This dataset contains information on German contacts provided to refugees. Information about the federal state of the SWAF location is included in the EU data edition. More localized information (county or municipality) is only available remotely or on site.

MORE_Local “Dataset on the Mentoring of Refugees Projekt”: A new dataset on the Mentoring of Refugees (MORE) project. Carried out in partnership with Start with a Friend (SWAF), this project aimed at bringing refugees and locals together to form friendships. This dataset contains information from the surveys of the locals in the project.

plueckel “Catch-up questionnaire” (*long*): The PLUECKEL dataset contains all waves of the \$PLUECKE datasets in SOEP-Core. Temporary drop-outs (“gaps”) can cause problems for longitudinal analyses. This has especially negative consequences for the employment and income data. That is why the SOEP tries to fill in at least some of the key missing information. PLUECKEL is a small questionnaire covering information on the year previous to which the

temporary drop-out occurred. It covers questions on job-related changes, employment calendar, income, education, and qualifications.

abroad “Questionnaire for respondents who have moved abroad” (long): With the pilot study “Life outside Germany” in 2008, the longitudinal SOEP study ventured into completely uncharted methodological territory by attempting to locate the addresses of former SOEP respondents who have since moved abroad and to survey these individuals with the help of a specially developed written questionnaire on the reasons for their move. The project was discontinued due to insufficient case numbers in 2014.

vpl “Questionnaire on the deceased individual” (long): The VPL dataset contains all waves of the \$VP datasets of SOEP-Core. The VPL file contains information about respondents who lost a relative in the previous year. It provides information about the deceased individual and the respondent who reported the death.

5.4.3 Survey Data

These datasets contain information on survey methodologies used in SOEP-Core. The various datasets contain detailed exit information provided by respondents and the household weighting factors that users need for representative analysis.

Dataset	Label	Format	Identifier (ID)	Special Identifier
design	Survey design	<i>long (time-constant)</i>	cid	intid
pbr_hhch	PBR_HHCH	<i>long</i>	pid, syear	hid, cid, pnralt, pnrneu, hhnrold

design “Survey design”: The dataset DESIGN provides information on the stratified sampling of the SOEP in the form of two variables. The variable STRAT identifies each of the discrete sampling groups described above. Altogether, the SOEP consists of 40 strata: one stratum in sample A, twenty-seven in sample B, one in sample C, three in sample D, one in sample E, two in sample F, four in sample G, and one in sample H. Each of these strata have unique inclusion probabilities. The variable design contains the inverse of this probability, i.e., the design weight.

pbr_hhch “PBR_HHCH”: The dataset pbr_hhch is a subfile of pbrutto that was used from 1984 to 2009 to identify individuals from households that underwent split-offs in subsamples A-H.

5.4.4 Generated Data

The SOEP team has prepared these datasets for easy use and subjects them to additional plausibility checks and quality controls prior to data release. In most cases, they consist of several variables and different survey instruments and are described in the documentation provided. As a result, these datasets cannot be assigned 1:1 to a single survey instrument.

Dataset	Label	Format	Identifier (ID)	Additional Identifier
pgen	Generated individual data	<i>long</i>	pid, syear	hid, cid, pgpartnr
hgen	Generated household data	<i>long</i>	hid, syear	cid
bioagel	Generated biographical information	<i>long</i>	pid, syear, persnre	hid, cid,
biopupil	Generated biographical information	<i>long</i>	pid, syear	hid, cid
kidlong	Data on children	<i>long</i>	pid, syear	hid, cid
pequiv	Cross National Equivalent File	<i>long</i>	pid, syear	hid, cid
biobirth	Generated biographical information	<i>wide</i>	pid	cid, kidpnr01- kidpnr15
bioedu	Generated biographical information	<i>long</i> (time-constant)	pid	cid
bioimmig	Generated biographical information	<i>long</i>	pid, syear	hid, cid
biojob	Generated biographical information	<i>long</i> (time-constant)	pid	cid
bioparen	Generated biographical information	<i>long</i> (time-constant)	pid	cid, fnr, mnr
biosib	Generated biographical information	<i>wide</i>	pid	cid, sibpnr1-sibpnr11
biotwin	Generated biographical information	<i>wide</i>	pid	cid, pnrtwin, pnrtrip, pnrquad
camces	Highest educational qualification, migrants sample M1 and M2	<i>long</i>	pid	hid, syear, cid
cogdj	Data on cognitive tests (Youth)	<i>long</i>	pid	syear, cid
cognit	Data on cognitive potential	<i>long</i>	pid	syear, cid, intid
cog_refu	Data on cognitive tests (Refugees)	<i>CS</i>	pid	syear, cid, hid
gripstr	Grip Strength Measures	<i>long</i>	pid, syear	cid, intid
hconsum	Household Consumption Module	<i>CS</i>	hid	syear, cid
health	Data on health indicators	<i>long</i>	pid, syear	cid
hwealth	Wealth module	<i>long</i>	hid, syear	cid
interviewer	Data on the SOEP interviewer	<i>long</i>	intid, syear	cid
mihinc	Multiple imputed data on monthly household income	<i>long</i>	hid, syear	cid
pflege	Persons needing care within the household	<i>long</i>	pid, syear	cid
pkal	Individual calendar	<i>long</i>	pid, syear	hid, cid
pwealth	Wealth module	<i>long</i>	pid, syear	hid
timepref	Experiment on time preferences	<i>CS</i>	pid	hid, syear, cid
trust	Experiment on trust	<i>long</i>	pid	hid, syear, cid

pgen “Generated individual data” (long): PGEN contains all waves of the \$PGEN datasets in SOEP-Core. The PGEN-file contains user-friendly data on the individual level that are consolidated from different sources. The plausibility is validated longitudinally in many respects, making the data superior to those in PL in most situations. The file contains one row for each individual (pid is unique) with a completed individual or youth questionnaire.

hgen “Generated household data” (long): HGEN contains all waves of the \$HGEN datasets in SOEP-Core. In order to minimize computational effort for the user, the SOEP provides yearly status variables on the household level. The HGEN data provide a set of time-invariant variables generated from the SOEP household questionnaire. They only include households that participated in the respective year.

bioagel “Generated biographical information” (long): The BIOAGEL data files are generated using information collected in the “Mother & Child” and “Parent” questionnaires. BIOAGEL is now provided in one dataset.

biopupil “Generated biographical information” (long): The BIOPUPIL data files are generated using information collected in the “Pre-Teen” and “Early-Youth” questionnaires. BIOPUPIL is provided in one dataset.

kidlong “Data on children” (long): The variables stored in the KIDLONG file are based on the information collected annually and contained in the wave-specific \$KIND files. The relevant information is not provided by children themselves but is obtained from answers to questions in the household questionnaire provided by the respondent within the household (usually the head of the household). This data is reaggregated at the individual level and stored as child-specific entries in the file \$KIND.

pequiv “Cross-National Equivalent File” (long): PEQUIV contains all waves of the \$PEQUIV datasets in SOEP-Core. The PEQUV-File is based on the Cross-National Equivalent File (CNEF) with extended income information for the SOEP. This file comprises not only the aggregated income figures from CNEF but also additional separate income components.

pkal “Individual calendar” (long): PKAL contains all waves of the \$PKAL datasets in SOEP-Core. The PKAL datasets contain calendar variables from the individual questionnaire. The dataset includes the individual’s employment or educational status on a monthly basis as well as the individual’s income status.

biobirth “Generated biographical information” (wide): The file BIOBIRTH provides information on fertility histories of adult respondents in the SOEP. Up to 2014 (version 30, wave BD), the data were stored in two separate files: BIOBIRTH containing female fertility histories, and BIOBRTHM providing male fertility histories. Fertility histories in BIOBIRTH provide information on every woman (as well as every man with panel entry since 2001) who has ever completed at least one SOEP interview.

Outdated bioedu “Generated biographical information” (long time constant): The SOEP contains a broad range of variables on early childhood education and care, educational participation, educational degrees, and related topics. The BIOEDU dataset is designed to provide ready-made variables on educational transitions and related topics for use in longitudinal analysis.

bioimmig “Generated biographical information” (long): The variables contained in BIOIMMIG relate to foreigners in (and migrants to) Germany. Questions deal with the desire to return to the home country, the presence of relatives in the home country, reasons for coming to Germany, and conditions upon initial arrival in Germany.

Outdated biojob “Generated biographical information” (long time constant): The purpose of BIOJOB is to provide a file that offers the user convenient access to biographical information on past job activities. BIOJOB consists of generated variables as well as plain questionnaire information. Up to now, all but two variables in BIOJOB are time-invariant. Information on occupational changes and on the age at the most recent change of occupation refer to the date of the respondent’s biography interview.

Outdated bioparen “Generated biographical information” (long time constant): The dataset BIOPAREN contains biographical entries on the parents’ and respondent’s background. The information in BIOPAREN is obtained from two sources: from proxy entries by children on their parents in the biography questionnaire and youth questionnaire, and from direct entries by parents when the respondent lives in the same household as the parents. Please note that BIOPAREN focuses on the social parent. Biological parent identifiers can be found in BIOBIRTH.

Outdated biosib “Generated biographical information” (wide): BIOSIB provides information on siblings living within SOEP households. The dataset contains the individual identifiers of all siblings in a SOEP household. It includes

information on the individual sibling's sex, year of birth, number of siblings, position in birth order, and relationship between siblings.

Outdated biotwin “Generated biographical information” (wide): The file BIOTWIN contains all twins that were ever identified within the SOEP. To be classified as a twin, a individual is required to have exactly the same age as his or her sibling (year & month of birth), have a relationship to the head of the household that indicates that he or her and a second individuals are siblings, and have the same mother (as far as a pointer to the mother is available). Furthermore, it is not only twins that are recorded in the BIOTWIN dataset, but also triplets or quadruple siblings.

camces “Highest educational qualification, migrant samples M1 and M2” (CS): The CAMCES-File provides information about computer-assisted measurement and coding of educational qualifications in surveys.

cogdj “Data from cognitive tests (Youth)” (CS): In SOEP 2006, a separate questionnaire with cognitive tests for adolescents was used for the first time: “Lust auf DJ”. The acronym “DJ” stands for “Denksport und Jugend” (mind sports and youth), but it was named for its more common association with “disc jockey”. The questionnaire “Lust auf DJ” was created for all respondents aged 16-17.

cognit “Data on cognitive potential” (long): In the 2006 survey year, for the first time, short cognitive tests were carried out with a subsample of the SOEP. The goal was to employ a robust set of instruments that could be administered easily by trained interviewers within just a few minutes. COGNIT06 provides the aggregated sum scores (total values for three time packages, so-called “parcels” of 30, 60 and 90 seconds).

cog_refu “Data on cognitive tests (refugees)” (CS): The dataset contains sum scores for two competence measurements (previous school knowledge and basic cognitive skills) of youths born in 2000, 2003 and 2005 surveyed in 2017.

gripstr “Measures of grip strength (left and right hand)” (long): The data on grip strength from the survey year 2012 is now included in the GRIPSTR dataset.

hconsum “HH consumption module” (CS): We were faced with three methodological challenges in generating the final consumption data. First, due to the design of the consumption module, inconsistent answers arose between the amounts give for monthly and annual consumption. Second, there was the common problem of missing data, here in particular item nonresponse. And third, consumption data are usually blurred by heaping. For researchers who do not want their consumption variables to include changes from all steps of data preparation, the new dataset “HCONSUM” contains not only the prepared consumption variables but also flag variables providing researchers the opportunity to select individual solutions.

health “Data on health indicators” (long): Starting in 2002, the SOEP health module in the individual questionnaire has been revised and replicated at two-year intervals. In the HEALTH file, users find, for instance, the generated variables on height and weight with imputation flags and a user-friendly longitudinal checked generated variable for Body Mass Index (BMI).

hwealth “Wealth module” (long): The generated SOEP wealth data is stored in two separate data files called PWEALTH for information at the individual level and HWEALTH for correspondingly aggregated data at the household level. HWEALTH contains all information on the household level; it is purely the result of aggregating the individual-level information in PWEALTH. However, for all individuals with valid household-level information who did not respond to the individual questionnaire (partial unit non-response), imputations have been carried out and the results are included in HWEALTH.

interviewer “Data on the SOEP interviewer” (long): The SOEP aims not only to collect high-quality data on the living conditions and well-being of households, but also to provide a valuable empirical source for survey research. The INTERVIEWER file provides users with easy access to all available longitudinal information on the SOEP interviewers.

mihinc “Multiple imputed data on monthly household income” (long): The dataset MIHINC contains the complete imputation results and is available separately. To be compatible with methods for analyzing multiply imputed data, MIHINC is constructed in the “stacked” or MIM data format. It contains the following variables: HHNRAKT, SVYYEAR, MJ, MI, IHINC and IMPFLAG. Since 1995 for every survey household in all survey years, there are ten imputed values for current household income.

pflege “Persons needing care within the household” (long): Since wave B (1985), the SOEP household questionnaire includes questions on household members in need of care. In order to support individual-level analysis, this information

has been restructured and is stored in the cumulative file PFLEGE.

pwealth “Wealth module” (long): For the first time in 2002, the individual questionnaire included a special module focusing on wealth. It included questions on seven different wealth components: owner-occupied property (including debt), other property (including debt), financial assets, private pensions (including life insurance and building savings contracts), business assets, tangible assets, and consumer credit. The generated SOEP wealth data are stored in two separate data files called PWEALTH for information at the individual level and HWEALTH for correspondingly aggregated data at the household level. Wealth-related variable names in the file PWEALTH consist of six digits. The first digit tells the user which wealth component is referred to, and the second to sixth digits provide more detailed information about possible filter information, the personal share, the gross amount, and the amount of any outstanding debt. In principle, a digit is coded “1” if a given variable does indeed contain this specific piece of information and “0” otherwise. The wealth information in the SOEP questionnaire is surveyed at the individual level and thus also imputed or edited at the individual level (although checked against household information for consistency).

timepref “Experiment on time preferences” (CS): Following the behavioral experiment on trust and trustworthiness carried out in the 2003, 2004, and 2005 SOEP surveys, the experiment “time preferences” was run in 2006. In this experiment on economic behavior, respondents were asked to decide how they would want to receive €200 in prize money: if they would want to receive it immediately by check or if they would want to wait and receive a larger amount later, that is, with interest.

trust “Experiment on trust” (long): The economic behavior experiment on trust and trustworthiness from survey years 2003, 2004, and 2005 served to measure trust based on an investment game, a one-off game for two players who interact anonymously. The first player receives a credit of ten points and can overwrite any number of points of the second player. Each overwritten point is doubled. The second player also receives a credit of ten points. After receiving the (doubled) points from the first player, the second player decides how much of her own credit she will transfer to the first player (zero to ten points). As with the first transfer, the recipient’s points are doubled. After the decision of the second player, the game ends and the other players are paid (one point corresponds to one euro, the total is paid by check a few days later). The trust dataset thus contains the information from all three waves in which the behavioral experiment was conducted.

5.4.5 Spell Data

Spell, duration, and event history data are used frequently in the social sciences. In the strict sense of the word, spell data are about time periods with a defined start and end. General information about the data structure of spell data can be found in the chapter *Data Structure in Spell Format (spell)*

Working with spell data:

[Working with spell data \(pdf\):](#)

[Working with spell data \(do-files\):](#)

How to generate spell data from data in wide format: Based on the migration biographies in the IAB-SOEP Migration Sample:

[Generating spell data:](#)

Dataset	Label	Format	Identifier (ID)	Additional Identifier
artkalen	Spell data from the activity calendar	<i>spell</i>	pid	cid
biocouplm	Generated biographical information	<i>spell</i>	pid	cid, coupid
biocouply	Generated biographical information	<i>spell</i>	pid	cid

continues on next page

Table 10 – continued from previous page

Dataset	Label	Format	Identifier (ID)	Additional Identifier
biomarsm	Generated biographical information	<i>spell</i>	pid	cid
biomarsy	Generated biographical information	<i>spell</i>	pid	cid
einkalen	[deprecated] Spell data on income	<i>spell</i>	pid	cid
lifespell	Spell Information on the pre- and post-survey history of SOEP respondents	<i>spell</i>	pid	cid
migspell	Migration history	<i>spell</i>	pid	cid
pbiospe	Generated biographical information	<i>spell</i>	pid	cid
refugspell	Migration history	<i>spell</i>	pid	cid
sozkalen	[deprecated] Spell data on social benefits	<i>spell</i>	hid, cid	

artkalen “Spell data from the activity calendar” (long): The ARTKALEN contains spells (monthly) for events starting in January 1983. This is in contrast to PBIOSPE, where spells were in yearly durations, and events previous to 1983 were included. The information on activity status is collected on a monthly basis in the yearly individual questionnaire and stored in the file ARTKALEN.

biocouplm “Generated biographical information” (long): With the BIOCPLM the SOEP provides consistent and continuous partnership histories for nearly all adult respondents. BIOCPLM is built on the prospective information at the time of each interview. The relationship histories are collected on a monthly basis from all adult SOEP participants since their entry into the SOEP.

biocouply “Generated biographical information” (long): With the BIOCPLY, the SOEP provides consistent and continuous partnership histories for nearly all adult respondents. BIOCPLY is built on retrospective and prospective information at the time of each interview. The relationship histories are provided on an annual basis.

biomarsm “Generated biographical information” (long): With BIOMARSM the SOEP provides consistent and continuous marital histories for nearly all adult respondents. BIOMARSM is built on the prospective information at the time of each interview. The marital histories are collected on a monthly basis from all adult SOEP participants since their entry into the SOEP.

biomarsy “Generated biographical information” (long): With BIOMARSY the SOEP provides consistent and continuous marital histories for nearly all adult respondents. BIOMARSY is built on retrospective and prospective information at the time of each interview. The marital histories are provided on an annual basis.

lifespell “Spell information on the pre- and post-survey history of SOEP respondents” The SOEP team regularly conducts follow-up studies to relocate attriters. These studies draw on official register data and allow us to determine whether a individual is still living in Germany, is deceased, or has moved abroad since the last SOEP interview. The information is combined in a spell file LIFESPELL. This dataset reports all available information on the pre- and post-survey history of all individuals who have ever been a member of a SOEP household.

Outdated migspell “Migration history” (long): MIGSPELL is derived from the migration biographies, which are collected from each new respondent of the IAB-SOEP migration samples M1 and M2. It contains data on moves by foreign-born migrants as well as on stays abroad by German-born respondents.

pbiospe “Generated biographical information” (long): The spell file PBIOSPE is based on the information on activity status over the life course, which is collected as a matrix from every respondent who completes the biographical

questionnaire. The observations start at the age of 15 and end at the current age (up to age 65). To update ongoing employment information in PBIOSPE, information from the yearly individual questionnaire is also used.

Outdated *refugspell* “Migration history” (long): For migration biographies in the refugee samples, we created the spell dataset REFUGSPELL. The variables in MIGSPELL and REFUGSPELL are derived from different instruments and only partially overlap. The data structure allows the dataset to be linked with MIGSPELL if desired.

Outdated *sozkalen* “Spell data on social benefits”: The file SOZKALEN provides spell data on households receiving social assistance, defining the beginning, end, and censoring status of any period of receiving 3 different types of assistance. This file is set up using information from the calendar that is collected for the previous year (between 1992–2000). Thus, it contains information on a monthly basis.

Last change: Jul 24, 2023

5.5 Data Processing

The following overview shows which datasets form the basis of each questionnaire, and the respective data processing process, from the questionnaire to the wave-specific datasets, to the prepared “long” datasets. Please note that not all datasets are based on questionnaires but that many have been prepared meticulously by our staff. The table therefore does not show the full range of datasets available.

Year	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Version	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	
Questionnaire / Wave	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	ba	bb	bc	bd	be	bf	bg	bh	bi	bj	bk	bl	long Dataset
Household	Shost																Sh, Skind												hl, kidl										
Individual	Spost																Sp, Spkal												pl, pkal										
Biography	biotelela																Stela*												biol										
Individual with Biography																	Sp, Stela*												pl, biol										
Mother-Child-Instruments	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Mother-Child (Newborns)																	Smuki*												bioagel										
Mother-Child (2-3-year-olds)																	Smuki2*												bioagel										
Mother-Child (5-6-year-olds)																	Smuki3*												bioagel										
Parents (7-8-year-olds)																	Selt*												bioagel										
Mother-Child (9-10-year-olds)																	Smuki5*												bioagel										
Youth Instruments	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Pre-Teen																	Sschool												biopupil										
Early Youth																	School2												biopupil										
"Lust auf DJ"																	cogdj																						
Youth																	Sjugend												jugendl										
Additional Instruments	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Catch-Up Individual (Re-Questioning)																	Spluecke												plueckel										
Deceased Individual																	Svp												vpl										
Abroad																													abroad										
Cognitive Test																													cognit										
Grip Strength Test																													gripstr										
SOEP CoV																													cov										
SOEP CoV 2																													cov										

In addition to the classic SOEP survey instruments, there are also a large number of sample-specific questionnaires whose information flows into other unlisted raw datasets (e.g. \$pausl, \$post, \$pkalost etc.). The chapter [Sample-Specific Questionnaires](#) explains why such special survey instruments exist, how they become raw datasets and in which long datasets these variables can be found.

Last change: Jul 24, 2023

5.6 Dataset Identifiers

Because of the overall data structure with data on different observational levels, any analysis requires the combination of data using matching or merging procedures. These merging procedures need identifiers such that a combination of datasets becomes feasible. The central individual identifier across time is **pid**, which is fixed over time (and of course datasets). Since a person might change the household in which he or she lives at any point in time, yearly household identifiers called **hid** are necessary, facilitating matching depending on the dataset used. Finally, each individual (respondents as well as children) can be traced back to be a member of or a split-off from an original household from the very first wave. This household's ID, which is fixed no matter how often a person changes households over time, is called **cid**. In addition, respondents in long data can be differentiated by survey year. The **syear** variable can be used to

identify a respondent's survey year. The SOEP provides additional identifiers in the various datasets in order to identify respondents and to provide further possibilities for merging datasets. A excerpt of these additional identifiers can be found here:

Please note that these are not all identifier variables. The name of the identifier variable can change depending on the dataset used.

- **parid** "Unchanging Individual identifier of Partner (PID)"
- **pgpartnr** "Individual Identifier of Partner"
- **coupid** "Couple Identifier"
- **intid** "Interviewer Identifier"
- **intid1** "Identifier of First Interviewer"
- **vpid** "Individual Identifier of Deceased Individual"
- **mnr** "Individual Identifier Mother"
- **fnr** "Individual Identifier Father"
- **kidpnr01-kidpnr19** "Individual Identifier nth Child"
- **sibpnr1-sibpnr11** "Individual Identifier, nth Sibling"
- **pnrtnw** "Individual Identifier 2nd Sibling"
- **pnrtrip** "Individual Identifier 3rd Sibling"
- **pnrquad** "Individual Identifier 4th Sibling"

5.6.1 Partner Identifier

Partner identification (parid and pgpartnr)

Partner indicators (parid from ppathl and pgpartnr from pgen) 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). 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 from pbrutto), the partnership biography in the lifehistory calendar reported by new respondents, 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. parid includes all persons that have ever participated in the SOEP. pgpartnr from the pgen dataset contains the same information, but is restricted to the pgen population and includes only persons with persons interview.

Monthly Couple Identifier (coupid)

The COUPID can only be found in the biocouplm dataset, because partnerships can change many times within a year. Multiple partnerships within a year can only be recorded correctly if the partnership is assigned to the exact month. So if a partnership for example ends in March, its joint COUPID also ends and a new one begins in June or so.

5.6.2 Family Identifier

Individual Identifier nth Child (kidpnr01-kidpnr19)

kidpnr01-kidpnr19 in the data set biobirth contains invariable individual identifier of biological children[nn] of the respondent (for the first child up to the 19th child), given it is identifiable in the SOEP. The sequence of children within biobirth is recorded with regards to the birth order in terms of age of the children. The order ranks from the oldest child specified under kidpnr01 to the youngest child. If the age is missing it is listed in the first record (kidpnr01), and in subsequent records following kidpnr01 if more than one child's personal identifier remains missing. kidpnr[nn] is "-1" if a child was reported in the birth biography who could not be assigned to a SOEP household (Children outside the parental household). If no child could be identified in the household context or in the birth biography, the code "-2" is assigned.

Individual Identifier Mother/Father (mnr fnr)

The personal ID of the parents (fnr and mnr) from bioparen is generated in three steps:

- The parents of the respondent are identified by the relationship to the head of the household (stell in pbrutto). Ideally, the children's parents are identified at the time of the first survey of the child. Furthermore, the social parents and not necessarily the biological parents are identified.
- The parents of the respondent are identified via the mother's ID as well as the mother's partner ID in \$\$kind. By using these variables the "oldest" parents are identified. Ideally, these are the parents at the time the child is 17 years old (one year before the first survey).
- The biological mother-ID and father -ID of the respondent can be identified in biobirth.

As bioparen aims at identifying the social parents that live in the household when the child is surveyed, the steps above are carried out in the hierarchy 1-3 with step 1 having the highest priority. If one is interested in only biological parents, please have a look at the information in biobirth

Individual Identifier, nth Sibling (sibpnr1-sibpnr11)

The variables provide the never changing person IDs for the siblings of the individual identified by PID. The sibling relationship is generated from the parent information in biobirth and bioparen. Two persons are defined as siblings if they report both, the same mother and father, only the same mother, or only the same father. This information on the sibling relationship is stored in sibdef1-sibdef11. In the case of inconsistent information on parents in biobirth and bioparen, bioparen was assigned the lowest priority. Please note, that bioparen uses a social definition of parenthood based on cohabitation. In contrast, biosib contains both biological (biobirth) and social siblings with a higher priority on biological relations.

Individual Identifier Twins, Triplets, Quadruplets, (pnrtwin, pnrtrip, pnrtquad)

The ids pnrtwin, pnrtrip and pnrtquad from biotwin contain all twins that were ever identified within the SOEP. pnrtwin and – in rare cases if available – pnrtrip or pnrtquad contain the individual identifier of second, and third or fourth sibling in the group. This means that every case in the data set consists of a group of twins (or triplets or quadruplets). The code “-2” is assigned to pnrtrip and/or pnrtquad if a third or fourth twin sibling doesn’t exist. PERSNR and PNRTWIN however should always contain valid codes.

Individual Identifier of Deceased Individual (vpid)

vpid in the vpl data set contains the individual identifier of a deceased individual and is difficult to interpret because

- SOEP respondents in a household may provide information on several deceased persons. These deceased persons may or may not have participated in SOEP.
- Non-SOEP respondents provide information about one (or even more) deceased SOEP person(s).

So the following scenarios can occur:

	Deceased person in SOEP	Deceased person NOT in SOEP
Respondent is interviewee in SOEP	PID, VPID	PID, -/-
Respondent is NOT interviewee in SOEP	-/-, VPID	-/-, -/-

This means a person with a pid=0 has not been part of the SOEP, but has completed a deceased person questionnaire for a deceased person from a SOEP household (vpid has a SOEP ID). When individuals with SOEP ID (pid) report a deceased person who was not part of SOEP, special vpid are assigned:

pid	syear	vpid
32634001	2016	32634099
32634001	2017	32634099
32683702	2016	32683799
32683702	2018	32683798
32683702	2018	32683797

90s numbers in vpid are assigned to “non-Soep participants”. E.g. if a mother is deceased, but she did not participate in the Soep. In this example a No-SOEP-Person died in 2016, another No-SOEP-Person in 2017. Different 90s (e.g. 98/99) are only assigned if more than one No-SOEP-Person died in a year. When a SOEP person dies and is reported by another person, the vpid is the pid of the deceased respondent.

5.6.3 Interviewer Identifier

Interviewer ID (intid and intid1)

Intid and intid1 are fixed IDs over time to identify interviewers across years, households and questionnaires within datasets. The interviewer ID is used to identify the respective interviewer of different respondents. Unlike most other datasets in the SOEP the interviewer dataset has no PID or HID to identify the observations, but you can merge the interviewer information to other datasets using the intid. Due to changing IDs in the SOEP raw data in this and past versions of the interviewer data set it may happen that the intid of an interviewer changes over time. This can happen at most once per interviewer and is unfortunately not flagged. When the interviewer is replaced or when the interviewer changes over time, intid1 references to the first interviewer who conducted an interview in the survey household.

Last change: Apr 04, 2023

5.7 Versioning and Harmonization

In some cases, variables in long format with the same content but collected in different ways need to be harmonized to ensure that they remain consistent and comparable over time. Starting with SOEP Core v.34, SOEP offers versioning and harmonization solutions for such variables in all *Original Data* in long format. These versions and harmonizations are recognizable in the variable name. The “_v” suffix indicates possible differences in a variable. Harmonization suggestions generated by SOEP from the different versions of these variables can be recognized with the “_h” suffix. In general, particular caution is required when using variables marked “_v” or “_h”:

1.) Differences in Response Options

Variables are versioned and harmonized because the response options have changed over time.

2.) Differences in Coding of Response Options

Variables are versioned and harmonized because the coding of the response options has changed over time. Since the values of certain response options can change, it is not possible to easily integrate the various wave-specific variables into a variable in long format. The variable must be appropriately harmonized to be useable.

3.) Content Differences in the Questions.

Variables are versioned and harmonized because the questions were asked differently in different years, but the content belongs together. If the content or wording of the question changes, the wave-specific variables cannot easily be integrated into a long variable.

4.) Changes in Question Type.

Variables are versioned and harmonized because the questions were asked differently in different years, for example, as a question with multiple response options and later as a question with a single response option. A possible multiple answer in certain years makes it difficult to easily integrate the wave-specific variables into a variable in long format.

5.) Euro Harmonization

Variables are versioned and harmonized because they are metric and were surveyed as DM amounts before the introduction of the euro. For the long version of the variable, metric variables based on different currencies in different years are harmonized as euro amounts.

6.) Differences in Metric Variables

Variables are versioned and harmonized if they contain a year and were provided in the wave-specific raw data with different numbers of digits. The years are standardized and presented in the harmonized version with four digits. In addition, possible problems with decimal digits in metric variables from the raw datasets are corrected for the long-format variable.

7.) Different Respondents

Variables are versioned and harmonized when different groups of respondents have received different survey instruments and the variables have not been integrated into the wave-specific raw datasets. Special samples or a specific filtering in the questionnaire can lead to certain groups of people receiving different questions that belong together in terms of content. Such different variables are harmonized in the long version of the variable.

A more detailed explanation of the versioning and harmonization concept can be found in the section [Working with harmonized Variables](#)

Last change: Mar 28, 2023

5.8 Missing Conventions

Survey variables might be missing, that is, lacking a valid code or value, for different reasons. In the SOEP, negative values are not valid for any variable, but are used instead to code different reasons for missing information. There are two possible origins of missing values: the respondent's answer or the survey design. In the first case, the respondent may refuse to answer or not know an answer or may report invalid values. In the second case, the interview design may exclude respondents with certain characteristics from some questions (e.g., men will never be asked if they are pregnant). The following codes are used:

Code	Label
-1	No answer / don't know
-2	Does not apply
-3	Implausible value
-4	Inadmissible multiple response
-5	Not included in this version of the questionnaire
-6	Version of questionnaire with modified filtering
-7	Only available in less restricted edition
-8	Question not part of the survey program this year ¹

¹ Only applicable to datasets in long format.

A person might decline to answer a question. This occurs mainly with sensitive questions (e.g., income-related questions) and when respondents simply do not know the answer. In such cases, the missing code is “-1” for “no answer / don't know”. Note that the SOEP does not distinguish between a refusal to answer and a true “don't know”. Information may be missing when a question is not asked because it is not relevant to a specific person, e.g., owner-occupiers will not be asked about the amount of rent they pay. In such cases, the question “does not apply” to this person, and the variable receives a code of “-2”. Sometimes invalid answers occur when respondents fill out a PAPI interview themselves or the interviewer mistypes an answer (e.g., working hours over 168 per week). In such cases, multiple checks are carried out, and if the inconsistency remains, the variable is recoded “-3 Implausible value”. Some questions contain multiple answer possibilities and respondents are asked to pick one answer. In the SOEP PAPI questionnaires, respondents sometimes ignore this request and give more than one answer (e.g., “very good” and “good” when asked about their current health status). In such cases, if the correct answer cannot be determined from the questionnaire itself, the code “-4 Invalid Multiple Answers” is assigned to this variable. With the extensions to the SOEP in recent years, entirely new samples have been added to SOEP-core. In these samples, questions are sometimes left out completely, e.g., to shorten the questionnaire or because the focus of the sample is different (as is the case with SOEP-related studies). In such cases, the variable will be set to “-5 Not included in this version of the questionnaire” for an entire subsample. With the use of CAPI, recent developments include an “integrated” individual questionnaire, i.e., the biography part and the “regular” part of the questionnaire are combined into one questionnaire. Some of the questions in the biography part are repeated in the regular part. Whereas the respondent will answer the same question twice the PAPI mode, the CAPI allows the respondent to filter around the question if it has already been asked. These cases are very rare, but if they occur, they receive a code “-6 Version of questionnaire with modified filtering”. SOEP-Core offers a variety of different *editions* of the data. Due to data protection regulations, some variables of these *editions* may not be made

accessible. Variables with increased restrictions are for example variables that provide federal state level information. Because the variable may not be made accessible in a specific edition the federal state level information still remains in the data but they are assigned a missingcode “-7 Only available in less restricted edition”.

Last change: Apr 04, 2023

WORKING WITH SOEP DATA

The following exercises are taken from our SOEP Campus Workshops, a service especially for young scholars in the disciplines of sociology, economics, and psychology. Here we provide introductions to the use of the SOEP data.

In order to familiarize yourself with the SOEP data as best as possible as a new user, you should first familiarize yourself with the tracking data.

Tracking data are the basis for linking your research-relevant variables. In addition to various demographic information, tracking data also provide information on how the interview was conducted. These datasets should be understood as initial data that you can use to merge your research-relevant variables via the individual and household numbers.

6.1 Working with Tracking Data (PPATHL)

For all years since 1984, the PPATHL dataset contains information on all persons who have ever lived in a SOEP household when a survey was conducted (i.e., all adult respondents as well as children under 17 years of age and household members who have never given an interview). PPATHL is important in distinguishing research units (persons), especially for longitudinal analysis.

Time-constant information on individuals:

- Never Changing Person ID (pid)
- ID Household (hid)
- Gender (sex), year of birth (gebjahr), year of death if applicable (todjahr)
- Migration Background (migback)
- Sample Member (psample)
- Year Moved to Germany (immiyear)
- Country Born In (corigin)




Time-varying information from individuals:

- Survey Year (syear)
- Survey Status (netto)
- Sample Membership (pop)
- Survey Region in 1989 (East or West Germany) (loc1989)

The dataset is explained in more detail in the following documentation:

[Documentation PPATHL:](#)

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/log
- H:/material/exercises/output
- H:/material/exercises/temp

These are used to store your script, log files, datasets and temporary datasets. Open an empty do-file and define your paths with globals:

```

1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:/material/exercises"
5 global MY_IN_PATH "//hume/rdc-prod/distribution/soep-core/soep.v37/eu/
↩Stata/"
6 global MY_DO_FILES "$AVZ/do/"
7 global MY_LOG_OUT "$AVZ/log/"
8 global MY_OUT_DATA "$AVZ/output/"
9 global MY_OUT_TEMP "$AVZ/temp/"

```

Attention: Please note that until version 33 (v33), PPATH was called PPFAD. The following exercises are done with version 37 (v37).

The global „AVZ“ defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to the data you ordered.

Based on the data in PPATHL, answer the following questions:

1. Look at the two people with the Person IDs (pid) 2102 and 19202

a) What is their gender? When were they born and when (if applicable) did they die?

Open the PPATHL dataset. Search the dataset for variables that describe survey year, sex, year of birth and year of death. Display the information from the variables for individuals 2102 and 19202.

```

1 use "${MY_IN_PATH}ppathl.dta", clear
2 list pid syear sex gebjahr todjahr if pid == 2102 | pid == 19202

```

	pid	syear	sex	gebjahr	todjahr
646.	2102	1984	[2] Female	1927	1999
647.	2102	1985	[2] Female	1927	1999
648.	2102	1986	[2] Female	1927	1999
649.	2102	1987	[2] Female	1927	1999
650.	2102	1988	[2] Female	1927	1999
651.	2102	1989	[2] Female	1927	1999
652.	2102	1990	[2] Female	1927	1999
653.	2102	1991	[2] Female	1927	1999
654.	2102	1992	[2] Female	1927	1999
655.	2102	1993	[2] Female	1927	1999
656.	2102	1994	[2] Female	1927	1999
657.	2102	1995	[2] Female	1927	1999
658.	2102	1996	[2] Female	1927	1999
659.	2102	1997	[2] Female	1927	1999
660.	2102	1998	[2] Female	1927	1999
7975.	19202	1985	[1] Male	1960	-2
7976.	19202	1986	[1] Male	1960	-2

Individual 2102 is female, was born in 1927 and died in 1999. She has participated annually since 1984 until 1998. Individual 19202 is male, was born in 1960 and participated twice, in 1985 and 1986. The value “-2” for the variable year of death (todjahr) stands for “Does not apply”. For more information on the values, see the [Missing Conventions](#).

b) Were these people and their parents born in Germany?

In the dataset, search for a variable that describes the migration background and the survey year. Display the information from the variables for individuals 2102 and 19202.

```
list pid syear migback if pid == 2102 | pid == 19202
```

	pid	year	migback
646.	2102	1984	[1] no migration background
647.	2102	1985	[1] no migration background
648.	2102	1986	[1] no migration background
649.	2102	1987	[1] no migration background
650.	2102	1988	[1] no migration background
651.	2102	1989	[1] no migration background
652.	2102	1990	[1] no migration background
653.	2102	1991	[1] no migration background
654.	2102	1992	[1] no migration background
655.	2102	1993	[1] no migration background
656.	2102	1994	[1] no migration background
657.	2102	1995	[1] no migration background
658.	2102	1996	[1] no migration background
659.	2102	1997	[1] no migration background
660.	2102	1998	[1] no migration background
7975.	19202	1985	[2] direct migration background
7976.	19202	1986	[2] direct migration background

Individual 2102 has no migration background. Individual 19202 has a direct migration background which means that he was not born in Germany.

c) If they immigrated to Germany, in which year and from what country?

Search the dataset for a variable that describes the country of birth, the year of moving to Germany and the survey year. Display the information from the variables on individuals 2102 and 19202.

```
list pid syear immiyear corigin if pid == 2102 | pid == 19202
```

	pid	syear	immiyear		corigin
646.	2102	1984	-2	[1]	Germany
647.	2102	1985	-2	[1]	Germany
648.	2102	1986	-2	[1]	Germany
649.	2102	1987	-2	[1]	Germany
650.	2102	1988	-2	[1]	Germany
651.	2102	1989	-2	[1]	Germany
652.	2102	1990	-2	[1]	Germany
653.	2102	1991	-2	[1]	Germany
654.	2102	1992	-2	[1]	Germany
655.	2102	1993	-2	[1]	Germany
656.	2102	1994	-2	[1]	Germany
657.	2102	1995	-2	[1]	Germany
658.	2102	1996	-2	[1]	Germany
659.	2102	1997	-2	[1]	Germany
660.	2102	1998	-2	[1]	Germany
7975.	19202	1985	1980	[2]	Turkey
7976.	19202	1986	1980	[2]	Turkey

Individual 2102 is born in Germany and has therefore no immigration year. Individual 19202 immigrated from Turkey in 1980.

d) Are these people from East or West Germany?

Search the dataset for a variable that tells whether respondents are from the East or West, the survey year and sample. Display the information from the variables for individuals 2102 and 19202.

```
list pid syear loc1989 psample if pid == 2102 | pid == 19202
```

	pid	year	loc1989				psample
646.	2102	1984	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
647.	2102	1985	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
648.	2102	1986	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
649.	2102	1987	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
650.	2102	1988	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
651.	2102	1989	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
652.	2102	1990	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
653.	2102	1991	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
654.	2102	1992	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
655.	2102	1993	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
656.	2102	1994	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
657.	2102	1995	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
658.	2102	1996	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
659.	2102	1997	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
660.	2102	1998	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
7975.	19202	1985	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)
7976.	19202	1986	[2]	West Germany (FRG)	incl. West Berlin	[1]	A 1984 Initial Sample (West)

The variable loc1989 shows where the individual lived in 1989. Individuals 2102 and 19202 lived in West Germany in 1989 and, accordingly, were from Sample A (West).

e) What sources provide the information on the migration background and year of death

Search the data set for variables that give you the sources of information for year of death and migration background. The variable miginfo contains the information about the usage of (grand-)parents' migration history in the SOEP. The variable todinfo gives the source of the information for all persons who have been identified as deceased over the course of SOEP. Display the information from the variables for individuals 2102 and 19202.

```
list pid syear miginfo todinfo if pid == 2102 | pid == 19202
```

	pid	year	miginfo		todinfo	
646.	2102	1984	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
647.	2102	1985	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
648.	2102	1986	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
649.	2102	1987	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
650.	2102	1988	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
651.	2102	1989	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
652.	2102	1990	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
653.	2102	1991	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
654.	2102	1992	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
655.	2102	1993	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
656.	2102	1994	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
657.	2102	1995	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
658.	2102	1996	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
659.	2102	1997	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
660.	2102	1998	[1]	No (grand-)parental information	[5]	Infratest drop-out study 2001
7975.	19202	1985	[1]	No (grand-)parental information	[-2]	Does not apply
7976.	19202	1986	[1]	No (grand-)parental information	[-2]	Does not apply

The information on the migration background for both individuals come from the respondents themselves. No further indicators are provided. For individual 2102, the information for the year of death comes from an Infratest Follow-Up Study of drop-outs in 2001. For individual 19202 the year of death is not provided.

2. How many people lived in a private household that was interviewed in 2016 and completed the individual questionnaire?

Search the dataset for variable that describe the population in the 2016 survey year. Display the characteristics of the population variable.

```
tab pop if sy==2016
```

Sample Membership	Freq.	Percent	Cum.
[1] Private HH, German HH-Head	31,696	55.33	55.33
[2] Private HH, Foreign HH-Head	13,911	24.28	79.61
[3] Institutional. HH, Collective accom	141	0.25	79.86
[4] Institutional. HH, Collective accom	3,007	5.25	85.11
[5] Not Compl. Private HH, German HH-He	5,947	10.38	95.49
[6] Not Compl. Private HH, Foreign HH-H	2,518	4.40	99.88
[7] Not Compl. Institutional. HH, Colle	31	0.05	99.94
[8] Not Compl. Institutional. HH, Colle	36	0.06	100.00
Total	57,287	100.00	

Values 1 and 2 are relevant to answer the question because they describe private households with completed interview.

Search the dataset for variable that describe the survey status in the 2016 survey year. Display the characteristics of the survey status:

```
fre netto if syyear==2016
```

netto — Current survey status

		Freq.	Percent	Valid	Cum.
Valid	10 [10] Interviewee With Succesful Interview (_P)	5562	9.71	9.71	9.71
	12 [12] Individual Questionnaire And Person Biography	8508	14.85	14.85	24.56
	14 [14] Individual Questionnaire And Other Questionnaires	30	0.05	0.05	24.61
	15 [15] Individual Questionnaire And Experiments, Test	14903	26.01	26.01	50.63
	17 [17] Youth Biography First Time Surveyed, Age 17	535	0.93	0.93	51.56
	19 [19] Individual Questionnaire Without Household Interview	113	0.20	0.20	51.76
	20 [20] Children in Succesfully Interviewed Households (_Kind)	10638	18.57	18.57	70.33
	21 [21] Children With Mother-Child Questionnaire_I, Age 0-1	349	0.61	0.61	70.94
	22 [22] Children With Mother-Child Questionnaire_II, Age 2-3	393	0.69	0.69	71.62
	23 [23] Children With Mother-Child Questionnaire_III, Age 5-6	685	1.20	1.20	72.82
	24 [24] Children age 7-8, with parental questionnaire	746	1.30	1.30	74.12
	25 [25] Children age 9-10, with parental questionnaire	538	0.94	0.94	75.06
	26 [26] Students Age 11-12	559	0.98	0.98	76.04
	28 [28] Youth questionnaire, Age 13-14	526	0.92	0.92	76.95
	29 [29] Youth from refugee sample, age 16-17	219	0.38	0.38	77.34
	30 [30] Persons In Successfully Interviewed HH Without Individual Interview	3965	6.92	6.92	84.26
	31 [31] Successful Gap Interview (_LUECKE)	284	0.50	0.50	84.75
	32 [32] Successfully Completed Biography Questionnaires	1	0.00	0.00	84.76
	34 [34] Successful Tests and Experiments	12	0.02	0.02	84.78
	35 [35] Part. Success, without HH interview	136	0.24	0.24	85.01
	40 [40] Person in non completed gross HH	5783	10.09	10.09	95.11
	44 [44] Test or experiment in non completed HH	1	0.00	0.00	95.11
	48 [48] Youth questionnaire (age 17) in non completed HH	124	0.22	0.22	95.33
	49 [49] Children (0-16) in non completed HH	2057	3.59	3.59	98.92
	90 [90] Individual Dropouts PBR_EXIT	306	0.53	0.53	99.45
	91 [91] Moved abroad	133	0.23	0.23	99.68
	99 [99] Has Died	181	0.32	0.32	100.00
	Total	57287	100.00	100.00	

Respondents with survey status between 10 and 15 or survey status 19 completed the individual questionnaire. These are all individuals 18 years and older.

Cross-tabulate the variables netto and pop with an appropriate restricting condition to answer the question.

```

1      tab netto pop if ((netto>=10 & netto<=15) | netto==19) & (pop==1 |
↪pop==2) & (syear==2016)

```

Current survey status	Sample Membership		Total
	[1] Priva	[2] Priva	
[10] Interviewee With	5,362	173	5,535
[12] Individual Quest	1,685	5,339	7,024
[14] Individual Quest	30	0	30
[15] Individual Quest	14,055	757	14,812
Total	21,132	6,269	27,401

In 2016, a total of 27,401 respondents completed the individual questionnaire for Sample Membership 1 and 2.

3. PPATHL allows you to see which populations can be viewed from a longitudinal perspective:

a) How many people who answered the individual questionnaire in 2000 also took part in the survey in 2014?

Generate a variable and limit the survey status to individuals who answered an individual questionnaire in 2000 and 2014. Note that the values 10,12,13,14,15,16,18,19 of the netto variable mean realized interviews. Sort the dataset by the variable pid and generate a second variable to calculate the sum of the first generated variable. Display the characteristics of the survey status under the condition that the individual questionnaire has been answered.

```

1      gen v1 = 1 if (netto>=10 & netto<=19 & syear==2000) | (netto>=10 & netto
↪<=19 & syear==2014)
2      bysort pid : egen v2 = sum(v1) if netto>=10 & netto<=19
3      tab netto syear if v2==2 & (syear==2014 | syear==2000)

```

Current survey status	Survey Year		Total
	2000	2014	
[10] Interviewee With	7,505	5,143	12,648
[12] Individual Quest	63	1	64
[15] Individual Quest	0	2,492	2,492
[16] Individual Quest	71	0	71
[19] Individual Quest	0	3	3
Total	7,639	7,639	15,278

A total of 7,639 respondents completed the individual questionnaire in 2000 and 2014.

b) How many people answered the individual questionnaire every year from 2000 to 2014?

Generate a variable that counts the number of waves of completed individual interviews and limit it to the years 2000 until 2014. If the generated variable takes the value 15, a person has completed a personal interview 15 years in a row. Display the survey status and the survey year with the newly created variable.


```

1      egen h1 = count(syear) if netto>=10 & netto<=19 & syear>=2000 & syear
2      <=2014, by(pid)
      tab netto syear if h1==15 & syear>=2000 & syear<=2014

```

Current survey status	Survey Year								Total
	2000	2001	2002	2003	2004	2005	2006	2007	
[10] Interviewee With	6,568	4,086	6,642	6,661	6,656	6,664	6,662	6,663	88,509
[12] Individual Quest	44	2,542	23	3	7	1	1	0	2,623
[13] Individual Quest	0	37	0	0	0	0	0	0	37
[15] Individual Quest	0	0	0	0	0	0	0	0	8,731
[16] Individual Quest	53	0	0	0	0	0	0	0	53
[19] Individual Quest	0	0	0	1	2	0	2	2	22
Total	6,665	6,665	6,665	6,665	6,665	6,665	6,665	6,665	99,975

Current survey status	Survey Year							Total
	2008	2009	2010	2011	2012	2013	2014	
[10] Interviewee With	4,564	6,665	4,590	6,663	4,459	6,665	4,301	88,509
[12] Individual Quest	2	0	0	0	0	0	0	2,623
[13] Individual Quest	0	0	0	0	0	0	0	37
[15] Individual Quest	2,096	0	2,072	0	2,201	0	2,362	8,731
[16] Individual Quest	0	0	0	0	0	0	0	53
[19] Individual Quest	3	0	3	2	5	0	2	22
Total	6,665	6,665	6,665	6,665	6,665	6,665	6,665	99,975

A total of 6,665 people completed the individual questionnaire every year from 2000-2014.

c) How many people who turned 15 in 2011 and spent at least part of their childhood in a SOEP household took part in the survey in 2016?

Generate a variable with people who turned 15 in 2011 and had lived in a survey household as a child. The age of the respondent can be determined with the year of birth, and you can limit children using the net code. Display the new generated variable and the year of birth.

```

1      gen a15kind = 1 if 2011-gebjahr==15 & netto>=20 & netto<30 & syear==2011
2      tab a15kind gebjahr

```

a15kind	Geburtsjah r -45Steller- 1996		Total
1	741		741
Total	741		741

A total of 741 people were 15 years old in 2011 and lived as children in a survey household.

To find out if these 741 people filled out a person questionnaire in 2016, we generate a second variable that fills up the value of one person for all remaining available years. Limit the net code and survey year to narrow down the cases appropriately.

```

1      bysort pid : egen a1 = max(a15kind)
2      tab netto if a1==1 & netto>=10 & netto<20 & syear==2016

```

Current survey status	Freq.	Percent	Cum.
[10] Interviewee With Succesful Intervi	70	22.65	22.65
[12] Individual Questionnaire And Perso	2	0.65	23.30
[15] Individual Questionnaire And Exper	227	73.46	96.76
[19] Individual Questionnaire Without H	10	3.24	100.00
Total	309	100.00	

A total of 309 people who were 15 years old at the time of the survey and had been part of a survey household as a child in 2011 completed an individual interview in 2016.

d) The individual with pid=588010 was born in 1984 in a panel household and was still part of the sample in 2009. The individual changed households twice during this time. In which years?

To identify how often and when a individual changed households, you must display all available household numbers in PPATHL for individual 588010.

```

1      list pid hid syear gebjahr if pid==588010

```

	pid	hid	syear	gebjahr
292415.	588010	58807	1985	1984
292416.	588010	58807	1986	1984
292417.	588010	73407	1987	1984
292418.	588010	73407	1988	1984
292419.	588010	73407	1989	1984
292420.	588010	73407	1990	1984
292421.	588010	73407	1991	1984
292422.	588010	73407	1992	1984
292423.	588010	73407	1993	1984
292424.	588010	73407	1994	1984
292425.	588010	73407	1995	1984
292426.	588010	73407	1996	1984
292427.	588010	73407	1997	1984
292428.	588010	73407	1998	1984
292429.	588010	73407	1999	1984
292430.	588010	73407	2000	1984
292431.	588010	73407	2001	1984
292432.	588010	73407	2002	1984
292433.	588010	73407	2003	1984
292434.	588010	73407	2004	1984
292435.	588010	73407	2005	1984
292436.	588010	73407	2006	1984
292437.	588010	73407	2007	1984
292438.	588010	132608	2008	1984
292439.	588010	132608	2009	1984
292440.	588010	132608	2010	1984

Individual 588010 has participated in the survey since 1985 as part of household 58807. From 1987 to 2007 the individual was in household 73407, from 2008 on, the individual was in household 132608.

Last change: Jul 24, 2023

To this tracking data, users should merge variables they want to analyze. For putting together cross-sectional data sets, these exercises are helpful:

6.2 Generating a Cross-Sectional Dataset

This example involves generating a dataset to analyze health satisfaction determinants in 2008, and you can either use the Paneldata.org syntax generator or write your own syntax file to perform this task. You can search for the variable names in Paneldata.org (or use the variables below directly).

1. Generate a cross-sectional dataset for the year 2008, which should contain all persons with the following characteristics:

- Respondents in 2008 **"ynetto"**
- Lived in a private household in 2008 **"ypop"**





The dataset should contain the following variables of interest.

- satisfaction with health **"yp0101"**
- smoking currently yes/no **"yp10601"**
- current employment status **"emplst08"**
- monthly household net income **"hinc08"**

In addition, the dataset should contain the following additional information for a 2008 cross-sectional analysis (these variables are automatically generated by paneldata.org):

- current cross-section weighting factor **"yphrf"**
- personal number **"pid"**
- original household number **"cid"**
- current household number **"hid_2008"**
- sample affiliation **"psample"**
- gender **"sex"**
- year of birth **"gebjahr"**

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

These are used to store commands, log files, datasets, and temporary datasets. Open an empty do file and define your created paths with globals:

```

1 *****
2 * Set relative paths to the working directory
3 *****
4 global MY_PATH_IN    "\\hume\rdc-prod\distribution\soep-core\soep.v37\eu\
↪Stata\raw\"
5 global MY_PATH_OUT   "H:\Exercise\"
6 global MY_FILE_OUT   "${MY_PATH_OUT}cross-sectional-exercise.dta"
7 global MY_LOG_FILE   "${MY_PATH_OUT}cross-sectional-exercise.log"
8 capture log close
9 log using "${MY_LOG_FILE}", text replace

```

The global “AVZ” defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to your data.

Use ppath as the source file together with the required variables. Keep all cases with completed interviews. In addition, your dataset should only contain respondents who can make a statement on the content of the question. For example, you can use the net code to identify and remove children from your dataset.

```

1 * * * PFAD * * *
2
3 use ypop cid sex hid_2008 pid psample ynetto gebjahr using "${MY_PATH_IN}
↪ppfad.dta", clear
4
5
6 * * * BALANCED VS UNBALANCED * * *
7
8 keep if ( (ynetto >= 10 & ynetto < 20) )
9
10
11 * * * PRIATVE VS ALL HOUSEHOLDS * * *
12
13 keep if ( (ypop == 1 | ypop == 2) )
14
15
16 * * * SORT PFAD * * *
17
18 save "${MY_PATH_OUT}ppfad.dta", replace
19 clear

```

Save the modified data temporarily. Now link your dataset with the weights of the SOEP and save your dataset as a master file.

```

1 * * * HRF * * *
2
3 use cid pid yphrf prgroup using "${MY_PATH_IN}phrf.dta"
4 sort pid
5 save "${MY_PATH_OUT}hrf.dta", replace
6 clear
7
8
9 * * * CREATE MASTER * * *
10

```

(continues on next page)

(continued from previous page)

```

11      use "${MY_PATH_OUT}pfad.dta", clear
12      merge 1:1 pid cid using "${MY_PATH_OUT}hrf.dta", keep(master match) nogen
13      save "${MY_PATH_OUT}master.dta", replace
14      clear

```

Now prepare the content variables. Search for the content variables you are looking for from the various datasets and temporarily save the datasets you have created.

```

1      * * * READ DATA * * *
2
3      use emplst08 pid using "${MY_PATH_IN}ypgen.dta", clear
4      save "${MY_PATH_OUT}ypgen.dta", replace
5
6
7      use yp0101 pid yp10601 using "${MY_PATH_IN}yp.dta", clear
8      save "${MY_PATH_OUT}yp.dta", replace
9
10
11     use hinc08 hid_2008 using "${MY_PATH_IN}yhgen.dta", clear
12     save "${MY_PATH_OUT}yhgen.dta", replace

```

Link the datasets you have created to your master file and save for analysis.

```

1      * * * MERGE DATA * * *
2
3      use "${MY_PATH_OUT}master.dta", clear
4      merge 1:1 pid using "${MY_PATH_OUT}ypgen.dta", keep(master match) nogen
5      merge 1:1 pid using "${MY_PATH_OUT}yp.dta", keep(master match) nogen
6      merge m:1 hid_2008 using "${MY_PATH_OUT}yhgen.dta", keep(master match) ↵
↵nogen
7
8
9      * * * DONE * * *
10
11     label data "paneldata.org"
12     save "${MY_FILE_OUT}", replace
13     desc
14
15     log close

```

You have successfully created a cross-sectional dataset for the year 2008.

2. Encode missing values into system missings (STATA)!

In SOEP, the missing codes of variables are described in detail with the values -1 to -8. To learn more about missing codes, see the section [Missing Conventions](#). For content analysis, it is not always necessary to differentiate missing codes. Therefore you should be able to convert missing codes:

```

1      use "${MY_FILE_OUT}", clear
2
3
4      *****
5      *** Exercise 2) ***

```

(continues on next page)

(continued from previous page)

```

6      * Encode missing values into missing values in system missings (STATA)!
7      ****
8
9      * mvdecode = Change missing values to numeric values and vice versa
10     mvdecode _all, mv(-1=. \ -2=.t \ -3=.x \ -5=.y \ -8=.z)

```

Open the dataset for your analysis and summarize all missing codes.

3. How does average health satisfaction differ a) by gender

Satisfaction was measured on a scale of 1 to 10. To compare average satisfaction with health between women and men, you should display the mean value for both genders.

```

1      *unweighted*
2      tabstat yp0101, by(sex)

```

```

. *a) by sex:
.      *unweighted*
.      tabstat yp0101, by(sex)

```

```

Summary for variables: yp0101
by categories of: sex (Sex)

```

sex	mean
[1] Male	6.616534
[2] Female	6.516729
Total	6.56428

Since you have previously added the SOEP weighting factors to the dataset for your analysis, you should use the weighting for a representative analysis.

```

1      *weighted*
2      tabstat yp0101 [aw=yphrf], by(sex)

```

```
.      *weighted*
.      tabstat yp0101 [aw=yphrf], by(sex)
```

```
Summary for variables: yp0101
by categories of: sex (Sex)
```

sex	mean
[1] Male	6.53008
[2] Female	6.407367
Total	6.467019

b) Employment status

Now proceed in a similar way when comparing satisfaction with health and employment status. Compare the mean values again:

```
*b) by job status:
*unweighted*
tabstat yp0101, by(emplst08)
```

```
. *b) by job status:
.      *unweighted*
.      tabstat yp0101, by(emplst08)

Summary for variables: yp0101
by categories of: emplst08 (Employment Status)
```

emplst08	mean
[1] Full-Time Em	6.931818
[2] Regular Part	6.805956
[3] Vocational T	7.792453
[4] Marginal, Ir	6.739879
[5] Not Employed	6.085035
[6] Sheltered wo	5.72
Total	6.56428

Since you have previously added the SOEP weighting factors to the dataset for your analysis, you should use the weighting for a representative analysis.

```
*weighted*
tabstat yp0101 [aw=yphrf], by(emplst08)
```



```
.      *weighted*
.      tabstat yp0101 [aw=yphrf], by(emplst08)

Summary for variables: yp0101
      by categories of: emplst08 (Employment Status)
```

emplst08	mean
[1] Full-Time Em	6.847115
[2] Regular Part	6.704637
[3] Vocational T	7.822574
[4] Marginal, Ir	6.615801
[5] Not Employed	5.987851
[6] Sheltered wo	4.937647
Total	6.467019

c) Age

Since you do not have a variable that represents age, you must generate a suitable age variable using the birth year variable. The year of birth is metric and should be categorized for analysis. Define categories for your age variable and assign suitable labels.

```
1      *c) by age in 2008 (<30, 30-64, 65+)
2
3      gen age=2008-gebjahr
4      gen age_3=age
5      recode age_3 (17/29=1) (30/64=2) (65/120=3)
6      label define age_3 1 "17-29" 2 "30-64" 3 "65+"
7      label values age_3 age_3
```

Create a mean value comparison with your age variable and health satisfaction in weighted and unweighted form.

```
1      *unweighted*
2      tabstat yp0101, by(age_3)
```

```
.      *unweighted*
.      tabstat yp0101, by(age_3)
```

```
Summary for variables: yp0101
by categories of: age_3
```

age_3	mean
17-29	7.640552
30-64	6.607247
65+	5.714101
Total	6.56428

```
1      *weighted*
2      tabstat yp0101 [aw=yphrf], by(age_3)
```

```
.      *weighted*
.      tabstat yp0101 [aw=yphrf], by(age_3)
```

```
Summary for variables: yp0101
by categories of: age_3
```

age_3	mean
17-29	7.595288
30-64	6.483365
65+	5.660658
Total	6.467019

d) Income

As with age, generate a categorized version of income for household net income:

```
1      *d) by monthly household net income (-1.999, 2.000-3.999, 4000+ Euro)
2      gen hinc08_3 = hinc08
3      recode hinc08_3 (0/1999=1) (2000/3999=2) (4000/99999=3)
4      label define hinc08_3 1 "<2000 Euro" 2 "2000-<4000 Euro" 3 "4000+ Euro"
5      label values hinc08_3 hinc08_3
```

Display the mean values in weighted and unweighted form:

```
1      *unweighted*
2      tabstat yp0101, by(hinc08_3)
```

```
.      *unweighted*
.      tabstat yp0101, by(hinc08_3)
```

```
Summary for variables: yp0101
      by categories of: hinc08_3
```

hinc08_3	mean
<2000 Euro	6.042256
2000-<4000 Euro	6.69125
4000+ Euro	7.11391
Total	6.551677

```
*weighted*
tabstat yp0101 [aw=yphrf], by(hinc08_3)
```

```
.      *weighted*
.      tabstat yp0101 [aw=yphrf], by(hinc08_3)
```

```
Summary for variables: yp0101
      by categories of: hinc08_3
```

hinc08_3	mean
<2000 Euro	5.988714
2000-<4000 Euro	6.6906
4000+ Euro	7.126235
Total	6.446908

e) Smoking

Since this variable is nominal, adjustments to this variable are not necessary. Display average satisfaction with health for smokers and non-smokers in weighted and unweighted form:

```
*e) by smoking yes/no

*unweighted*
tabstat yp0101, by(yp10601)
```

```
.      *unweighted*  
.      tabstat yp0101, by(yp10601)
```

```
Summary for variables: yp0101  
by categories of: yp10601 (Currently Smoke)
```

yp10601	mean
[1] Yes	6.551121
[2] No	6.570124
Total	6.564997

```
1      *weighted*  
2      tabstat yp0101 [aw=yphrf], by(yp10601)
```

```
.      *weighted*  
.      tabstat yp0101 [aw=yphrf], by(yp10601)
```

```
Summary for variables: yp0101  
by categories of: yp10601 (Currently Smoke)
```

yp10601	mean
[1] Yes	6.448555
[2] No	6.476664
Total	6.468664

Last change: May 30, 2023

6.3 Syntax Generator on paneldata.org

[Paneldata](#) allows registered users to collect and save variables relevant to their research in a variable basket. These variables can be simply written into a single dataset with the script generator. The script generator helps you with data management and can save valuable working time.

Open [Paneldata](#)

paneldata.org Studies Search Register / log in

Update June 2022

- SOEP-Core v37 is now available.
- TwinLife v6-0-0 is now available.



paneldata.org

soep-base
SOEP-Core
SOEP-IS

Click on “Register / login” to log in to paneldata.org.

paneldata.org Studies Search Register / log in

User login

Username:

Password:

[Login](#)

New user? [Register here.](#)

Forgot your password? [Create a new one here.](#)

[Contact / feedback](#)

[German Socio-economic Panel \(SOEP\) Imprint](#)

If you have already registered, go to “user login”. As a new user, you can register under “register here”. Once you have logged in, you have access to the variable basket and the syntax generator.



soep-base
SOEP-Core
SOEP-IS
IAB-SOEP-MIG
TwinLife
German Internet Panel (GIP)

To access the activated functions, click on the button “my baskets”. You will be taken to your personal workspace on paneldata.org.

“My baskets” displays your variable baskets. If you click on “create basket”, you can create a new basket.

Create basket

Name*

Name of the basket (Lowercase)

Label

Label of the basket

Description (Markdown)

Description of the basket (Markdown)

Study*

Foreign key to studies.Study

Create basket

When creating the basket, first define the name of the variable basket. The name must be lower case to be accepted by Paneldata. Optionally, you can assign a label and enter a description. Finally, you select the study that you want to use as a database for your research. Now click on “Create basket” and your newly created variable basket appears in the interface.

Basket successfully created.

Baskets

Create basket

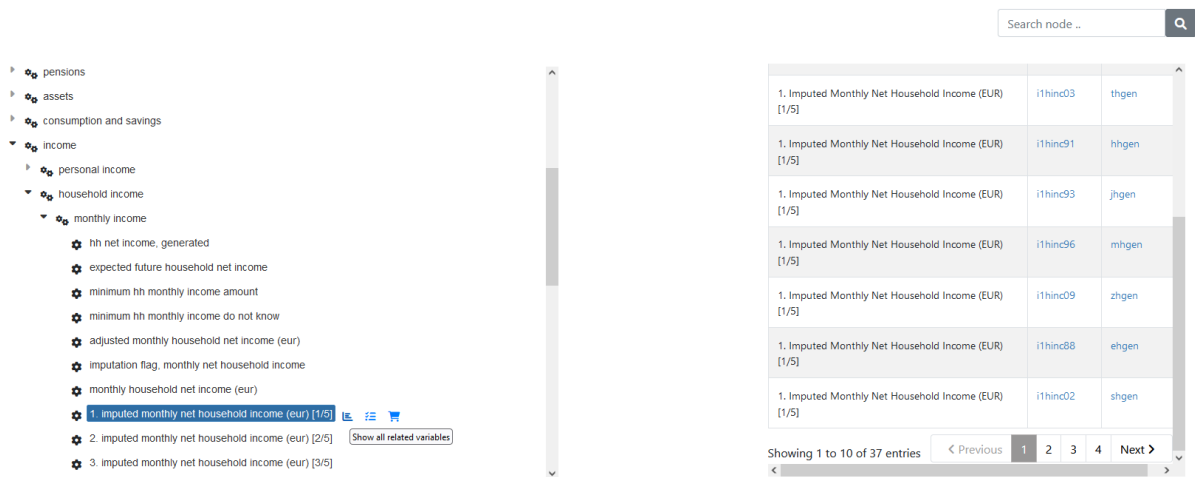
willingness to take risks [basket: willingness to take risks]
SOEP/willingness to take risks



Now search for the relevant variables on paneldata.org and add them to your individual basket. For example, you are interested in monthly net household income. If you do not know the variable name, you can find the overarching concept using the topic search. Click on “paneldata.org” to get to the main page. Select the study SOEP-Core and click on “topics” at the top of the page.



Check the different topics for income-relevant concepts and select “income, taxes, and social security”.



Browse the topic list and you will reach the sub-topic “income” → “household income” → “monthly income”. There you will find the variables you are looking for. Click on “show all related variables” and you will see the history of variables on the right hand side of the window.

Select the variable of your desired wave and you will reach the variable overview with important information about the variable.

mhgen/i1hinc96: 1. Imputed Monthly Net Household Income (EUR) [1/5]

Related variables **37**
Input variables **0**
Output variables **1**

0: hgen/hgi1hinc	1984: ahgen/i1hinc84	1985: bhgen/i1hinc85	1986: chgen/i1hinc86
1987: dhgen/i1hinc87	1988: ehgen/i1hinc88	1989: fhgen/i1hinc89	1990: ghgen/i1hinc90
1991: hhgen/i1hinc91	1992: ihgen/i1hinc92	1993: jhgen/i1hinc93	1994: khgen/i1hinc94
1995: lhgen/i1hinc95	1996: mhgen/i1hinc96	1997: nhgen/i1hinc97	1998: ohgen/i1hinc98
1999: phgen/i1hinc99	2000: qhgen/i1hinc00	2001: rhgen/i1hinc01	2002: shgen/i1hinc02
2003: thgen/i1hinc03	2004: uhgen/i1hinc04	2005: vhgen/i1hinc05	2006: whgen/i1hinc06
2007: xhgen/i1hinc07	2008: yhgen/i1hinc08	2009: zhgen/i1hinc09	2010: bhgen/i1hinc10
2011: bbhgen/i1hinc11	2012: bchgen/i1hinc12	2013: bdhgen/i1hinc13	2014: behgen/i1hinc14
2015: bfhgen/i1hinc15	2016: bghgen/i1hinc16	2017: bhhgen/i1hinc17	2018:

Basket
willingness to take risks Add variable
Create a new basket

Info
i1hinc96
1. imputed monthly net household income (eur) [1/5]
English label: 1. Imputed Monthly Net Household Income (EUR) [1/5]
German label: 1.Imput.monatl.HH-Nettoeinkommen (EUR) [1/5]
mhgen – Generated Household Data 1996
Analysis Unit: household
Dataset Type: Generated (raw folder)
Period: 1996

When logged in, the basket area appears in the overview of variables. Your baskets are listed there. If you want to add the variable to a basket, click on “add to basket”. If the variable is already in the basket and you want to remove it, select “remove from basket”. If you want to create a new basket within the overview of variables, click on “create a new basket” and your variable will automatically be placed in the new basket. You can access the basket overview by clicking on the name of your basket in the “basket” section. Alternatively, you can click on the button “my baskets” and you will also return to the basket overview.

Title: willingness to take risks
Study: SOEP-Core
Note: The script-generators help you especially to merge variables from different year-specific datasets (e.g. bhp, bgp) to one wide file. This (and more) work is already done in the long files (e.g. pl, hl), which you find in the top-level directory. If you still want to use the script-generator, please note that only the files in the “raw” subdirectory can be processed. Please specify the complete address of the “raw” subdirectory (e.g. D:\v35\raw) as your “input path”.

Actions
Export to CSV
Warning: The generator for Stata and R scripts is working for v37 data but is not fully tested yet. (Feedback to kwenzig@diw.de is welcome!) The generator for the SPSS script will not be updated and will be removed.
New script using the **soep-stata** generator
New script using the **soep-spss** generator
New script using the **soep-r** generator
Add variables

Created scripts

Concept	0	1984	1985	1986	1987	1988	1989	1990	1991	1992
_hgen_hgi1hinc 1. imputed	Add all 🛒 hgi1hinc	🛒 i1hinc84	🛒 i1hinc85	🛒 i1hinc86	🛒 i1hinc87	🛒 i1hinc88	🛒 i1hinc89	🛒 i1hinc90	🛒 i1hinc91	🛒 i1hinc92

Click on the basket with your added variable and you will get an overview of all variables in your basket. With “add all”, you add the variables of all survey waves and the shopping cart is highlighted in green. If you are interested in a specific survey period, you can select the wave-specific variables by clicking on the shopping cart. Click on “remove all” to remove the variable from your basket. Furthermore, you can export your chosen variables to a CSV-File (Comma Separated Value-File) and, for example, import them in STATA.

Once you have filled your basket and selected the desired survey waves, you can merge all variables into one dataset. To do this, click on “new script using the soep-xxx generator” in the “actions” area. You can choose between different statistical programs.

paneldata.org
Studies ▼
Search

My baskets
My account
Logout

willingness to take risks
No more links at the moment...

Script: willingness to take risks

Configure basket

Name

Label

Script generator

soep-stata ▼

Input path

Output path

Analysis Unit

Individual ▼

Private households

Private households only ▼

raw script

```

* -----
* This command file was generated by paneldata.org
* -----
* !!! IMPORTANT - WARNING !!!
* You alone are responsible for contents and appropriate.
* usage by accepting the usage agreement.
* -----
* Please report any errors of the code generated here
* to soepmail@diw.de
* -----

* * * 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 * * *
```

In the script generator, you can create a script that matches your preferred variables. Specify the name of your script, select the statistics program you are using. Enter the (local) addresses of the folders in which the data is to be found or the result (and temporary intermediate results) is to be written in the two path fields (“input” and “output” path).

paneldata.org
Studies ▼
Search

My baskets
My account
Logout

willingness to take risks
No more links at the moment...

Analysis Unit

Individual ▼

Private households

Private households only ▼

Sample composition

balanced ▼

Age group

All adult respondents ▼

Update script

```

set more off

* * * NOT PROCESSED * * *.

* * * PFAD * * *

use hhnr persnr sex gebjahr psample thnr tnetto tpop bfhhnr bfnetto bfpop rhnr rnetto rpop zhnr znetto

* * * BALANCED VS UNBALANCED * * *

keep if ( (tnetto >= 10 & tnetto < 20) & (bfnetto >= 10 & bfnetto < 20) & (rnetto >= 10 & rnetto < 20) & (

* * * PRIVATE VS ALL HOUSEHOLDS * * *

keep if ( (tpop == 1 | tpop == 2) | (bfpop == 1 | bfpop == 2) | (rpop == 1 | rpop == 2) | (zpop == 1 | zpc
```

In the “analysis unit” section, you decide whether all persons are considered individually within the household (“individual”) or whether you are interested solely in the household as a whole (“household”). With “sample composition” you can choose between “balanced” and “unbalanced”. If you select “balanced”, you will receive a dataset without missing codes. The respondents provided information on all variables. For more information about balanced and un-

balanced datasets, see the section *Principles of Data Analysis*. Under “age group”, you can limit the respondents. When you are satisfied with your settings, click on “Update Script” and your script will be created.

Script: willingness to take risks

Configure basket

Name

Willingness to take risks

Label

Willingness to take risks

Script generator

soep-stata

Input path

Z:\input\DATA\soep33.1_de\stata

Output path

H:\output\Willingness to take risks\

Analysis Unit

Individual

raw script

```

* -----
* This command file was generated by paneldata.org
* -----
* !!! I M P O R T A N T - W A R N I N G !!!
* You alone are responsible for contents and appropriate.
* usage by accepting the usage agreement.
* -----
* Please report any errors of the code generated here
* to soepmail@diw.de
* -----

* * * 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 * * *.

```

If you click on the “raw script” button, the script is displayed in text form. Copy it to your statistical software. The result dataset has the name new(.dta, .sav). If you want to change it, you have to do it in the script. Execute the script with your statistical software and you will receive your dataset with all your chosen variables.

Last change: May 30, 2023




The SOEP team recommends using the long data sets. How to merge longitudinal datasets and what user service these datasets provide can be seen in these exercises:

6.4 Generating a Longitudinal Dataset

This example focuses on generating a dataset to analyze determinants of health satisfaction. You can either use the syntax generator in paneldata.org or write a syntax file yourself. You can search for variable names in Paneldata.org.

In the previous examples, you created an exercise path with four subfolders as well as corresponding globals in the STATA do-file. You can use the same folders and globals for this exercise.

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do

- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

1. Generate an unbalanced panel dataset for the years 2006 to 2008 using paneldata.org if you wish. The dataset should contain all respondents in private households:

The data set should contain the following variables of interest:

- health satisfaction **"wp0101" "xp0101" "yp0101"**
- currently smoking yes/no **"wp9301" "yp10601"**
- current employment status **"emplst06" "emplst07" "emplst08"**
- monthly household net income **"hinc06" "hinc07" "hinc08"**

In addition, the dataset should include the following additional information for analysis from 2006 to 2008:

- cross-sectional weighting factors for all relevant years **"wphrf" "xphrf" "yphrf"**
- individual identifier **"pid"**
- original household number **"cid"**
- household number for all relevant years **"hid_2006" "hid_2007" "hid_2008"**
- sample membership **"psample"**
- sex **"sex"**
- year of birth **"gebjahr"**
- population membership **"wpop" "xpop" "ypop"**

If you need detailed instructions on how the script generator works in paneldata.org, you can find them in the chapter *Syntax Generator on paneldata.org*.

If you would like to assemble your dataset yourself, you can do this with the datasets you have assembled. From the previous exercise with tracking data, you may already have an idea where to get most of the variables.

Since we want to have an unbalanced panel set, the \$netto variable for the years 2006 to 2008 must also be used. In addition, our analysis must limit population membership, as we are only interested in household respondents.

1.1. Create a Master File

Use ppfad as the source file together with the required variables that you may have already found in [Paneldata.org](https://paneldata.org) or identified from the variable label in the dataset. Note that only variables from the years to be analyzed should be used.

```
1      use ypop hid_2007 cid xpop sex wnetto wpop hid_2008 xnetto pid hid_2006
2      ↪psample ynetto gebjahr ///
3      using "${MY_PATH_IN}ppfad.dta", clear
4
```

Since we want to obtain an unbalanced data set, i.e., individuals who have completed an individual questionnaire at least once within the last three years, you must restrict the variable \$netto (survey status). Also, we only want to analyze private households, so we need a further restriction of the \$pop (sample membership) variable.

```
1      * * * BALANCED VS UNBALANCED * * *
2
3      keep if ( (xnetto >= 10 & xnetto < 20) | (ynetto >= 10 & ynetto < 20) |
```

(continues on next page)

(continued from previous page)

```

4  ↪(wnetto >= 10 & wnetto < 20) )
5
6      * * * ONLY PRIVATE HOUSEHOLDS * * *
7
8      keep if ( (xpop == 1 | xpop == 2) | (ypop == 1 | ypop == 2) | (wpop == 1
9  ↪ | wpop == 2) )
10
11     * * * GENDER ( male = 1 / female = 2) * * *
12
13     /* all genders */
14
15     * * * SAVE PFAD * * *
16
17     save "${MY_PATH_OUT}pfad.dta", replace
18     clear
19

```

What is still missing is the cross-sectional weighting factor and the variables of interest for the analysis.

```

1      * * * HRF * * *
2
3      use cid wphrf pid yphrf xphrf prgroup ///
4      using "${MY_PATH_IN}phrf.dta"
5      save "${MY_PATH_OUT}hrf.dta", replace
6      clear

```

Now we come to the content variables. In order not to have to click through all of the datasets in the data release, it is recommended that the label be entered for the variable of interest from paneldata.org.

Use the filter to narrow your search. Select our main study SOEP-Core, the search type “variable”, the conceptual dataset “Original (raw folder)”, analysis unit and the corresponding year. Once you have clicked on the year of interest, a variable history is displayed. You can use this to see which years the variable was collected and what the variable is called.

Example: Variable Label “Satisfaction Health”

AllVariablesConceptsQuestionsPublicationsTopics

satisfaction health

Search: satisfaction healthStudy: SOEP-CoreConceptualDataset: Original (raw folder)AnalysisUnit: individualClear All

results found in 47ms

Relevance

Study

Select all

SOEP-Core (2175)

Conceptual Dataset

Select all

Original (raw folder) (2175)

Generated (raw folder) (646)

Original (top level folder) (17)

Generated (top level folder)

Gen. Biographical (top level)

Analysis Unit

Select all

individual (2175)

household (24)

[np0101] Satisfaction With Health

Variable in study: SOEP-Core | dataset: np | period: 1997 | analysis unit: individual

[ip9801] Satisfaction With Health

Variable in study: SOEP-Core | dataset: ip | period: 1992 | analysis unit: individual

[jp0101] Satisfaction With Health

Variable in study: SOEP-Core | dataset: jp | period: 1993 | analysis unit: individual

[op0101] Satisfaction With Health

Variable in study: SOEP-Core | dataset: op | period: 1998 | analysis unit: individual

[up0101] Satisfaction With Health

Example: Variable Label “currently smoking yes/no”

[All](#)
[Variables](#)
[Concepts](#)
[Questions](#)
[Publications](#)
[Topics](#)

×

Search: currently smoke ×
 Study: SOEP-Core ×
 ConceptualDataset: Original (raw folder) ×
 AnalysisUnit: individual ×
 Clear All

results found in 41ms Relevance ⬆

Study

☐ Select all

☒ SOEP-Core (1613)

Conceptual Dataset

☐ Select all

☒ Original (raw folder) (1613)

☐ Generated (raw folder) (92)

☐ Gross (raw folder) (199)

☐ Tracking (raw folder) (129)

☐ Original (top level folder)

☐ Generated (top level folder)

☐ Gen. Biographical (top level)

Analysis Unit

☐ Select all

☒ individual (1613)

☐ household (197)

[wp9301] **Currently** **Smoke**

Variable in study: SOEP-Core | dataset: wp | period: 2006 | analysis unit: individual

[sp9401] **Currently** **Smoke**

Variable in study: SOEP-Core | dataset: sp | period: 2002 | analysis unit: individual

[bepm_p_20001] **Currently** **Smoke**

Variable in study: SOEP-Core | dataset: bep_mig | period: 2014 | analysis unit: individual

[bep9401] **Currently** **Smoke**

Variable in study: SOEP-Core | dataset: bep | period: 2014 | analysis unit: individual

[bcp9701] **Currently** **Smoke**

Variable in study: SOEP-Core | dataset: bcp | period: 2012 | analysis unit: individual

Example: Variable Label “current employment status”

AllVariablesConceptsQuestionsPublicationsTopics

current employment status

Search: current employment statusStudy: SOEP-CoreConceptualDataset: Generated (raw folder)AnalysisUnit: individualClear All

results found in 51ms

Relevance

Study

Select all

SOEP-Core (2204)

Conceptual Dataset

Select all

Original (raw folder) (3554)

Generated (raw folder) (2204)

Gross (raw folder) (282)

Original (top level folder) (102)

Generated (top level folder) (102)

Tracking (raw folder) (131)

Tracking (top level folder) (102)

Analysis Unit

Select all

individual (2204)

household (102)

[emplst94] Employment Status

Variable in study: SOEP-Core | dataset: kpgen | period: 1994 | analysis unit: individual

[emplst13] Employment Status

Variable in study: SOEP-Core | dataset: bdpgen | period: 2013 | analysis unit: individual

[emplst12] Employment Status

Variable in study: SOEP-Core | dataset: bcpgen | period: 2012 | analysis unit: individual

[emplst11] Employment Status

Variable in study: SOEP-Core | dataset: bbpgen | period: 2011 | analysis unit: individual

[emplst99] Employment Status

Variable in study: SOEP-Core | dataset: ppgen | period: 1999 | analysis unit: individual

Example: Variable Label “monthly net household income”

All
Variables
Concepts
Questions
Publications
Topics

monthly net household income

Search: monthly net household income
Study: SOEP-Core
ConceptualDataset: Generated (raw folder)
AnalysisUnit: household
Clear All

Study
☐ Select all
☒ SOEP-Core (453)

Conceptual Dataset
☐ Select all
☐ Original (raw folder) (702)
☐ Gross (raw folder) (613)
☒ Generated (raw folder) (453)
☐ Original (top level folder) (1)
☐ Generated (top level folder) (1)
☐ Tracking (raw folder) (39)
☐ Gross (top level folder) (36)

Analysis Unit
☐ Select all
☐ individual (2614)
☒ household (453)

results found in 55ms
Relevance

[fhinc97] Imputation Flag, Monthly Net Household Income

Variable in study: SOEP-Core | dataset: nhgen | period: 1997 | analysis unit: household

[fhinc96] Imputation Flag, Monthly Net Household Income

Variable in study: SOEP-Core | dataset: mhgen | period: 1996 | analysis unit: household

[fhinc95] Imputation Flag, Monthly Net Household Income

Variable in study: SOEP-Core | dataset: lhgen | period: 1995 | analysis unit: household

[fhinc92] Imputation Flag, Monthly Net Household Income

Variable in study: SOEP-Core | dataset: ihgen | period: 1992 | analysis unit: household

[fhinc94] Imputation Flag, Monthly Net Household Income

Variable in study: SOEP-Core | dataset: khgen | period: 1994 | analysis unit: household

To merge the data, you can either use the script generator on paneldata.org or write the syntax manually into a do-file. We now have all the information we need to create a master file.

```

1      using "${MY_PATH_IN}phrf.dta"
2      save "${MY_PATH_OUT}hrrf.dta", replace
3      clear
4
5
6
7      * * * CREATE MASTER * * *
8
9      use "${MY_PATH_OUT}pfad.dta", clear
10     merge 1:1 pid cid using "${MY_PATH_OUT}hrrf.dta", keep(master match) nogen
11     save "${MY_PATH_OUT}master.dta", replace
12
13
14     * * * READ DATA * * *
15
16     use hinc07 hid_2007 using "${MY_PATH_IN}xhgen.dta", clear
17     save "${MY_PATH_OUT}xhgen.dta", replace
18
19
20     use xp0101 pid using "${MY_PATH_IN}xp.dta", clear
21     save "${MY_PATH_OUT}xp.dta", replace

```

(continues on next page)

(continued from previous page)

```

22
23
24     use yp0101 pid yp10601 using "${MY_PATH_IN}yp.dta", clear
25     save "${MY_PATH_OUT}yp.dta", replace
26
27
28     use wp9301 pid wp0101 using "${MY_PATH_IN}wp.dta", clear
29     save "${MY_PATH_OUT}wp.dta", replace
30
31
32     use emplst07 pid using "${MY_PATH_IN}xpgen.dta", clear
33     save "${MY_PATH_OUT}xpgen.dta", replace
34
35
36     use hinc08 hid_2008 using "${MY_PATH_IN}yhgen.dta", clear
37     save "${MY_PATH_OUT}yhgen.dta", replace
38
39
40     use pid emplst06 using "${MY_PATH_IN}wpngen.dta", clear
41     save "${MY_PATH_OUT}wpngen.dta", replace
42
43
44     use hid_2006 hinc06 using "${MY_PATH_IN}whngen.dta", clear
45     save "${MY_PATH_OUT}whngen.dta", replace
46
47
48     use emplst08 pid using "${MY_PATH_IN}ypngen.dta", clear
49     save "${MY_PATH_OUT}ypngen.dta", replace

```

With the help of a unique identifier, which is either the household (`hid_`) or individual identifier (`pid`), you can now merge all datasets or individual variables to `ppfad`. Which identifier to use when depends on the unit of analysis. Since we are on the individual level, our indicator is `pid` (individual identifier).

We load the dataset `ppfad` and merge our datasets or variables to `ppfad`.

```

1      * * * MERGE DATA * * *
2
3      use "${MY_PATH_OUT}master.dta", clear
4      merge m:1 hid_2007 using "${MY_PATH_OUT}xhngen.dta", keep(master match)
5      ↪nogen
6      merge 1:1 pid using "${MY_PATH_OUT}xp.dta", keep(master match) nogen
7      merge 1:1 pid using "${MY_PATH_OUT}yp.dta", keep(master match) nogen
8      merge 1:1 pid using "${MY_PATH_OUT}wp.dta", keep(master match) nogen
9      merge 1:1 pid using "${MY_PATH_OUT}xpngen.dta", keep(master match) nogen
10     merge m:1 hid_2008 using "${MY_PATH_OUT}yhngen.dta", keep(master match)
11     ↪nogen
12     merge 1:1 pid using "${MY_PATH_OUT}wpngen.dta", keep(master match) nogen
13     merge m:1 hid_2006 using "${MY_PATH_OUT}whngen.dta", keep(master match)
14     ↪nogen
15     merge 1:1 pid using "${MY_PATH_OUT}ypngen.dta", keep(master match) nogen

```

(continues on next page)

(continued from previous page)

```

15      * * * DONE * * *
16
17      label data "paneldata.org"
18      save "${MY_FILE_OUT}", replace
19      desc
20
21      log close

```

2. Encode missing values in system failings (STATA)!

After the master file has been created with all required information, the missing values, which can take between -1 to -8 in SOEP, must be recoded to missings. This step is important for converting a wide-format data set to a long format.

```

1      mvdecode _all, mv(-1=. \ -2=.t \ -3=.x \ -5=.y \ -8=.z)

```

3. The data set is in “wide” format, i.e., additional years are displayed as additional variables (columns). For many analyses, it makes sense to convert datasets into the “long” format. In long format, additional years are displayed as additional lines. If the dataset covers three years, as in this example, there are three lines for each person. Convert the data set to long format using the STATA command reshape.!

Since these are cross-sectional variables, it can be assumed that each variable has at least one wave abbreviation, which makes the variable unique. Conversely, this means that the variables must be renamed before the reshape command.

Before renaming all original variables (e.g., from \$P data sets) it must be checked whether the question and the answer categories were the same in all years (you can also look up the exact wording of the question in the corresponding questionnaire). If changes are made, the variables may have to be recoded.

```

1      *Check if original variable have changed over time
2      tab1 wp0101 xp0101 yp0101
3      tab1 wp9301 yp10601
4      /*additionally check questionnaires for exact wording*/

```

How you rename the variables is largely up to you. However, you should ensure that the name remains consistent over time and that the variable only differs according to the year (variable name + four-digit year suffix, e.g., zufr2006, zufr2007, zufr2008). You can rename the variables either manually, line by line, or for advanced users using a loop.

Example of manual renaming:

```

1      *rename time-variant variables
2      *with examples how to use loops (but can also be done "manually")
3      rename wp9301 smoke2006
4      rename yp10601 smoke2008
5      rename wp0101 health2006
6      rename xp0101 health2007
7      rename yp0101 health2008
8      * * *

```

Example of a loop:

```

1      foreach x in 6 7 8 {
2          rename hinc0`x' hinc200`x'
3          rename emplst0`x' emplst200`x'
4      }
5

```

(continues on next page)

(continued from previous page)

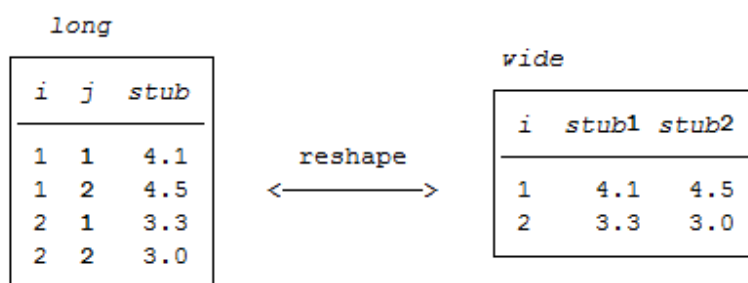
```

6      local y=2006
7      foreach w in w x y {
8          rename `w'netto netto`y'
9          rename `w'pop pop`y'
10         rename `w'phrf phrf`y'
11         local y=`y'+1
12     }
13

```

3.1. The reshape command

Now that we have made all relevant preparations, you can start to convert the dataset. If you want to convert a dataset, you can do this in both directions:



In our case, we reshape from wide to long. This means that a new variable name must be assigned for the year of the survey (j). The variable is then generated automatically. Currently, each person is assigned a line in Stata.

pid	cid	wave	sex	smoke2006	smoke2008
12345	123	x	m	yes	yes
54321	211	x	m	no	no

```

1      *reshape dataset to long-format
2      reshape long health smoke emplst hinc netto pop phrf, i(pid) j(year)
3      bys pid: gen waves=_N                               /*additional information: count_
↳number of waves per person*/
4      tab waves

```

After the reshape command, you have one line per year for each person:

pid	cid	wave	year	sex	smoke
12345	123	x	2006	m	yes
12345	123	y	2007	m	.
12345	123	z	2008	m	yes

4. Perform analyses based on the data. Try to answer the following questions:

a. Has men's and women's average satisfaction with health changed over the three years?

Satisfaction with health was measured on a scale from 1 to 10, with a value of 10 representing the highest possible level of satisfaction. To compare the average satisfaction with health between women and men, you should display the mean value for both sexes. The mean value is displayed weighted here.

```
mean health [pw=phrf], over(sex year)
```

```
.      mean health [pw=phrf], over(sex year)
```

Mean estimation Number of obs = 61,786

	Mean	Std. err.	[95% conf. interval]	
c.health@sex#year				
[1] maennlich#2006	6.584275	.0297547	6.525956	6.642594
[1] maennlich#2007	6.568137	.0302209	6.508904	6.62737
[1] maennlich#2008	6.544452	.0309931	6.483706	6.605199
[2] weiblich#2006	6.450992	.0282566	6.395609	6.506375
[2] weiblich#2007	6.406431	.02898	6.34963	6.463232
[2] weiblich#2008	6.402025	.0303387	6.342561	6.461489

The output shows the average values for men and women for all three years. The first three values show men's average satisfaction with health between 2006 and 2008, while the last three values show women's average satisfaction with health.

b. What is the proportion of people for whom health satisfaction has increased from 2006 to 2007?

To answer this question, the difference between 2006 and 2007 should be displayed. You should make sure that the analysis is conducted only within one persnr (individual identifier) and only for satisfaction in the following year.

```
sort pid year
gen diff=health-health[_n-1] if pid==pid[_n-1] & year==year[_n-1]+1
tab diff if year==2007 /*unweighted*/
```

```
.      tab diff if year==2007                                /*unweighted*/
```

diff	Freq.	Percent	Cum.
-10	5	0.03	0.03
-9	4	0.02	0.05
-8	21	0.11	0.15
-7	42	0.21	0.36
-6	98	0.49	0.86
-5	224	1.13	1.99
-4	402	2.03	4.02
-3	843	4.25	8.27
-2	1,738	8.77	17.05
-1	3,765	19.00	36.05
0	6,327	31.94	67.98
1	3,387	17.10	85.08
2	1,645	8.30	93.38
3	721	3.64	97.02
4	323	1.63	98.65
5	154	0.78	99.43
6	54	0.27	99.70
7	37	0.19	99.89
8	9	0.05	99.93
9	9	0.05	99.98
10	4	0.02	100.00
Total	19,812	100.00	

Since you have previously added the SOEP weighting factors to the dataset for your analysis, you should use the weighting for a representative analysis.

```
1      tab diff if year==2007 [aw=phrf]                        /*weighted*/
```

```
. tab diff if year==2007 [aw=phrf] /*weighted*/
```

diff	Freq.	Percent	Cum.
-10	6.4287963	0.03	0.03
-9	4.18069197	0.02	0.05
-8	26.4363266	0.14	0.19
-7	43.6171928	0.22	0.41
-6	105.6233267	0.54	0.95
-5	228.674817	1.17	2.12
-4	390.877897	2.00	4.12
-3	929.069666	4.75	8.87
-2	1,730.3911	8.85	17.72
-1	3,682.8975	18.83	36.55
0	6,042.0943	30.89	67.45
1	3,285.9198	16.80	84.25
2	1,642.6066	8.40	92.65
3	760.494745	3.89	96.53
4	359.745177	1.84	98.37
5	190.084164	0.97	99.35
6	58.9395416	0.30	99.65
7	45.1908695	0.23	99.88
8	5.78484079	0.03	99.91
9	14.9852049	0.08	99.98
10	2.957443338	0.02	100.00
Total	19,557	100.00	

The values less than 0 show a deterioration in health satisfaction. The value 0 means constant health satisfaction, and all values above 0 show a positive change in satisfaction with their health. With a value of 10, it can be assumed that these people were interviewed for the first time in 2007 or 2008.

c. In what direction and how much has satisfaction with health changed from 2006 to 2008 among people who quit smoking after 2006?

The procedure is similar to the previous question, except that the element “smoke yes/no” is added.

```

1      gen diff2=health-health[_n-2] if pid==pid[_n-2] & year==year[_n-2]+2 &
  ↪ year==2008
2      gen quit=.
3      replace quit=0 if smoke==1 & smoke[_n-2]==1 & pid==pid[_n-2] &
  ↪ year==year[_n-2]+2 & year==2008
4      replace quit=1 if smoke==2 & smoke[_n-2]==1 & pid==pid[_n-2] &
  ↪ year==year[_n-2]+2 & year==2008
5      replace quit=2 if smoke==2 & smoke[_n-2]==2 & pid==pid[_n-2] &
  ↪ year==year[_n-2]+2 & year==2008
6      replace quit=3 if smoke==1 & smoke[_n-2]==2 & pid==pid[_n-2] &
  ↪ year==year[_n-2]+2 & year==2008
7      label define quit 0 "smoker" 1 "quit" 2 "non-smoker" 3 "begin"
8      label values quit quit

```

(continues on next page)

(continued from previous page)

```
tabstat diff2, by(quit)
```

```
.      tabstat diff2, by(quit)
```

Summary for variables: diff2
Group variable: quit

quit	Mean
smoker	-.1527391
quit	-.3497076
non-smoker	-.1603457
begin	-.1383538
Total	-.1667398

To obtain a weighted mean value, address the analysis weight after the generated variable.

```
tabstat diff2 [aw=phrf], by(quit)      /*weighted*/
```

```
.      tabstat diff2 [aw=phrf], by(quit)      /*weighted*/
```

Summary for variables: diff2
Group variable: quit

quit	Mean
smoker	-.1501039
quit	-.4083783
non-smoker	-.1557582
begin	-.205733
Total	-.1677405

This illustration shows the mean of the health variable under the condition of the quit variable that we generated beforehand. With a mean of -0.24 (weighted -0.35), the biggest change in health satisfaction is seen in people who quit smoking after 2006. For example, if a person smoked in 2006 and indicated a satisfaction value of 8, the person indicates a satisfaction value of 7.76 after he/she stopped smoking in 2008. So you can assume that when a person stops smoking, their perceived health state deteriorates. Now we have to test if the assumption is correct.

d. Does quitting smoking make your health worse? To what extent could the result of the analysis “stop smoking” be distorted?

In order to establish a connection between health satisfaction and stopping smoking, one should use the t-test or to be more specific, the one-sample t-test. It checks whether the mean value of a sample deviates significantly from a known expected value (specified in the null hypothesis).


```

1      * Notes: So far we have not tested whether the difference is
2      ↪ statistically significant
      ttest diff2==0 if quit==1

```

```
.      ttest diff2==0 if quit==1
```

One-sample t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
diff2	855	-.3497076	.0715921	2.093381	-.4902247	-.2091905

```

      mean = mean(diff2)                                t =  -4.8847
H0: mean = 0                                           Degrees of freedom =    854

```

```

      Ha: mean < 0          Ha: mean != 0          Ha: mean > 0
Pr(T < t) = 0.0000      Pr(|T| > |t|) = 0.0000      Pr(T > t) = 1.0000

```

H0 Hypothesis: If one stops smoking, it has no effect on health.

For this test we assume a 95% probability. What we want to check now is whether the H0 hypothesis can be rejected or not. If you look at the output of the test, you first see the mean value of 1 (quit smoking) of the variable quit. The last line of the output shows the significance level. If it falls below the value 0.05, one can speak of a statistically significant result. In our example, the null hypothesis can be discarded because its value is less than 0.05 percent. So quitting smoking has a significant impact on a person's perceived health.

Last change: May 30, 2023

6.5 Working with harmonized Variables

This exercise shows you how to work effectively with versioned and harmonized SOEP variables. Please note that the new SOEP versioning and harmonizing concept has only been available since SOEP-Core v34 and only applies to the original SOEP-Core data in long format.

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/output
- H:/material/exercises/temp

- H:/material/exercises/log

These are used to store your script, log files, datasets, and temporary datasets. Open an empty do-file and define your paths with globals:

```

1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:\material\exercises"
5 global MY_IN_PATH "\\hume\rdc-gen\consolidated\soep-long\soep.v34"
6 global MY_DO_FILES "$AVZ\do\"
7 global MY_LOG_OUT "$AVZ\log\"
8 global MY_OUT_DATA "$AVZ\output\"
9 global MY_OUT_TEMP "$AVZ\temp\"

```

The global "AVZ" defines the main path. The main paths are subdivided using the globals "MY_IN_PATH", "MY_DO_FILES", "MY_LOG_OUT", "MY_OUT_DATA", "MY_OUT_TEMP". The global "MY_IN_PATH" contains the path to your ordered data.

1.) Differences in Response Options

Variables are versioned and harmonized because the response options have changed over time.

Is this a job creation scheme (ABM) or structural adjustment measure (SAM)?

Yes ☐ No ☐

42. Is it an "ABM" Job (created through the government employment program) or a "1 Euro Job" (for non-profit work)?

Yes, an ABM job (government employment program) ☐
 Yes, a 1 Euro job (non-profit work) ☐
 No ☐

The variable plb0038_v1 was obtained from a simple yes/no question between 1992 and 2004. Since 2005, new response options have been added. The individual questionnaires from 2004 and 2005 show these differences. Through the versioning of the variable plb0038, this difference is recognizable to the data user when tabulating the variable. The variable label also shows the beginning and end of the period in which the question was asked differently.

```

1 use "$MY_IN_PATH\pl.dta"
2 tab plb0038_v1
3 tab plb0038_v2

```

```
. tab plb0038_v1
```

Job Creation Measure Job (1992-2004)	Freq.	Percent	Cum.
[-8] Question this year not part of the	440,893	67.61	67.61
[-2] Does not apply	191,409	29.35	96.96
[-1] No answer / don't know	927	0.14	97.10
[1] Yes	1,060	0.16	97.26
[2] No	17,852	2.74	100.00
Total	652,141	100.00	

Job Creation Measure Job (2005-2014)	Freq.	Percent	Cum.
[-8] Question this year not part of the	405,647	62.18	62.18
[-6] Version of questionnaire with modi	462	0.07	62.26
[-2] Does not apply	210,244	32.23	94.48
[-1] No answer / don't know	1,211	0.19	94.67
[1] Yes, Job Creation Measure	276	0.04	94.71
[2] Yes, Community Service	368	0.06	94.77
[3] No	34,121	5.23	100.00
Total	652,329	100.00	

The variable plb0038_v1 is recoded during the harmonization process and written into a new variable, plb0038_h, together with plb0038_v2. The harmonized version of the variable should cover the survey period from 1992 to 2014 and should be usable.

```
1      tab plb0038_h
2      tabstat plb0038_v1 plb0038_v2 plb0038_h, by(syear)
```

Job Creation Measure Job (harmonized)	Freq.	Percent	Cum.
[-8] Question this year not part of the	194,399	29.80	29.80
[-6] Version of questionnaire with modi	462	0.07	29.87
[-2] Does not apply	401,653	61.57	91.44
[-1] No answer / don't know	2,138	0.33	91.77
[1] Yes, Job Creation Measure	1,336	0.20	91.98
[2] Yes, Community Service	368	0.06	92.03
[3] No	51,973	7.97	100.00
Total	652,329	100.00	

```
. tabstat plb0038_v1 plb0038_v2 plb0038_h, by(syear)
```

Summary statistics: mean

by categories of: syear (Erhebungsjahr (SurveyYear))


syear	plb003..	plb003..	plb~38_h
1984	-8	-8	-8
1985	-8	-8	-8
1986	-8	-8	-8
1987	-8	-8	-8
1988	-8	-8	-8
1989	-8	-8	-8
1990	-8	-8	-8
1991	-8	-8	-8
1992	-1.333284	-8	-1.17489
1993	-1.326277	-8	-1.163973
1994	-1.38004	-8	-1.231945
1995	-1.419015	-8	-1.280506
1996	-8	-8	-8
1997	-1.73387	-8	-1.67417
1998	-1.741104	-8	-1.681663
1999	-1.713454	-8	-1.64984
2000	-1.731038	-8	-1.667521
2001	-1.732585	-8	-1.669634
2002	-1.744852	-8	-1.684078
2003	-1.748795	-8	-1.688559
2004	-1.743176	-8	-1.681684
2005	-8	-1.277754	-1.277754
2006	-8	-1.255703	-1.255703
2007	-8	-1.238581	-1.238581
2008	-8	-1.245834	-1.245834
2009	-8	-1.238217	-1.238217
2010	-8	-1.224139	-1.224139
2011	-8	-1.325479	-1.325479
2012	-8	-1.264446	-1.264446
2013	-8	-1.433777	-1.433777
2014	-8	-1.432951	-1.432951
2015	-8	-8	-8
2016	-8	-8	-8
2017	-8	-8	-8
Total	-5.941218	-5.466984	-3.380835

2.) Differences in Coding of Response Options

Variables are versioned and harmonized because the coding of the response options has changed over time. Since the

values of certain response options can change, the various wave-specific variables cannot be integrated easily into a variable in long format. The variable must be appropriately harmonized to be useable.

21. What type of an employment change was that?

 *In the case that you have changed positions several times,
please pick the appropriate reason for the most recent change.*

I have entered employment for the first time in my life ☐ ☒

Skip to question 24!

I have started up with paid employment again
after not having been employed for a while ☐


I have started a new position with a different employer
(for temporary workers this includes working
in an temporary workplace) ☐


I have become self-employed ☐

I have been taken on by the company in which I did my
apprenticeship / worked as part of a state employment
program / was employed on a free-lance basis ☐

I have changed positions within the same company ☐

26. What type of an employment change was that?

 *In the case that you have changed positions several times,
please pick the appropriate reason for the most recent change.*

I have entered employment for the first time in my life ☐ 


Skip to question 29!

I have started up with paid employment again
after not having been employed for a while ☐

I have started a new position with a different employer
(for temporary workers this includes working
in an temporary workplace) ☐

I have been taken on by the company in which I did my
apprenticeship / worked as part of a state employment
program / was employed on a free-lance basis ☐

I have changed positions within the same company ☐

I have become self-employed ☐ 

**Did you receive funds from any
government programs to start
your own business?**

From 1994 to 2004, the question about “job change” was asked in the individual questionnaire as a category question with six response options. The order of the response options changed in 2005.

```

1      tab plb0284_v1
2      tab plb0284_v2

```

```
. tab plb0284_v1
```

Type Of Job Change (1994–2004)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	453,958	69.61	69.61
[-2] Does not apply	176,060	27.00	96.61
[-1] no answer	262	0.04	96.65
[1] First Time Employed	3,036	0.47	97.11
[2] Job After Break	6,571	1.01	98.12
[3] Job With New Employer	8,852	1.36	99.48
[4] New Job-Self Employed	1,319	0.20	99.68
[5] Company Taken Over	451	0.07	99.75
[6] Changed Job, Same Firm	1,632	0.25	100.00
Total	652,141	100.00	

```
. tab plb0284_v2
```

Type Of Job Change (2005–2017)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	316,675	48.56	48.56
[-5] Not Included In Questionnaire Vers	15,139	2.32	50.88
[-2] Does not apply	281,276	43.13	94.01
[-1] no answer	115	0.02	94.03
[1] First Job	4,342	0.67	94.70
[2] Returned to Past Employer After Bre	6,539	1.00	95.70
[3] New Position Different Employer	21,628	3.32	99.01
[4] Taken On Ba Company	1,519	0.23	99.25
[5] Changed Position Within Company	2,664	0.41	99.66
[6] New Job Self-Employed	2,244	0.34	100.00
Total	652,141	100.00	

In addition to the different order of the response options, the coding order also changed. The data are stored in the wave-specific “raw” datasets with different coding and are contained in the variables plb0284_v1 and plb0284_v2. To use the variable for all survey years, it is necessary to harmonize the different versions. The variable plb0284_v1 is recoded (recode (1=1)(2=2)(3=3)(4=6)(5=4)(6=5)) and then written together with plb0284_v2 as plb0284_h. The new variable plb0284_h is created by the harmonization process.

```

1      tab plb0284_h
2      tabstat plb0284_v1 plb0284_v2 plb0284_h, by(syear)

```

. tab plb0284_h

Type Of Job Change (harmonized)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	118,492	18.17	18.17
[-5] Not Included In Questionnaire Vers	15,139	2.32	20.49
[-2] Does not apply	457,336	70.13	90.62
[-1] no answer	377	0.06	90.68
[1] First Job	7,378	1.13	91.81
[2] Returned to Past Employer After Bre	13,110	2.01	93.82
[3] New Position Different Employer	30,480	4.67	98.49
[4] Taken On Ba Company	1,970	0.30	98.79
[5] Changed Position Within Company	4,296	0.66	99.45
[6] New Job Self-Employed	3,563	0.55	100.00
Total	652,141	100.00	

```
. tabstat plb0284_v1 plb0284_v2 plb0284_h, by(syear)
```

Summary statistics: mean

by categories of: syear (Erhebungsjahr (SurveyYear))

syear	p~284_v1	p~284_v2	plb028~h
1984	-8	-8	-8
1985	-8	-8	-8
1986	-8	-8	-8
1987	-8	-8	-8
1988	-8	-8	-8
1989	-8	-8	-8
1990	-8	-8	-8
1991	-8	-8	-8
1992	-8	-8	-8
1993	-8	-8	-8
1994	-1.499068	-8	-1.496609
1995	-1.430709	-8	-1.426859
1996	-1.503738	-8	-1.500259
1997	-1.474215	-8	-1.471806
1998	-1.471506	-8	-1.471097
1999	-1.396592	-8	-1.397586
2000	-1.432495	-8	-1.428792
2001	-1.416805	-8	-1.413583
2002	-1.483425	-8	-1.480119
2003	-1.533944	-8	-1.531953
2004	-1.560425	-8	-1.555384
2005	-8	-1.564937	-1.564937
2006	-8	-1.548081	-1.548081
2007	-8	-1.492387	-1.492387
2008	-8	-1.480898	-1.480898
2009	-8	-1.473259	-1.473259
2010	-8	-1.394461	-1.394461
2011	-8	-1.482337	-1.482337
2012	-8	-1.418576	-1.418576
2013	-8	-1.937847	-1.937847
2014	-8	-1.311269	-1.311269
2015	-8	-1.540338	-1.540338
2016	-8	-1.865138	-1.865138
2017	-8	-1.635675	-1.635675
Total	-6.017658	-4.686554	-2.703361



3.) Content Differences in the Questions.

Variables are versioned when questions were asked differently in different years but the content belongs together. If the

content or wording of the question changes, the wave-specific variables cannot easily be integrated into a long variable.

108. Have you yourself ever inherited something or received a gift of great value?

We are referring to gifts or inheritance of house or land, securities, investments, other forms of wealth or large amounts of money.

Yes ☐  No ☐  Skip to question 109!

155. Have you personally received an inheritance or larger endowment in the last 15 years?

We are referring mainly to transfers of home or property ownership, securities, participating interests, and other assets or larger sums of money.

Yes..... ☐  No ☐  Question 157!

In the 2001 individual questionnaire, respondents were asked whether they had ever received an inheritance. In 2017, this question was worded differently: respondents were asked whether they had received an inheritance in the last 15 years. The questions are similar but cover different time periods. Therefore, the variable is not harmonized but made available as versioned variables. Data users have to decide whether or not to use the variables in the same way.

tab plc0375_v1

tab plc0375_v2

. tab plc0375_v1

Erbschaft (jemals) (2001)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	629,790	96.57	96.57
[-1] no answer	140	0.02	96.59
[1] Yes	3,307	0.51	97.10
[2] No	18,904	2.90	100.00
Total	652,141	100.00	


. tab plc0375_v2

Erbschaft (letzte 15 Jahre) (2017)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	619,656	95.02	95.02
[-5] Not Included In Questionnaire Vers	5,703	0.87	95.89
[-1] no answer	187	0.03	95.92
[1] Yes	2,683	0.41	96.33
[2] No	23,912	3.67	100.00
Total	652,141	100.00	

4.) Change of Question Type.

Variables are versioned and harmonized when questions were asked differently in different years, for example, first as a question with multiple response options and later as a question with a single response option. The possibility to provide multiple answers in certain years makes it difficult to integrate the wave-specific variables into a variable in long format.

Do you receive a scholarship to pay for your undergraduate or graduate studies?


 *If so, from what organization?*

No ☐

Yes, BAföG ☐

Yes, other ☐

Do you receive a grant/scholarship to pay for your undergraduate or graduate studies?

 *If so, from what organization?*

No ☐

Yes, BAföG ☐

Yes. other ☐

When comparing the question on scholarships in the individual questionnaires from 2011 and 2012, it appears that there should be no differences in the variables. Nevertheless, the two questions seem to have been asked differently and stored differently in the raw datasets. This results in several versioned variables.

```
1      tab plg0015_v1
2      tab plg0015_v2
3      tab plg0015_v3
4      tab plg0015_v4
```

Studium: Stipendium - Einfachnennung (2007-2011)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	535,326	82.09	82.09
[-2] Does not apply	112,989	17.33	99.41
[-1] no answer	352	0.05	99.47
[1] No student aid, stipend	2,670	0.41	99.88
[2] Yes, student aid	651	0.10	99.98
[3] Yes, other	153	0.02	100.00
Total	652,141	100.00	

. tab plg0015_v2

Studium: Stipendium - Kein Stipendium (2012-2017)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	476,953	73.14	73.14
[-5] Not Included In Questionnaire Vers	11,825	1.81	74.95
[-2] Does not apply	158,864	24.36	99.31
[-1] no answer	337	0.05	99.36
[1] No student aid, stipend	4,162	0.64	100.00
Total	652,141	100.00	

. tab plg0015_v3

Studium: Stipendium - Ja, BAfoeG (2012-2017)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	476,953	73.14	73.14
[-5] Not Included In Questionnaire Vers	11,825	1.81	74.95
[-2] Does not apply	162,090	24.86	99.80
[1] Yes, student aid	1,273	0.20	100.00
Total	652,141	100.00	

```
. tab plg0015_v4
```

Studium: Stipendium - Ja, sonstiges Stipendium (2012-2017)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	476,953	73.14	73.14
[-5] Not Included In Questionnaire Vers	11,825	1.81	74.95
[-2] Does not apply	163,052	25.00	99.95
[-1] no answer	2	0.00	99.95
[1] Yes, other	309	0.05	100.00
Total	652,141	100.00	

As you can see, the variable was asked from 2007 to 2011 as a category question with three response options. As a result, respondents could only give one answer. Since 2012, the question has used binary items. It is quite possible that a respondent gave more than one answer. The harmonized version of the variable integrates the binary items from plg0015_v2, plg0015_v3, and plg0015_v4 into the harmonized version plg0015_h. The coding of the variable plg0015_v1 is used as the generation framework. In addition, the harmonization proposal takes into account the problematic multiple answers with the value four.

```
1 plg0015_h
2 tabstat plg0015_v1 plg0015_v2 plg0015_v3 plg0015_v4 plg0015_h, by(syear)
```

```
. tab plg0015_h
```

University: Scholarship (harmonized)	Freq.	Percent	Cum.
[-8] Question this year not part of sur	360,138	55.22	55.22
[-2] Does not apply	282,456	43.31	98.54
[-1] no answer	352	0.05	98.59
[1] No student aid, stipend	6,832	1.05	99.64
[2] Yes, student aid	1,901	0.29	99.93
[3] Yes, other	439	0.07	100.00
[4] Multiple Answers	23	0.00	100.00
Total	652,141	100.00	

```
. tabstat plg0015_v1 plg0015_v2 plg0015_v3 plg0015_v4 plg0015_h, by(syear)
```

Summary statistics: mean

by categories of: syear (Erhebungsjahr (SurveyYear))

syear	plg001..	plg001..	pl~15_v3	pl~15_v4	plg~15_h
1984	-8	-8	-8	-8	-8
1985	-8	-8	-8	-8	-8
1986	-8	-8	-8	-8	-8
1987	-8	-8	-8	-8	-8
1988	-8	-8	-8	-8	-8
1989	-8	-8	-8	-8	-8
1990	-8	-8	-8	-8	-8
1991	-8	-8	-8	-8	-8
1992	-8	-8	-8	-8	-8
1993	-8	-8	-8	-8	-8
1994	-8	-8	-8	-8	-8
1995	-8	-8	-8	-8	-8
1996	-8	-8	-8	-8	-8
1997	-8	-8	-8	-8	-8
1998	-8	-8	-8	-8	-8
1999	-8	-8	-8	-8	-8
2000	-8	-8	-8	-8	-8
2001	-8	-8	-8	-8	-8
2002	-8	-8	-8	-8	-8
2003	-8	-8	-8	-8	-8
2004	-8	-8	-8	-8	-8
2005	-8	-8	-8	-8	-8
2006	-8	-8	-8	-8	-8
2007	-1.893852	-8	-8	-8	-1.893852
2008	-1.904288	-8	-8	-8	-1.904288
2009	-1.897845	-8	-8	-8	-1.897845
2010	-1.900711	-8	-8	-8	-1.900711
2011	-1.900707	-8	-8	-8	-1.900707
2012	-8	-1.934174	-1.978987	-1.994318	-1.898939
2013	-8	-1.925927	-1.975288	-1.994218	-1.885709
2014	-8	-1.920044	-1.975969	-1.994539	-1.881413
2015	-8	-2.122834	-2.175919	-2.19159	-1.889821
2016	-8	-2.391366	-2.439758	-2.45542	-1.89875
2017	-8	-2.438633	-2.49509	-2.509958	-1.892627
Total	-6.907259	-6.422924	-6.436731	-6.441162	-5.266231

5.) Euro harmonisation

Variables are versioned and harmonized because they are metric and were asked as DM amounts before the introduction

of the euro. For the long version of the variable, metric variables based on different currencies in different years are harmonized as euro amounts.

Most of the variables harmonized in the long datasets are amounts of money. Before the introduction of the euro, such information was collected in DM.

57. How high was your income from employment last month?

- ☞ If you received extra income such as vacation pay or back pay, please do **not** include this. Please do include overtime pay.
- ☞ Please do **not** include "Kindergeld", even if this is paid by the employer.

Please fill in both:

- **gross** income, which means wages or salary before deduction of taxes and social security
- **net** income, which means the sum after deduction of taxes, social security, and unemployment and health insurance.

My income was:

gross

DM

58. How high was your income from employment last month?

- ☞ If you received extra income such as vacation pay or back pay, please do **not** include this. Please do include overtime pay.
- ☞ Please do **not** include "Kindergeld", even if this is paid by the employer.
- ☞ If you are self-employed: Please estimate your monthly income before and after tax.

Please fill in both:

- gross income, which means wages or salary before deduction of taxes and social security
- net income, which means the sum after deduction of taxes, social security, and unemployment and health insurance.

My income was:

gross

EURO

Euro harmonisation involves DM amounts being multiplied by the exchange rate so that the harmonized version of the variable represents euro amounts.

```
1 list pid syeay plc0013_v1 plc0013_h if pid==7006001 & syeay==2001
2 tabstat plc0013_v1 plc0013_v2 plc0013_h, by(syeay)
```

```
. list pid syeay plc0013_v1 plc0013_h if pid==7006001 & syeay==2001
```

	pid	syeay	plc001..	plc~13_h
478114.	7006001	2001	4200	2147

```
. tabstat plc0013_v1 plc0013_v2 plc0013_h, by(syear)
```

Summary statistics: mean

by categories of: syear (Erhebungsjahr (SurveyYear))

syear	plc001..	pl~13_v2	plc~13_h
1984	1307.255	-8	667.9481
1985	1345.458	-8	687.482
1986	1418.1	-8	724.6339
1987	1467.424	-8	749.8613
1988	1521.479	-8	777.4971
1989	1620.861	-8	828.3211
1990	1143.434	-8	583.5691
1991	1513.081	-8	773.2402
1992	1651.645	-8	844.0551
1993	1807.099	-8	923.5413
1994	1872.348	-8	956.8878
1995	1938.037	-8	990.4832
1996	2003.138	-8	1023.763
1997	1990.75	-8	1017.419
1998	1952.576	-8	997.8835
1999	2046.094	-8	1045.72
2000	2002.836	-8	1023.575
2001	2055.956	-8	1050.741
2002	-8	1307.774	1307.774
2003	-8	1251.525	1251.525
2004	-8	1256.811	1256.811
2005	-8	1224.242	1224.242
2006	-8	1208.506	1208.506
2007	-8	1253.102	1253.102
2008	-8	1265.268	1265.268
2009	-8	1248.049	1248.049
2010	-8	1248.695	1248.695
2011	-8	1366.912	1366.912
2012	-8	1399.632	1399.632
2013	-8	1402.34	1402.34
2014	-8	1456.721	1456.721
2015	-8	1457.494	1457.494
2016	-8	1289.785	1289.785
2017	-8	1346.437	1346.437
Total	659.0953	814.6312	1157.022

Last change: Mar 28, 2023

To get an idea of the analysis potential offered by the SOEP, we recommend the following exercises to our users:

6.6 Longitudinal Data Analysis

Simple cross-sectional analyses show that married people have higher life satisfaction than singles. You want to check this on the basis of longitudinal analysis with the SOEP.

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

These are used to store your script, log files, datasets, and temporary datasets. Open an empty do-file and define the paths you created with globals:

```

1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:\Exercise\"
5 global MY_IN_PATH "\\hume\rdc-prod\distribution\soep-core\soep.v37\eu\
↩Stata\"
6
7 global MY_DO_FILES "$AVZ\do\"
8 global MY_LOG_OUT "$AVZ\log\"
9 global MY_OUT_DATA "$AVZ\output\"
10 global MY_OUT_TEMP "$AVZ\temp\"

```

The global “AVZ” defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to your ordered data.

Create a master file that uses the important variables from ppathl.

You should always add some variables from PPATHL to your dataset by default. Download the following information from PPATHL:

- Individual identifier "**pid**"
- Household identifier "**pid**"
- Survey year "**syear**"
- The net variable with information on the interview type "**netto**"
- The weighting variable "**phrf**"
- The gender of the person "**sex**"
- The migration background "**migback**"


```

1      *** Step 1) Start with basic information from PPFADL ***
2
3      use pid hid syear netto phrf migback sex using ${MY_IN_PATH}\ppathl.dta

```

Search for matching variables and add them to your dataset

To perform your analysis, you need different SOEP variables. The SOEP offers various options for a variable search:

- Search the questionnaires for useful variables. (for more information, see the section [Variable Search with Questionnaires](#))
- Find a suitable variable via the topic list of paneldata.org (for more information, see the section [Topic Search with paneldata.org](#))
- Search for a suitable variable using a search term in paneldata.org (for more information, see the section [Variable Search with paneldata.org](#))
- Use the documentation provided on the generated variables (for more information, see the section [Documentation on Generated Data](#))

In this case, we use the variables **"pgfamstd"** (marital status) and **"plh0182"** (life satisfaction).

```

1      *** Step 2) Add the relevant variables: here: family status and life
2      ←satisfaction ***
3      merge 1:1 pid syear using ${MY_IN_PATH}\pgen, keepusing(pgfamstd) keep(1
4      ←3) nogen
5
6      merge 1:1 pid syear using ${MY_IN_PATH}\pl, keepusing(plh0182) keep(1 3)
7      ←nogen
8
9      save ${MY_OUT_DATA}\ppathl.dta, replace

```

6.6.1 Clean and inspect the data

Encode all missing values to system missing. Since you are interested in individual characteristics in your analysis: Delete all measurements that are not based on successful individual interviews.

```

1      mvdecode _all, mv(-8/-1)
2
3      tab netto
4      drop if netto>19

```

```
. tab netto
```

Current Wave Survey Status	Freq.	Percent	Cum.
[10] Interviewee With Successful Interv	514,447	52.79	52.79
[12] Individual Questionnaire And Perso	59,730	6.13	58.92
[13] Individual Questionnaire And Youth	318	0.03	58.95
[14] Individual Questionnaire And Other	32	0.00	58.96
[15] Individual Questionnaire And Exper	38,663	3.97	62.92
[16] Individual Questionnaire, First Ti	5,946	0.61	63.53
[17] Youth Biography First Time Surveye	4,859	0.50	64.03
[18] Individual Questionnaire And Child	8	0.00	64.03
[19] Individual Questionnaire Without H	538	0.06	64.09
[20] Children in Successfully Interviewe	169,841	17.43	81.52
[21] Children With Mother-Child Questio	5,318	0.55	82.06
[22] Children With Mother-Child Questio	5,792	0.59	82.66
[23] Children With Mother-Child Questio	5,457	0.56	83.22
[24] Children age 7-8, with parental qu	4,875	0.50	83.72
[25] Children age 9-10, with parental q	4,097	0.42	84.14
[26] Students Age 11-12	1,759	0.18	84.32
[27] Children with Mother-Child Questio	2,186	0.22	84.54
[28] Youth questionnaire, Age 13-14	526	0.05	84.60
[29] Jugendliche 16-17 Jahre (ohne Juge	222	0.02	84.62
[30] Persons In Successfully Interviewe	128,343	13.17	97.79
[31] Successful Gap Interview (_LUECKE)	8,401	0.86	98.65
[32] Successfully Completed Biography Q	35	0.00	98.65
[33] Successful Youth Questionnaire	22	0.00	98.66
[34] Successful Tests and Experiments	122	0.01	98.67
[61] Gap Interview without HH reference	35	0.00	98.67
[62] Gap Interview with drop out	5	0.00	98.67
[80] Individual Without Any Current Inf	642	0.07	98.74
[81] Prior Interviewee Without Any Curr	359	0.04	98.78
[88] Repatriate - (moved abroad before	75	0.01	98.78
[89] Repatriate - (was drop out [90])	256	0.03	98.81
[90] Individual Dropouts PBR_EXIT	3,835	0.39	99.20
[91] Moved abroad	2,158	0.22	99.42
[92] Moved abroad (abroad)	177	0.02	99.44
[93] Moved abroad (exit)	65	0.01	99.45
[97] advice to dead person (exit)	981	0.10	99.55
[98] advice to dead person (_VP)	122	0.01	99.56
[99] Has Died	4,262	0.44	100.00
Total	974,509	100.00	

How many people contribute measurements and what is the proportion of people contributing at least 10 waves in a row?

Define the dataset as a panel dataset.

```
xtset pid syer
```

(continues on next page)

(continued from previous page)

xtdes

```
. xtdes

      pid: 101, 102, ..., 55404001      n =      105068
     syear: 1984, 1985, ..., 2020      T =         37
      Delta(syear) = 1 unit
      Span(syear)  = 37 periods
      (pid*syear uniquely identifies each observation)

Distribution of T_i:  min      5%      25%      50%      75%      95%      max
                   1         1         2         4        10        24        37

      Freq.  Percent   Cum.   |   Pattern
      -----|-----
      4444      4.23    4.23   |   .....1
      3531      3.36    7.59   |   .....1111
      2779      2.64   10.24   |   .....11
      2171      2.07   12.30   |   .....1111111111
      2049      1.95   14.25   |   .....1.....
      1889      1.80   16.05   |   .....1.....
      1860      1.77   17.82   |   .....1111111111
      1740      1.66   19.48   |   .....11.....
      1610      1.53   21.01   |   .....1...
      82995     78.99  100.00   |   (other patterns)
      -----|-----
      105068   100.00         |   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

105,068 respondents have contributed information in waves a (1984) to bk (2020) and 75% of the 105,068 respondents have provided information for at least 10 waves.

How many people took part in the survey in 2010 and contributed to continuous measurements up to 2014?

```
xtdes if syear>=2010 & syear<=2014
```

```
. xtset if syear>=2010 & syear<=2014
```

```
      pid: 602, 901, ..., 35033302      n =      45438
    syear: 2010, 2011, ..., 2014      T =         5
      Delta(syear) = 1 unit
      Span(syear)  = 5 periods
      (pid*syear uniquely identifies each observation)
```

```
Distribution of T_i:  min      5%      25%      50%      75%      95%      max
                     1         1         2         3         5         5         5
```

Freq.	Percent	Cum.	Pattern
14673	32.29	32.29	11111
4992	10.99	43.28	1....
4342	9.56	52.83	.1111
4234	9.32	62.15	...11
2669	5.87	68.03	11...
2307	5.08	73.10	..111
1924	4.23	77.34	1111.
1742	3.83	81.17	...1.
1548	3.41	84.58	111..
7007	15.42	100.00	(other patterns)
45438	100.00		XXXXXX

14,673 respondents provided continuous information from 2010 to 2014.

6.6.2 Univariate inspection & analysis

How does the mean of life satisfaction change over time?

```
1      *** Step 4) univariate inspection & analysis
2      table syear, statistic (mean plh0182)
```

```
. table syear, statistic (mean plh0182)
```

	Mean
Erhebungsjahr	
1984	7.425771
1985	7.237014
1986	7.285553
1987	7.137283
1988	7.082565
1989	7.101457
1990	7.049266
1991	6.948061
1992	6.915608
1993	6.884618
1994	6.857764
1995	6.887923
1996	6.900363
1997	6.792789
1998	6.949559
1999	6.968906
2000	7.088658
2001	7.104758
2002	7.045965
2003	6.963976
2004	6.800537
2005	6.948051
2006	6.914468
2007	6.94629
2008	6.981673
2009	6.976511
2010	7.246195
2011	7.178485
2012	7.192234
2013	7.314208
2014	7.247232
2015	7.380125
2016	7.357209
2017	7.335349
2018	7.403988
2019	7.539315
2020	7.589828
Total	7.147626

What proportion of people are a) married in 2014 or b) have a migration background? Compare weighted with unweighted frequency tables: Who is overrepresented in SOEP?

```

1      tab1 pgfamstd migback if syeare==2014
2      tab pgfamstd [aw=phrf] if syeare==2014
3      tab migback [aw=phrf] if syeare==2014

```

```
. tab1 pgfamstd migback if syeare==2014
```

```
-> tabulation of pgfamstd if syeare==2014
```

Marital Status	Freq.	Percent	Cum.
[1] Verheiratet, mit Ehepartner zusammen	16,130	57.70	57.70
[2] Verheiratet, dauernd getrennt leben	674	2.41	60.11
[3] Ledig	7,118	25.46	85.57
[4] Geschieden / eingetragene gleiches	2,477	8.86	94.43
[5] Verwitwet / Lebenspartner/-in aus e	1,469	5.25	99.69
[6] Ehepartner im Ausland	14	0.05	99.74
[7] Eingetragene gleichgeschlechtliche	57	0.20	99.94
[8] Eingetragene gleichgeschlechtliche	17	0.06	100.00
Total	27,956	100.00	

```
. tab pgfamstd [aw=phrf] if syeare==2014
```

Marital Status	Freq.	Percent	Cum.
[1] Verheiratet, mit Ehepartner zusammen	14,044.077	50.70	50.70
[2] Verheiratet, dauernd getrennt leben	655.879262	2.37	53.06
[3] Ledig	8,039.8636	29.02	82.08
[4] Geschieden / eingetragene gleiches	2,637.7814	9.52	91.61
[5] Verwitwet / Lebenspartner/-in aus e	2,214.2843	7.99	99.60
[6] Ehepartner im Ausland	26.5442746	0.10	99.69
[7] Eingetragene gleichgeschlechtliche	57.1027568	0.21	99.90
[8] Eingetragene gleichgeschlechtliche	27.4673885	0.10	100.00
Total	27,703	100.00	

The data show that married people are overrepresented in the SOEP and single people are underrepresented. The weighting makes it representative again for Germany.

-> tabulation of migback if syeare==2014

Migrationshintergrund	Freq.	Percent	Cum.
[1] kein Migrationshintergrund	20,846	74.34	74.34
[2] direkter Migrationshintergrund	5,241	18.69	93.03
[3] indirekter Migrationshintergrund	1,955	6.97	100.00
Total	28,042	100.00	

. tab migback [aw=phrf] if syeare==2014

Migrationshintergrund	Freq.	Percent	Cum.
[1] kein Migrationshintergrund	22,410.1885	80.65	80.65
[2] direkter Migrationshintergrund	3,921.7779	14.11	94.77
[3] indirekter Migrationshintergrund	1,454.0336	5.23	100.00
Total	27,786	100.00	

In the SOEP sample, respondents with a direct or indirect migration background are overrepresented.

How many of those persons who reported a life satisfaction scale value of 7 in one survey year also indicated the scale value of 7 in the following survey year?

xttrans plh0182

. xttrans plh0182

Current Life Satisfaction	Current Life Satisfaction											Total
	0	1	2	3	4	5	6	7	8	9	10	
0	20.30	8.31	10.61	11.19	7.47	19.50	5.71	5.84	6.37	1.95	2.74	100.00
1	8.61	10.60	15.58	13.55	9.53	17.08	6.58	6.77	6.58	3.39	1.74	100.00
2	3.77	5.18	14.47	16.75	11.29	19.24	8.82	8.79	7.98	2.43	1.26	100.00
3	1.86	2.45	7.79	16.11	14.66	23.04	11.64	11.24	8.34	2.00	0.87	100.00
4	0.89	1.24	4.19	10.55	15.47	26.02	15.54	14.43	8.92	1.94	0.81	100.00
5	0.75	0.66	2.06	5.10	7.86	32.32	16.97	17.50	12.70	2.46	1.60	100.00
6	0.24	0.32	1.07	2.81	4.98	18.20	22.66	27.74	17.53	3.08	1.37	100.00
7	0.13	0.14	0.54	1.53	2.42	9.53	14.20	34.57	29.86	5.42	1.66	100.00
8	0.10	0.11	0.36	0.79	1.11	5.20	6.57	21.77	46.31	14.03	3.65	100.00
9	0.10	0.12	0.25	0.45	0.63	2.69	3.15	10.34	36.80	36.06	9.40	100.00
10	0.29	0.13	0.30	0.61	0.68	4.09	2.90	7.51	23.21	23.28	37.01	100.00
Total	0.44	0.43	1.23	2.58	3.48	11.67	10.92	21.61	30.36	12.03	5.25	100.00

34.57% of the respondents who reported a life satisfaction of 7 again reported a value of 7 in the following year.

Is it more likely that a highly dissatisfied person (value: 0) will be less dissatisfied the following year or that a very satisfied (value: 10) person will be less satisfied the following year?

xttrans plh0182

```
. xttrans plh0182
```

Current Life Satisfacti on	Current Life Satisfaction											Total
	0	1	2	3	4	5	6	7	8	9	10	
0	20.30	8.31	10.61	11.19	7.47	19.50	5.71	5.84	6.37	1.95	2.74	100.00
1	8.61	10.60	15.58	13.55	9.53	17.08	6.58	6.77	6.58	3.39	1.74	100.00
2	3.77	5.18	14.47	16.75	11.29	19.24	8.82	8.79	7.98	2.43	1.26	100.00
3	1.86	2.45	7.79	16.11	14.66	23.04	11.64	11.24	8.34	2.00	0.87	100.00
4	0.89	1.24	4.19	10.55	15.47	26.02	15.54	14.43	8.92	1.94	0.81	100.00
5	0.75	0.66	2.06	5.10	7.86	32.32	16.97	17.50	12.70	2.46	1.60	100.00
6	0.24	0.32	1.07	2.81	4.98	18.20	22.66	27.74	17.53	3.08	1.37	100.00
7	0.13	0.14	0.54	1.53	2.42	9.53	14.20	34.57	29.86	5.42	1.66	100.00
8	0.10	0.11	0.36	0.79	1.11	5.20	6.57	21.77	46.31	14.03	3.65	100.00
9	0.10	0.12	0.25	0.45	0.63	2.69	3.15	10.34	36.80	36.06	9.40	100.00
10	0.29	0.13	0.30	0.61	0.68	4.09	2.90	7.51	23.21	23.28	37.01	100.00
Total	0.44	0.43	1.23	2.58	3.48	11.67	10.92	21.61	30.36	12.03	5.25	100.00

The rows reflect the initial values, and the columns reflect the final values. Around 20% of those who were completely dissatisfied (value: 0) in the base year remained completely dissatisfied in the following year. About 80% of these completely dissatisfied people from the base year were more satisfied in the following year. Of the completely satisfied persons (value: 10), about 37% remained just as satisfied in the following year, but 63% became less satisfied. It is more likely that a completely dissatisfied person will become more satisfied in the following year than that a completely satisfied person will become less satisfied.

Which transitions in marital status can be observed particularly frequently in the data?

xttrans pgfamstd

```
. xttrans pgfamstd
```

Marital Status In Survey Year	Marital Status In Survey Year								Total
	1	2	3	4	5	6	7	8	
1	98.49	0.90	0.00	0.10	0.50	0.01	0.00	0.00	100.00
2	4.09	74.86	0.00	18.55	1.43	1.07	0.00	0.00	100.00
3	4.09	0.15	95.63	0.02	0.00	0.06	0.04	0.01	100.00
4	4.08	0.25	0.00	95.62	0.00	0.00	0.03	0.01	100.00
5	0.36	0.07	0.00	0.00	99.57	0.00	0.00	0.00	100.00
6	12.44	25.84	0.00	0.16	0.00	61.56	0.00	0.00	100.00
7	0.00	0.00	0.00	0.32	0.00	0.00	95.82	3.86	100.00
8	0.00	0.00	0.00	3.92	1.96	0.00	5.88	88.24	100.00
Total	62.00	2.17	22.53	6.83	6.27	0.11	0.07	0.01	100.00

Survey respondents who were married but lived separated [value 2] in the base year and reported divorce as their family status in the following year [value 4] can be observed particularly frequently (about 19%).

6.6.3 Simple cross sectional analyses

You now want to find the correlation between marital status and life satisfaction. Is there an effect of marriage on life satisfaction? And if so, is it a sustained effect?

First, calculate the correlation between family status and life satisfaction from a cross-sectional perspective for 2010: Are married people happier than singles?

```
1 table pgfamstd if syear==2010, statistic (mean plh0182)
```

```
. table pgfamstd if syear==2010, statistic (mean plh0182)
```

	Mean
Marital Status In Survey Year	
[1] Married	7.39547
[2] Married, But Separated	6.757911
[3] Single	7.206073
[4] Divorced	6.704758
[5] [de] Verwitwet / Lebenspartner/-in aus eingetragener gleichgeschlechtlicher Partnerschaft verstorben	6.786533
[6] husband/wife abroad	7.666667
[7] Registered Same-Sex Partnership, Living Together	7.269231
[8] Registered Same-Sex Partnership, Living Apart	7
Total	7.24684

At first glance, married couples seem happier than singles.

Now generate a variable that indicates a transition from “single” to “married”.

How many such transitions can you find in the data?

```
1 **define event: transition to marriage
2 generate to_mar=1 if pgfamstd==1 & l.pgfamstd==3
3 tab to_mar
```

```
. tab to_mar
```

to_mar	Freq.	Percent	Cum.
1	5,559	100.00	100.00
Total	5,559	100.00	

A total of 5,559 people can be observed changing status from single to married.

What is the average level of life satisfaction immediately after the transition to marriage (i.e., in the first survey in which the transition can be observed) and how high is life satisfaction immediately before the transition to marriage?

```
1 **standard way of life-event analysis
2 sum plh0182 if to_mar==1
3 sum l.plh0182 if to_mar==1
```

(continues on next page)

(continued from previous page)

```

**alternative way
generate dif_sat= plh0182- l.plh0182

```

```
. sum plh0182 if to_mar==1
```

Variable	Obs	Mean	Std. dev.	Min	Max
plh0182	5,548	7.692682	1.501938	0	10

```
. sum l.plh0182 if to_mar==1
```

Variable	Obs	Mean	Std. dev.	Min	Max
plh0182 l1.	5,527	7.592727	1.519223	0	10

Before the transition to marriage, the average life satisfaction of the respondents is 7.59. In the following year, that is, after the transition to marriage, the average life satisfaction of the respondents is 7.69. It can be seen that with the transition to marriage, average life satisfaction rises slightly by 0.10.

Map the complete satisfaction history around the “marriage entry” event [3 years before; 3 years after].

```

**preparing illustration of trajectory
generate t=0 if to_mar==1 & l.to_mar~=1 & l2.to_mar~=1 & l3.to_mar~=1 &
l4.to_mar~=1 & l5.to_mar~=1 & l6.to_mar~=1 & l7.to_mar~=1 & l8.to_mar~=1 & l9.to_mar~
=1 & l10.to_mar~=1 & l11.to_mar~=1 & l12.to_mar~=1 & l13.to_mar~=1 & l14.to_mar~=1
replace t=1 if l.t==0
replace t=2 if l2.t==0
replace t=3 if l3.t==0
replace t=-1 if f.t==0
replace t=-2 if f2.t==0
replace t=-3 if f3.t==0

table t, statistic (mean plh0182 n plh0182)

```

```
. table t, statistic (mean plh0182 n plh0182)
```

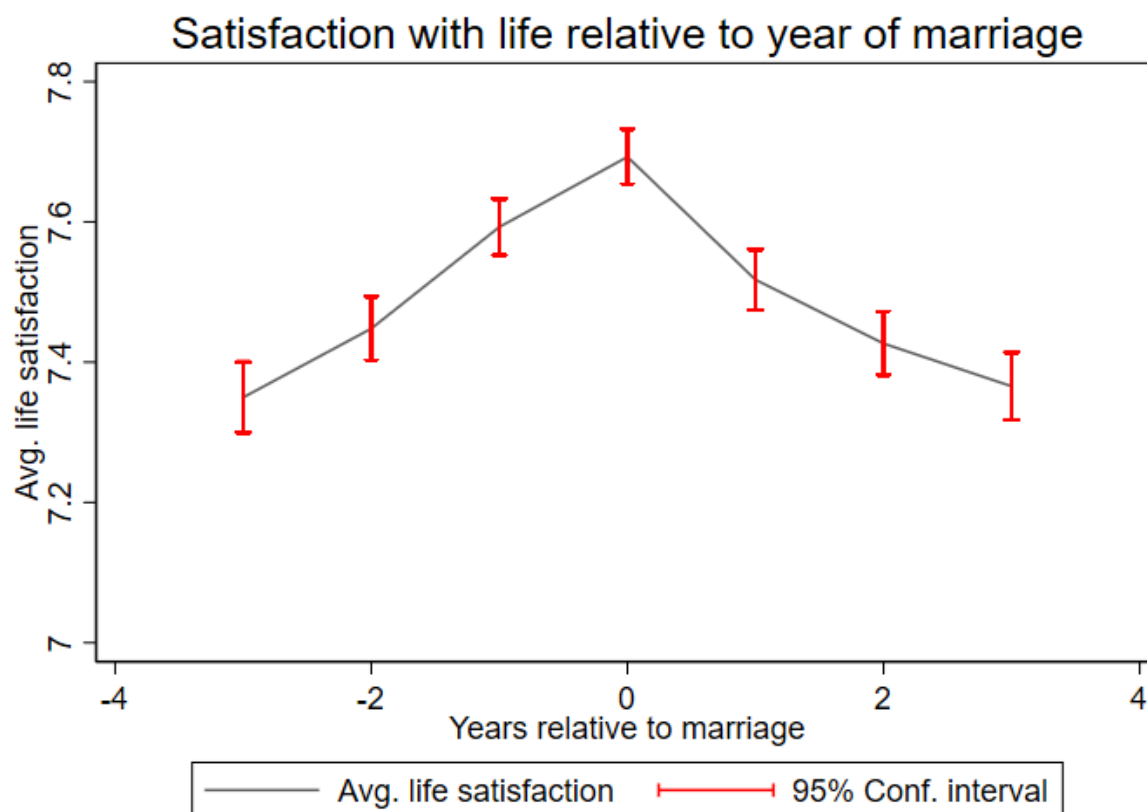
	Current Life Satisfaction	Current survey status
t		
-3	7.34979	10.58559
-2	7.447857	10.55619
-1	7.592727	10.58032
0	7.692682	10.43713
1	7.51758	10.37189
2	7.426553	10.38541
3	7.365672	10.37767
Total	7.499756	10.47109

Choose a suitable presentation for your results and let Stata create a graphic.

```

1      ** Preparing graph of event
↳analysis
2      sort t
3      cap drop meanplh0182
4      by t: egen meanplh0182 = mean(plh0182)
5
6      cap drop upper
7      gen upper = .
8      forval i = -3/3{
9      su plh0182 if t == `i'
10     replace upper = r(mean) + 1.96 * r(sd)/sqrt(r(N)) if t == `i'
11     }
12
13     cap drop lower
14     gen lower = .
15     forval i = -3/3{
16     su plh0182 if t == `i'
17     replace lower = r(mean) - 1.96 * r(sd)/sqrt(r(N)) if t == `i'
18     }
19
20     twoway (line meanplh0182 t) (rcap upper lower t, lcolor("red")) , title(
↳ "Satisfaction with life relative to year of marriage") legend(label(1 "Avg. life
↳ satisfaction") label(2 "95% Conf. interval")) scheme(slmono) xtitle("Years relative to
↳ marriage") ytitle("Avg. life satisfaction")

```



The graph shows that a positive effect on life satisfaction can be observed when family status changes from single to married. In the following years of the existing marriage, life satisfaction decreases again and approaches the initial satisfaction before the marriage.

Last change: Jun 01, 2023

6.7 Working with Migration Data (BIOIMMIG)

With its migration and refugee samples, SOEP provides a wide range of information on people with a history of migration or forced migration.

In the BIOIMMIG dataset, you will find relevant information on the history of migration or forced migration, including refugees' and migrants' motives for leaving their country of origin, their living conditions upon arrival in Germany, as well as information in edited form on any relatives in the country of origin and the desire to return to the country of origin. For more information about this dataset and a list of the variables it contains, see: [BIOIMMIG Documentation](#).

In the following, we will use this record and other information from the SOEP to create a status variable that you can use to distinguish whether or not people with a migration background also have a background of forced migration, that is, whether migrants are also refugees.

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

These are used to store commands, log files, datasets, and temporary datasets. Open an empty do-file and define your paths with globals:

```

1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:/material/exercises"
5 global MY_IN_PATH "//hume/rdc-prod/distribution/soep-core/soep.v37/eu/
↪Stata/"
6 global MY_DO_FILES "$AVZ/do/"
7 global MY_LOG_OUT "$AVZ/log/"
8 global MY_OUT_DATA "$AVZ/output/"
9 global MY_OUT_TEMP "$AVZ/temp/"

```

The global “AVZ” defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to the data you ordered.

Task 1: Preperation of the Data

Use the *Tracking Data* PPATHL as a master file and merge the appropriate migration dataset. Open the record or browse the documentation and search for a variable describing the immigration status. The biimgrp variable from the BIOIMMIG data set is the appropriate variable.

a) Which variable contains information about the status of each person when they immigrated to Germany?

Familiarize yourself with your variable and check the coding and case numbers.

```
1 /*
```

Familiarize yourself with your variable and check the coding and case numbers.

```
. tab biimgrp, m //Characteristics of the variable are examined.
```

BI: Immigration Group	Freq.	Percent	Cum.
[-5] Not included in this version of th	5,848	3.14	3.14
[-2] Does not apply	113,969	61.23	64.37
[-1] No Answer	1,373	0.74	65.11
[1] East German	3,687	1.98	67.09
[2] Person Of German Descent From Easte	28,029	15.06	82.15
[3] German Who Lived Abroad	1,195	0.64	82.79
[4] Citizen Of EU Country (up to 2009 E	6,935	3.73	86.52
[5] Asylum seeker, refugee	9,419	5.06	91.58
[6] Other Foreigner	15,681	8.42	100.00
Total	186,136	100.00	

b) On the basis of this variable, generate the variable “escape”, which only distinguishes among three groups:

- 0 = Cases where no information is available
- 1 = All persons without escape background
- 2 = Asylum seekers / refugees

After you have familiarized yourself with the variable, recode it to fit your project. Then check the case numbers of your generated variable with the source variable.

```
1 Total | 63,755 100.00
2 */
```

```
. tab biimgrp Escape, m // biimgrp and escape are compared.
```

BI: Immigration Group	RECODE of biimgrp (BI: Immigration Group)			Total
	No Answer	no Escape	Escape	
[-5] Not included in	5,848	0	0	5,848
[-2] Does not apply	113,969	0	0	113,969
[-1] No Answer	1,373	0	0	1,373
[1] East German	0	3,687	0	3,687
[2] Person Of German	0	28,029	0	28,029
[3] German Who Lived	0	1,195	0	1,195
[4] Citizen Of EU Cou	0	6,935	0	6,935
[5] Asylum seeker, re	0	0	9,419	9,419
[6] Other Foreigner	0	15,681	0	15,681
Total	121,190	55,527	9,419	186,136

c) It may be that initially there is no information on the immigration status but this will change one year later.

Limit the dataset to the last observation available on the respective person, since this gives you the most comprehensive information.

```
*/
```

f) Save the generated data temporarily on your personal drive.

```
f)      Save the generated data record on your personal drive.
↳temporarily
      */
      save $MY_OUT_TEMP\biimgrp.dta, replace
```

Task 2: Add basic variables from PPATH and weights

Attention: Please note that since version 34 (v34), PPFAD can be found in the subdirectory “Raw” of the data distribution file. The following exercises are done with version 33.1 (v33.1), where the tracking file was named PPFAD.

a) Load the following information from PPATH:

- Permanent Individual Identifier "**persnr**"
- Household Identifier "**hhnr**" and the current household number "**bghhnr**"
- The net variable with information about the interview type "**bgnetto**"
- The gender of the person "**sex**"
- The year of birth "**gebjahr**"
- Variables on the migration background "**migback**", "**germborn**", "**corigin**", "**immiyear**"
- Information about the survey status: "**psample**"

If you want to familiarize yourself with the PPATH dataset, see the section *Working with Tracking Data (PPATHL)*.

```
/*
a)      Use the following information from PPFAD:
- Unchanging Person ID „persnr“
- Household number "hhnr" and the current household number "bghhnr".
- the net variable with information about the interview type "bgnetto".
- the gender of the person "sex"
- the year of birth "semester"
- Variables on the migration background "migback", "germborn" "corigin"
↳"immiyear"
- Information about the survey status: "bgnetto" and "psample".
*/
use persnr hhnr bghhnr bgnetto psample sex gebjahr germborn corigin_
↳immiyear migback using $MY_IN_PATH\ppfad.dta, clear
```

b) Merge the previously generated data using the individual identifier.

If you don't understand how to create your own cross-sectional dataset, see the chapter *Generating a Cross-Sectional Dataset*.

```
1      /*  
2      b)      Merge the previously generated data set using the person_  
↳number.  
3      */  
4  
5      merge 1:1 persnr using $MY_OUT_TEMP\biimgrp.dta, nogen
```

c) Add the corresponding individual extrapolation factors to the data.

```
1      c)      Add the corresponding data using the individual identifier.  
2      */  
3  
4      merge 1:1 persnr using $MY_IN_PATH\phrf.dta, keepus(bgphrf) nogen
```

d) Only keep respondents who completed a youth or individual questionnaire in 2016.

For example, to exclude children who have not provided immigration status information, use the net code from PPATH. Only keep individuals who completed an individual or youth interview.

```
1      /*  
2      d)      Only keep respondents who completed a youth or individual_  
↳questionnaire in 2016.  
3      */  
4  
5      tab bgnetto, m //Variable values are displayed  
6  
7      keep if inrange(bgnetto, 10, 19) // People who have a code between 10_  
↳and 19 will be kept.
```



```
. tab bgnetto, m //Variable values are displayed
```

Survey Status 2016	Freq.	Percent	Cum.
[-2] Does not apply	68,743	54.49	54.49
[10] Interviewee With Successful Interv	5,562	4.41	58.90
[12] Individual Questionnaire And Perso	8,570	6.79	65.70
[14] Individual Questionnaire And Other	30	0.02	65.72
[15] Individual Questionnaire And Exper	14,903	11.81	77.53
[17] Youth Biography First Time Surveye	535	0.42	77.96
[19] Individual Questionnaire Without H	113	0.09	78.05
[20] Children in Successfully Interviewe	10,682	8.47	86.51
[21] Children With Mother-Child Questio	349	0.28	86.79
[22] Children With Mother-Child Questio	393	0.31	87.10
[23] Children With Mother-Child Questio	685	0.54	87.64
[24] Children age 7-8, with parental qu	746	0.59	88.24
[25] Children age 9-10, with parental q	538	0.43	88.66
[26] Students Age 11-12	559	0.44	89.11
[28] Youth questionnaire, Age 13-14	526	0.42	89.52
[29] Youth from refugee sample, age 16-	222	0.18	89.70
[30] Persons In Successfully Interviewe	12,361	9.80	99.50
[32] Successfully Completed Biography Q	1	0.00	99.50
[34] Successful Tests and Experiments	13	0.01	99.51
[90] Individual Dropouts PBR_EXIT	306	0.24	99.75
[91] Moved abroad	133	0.11	99.86
[99] Has Died	181	0.14	100.00
Total	126,151	100.00	

Task 3: Generate a status variable with the following categories:.

- No migration background
- Migrant, 2nd generation
- Migrant, no information
- Migrant, not refugee
- Migrant, refugee

To generate this status variable, check the contents of the existing migration variables from PPATH (migback germborn).

```
1      /*
2      Generate a status variable with the following categories:
3      */
4
5      tab migback
```

```
. tab migback
```

Migration background	Freq.	Percent	Cum.
[1] no migration background	18,099	60.91	60.91
[2] direct migration background	9,456	31.82	92.74
[3] indirect migration background	2,158	7.26	100.00
Total	29,713	100.00	

```
tab germborn
```

Born in Germany	Freq.	Percent	Cum.
[1] born in Germany or immigr.<1950	20,257	68.18	68.18
[2] not born in Germany	9,456	31.82	100.00
Total	29,713	100.00	

Use the migration variables from PPATH (migback, germborn) and link this information with your previously generated refugee variable to build the described status variable from Task 3.

```

1      gen Status = 0 // All persons will first receive the missing code for
↳ "no info".
2      replace Status = 1 if migback == 1 & germborn == 1 // "no migback"
3      replace Status = 2 if migback == 3 // "2nd generation"
↳ (2nd generation migrants born by definition in Germany, therefore "& germborn == 1"
↳ here unnecessary
4      replace Status = 3 if germborn == 2 & Escape == 0 // "Immigrants
↳ without information"
5      replace Status = 4 if germborn == 2 & Escape == 1 // "Immigrants, no
↳ escape"
6      replace Status = 5 if germborn == 2 & Escape == 2 // "Immigrant, escape"
7
8      label def Statuslbl 0"no info" 1"no migback" 2"2. Generation" 3
↳ "Immigrants without information" 4"Immigrants, no escape" 5"Immigrant, escape"
9      label val Status Statuslbl // Values of the status variable receive label

```

Task 4: Content analysis:

a) How many refugees (foreign-born with refugee/asylum status) are now in your file?

Look at your status variable previously generated in task 3 to answer the question.

```

1      *** Exercise 4
↳ *****
2
3      /*
4      a)      How many refugees (foreign-born with refugee/asylum status)

```

(continues on next page)

(continued from previous page)

```

are now in your file?
*/
tab Status, m //Display Generated Status Variable

```

```
. tab Status, m //Display Generated Status Variable
```

Status	Freq.	Percent	Cum.
no info	18	0.06	0.06
no migback	18,099	60.91	60.97
2. Generation	2,158	7.26	68.24
Immigrants without information	826	2.78	71.02
Immigrants, no escape	4,098	13.79	84.81
Immigrant, escape	4,514	15.19	100.00
Total	29,713	100.00	

All 4,514 respondents who received the value 5 for the generated status variable have a direct migration background (migback==2), were not born in Germany (germborn==2), and fled their country of origin (flight==2 and biimgrp==5).

b) How many are there if you take the individual extrapolation factors into account? Interpret the results.

Look at the status variable generated in task 3 to answer the question.

```

/*
b) How many are there if you take the individual extrapolation
factors into account? Interpret the results.
*/
tab Status [aw=bgphrf], m //Display generated status variable weighted
with analytic weights

```

```
. tab Status [aw=bgphrf], m //Display generated status variable weighted with analytic weights
```

Status	Freq.	Percent	Cum.
no info	17.1538018	0.06	0.06
no migback	22,182.267	75.23	75.29
2. Generation	2,161.9832	7.33	82.63
Immigrants without information	622.927131	2.11	84.74
Immigrants, no escape	3,824.1688	12.97	97.71
Immigrant, escape	675.499938	2.29	100.00
Total	29,484	100.00	

After weighting, there are approximately 675 refugees in the dataset. The weighting thus corrected the number of refugees downwards.

c) How many persons are represented in the sample, taking the extrapolation factors into account?

To use frequency weights in STATA, integer weights are required. Create an integer frequency weight from the weighting factor provided so that you can make representative statements. Then take a look at the new results.

```

1      /*
2      c)      How many persons are represented when the sample taking the
↳extrapolation factors into account?
3      */
4
5      gen fweight = round(bgphrf) //Frequency weights for stata require
↳integer weight
6      tab Status [fw=fweight], m //Display generated status variable weighted
↳with frequency weights

```

```
. tab Status [fw=fweight], m //Display generated status variable weighted with frequency weights
```

Status	Freq.	Percent	Cum.
no info	40,818	0.06	0.06
no migback	52,781,778	75.23	75.29
2. Generation	5,144,356	7.33	82.63
Immigrants without information	1,482,236	2.11	84.74
Immigrants, no escape	9,099,488	12.97	97.71
Immigrant, escape	1,607,336	2.29	100.00
Total	70,156,012	100.00	

Around 1,600,000 people are represented.

d) What is the proportion of people over 40 years of age among the refugees?

Since the data in this exercise come from the wave “bg”, we are currently in the survey year 2016; if you need a description of the wave designations, please refer to the chapter Label. To generate a suitable age variable, you can use the year of birth (year of birth). If we look at the survey year 2016, all persons born in 1976 or earlier were over 40 years old. Generate a suitable age variable and look at the proportion of refugees over 40 years of age in weighted form:

```

1      /*
2      d)      What is the proportion of people over 40 years of age among
↳the refugees?
3      */
4
5      gen ue_40 = 0
6      replace ue_40 = 1 if gebjahr <= 1976 // Persons receive proficiency 1 if
↳they were born before 1975.
7
8      tab Status ue_40 [aw=bgphrf], m row nofreq

```

```
. tab Status ue_40 [aw=bgphrf], m row nofreq
```

Status	ue_40		Total
	0	1	
no info	57.54	42.46	100.00
no migback	28.83	71.17	100.00
2. Generation	59.22	40.78	100.00
Immigrants without in	8.91	91.09	100.00
Immigrants, no escape	37.10	62.90	100.00
Immigrant, escape	53.04	46.96	100.00
Total	32.28	67.72	100.00





The proportion of refugees over 40 years of age is about 47%.

Last change: Jun 30, 2023

6.8 Fixed Effects Estimation

Let's say you want to find out whether certain variables relevant to the labor market, such as work experience or time in education, influence a person's hourly wage. Other variables such as gender or marital status should also be taken into account. You decide to use the SOEP data to set up a fixed effects estimation model.

Create a path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

These are used to store your script, log files, datasets, and temporary datasets. Open an empty do-file and define your paths with globals:

```

1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:\material\exercises"
5 global MY_IN_PATH "\\hume\rdc-prod\distribution\soep-long\soep.v33.1\"

```

(continues on next page)

(continued from previous page)

```

↪ stata_en\"
6      global MY_DO_FILES "$AVZ\do\"
7      global MY_LOG_OUT "$AVZ\log\"
8      global MY_OUT_DATA "$AVZ\output\"
9      global MY_OUT_TEMP "$AVZ\temp\"

```

The global “AVZ” defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to your data.

a) Generate your own SOEPwage.dta dataset. The dataset should contain information on gross monthly wage, marital status, and other personal characteristics.

To perform your analysis, you need different SOEP variables. The SOEP offers various options for a variable search:

- Search the questionnaires for useful variables. (For more information, see the section [Variable Search with Questionnaires](#))
- Find a suitable variable in the topic list at [paneldata.org](#) (for more information, see the section [Topic Search with paneldata.org](#))
- Search for a suitable variable using a search term in [paneldata.org](#) (for more information, see the section [Variable Search with paneldata.org](#))
- Use the documentation provided for the generated variables (for more information, see the section [Documentation on Generated Data](#))

Use the various important variables of the ppfadl.dta dataset as your start file. Your source file should contain the following variables:

- Individual identifier "**pid**"
- Survey year "**syear**"
- Birth Year "**gebjahr**"
- The net variable with information on the interview type "**netto**"
- The weighting variable "**phrf**"
- The gender of the person "**sex**"
- Sample membership "**pop**"

```

1      use pid syear sex gebjahr netto pop phrf using "${MY_IN_PATH}/ppfadl.dta
↪", clear

```

Attention: Please note that since version 34 (v34), PPFADL has been renamed PPATHL. The following exercises are done with version 33.1 (v33.1), where the tracking file was named PPFADL.

Apply the necessary content variables to your starting dataset. You need the following variables for your analysis:

- Employment status **plb0022_h**
- Current gross income in euros "**pglabgro**"
- Actual weekly working hours "**pgtatzeit**"
- Full-time work experience "**pgexpft**"

- Years of education or training "**pgbilzeit**"
- Marital status in survey year "**pgfamstd**"

```

1      merge 1:1 pid syear using "${MY_IN_PATH}/pl.dta", keepus(plb0022)
↳ keep(master match) nogen
2      merge 1:1 pid syear using "${MY_IN_PATH}/pgen.dta", keepus(pglabgro
↳ pgtatzeit pgexpft pgbilzeit pgfamstd) keep(master match) nogen

```

Only keep people who have completed an interview and who live in a private household.

```

1      * Only select people with completed interviews
2      keep if inrange(netto, 10, 19)
3
4      * Only private households
5      keep if pop==1 | pop==2

```

Since you are only interested in the period from 2012 to 2016, remove all survey information that does not fall within this period. To finish, save your dataset.

```

1      * Period from 2012 to 2016
2      keep if syear>=2012 & syear<=2016

```

Exercise 1: Prepare your dataset

a) Load your created SOEPWage.dta dataset. It contains information on gross monthly wage, marital status, and other personal characteristics.

```

1      *** Exercise 1: Prepare your dataset
2      * a) Load data set
3      use "${MY_OUT_DATA}/SOEPWage.dta", clear

```

b) Recode all missing values in systemmissings (.)

```

1      * b) Recode Missings
2      mvdecode _all, mv(-8/-1 = .)

```

For more information about the missing codes for SOEP data, see the chapter *Missing Conventions*

c) Generate the variables “hourly wage” (gross monthly wage/4.33*working time) for persons who have earned at least 1 euro and have worked at least one hour, “Married vs. Unmarried” and age.

```

1      * c) Generate Variables
2      gen wage = pglabgro/(4.33*pgtatzeit) if pglabgro>=1 & pgtatzeit>=1
3
4      gen married = 1 if pgfamstd==1 | pgfamstd==6 | pgfamstd==7 | pgfamstd==8
5      replace married = 0 if inrange(pgfamstd, 2, 5)
6
7      gen age = syear - gebjahr

```

d) Adjust the variable “hourly wage” from outlier values by setting values smaller than the first percentile to the same value. Set values greater than 3 times the 99th percentile to 3*99th percentile. Then generate the variable lwage = log(wage).

```
1      * d) Adjust wage variable
2      sum wage, detail
3      replace wage = 1/3*r(p1) if wage<1/3*r(p1)
4      replace wage = 3*r(p99) if wage>3*r(p99) & wage<.
5
6      gen lwage = log(wage)
7      label variable lwage "Log hourly wage"
8
9      save "${MY_OUT_DATA}/SOEPWage_temp.dta", replace
```

Exercise 2: Descriptive statistics

a) Define the dataset as a panel dataset.

```
1      *** Exercise 2: Descriptive statistics
2      * a)
3      xtset pid syear // Declaring data as panel data
```

b) What percentage of people participated in all five waves (xtdescribe)

```
1      * b)
2      xtdescribe, patterns(16) // -> unbalanced panel
```



```
. * b)
. xtdescribe, patterns(16) // -> unbalanciertes Panel

      pid: 602, 901, ..., 38647702      n =      42808
      sye: 2012, 2013, ..., 2016      T =          5
      Delta(syear) = 1 unit
      Span(syear)  = 5 periods
      (pid*syear uniquely identifies each observation)
```

Distribution of T_i:

	min	5%	25%	50%	75%	95%	max
	1	1	2	4	5	5	5

Freq.	Percent	Cum.	Pattern
17069	39.87	39.87	11111
3941	9.21	49.081
3044	7.11	56.19	1....
2810	6.56	62.75	.1111
2581	6.03	68.78	11...
2040	4.77	73.55	1111.
1895	4.43	77.98	111..
1695	3.96	81.94	...11
1688	3.94	85.88	.1...
925	2.16	88.04	.11..
923	2.16	90.20	...1.
678	1.58	91.78	..111
671	1.57	93.35	.111.
425	0.99	94.34	11.11
402	0.94	95.28	111.1
289	0.68	95.95	1.111
1732	4.05	100.00	(other patterns)
42808	100.00		XXXXX

42808 respondents have contributed information within waves bc (2012) - bg (2016) and about 40% (17069) of the 42808 respondents have provided information for all waves.

c) Describe the variable “Married” with xttab and xttrans. Take a look at some individual wage (pid=30320901, pid=30932501, pid==3101602, pid==3101801) developments with xtline.

```
* c)
* Stability of the relationship status
xttab married
```

```
. xttab married
```

married	Overall		Between		Within
	Freq.	Percent	Freq.	Percent	Percent
0	58906	41.37	19717	46.23	94.69
1	83474	58.63	25014	58.65	95.88
Total	142380	100.00	44731	104.87	95.35

(n = 42652)

You can observe 41.37 percent of person-year observations with “married==no”. Within the period from 2012 to 2016, 19717 people responded at least once that they were not married. During the same period, 25014 persons reported at least once that they were married. Those who were not married for at least one year responded with “married==no” in 94.69% of the observations, whereas those who were married at least once responded in 95.88 percent of the observations with “married==yes”. This indicates very stable response behavior.

1 * Transition probabilities

2 xttrans married, freq

```
. xttrans married, freq
```

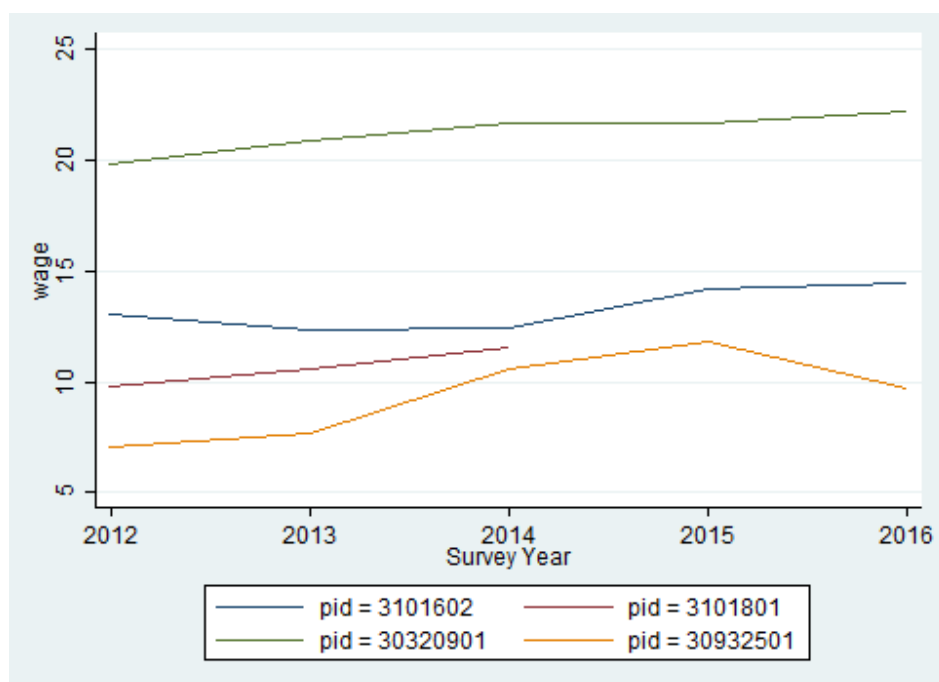
married	married		Total
	0	1	
0	39,112 96.87	1,264 3.13	40,376 100.00
1	881 1.49	58,428 98.51	59,309 100.00
Total	39,993 40.12	59,692 59.88	99,685 100.00

96.87 percent of the person-year observations with “married==no” are still not married in the next period. 98.51 percent of the persons who are married indicate that they will also be married in the following period. This is evidence of stable response behavior.

1 * Individual sequences of "wage"

2 xtline wage if pid==30320901 | pid==30932501 | pid==3101602 |

pid==3101801, overlay



The graphic shows a comparison of the hourly wage for four different respondents.

Exercise 3: Pooled OLS Regression

a) Execute a pooled OLS regression with “log hourly wage” as dependent variable and “married”, “gender”, “work experience” and “training time” as independent variables. Interpret the coefficients for “married”, “gender” and “length of training”. Why are these not causal effects?

```

1      *** Exercise 3: Pooled OLS Regression
2      * a) Pooled OLS
3      reg lwage married sex pgexpft pgbilzeit

```

```
. reg lwage married sex pgexpft pgbilzeit
```

Source	SS	df	MS	Number of obs = 78234		
Model	9531.59732	4	2382.89933	F(4, 78229) = 8027.72		
Residual	23221.0303	78229	.296834042	Prob > F = 0.0000		
				R-squared = 0.2910		
				Adj R-squared = 0.2910		
Total	32752.6276	78233	.418654885	Root MSE = .54482		

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
married	.1443034	.0041241	34.99	0.000	.1362203	.1523865
sex	-.1203015	.0041704	-28.85	0.000	-.1284754	-.1121276
pgexpft	.0143396	.0001791	80.08	0.000	.0139886	.0146906
pgbilzeit	.0988842	.0007078	139.71	0.000	.0974969	.1002714
_cons	1.19645	.0121292	98.64	0.000	1.172677	1.220224

The variables married, sex, and pgbilzeit most likely correlate with other disregarded/unobserved variables that have an effect on the wage. For example, women more often work in occupations with lower wages.

b) Run the regression again with the option “vce(cluster persnr)” to get clustered standard errors. How do the standard errors of the coefficients change?

```
1 * b) Pooled OLS with cluster standard errors
2 reg lwage married sex pgexpft pgbilzeit, vce(cluster pid)
```

```
. reg lwage married sex pgexpft pgbilzeit, vce(cluster pid)
```

```
Linear regression                                Number of obs = 78234
                                                F( 4, 25133) = 2415.06
                                                Prob > F      = 0.0000
                                                R-squared     = 0.2910
                                                Root MSE     = .54482
```

(Std. Err. adjusted for 25134 clusters in pid)

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
married	.1443034	.0066788	21.61	0.000	.1312126	.1573941
sex	-.1203015	.0070382	-17.09	0.000	-.1340967	-.1065063
pgexpft	.0143396	.0003257	44.03	0.000	.0137013	.014978
pgbilzeit	.0988842	.0012169	81.26	0.000	.096499	.1012693
_cons	1.19645	.0211759	56.50	0.000	1.154944	1.237956

The standard errors are getting bigger.

Exercise 4: Fixed Effects

a) Subtract the person-specific mean value from each variable of the model. Use the “egen” function. Ideally you should also use a loop.

```

1      *** Exercise 4: Fixed Effects
2      * a) Subtract person-specific averages
3
4      gen sample = 1
5      foreach var in lwage married sex pgexpft pgbilzeit {
6
7          bysort pid: egen `var'Mean = mean(`var')
8          replace `var'Mean = . if `var'==.
9          gen `var'Demeaned = `var' - `var'Mean
10         replace sample = 0 if `var'==.
11     }
12     bysort pid (sample): replace sample = sample[1]

```

b) Estimate the fixed effects model with the previously generated variables. Why isn't a coefficient estimated for “gender”? How do the coefficients change compared to the pooled OLS estimate? Is the effect of “married” now causally interpretable?

```

1      reg lwageDemeaned marriedDemeaned sexDemeaned pgexpftDemeaned_
      ↪pgbilzeitDemeaned, vce(cluster pid) nocons

```

```

. * b) Fixed Effects Modell
. reg lwageDemeaned marriedDemeaned sexDemeaned pgexpftDemeaned pgbilzeitDemeaned, vce(cluster pid) nocons
note: sexDemeaned omitted because of collinearity

```

Linear regression		Number of obs = 78234				
		F(3, 25133) = 645.95				
		Prob > F = 0.0000				
		R-squared = 0.0369				
		Root MSE = .24298				
(Std. Err. adjusted for 25134 clusters in pid)						
lwageDemeaned	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
marriedDemeaned	.0197547	.0098598	2.00	0.045	.0004289	.0390805
sexDemeaned	0	(omitted)				
pgexpftDemeaned	.0435521	.0010848	40.15	0.000	.0414259	.0456783
pgbilzeitDemeaned	.0660986	.0042643	15.50	0.000	.0577404	.0744568

No coefficient was estimated for gender because gender was stable over time for all observations. The coefficient of married is now significant at the 5% level!

c) Now estimate the fixed effects model using the command “xtreg lwage married sex pgexpft pgbilzeit, fe”. What do you notice about the coefficients compared to task 4 b)? And with the standard errors?

```

1      * c) xtreg, fe
2      xtreg lwage married pgexpft pgbilzeit, fe vce(cluster pid)

```

```
. xtreg lwage married pgexpft pgbilzeit, fe vce(cluster pid)
```

```
Fixed-effects (within) regression      Number of obs   =    78234
Group variable: pid                   Number of groups =    25134

R-sq:  within  = 0.0394                Obs per group: min =     1
      between  = 0.2228                      avg   =    3.1
      overall  = 0.1957                      max   =     5

                                F(3,25133)    =    643.92
corr(u_i, Xb)  = -0.4631                Prob > F      =    0.0000
```

(Std. Err. adjusted for 25134 clusters in pid)

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
married	.0224308	.0108103	2.07	0.038	.001242	.0436196
pgexpft	.0443073	.001109	39.95	0.000	.0421335	.0464811
pgbilzeit	.0765046	.0049426	15.48	0.000	.0668168	.0861924
_cons	.9253963	.0644086	14.37	0.000	.7991517	1.051641
sigma_u	.62923787					
sigma_e	.29340975					
rho	.8214025	(fraction of variance due to u_i)				

The coefficients are not identical to 4 b) and the standard errors become larger because model b) does not take into account the estimation of mean values in the standard errors.

d) Now add dummy variables for the years (i.year). What happens to the effect of “labor market experience”?

```
1      * d) xtreg with dummy
2      xtreg lwage married pgexpft pgbilzeit i.year, fe vce(cluster pid)
```

```
. * d) xtreg mit Jahres-Dummies
```

```
. xtreg lwage married pgexpft pgbilzeit i.syear, fe vce(cluster pid)
```

```
Fixed-effects (within) regression      Number of obs   =    78234
Group variable: pid                   Number of groups  =    25134

R-sq:  within  = 0.0599                Obs per group: min =     1
      between  = 0.0065                      avg   =    3.1
      overall  = 0.0152                      max   =     5

                                         F(7,25133)       =    344.67
corr(u_i, Xb)  = -0.2578                Prob > F         =    0.0000
```

(Std. Err. adjusted for 25134 clusters in pid)

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
married	.021538	.0106165	2.03	0.042	.0007292	.0423469
pgexpft	-.0124634	.0024322	-5.12	0.000	-.0172306	-.0076961
pgbilzeit	.0606128	.0048847	12.41	0.000	.0510384	.0701872
syear						
2013	.0552667	.0036671	15.07	0.000	.0480789	.0624545
2014	.0980733	.0047304	20.73	0.000	.0888014	.1073451
2015	.1545752	.0063392	24.38	0.000	.14215	.1670005
2016	.2026541	.0080508	25.17	0.000	.1868742	.2184341
_cons	1.882517	.0712664	26.42	0.000	1.742831	2.022203
sigma_u	.66907886					
sigma_e	.29027579					
rho	.8415946	(fraction of variance due to u_i)				

Effects on the variables remain significant. The model could possibly be specified on a case-by-case basis. The Mincer equation is based on (potential) labor market experience squared.

e) Now you can also square labor market experience into the model. To what extent does the effect of labor market experience change compared to task 5d)?

```
* e) expft squared
xtreg lwage married c.pgexpft##c.pgexpft pgbilzeit i.syear, fe_
vce(cluster pid)
```

```
. * e) expft auch als Quadrat
. xtreg lwage married c.pgexpft##c.pgexpft pgbilzeit i.syear, fe vce(cluster pid)
```

```
Fixed-effects (within) regression      Number of obs   =    78234
Group variable: pid                   Number of groups =    25134
```

```
R-sq:  within = 0.0648                Obs per group: min =      1
        between = 0.0776                avg =      3.1
        overall = 0.0811                max =      5
```

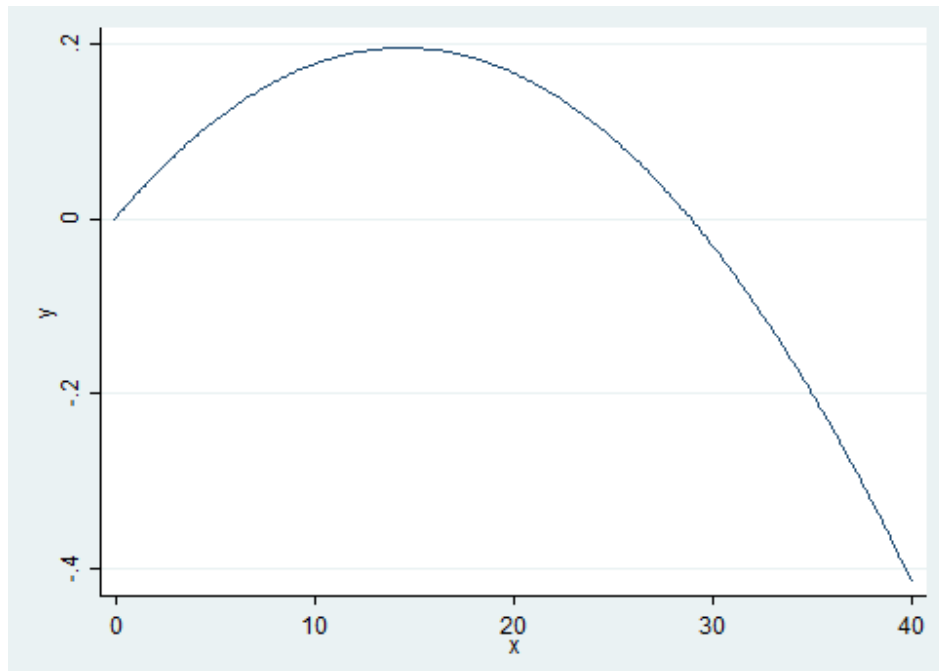
```
corr(u_i, Xb) = -0.1012                F(8,25133)       =    321.03
                                                Prob > F         =    0.0000
```

(Std. Err. adjusted for 25134 clusters in pid)

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
married	.0117953	.0106245	1.11	0.267	-.0090293	.0326199
pgexpft	.027049	.0035366	7.65	0.000	.0201171	.0339809
c.pgexpft#c.pgexpft	-.0009356	.0000582	-16.07	0.000	-.0010497	-.0008215
pgbilzeit	.0564758	.004831	11.69	0.000	.0470068	.0659449
syear						
2013	.0543771	.0036633	14.84	0.000	.0471967	.0615575
2014	.0971777	.0047248	20.57	0.000	.0879167	.1064386
2015	.1519717	.0063321	24.00	0.000	.1395605	.1643829
2016	.1980514	.0080426	24.63	0.000	.1822874	.2138155
_cons	1.692927	.0723071	23.41	0.000	1.551201	1.834653
sigma_u	.62325551					
sigma_e	.28951511					
rho	.82251756	(fraction of variance due to u_i)				

The coefficients of pgexpft and pgexpft^2 remain significant, whereas the coefficient for married is no longer significant.

```
graph twoway (func y = _b[pgexpft]*x + _b[c.pgexpft#c.pgexpft]*x*x,
↪ range(0 40))
```

The graph shows that the effects of the labor market experience decrease after approximately 15 years of professional experience.

f) Now estimate the model from task 5e) with longitudinal section weights. Why is the number of cases now significantly smaller? Why could the coefficient of “pgbilzeit” have changed?

Tip: Create your own longitudinal person weights, e.g., longitudinal person weight from wave A to wave D. Take the starting wave cross-sectional weight (aphrf) and multiply through by each following wave staying factor, as in the following example: `gen adphrf=aphrf*bpbleib*cpbleib*dpbleib`

Since you are looking at the period 2012-2016, you must create a suitable longitudinal weight. To do this, use the phrf dataset from the RAW subdirectory. Apply the required variables to your analysis dataset and generate your period-related longitudinal section weight. To understand the structure of the data distribution file and the location of the different datasets, visit the section [Data Distribution File](#). For more information about the weighting datasets and other survey datasets, see the section [Survey Data](#).

```

1      * f) Fixed Effects weighted
2      global MY_IN_PATH2 "\\hume\rdc-prod\complete\soep-core\soep.v33.2\stata_
en\"
3      rename pid persnr
4      merge m:1 persnr using "${MY_IN_PATH2}/phrf.dta", nogen keep(master_
match) keepus(bcphrf bdpbleib bepbleib bfpbleib bgpbleib)
5      gen wlong = bcphrf*bdpbleib*bepbleib*bfpbleib*bgpbleib
6      label variable wlong "Weighting BC-BG"
7      rename persnr pid

```

Now estimate the model from 5e) and use the created weight.

```

1      xtreg lwage married c.pgexpft##c.pgexpft pgbilzeit i.syear [pw=wlone],
fe vce(cluster pid)

```

```
. xtreg lwage married c.pgexpft#c.pgexpft pgbilzeit i.syear [pw=wlong], fe vce(cluster pid)
```

```
Fixed-effects (within) regression      Number of obs   =    48949
Group variable: pid                   Number of groups =    11790

R-sq:  within = 0.0880                Obs per group: min =     1
      between = 0.1275                  avg =     4.2
      overall  = 0.1290                  max =     5

                                         F(8,11789)      =    96.01
corr(u_i, Xb) = -0.3604                Prob > F        =    0.0000
```

(Std. Err. adjusted for 11790 clusters in pid)

lwage	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
married	.0050783	.0180717	0.28	0.779	-.0303453	.0405018
pgexpft	.0237458	.0067916	3.50	0.000	.0104331	.0370584
c.pgexpft#c.pgexpft	-.0008416	.0000986	-8.54	0.000	-.0010348	-.0006484
pgbilzeit	.1392754	.0176388	7.90	0.000	.1047005	.1738503
syear						
2013	.0471116	.0076671	6.14	0.000	.0320828	.0621404
2014	.0962616	.0098515	9.77	0.000	.0769511	.1155721
2015	.1490648	.013773	10.82	0.000	.1220674	.1760623
2016	.1960915	.0171793	11.41	0.000	.1624172	.2297658
_cons	.6993781	.2279552	3.07	0.002	.2525483	1.146208
sigma_u	.63332729					
sigma_e	.29092777					
rho	.8257534	(fraction of variance due to u_i)				

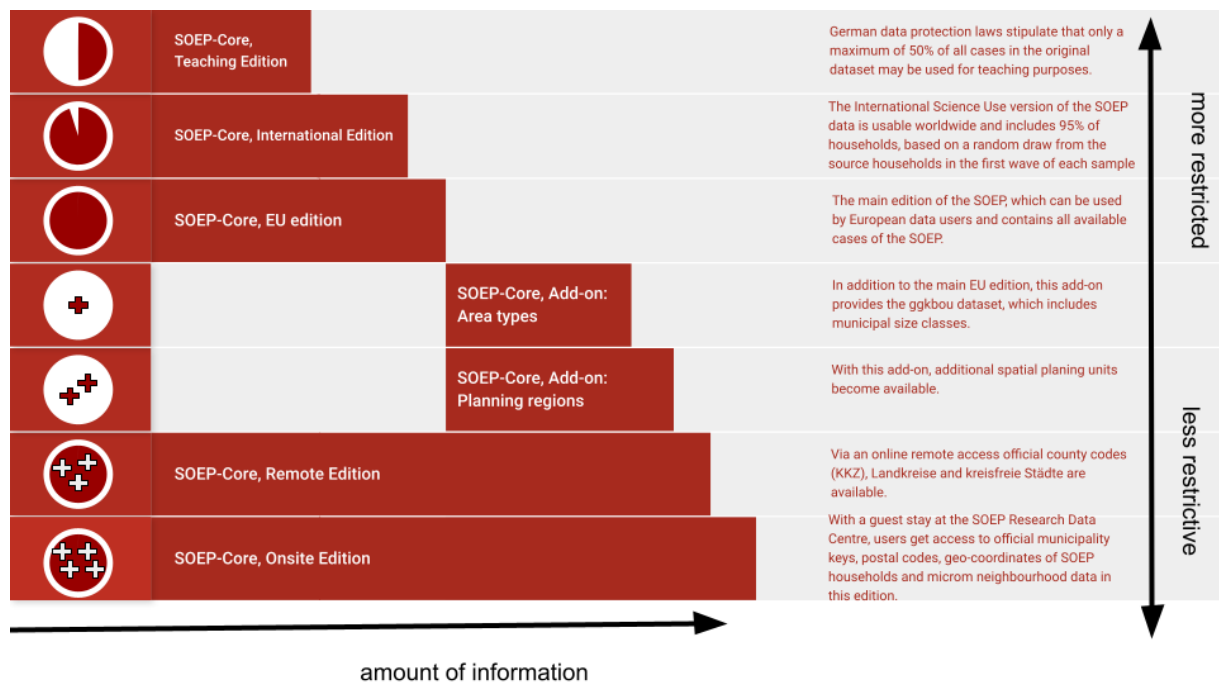
The number of observations is now much smaller. The effect of pgbilzeit is greater than before. Pgbilzeit has a lower effect in the wlong==0 group, where the return is different for each additional educational year. People in the wlong===0 group may not get the returns on additional education they expected on the local labor market and may therefore move -> higher dropout probability.

Last change: Mar 28, 2023

In order to gain the best possible insight into how to work with the various regional data on the SOEP, we recommend the following exercises:

6.9 Working with SOEP Regional Data

SOEP offers diverse possibilities for regional and spatial analysis. With the anonymized regional information on SOEP respondents' (households' and individuals') place of residence, it is possible to link numerous regional indicators on the levels of the federal states (Bundesländer), spatial planning regions, districts, and postal codes with the data on the SOEP households. However, specific security provisions must be made due to the sensitivity of the data under data protection law. Accordingly, data users are not allowed to give any information in their analyses that could indicate, for instance, the city or district in which respondents reside. The data nevertheless provide valuable background information for regional analysis.



For more information and to access the data, see [Regional Data](#)

Assume that for your research project, you want to measure current (2016) urban-rural differences in the population. You are particularly interested in the differences in interest in politics and the different satisfaction variables provided by the SOEP. You also want to take into account demographic differences in gender and age. To be able to evaluate the potential of the data for your project, you first need an overview. For regional analysis, for example, the municipal size classes from the regional data are suitable.

Create an exercise path with four subfolders:

do	07.05.2018 16:02	Dateiordner
log	12.04.2018 10:06	Dateiordner
output	21.06.2018 13:14	Dateiordner
temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do

- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

These are used to store your script, log files, datasets, and temporary datasets. Open an empty do-file and define your paths with globals:

```
1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:\material\exercises"
5 global MY_IN_PATH "\\hume\rdc-prod\complete\soep-core\soep.v33.2\stata_
  en\"
6
7 global region "\\hume\soep-region\DATA\soep33_de\"
8 global MY_DO_FILES "$AVZ\do\"
9 global MY_LOG_OUT "$AVZ\log\"
10 global MY_OUT_DATA "$AVZ\output\"
    global MY_OUT_TEMP "$AVZ\temp\"
```

The global “AVZ” defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to the data you ordered.

a) Prepare a dataset for cross-sectional analysis covering the survey year 2016 (wave bg).

To perform your analysis, you need different SOEP variables. The SOEP offers various options for a variable search:

- Search the questionnaires for useful variables (for more information, see the section [Variable Search with Questionnaires](#))
- Find a suitable variable in the topic list on paneldata.org (for more information, see the section [Topic Search with paneldata.org](#))
- Search for a suitable variable using a search term in paneldata.org (for more information, see the section [Variable Search with paneldata.org](#))
- Use the documentation provided by the generated variables (for more information, see the section [Documentation on Generated Data](#))

Your source file should contain the following variables:

- Permanent Individual ID "**persnr**"
- Original Household Number "**hhnr**"
- Current Wave Household Number "**bghhnr**"
- The Sex of the Person "**sex**"
- Year of Birth "**gebjahr**"
- Survey Status 2016 "**bgnetto**"
- Sample Membership 2016 "**bgpop**"
- Weighting Factor 2016 "**bgphrf**"
- Satisfaction With Health "**bgp0101**"
- Satisfaction With Sleep "**bgp0102**"
- Satisfaction With Work "**bgp0103**"

- Satisfaction With Housework "bgp0104"
- Satisfaction With Household Income "bgp0105"
- Satisfaction With Personal Income "bgp0106"
- Satisfaction With Dwelling "bgp0107"
- Satisfaction With Amount Of Leisure Time "bgp0108"
- Satisfaction With Child Care "bgp0109"
- Satisfaction With Family Life "bgp0110"
- Satisfaction With Social Life "bgp0111"
- Satisfaction with Democracy "bgp0112"
- Political Interest "bgp143"
- Current Sample Region "bgsampreg"
- Federal State "bgbula"
- Spatial Category by BBSR "bgregtyp"
- Municipal Class Sizes "ggk"

Use the key variables from the ppath.dta dataset as your starting file.

```
1      use hhnr persnr bghhnr sex gebjahr bgnetto bgpop using ${MY_IN_PATH}\
↪ppfad.dta, clear
```

Attention: Please note that since version 34 (v34), PPFAD can be found in the subdirectory "Raw" of the data distribution file. The following exercises are done with version 33.1 (v33.1), where the tracking file was named PPFAD.

Keep people who completed a questionnaire in 2016 and lived in a private household.

```
1      * Keep people who completed a questionnaire in 2016 and live in a
↪private household
2      keep if bghhnr>0 & inrange(bgnetto, 10, 29) & inlist(bgpop, 1, 2)
3      keep hhnr persnr bghhnr sex gebjahr bgnetto bgpop
4      merge 1:1 persnr using ${MY_IN_PATH}\phrf.dta, keep(match master)
↪keepusing (bgphrf) nogenerate
5      tempfile ppfad
6      save `ppfad'
```

Prepare the different datasets bgp, bghbrutto, regionl

```
1      * Prepare dataset bgp
2      use ${MY_IN_PATH}\bgp.dta, replace
3      keep persnr hhnr bghhnr bgp01* bgp143
4      tempfile bgp
5      save `bgp'
6
7      * Prepare dataset bghbrutto
8      use ${MY_IN_PATH}\bghbrutto.dta, replace
9      keep hhnr bghhnr bgsampreg bgbula bgregtyp
```

(continues on next page)

(continued from previous page)

```

10      tempfile bghbrutto
11      save `bghbrutto'
12
13      * Prepare dataset regionl
14      use ${region}\regionl_v33.dta, replace
15      keep if syear==2016
16      keep syear hhnr hhnrakt ggk
17      rename hhnrakt bghhnr
18      tempfile regionl
19      save `regionl'

```

Merge all datasets.

```

1      * Merge all datasets
2      use `ppfad'
3      merge 1:1 persnr using `bgp', keep(match master) nogenerate
4      merge m:1 bghhnr hhnr using `regionl', keep(match master) nogenerate
5      merge m:1 bghhnr hhnr using `bghbrutto', keep(match master) nogenerate

```

Recode negative values as missings.

```

1      * Recode negative values into missings
2      mvdecode sex gebjahr bgp01* bgp143,mv(-5/-1)

```

Categorize the municipal class sizes from the SOEP regional dataset.

```

1      * Categorize community class size
2      gen ggk_cat=.
3      replace ggk_cat=-1 if ggk== -1
4      replace ggk_cat=1 if ggk==1 | ggk==2
5      replace ggk_cat=2 if ggk==3
6      replace ggk_cat=3 if ggk==4 | ggk==5
7      replace ggk_cat=4 if ggk>5 & ggk<=7
8
9      lab var ggk_cat "Community Size categorised"
10     lab def ggk_cat -1 "No information" 1 "<=5000" 2 "5001 - 20000" 3 "20001_
↪ - 100000" ///
11     4 ">100000"
12     lab val ggk_cat ggk_cat

```

Generate an age variable.

```

1      * Generate age variable
2      gen alter= 2016-gebjahr if gebjahr > 0
3      gen alter_cat=1 if alter<=20
4      replace alter_cat=2 if alter>20 & alter<=30
5      replace alter_cat=3 if alter>30 & alter<=65
6      replace alter_cat=4 if alter>65 & alter<=120
7
8      lab var alter "age"
9      lab var alter_cat "age categorized"
10     lab def alter_cat 1 "<=20" 2 "21-30" 3 "31-65" 4 ">65"
11     lab val alter_cat alter_cat

```

Categorize a federal states variable.

```

1      * Categorize federal states
2      gen bgbula_cat=.
3      * Schleswig-Holstein + Hamburg
4      replace bgbula_cat=1 if bgbula==1 | bgbula==2
5      * Lower Saxony + Bremen
6      replace bgbula_cat=2 if bgbula==3 | bgbula==4
7      * Mecklenburg Western Pomerania + Brandenburg
8      replace bgbula_cat=3 if bgbula==13 | bgbula==12
9      * Saarland + Rhineland Palatinate
10     replace bgbula_cat=4 if bgbula==7 | bgbula==10
11     * Northrhine-Westphalia
12     replace bgbula_cat=5 if bgbula==5
13     * Hesse
14     replace bgbula_cat=6 if bgbula==6
15     * Baden-Württemberg
16     replace bgbula_cat=7 if bgbula==8
17     * Bavaria
18     replace bgbula_cat=8 if bgbula==9
19     * Berlin
20     replace bgbula_cat=9 if bgbula==11
21     * Saxony
22     replace bgbula_cat=10 if bgbula==14
23     * Saxony-Anhalt
24     replace bgbula_cat=11 if bgbula==15
25     * Thuringia
26     replace bgbula_cat=12 if bgbula==16
27
28     lab var bgbula_cat "Federal states categorized"
29     lab def bgbula_cat 1 "Schleswig-Holstein/Hamburg" 2 "Lower Saxony/Bremen"
30     ↪ 3 "Mecklenburg Western Pomerania/Brandenburg" ///
31     4 "Saarland/Rhineland Palatinate" 5 "Northrhine-Westphalia" 6 "Hesse" ///
32     ↪ 7 "Baden-Wuerttemberg" 8 "Bavaria" 9 "Berlin" 10 "Saxony" 11 "Saxony-
33     ↪ Anhalt" 12 "Thuringia"
34     lab val bgbula_cat bgbula_cat
35     drop bgbula
36     rename bgbula_cat bgbula

```

Put the variables in your preferred order and save your dataset.

```

1      * Order demography and identifiers first
2      order persnr hhnr bghhnr syear sex gebjahr alter alter_cat bgsampreg_
3      ↪ bgbula ggk ///
4      ggk_cat bgregtyp
5
6      save ${MY_OUT_DATA}\zeit_online.dta, replace

```

b) You want to get an initial overview of regional differences in satisfaction with various aspects of life. Use the variable `bgsampreg` and cross-stabilize the variable with all satisfaction variables to identify differences between East and West Germany, display the absolute and relative frequencies.

To save the tables, save them in a log file.

```

*****
capture log close
log using "${MY_LOG_OUT}\satisfaction.log", replace

* Life satisfaction

local varlist bgp0101 bgp0102 bgp0103 bgp0104 bgp0105 bgp0106 bgp0107
↪bgp0108 ///
bgp0109 bgp0110 bgp0111 bgp0112
foreach x of local varlist {
  tab bgsampreg `x' [aw= bgphrf] , row
}

```

Current Sample Region	Satisfaction With Health											Total
	[0] 0 Sat	[1] 1 Sat	[2] 2 Sat	[3] 3 Sat	[4] 4 Sat	[5] 5 Sat	[6] 6 Sat	[7] 7 Sat	[8] 8 Sat	[9] 9 Sat	[10] 10 S	
[1] West Germany	256.82471 1.15	260.7491 1.17	623.17631 2.79	1,180.878 5.28	1,226.948 5.49	2,717.234 12.16	2,324.916 10.40	4,208.661 18.83	5,384.689 24.09	2,623.874 11.74	1,546.069 6.92	22,354.019 100.00
[2] East Germany	67.27909 1.41	65.784226 1.38	175.10943 3.68	332.81232 6.99	283.75315 5.96	686.88971 14.43	548.93801 11.53	900.241273 18.92	999.234017 21.00	454.15063 9.54	244.78919 5.14	4,758.981 100.00
Total	324.1038 1.20	326.533325 1.20	798.28574 2.94	1,513.69 5.58	1,510.701 5.57	3,404.124 12.56	2,873.854 10.60	5,108.902 18.84	6,383.923 23.55	3,078.025 11.35	1,790.858 6.61	27,113 100.00

Current Sample Region	satisfaction with sleep											Total
	[0] 0 Sat	[1] 1 Sat	[2] 2 Sat	[3] 3 Sat	[4] 4 Sat	[5] 5 Sat	[6] 6 Sat	[7] 7 Sat	[8] 8 Sat	[9] 9 Sat	[10] 10 S	
[1] West Germany	159.40597 0.80	235.97229 1.19	644.89468 3.24	1,096.823 5.52	1,293.988 6.51	2,220.017 11.17	2,201.258 11.07	3,256.5262 16.38	4,299.5147 21.63	2,566.6674 12.91	1,903.7127 9.58	19,878.78 100.00
[2] East Germany	26.18853 0.62	37.661261 0.89	147.66919 3.48	280.15784 6.61	312.65151 7.38	589.20671 13.90	483.1413 11.40	627.41268 14.80	877.71191 20.71	505.79602 11.93	350.62287 8.27	4,238.22 100.00
Total	185.5945 0.77	273.633552 1.13	792.56387 3.29	1,376.981 5.71	1,606.6397 6.66	2,809.224 11.65	2,684.399 11.13	3,883.939 16.10	5,177.227 21.47	3,072.463 12.74	2,254.336 9.35	24,117 100.00

Current Sample Region	Satisfaction With Work											Total
	[0] 0 Sat	[1] 1 Sat	[2] 2 Sat	[3] 3 Sat	[4] 4 Sat	[5] 5 Sat	[6] 6 Sat	[7] 7 Sat	[8] 8 Sat	[9] 9 Sat	[10] 10 S	
[1] West Germany	108.18696 0.86	101.53684 0.81	226.91136 1.81	408.98276 3.26	421.367929 3.36	1,161.145 9.26	1,260.616 10.05	2,377.968 18.95	3,392.8893 27.04	1,994.521 15.90	1,091.584 8.70	12,545.71 100.00
[2] East Germany	27.931559 1.06	21.235589 0.81	38.931778 1.48	84.358325 3.21	121.25058 4.61	286.25775 10.89	240.69159 9.15	545.04361 20.73	730.78802 27.79	333.42068 12.68	199.38207 7.58	2,629.292 100.00
Total	136.11852 0.90	122.77242 0.81	265.84314 1.75	493.34109 3.25	542.61851 3.58	1,447.403 9.54	1,501.308 9.89	2,923.011 19.26	4,123.677 27.17	2,327.942 15.34	1,290.9661 8.51	15,175 100.00

To view all tables, look at your generated log file.

c) Now take a closer look at satisfaction with various aspects of life with the help of SOEP regional data. Use the municipal size classes. Create a table showing satisfaction with different aspects of life and highlighting differences by sex, age, municipal size class, and federal state.

```

foreach x of local varlist {
  * Tabulation of satisfaction by municipal size class and federal state
  table `x' sex alter_cat, by(bgbula ggk_cat) contents(freq) column row
↪stubwidth(20) cellwidth(8) cseppwidth(2) nomissing
  * Tabulation of satisfaction by municipal size class
  table `x' sex alter_cat, by(ggk_cat) contents(freq) column row
↪stubwidth(20) cellwidth(8) cseppwidth(2) nomissing
  * Tabulation of satisfaction by federal state

```

(continues on next page)

(continued from previous page)

```

7      table `x' sex alter_cat, by(bgbula) contents(freq) column row
8      ↪ stubwidth(20) cellwidth (8) csepxwidth(2) nomissing
      }

```

Federal states categorized, Community Size categorized and Satisfaction With Social Life	age categorized and Sex											
	<=20			21-30			31-65			>65		
	[1] Male	[2] Fema	Total	[1] Male	[2] Fema	Total	[1] Male	[2] Fema	Total	[1] Male	[2] Fema	Total
Schleswig-Holstein/H <=5000												
[0] Completely unsat							1		1			
[1] 1 On Scale 0-Low												
[2] 2 On Scale 0-Low												
[3] 3 On Scale 0-Low				1		1	1	1	2	1	1	2
[4] 4 On Scale 0-Low							2	2	4		1	1
[5] 5 On Scale 0-Low				1		1	5	7	12	4	1	5
[6] 6 On Scale 0-Low		1	1	2		2	5	10	15	1	2	3
[7] 7 On Scale 0-Low	3		3		2	2	17	16	33	4	4	8
[8] 8 On Scale 0-Low	1	2	3	3	4	7	21	32	53	10	8	18
[9] 9 On Scale 0-Low	1	3	4		4	4	18	22	40	9	7	16
[10] Completely sati	2	3	5	3	3	6	4	15	19	6	8	14
Total	7	9	16	10	13	23	74	105	179	35	32	67
Schleswig-Holstein/H 5001 - 20000												
[0] Completely unsat								1	1			
[1] 1 On Scale 0-Low												
[2] 2 On Scale 0-Low							3		3	1		1
[3] 3 On Scale 0-Low							3	1	4			
[4] 4 On Scale 0-Low				1		1	1	1	2	1		1
[5] 5 On Scale 0-Low							4	3	7		1	1
[6] 6 On Scale 0-Low				1	1	2	4	3	7	1	2	3
[7] 7 On Scale 0-Low		3	3		2	2	10	10	20	4	1	5
[8] 8 On Scale 0-Low	3	1	4	6	2	8	19	30	49	5	5	10
[9] 9 On Scale 0-Low	2	1	3	2	4	6	12	10	22	2	2	4
[10] Completely sati	3		3	2	1	3	4	10	14		1	1
Total	8	5	13	12	10	22	60	69	129	14	12	26
Schleswig-Holstein/H 20001 - 100000												
[0] Completely unsat								1	1			
[1] 1 On Scale 0-Low												
[2] 2 On Scale 0-Low					1	1			1	1		
[3] 3 On Scale 0-Low	1		1									
[4] 4 On Scale 0-Low				1		1		1	1			
[5] 5 On Scale 0-Low							1	7	8	3	4	7
[6] 6 On Scale 0-Low				1	1	2	4	4	8	3	2	5
[7] 7 On Scale 0-Low	2		2	3	4	7	15	13	28	1		1
[8] 8 On Scale 0-Low	1	1	2	3	4	7	22	25	47	4	4	8
[9] 9 On Scale 0-Low	1		1	6	4	10	13	23	36	3	5	8
[10] Completely sati		4	4	3	5	8	10	18	28	1	1	2
Total	5	5	10	17	19	36	65	93	158	15	16	31
Schleswig-Holstein/H >100000												
[0] Completely unsat							1		1			
[1] 1 On Scale 0-Low							1		1			
[2] 2 On Scale 0-Low							1	3	4	1		1
[3] 3 On Scale 0-Low							5	2	7	1	1	2
[4] 4 On Scale 0-Low							6		6	7	7	14
[5] 5 On Scale 0-Low					2	2						
[6] 6 On Scale 0-Low	1		1	1	1	2	13	17	30	8	7	15
[7] 7 On Scale 0-Low	3	2	5	3	10	13	25	32	57	3	9	12
[8] 8 On Scale 0-Low	2	2	4	10	8	18	44	60	104	14	20	34
[9] 9 On Scale 0-Low	8	4	12	7	10	17	25	37	62	12	15	27
[10] Completely sati	2	1	3	8	9	17	18	24	42	9	11	20
Total	16	9	25	29	40	69	139	175	314	55	70	125

To view all tables, look at your generated log file. As you can see, SOEP regional data can be used to analyze variables at the lowest regional levels.

d) Create a table that shows political interest differentiated by age, sex, and municipal size class in Bavaria

```

*****
capture log close
log using "${MY_LOG_OUT}\political_interest.log", replace

* Political interest
* Tabulation of political interest by municipal size class for Bavaria
table bgp143 sex alter_cat if bgbula==8, by(ggk_cat) contents(freq)
column row stubwidth(20) cellwidth (8) csepcwidth(2) nomissing

```

```
. table bgp143 sex alter_cat if bgbula==8, by(ggk_cat) contents(freq) column row stubwidth(20) cellwidth (8) csepcwidth(2) nomissing
```

Community Size categorised and Political Interests	age categorized and Sex											
	<=20			21-30			31-65			>65		
	[1] Male	[2] Fema	Total	[1] Male	[2] Fema	Total	[1] Male	[2] Fema	Total	[1] Male	[2] Fema	Total
<=5000												
[1] Very Strong				2	3	5	33	8	41	8	9	17
[2] Strong	11	3	14	9	15	24	124	97	221	33	24	57
[3] Not Much	13	17	30	31	32	63	129	202	331	35	36	71
[4] Not at All	12	17	29	15	19	34	42	78	120	1	6	7
Total	36	37	73	57	69	126	328	385	713	77	75	152
5001 - 20000												
[1] Very Strong	2	3	5	11	4	15	43	15	58	24	19	43
[2] Strong	10	7	17	26	14	40	138	128	266	72	45	117
[3] Not Much	21	17	38	28	38	66	187	281	468	55	74	129
[4] Not at All	14	17	31	18	31	49	68	120	188	6	13	19
Total	47	44	91	83	87	170	436	544	980	157	151	308
20001 - 100000												
[1] Very Strong	2		2	6		6	18	11	29	13	4	17
[2] Strong	6	3	9	11	10	21	56	48	104	30	26	56
[3] Not Much	11	7	18	25	34	59	85	127	212	22	26	48
[4] Not at All	9	6	15	16	27	43	53	69	122	2	4	6
Total	28	16	44	58	71	129	212	255	467	67	60	127
>100000												
[1] Very Strong	2		2	5	2	7	29	18	47	12	9	21
[2] Strong	1	5	6	25	22	47	101	85	186	40	29	69
[3] Not Much	6	13	19	26	31	57	85	142	227	22	26	48
[4] Not at All	1	4	5	12	20	32	37	50	87	3	12	15
Total	10	22	32	68	75	143	252	295	547	77	76	153

As you have seen here, the SOEP offers a wide range of possibilities for regional analysis. It is possible to allocate a multitude of regional indicators at the level of federal states, regional planning regions, districts, and postal codes.

Last change: Mar 28, 2023

6.10 Working with spatial data in R

6.10.1 Prerequisites

Data Access

To work with SOEP's spatial data you need to have a data distribution contract together with an application for stationary access to a terminal in the Research Data Center (RDC). The forms can be downloaded from [here](#). Having data access granted, you can use the IGEL workstations in the SOEP RDC to get access to the corresponding server infrastructures MORAN and HAUSER. The MORAN server hosts the coordinates of the SOEP households together with some fake coordinates that cannot be distinguished. On this server you can, for example, compute distances or nearest points as shown later in this tutorial. On the HAUSER server you are provided your computational outcome from the MORAN

server but without the coordinates. The coordinates of the SOEP households **must never** be available together with SOEP survey data. More information on how to use the SOEP IGEL technology can be found [here](#).

R

When working with SOEP's spatial data at the RDC using SOEP IGEL we provide you with the software you need. If you want to work with spatial test data before, you need a certain software setup. To be able to work yourself through the following examples you need to have [R](#) and [RStudio](#) installed. Further, for working the spatial data in R, you need to have GEOS, GDAL, and PROJ installed, too. When running a Windows OS, simply install the Rtools for Windows from [here](#). MacOS and Linux users are referred to the website of the [sf](#) package for installation instructions. You can find them [here](#).

If you are not yet familiar with R we recommend the book *R for Data Science* by Wickham and Grolemund which is available [here](#) for free. Besides that, there two freely available books which are helpful to get you started with spatial data in R:

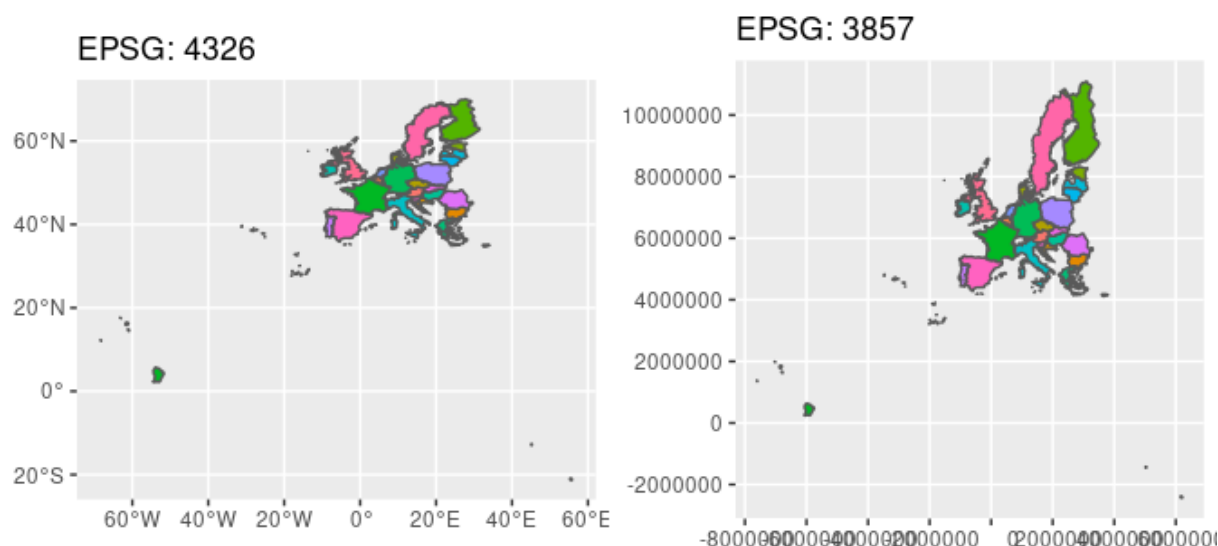
- *Spatial Data Science* by Pebesma and Bivand which is available [here](#)
- *Geocomputation with R* by Lovelace, Nowosad, and Muenchow which is available [here](#)

Coordinate Reference Systems

Any location of an object can be characterised by its coordinates. Besides a coordinate the altitude of an object might be of interest. So locations can be given in 2-dimensional or 3-dimensional space. A 2-dimensional coordinate referenced to two orthogonal axes (say x and y) can be given by either a pair of values (x,y) or by an angle (with respect to one of the axes) and a distance. The pair of coordinates (x,y) is referred to as cartesian coordinates and the latter as polar or spherical coordinates. In a 3-dimensional world, where you define a location on a globe, things become a little more complicated, because earth is an ellipsoid rather than a perfect sphere. In defining a location in a 3-dimensional coordinate system, the *coordinate reference system* (CRS) is very important. The CRS combines the information about the coordinate system itself and a geodetic datum. The geodetic datum describes how the location of the coordinate system is related to earth. So different CRS will provide different values of coordinates for the same location. Finally, the projection method provides information on how to flatten objects that are on a round surface so they can be viewed on a flat surface. The CRS is often summarised by an EPSG-code (European Petroleum Survey Group Geodesy) allowing for convenient transformations between different CRS. For more details we refer you to Chapter 2 of *Spatial Data Science* or Chapter 6 of *Geocomputation with R*. The most frequently used CRS are displayed in the following table.

EPSG code	Note	PRJ4-String	Projection
3857	WGS 84 / Pseudo-Mercator	+proj=merc +a=6378137 +b=6378137 +lat_ts=0 +lon_0=0 +x_0=0 +y_0=0 +k=1 +units=m +nadgrids=@null +wktext +no_defs +type=crs	Popular Visualisation Pseudo Mercator
4326	WGS 84	+proj=longlat +datum=WGS84 +no_defs +type=crs	(null)
25832	ETRS89 / UTM zone 32N	+proj=utm +zone=32 +ellps=GRS80 +units=m +no_defs +type=crs	Transverse Mercator
25833	ETRS89 / UTM zone 33N	+proj=utm +zone=33 +ellps=GRS80 +units=m +no_defs +type=crs	Transverse Mercator
32632	WGS 84 / UTM zone 32N	+proj=utm +zone=32 +datum=WGS84 +units=m +no_defs +type=crs	Transverse Mercator
32633	WGS 84 / UTM zone 33N	+proj=utm +zone=33 +datum=WGS84 +units=m +no_defs +type=crs	Transverse Mercator

The differences in the maps output can be seen in the following figure. The maps show Europe (provided by <https://data.opendatasoft.com>) in two different CRS.



The CRS of the fake SOEP households is defined by the EPSG code 32632. The CRS usually used by Google Maps / OpenStreetMap is defined by EPSG code 4326. To check you can use for example <https://coordinates-converter.com/>.

If you already know about all of the prerequisites and are familiar with R, the `sf` package as well as the `tidyverse`, you can directly switch to the complete example provided in section *Complete Example*. If you are not familiar with all of this, we recommend you go through the following text first.

6.10.2 Reading data

Spatial data comes in a variety of formats, ranging from ASCII-files to more sophisticated data structures like shapefiles or XML-dialects. Fortunately the R package `sf` provides functions to conveniently read most of the data formats. Besides that, the package and its use is well documented [here](#). The package can easily be installed and loaded by using the following lines of code in R.

```
install.packages('sf') # installing the package
library(sf) # loading the package

# further packages
install.packages('tidyverse') # a selection of packages
library(tidyverse) # (mainly using dplyr, ggplot2)
```

(continues on next page)

(continued from previous page)

```
install.packages('here') # for working with .RProj and paths
library(here)
```

Further packages that are useful when working with R are the tidyverse (a selection of several packages) and here (makes working in projects easier). A complete list of packages used to create this page is provided in Appendix at the end of this page.

When working with spatial data in the SOEP we provide you with fake coordinates of the SOEP households in a shapefile format. This format is easy to read using the sf package in R and provides you with all the details you need. The data you bring along is likely to be in a different format. Formats that can be read using `st_read` are listed in the [GDAL Documentation](#). If you bring along data in other formats consider the `haven` package for SAS, SPSS, or Stata formats, documented [here](#). Non-spatial data formats however, mostly include no information on the CRS. Thus you need to know and assign it to the data. An example when reading data from spreadsheets is provided below.

Shapefiles

Shapefiles consist of at least three files, namely

- `.shp` containing the geometries for example for a point (address) or a polygon (state)
- `.dbf` containing the attribute data for each geometry, for example, the address data (street, zip code, city) or names of the states
- `.shx` containing the indices linking geometries and attribute data

There are usually some more files provided containing further information. One example of shapefiles is provided by the Federal Agency for Cartography and Geodesy (BKG, GeoBasis-DE / BKG 2019). The BKG provides public available data on administrative areas in Germany. The dataset VG250 includes these administrative units of the hierarchical levels from the country down to municipalities and is available [here](#). Unzipping the folder will provide you with the shapefiles and the documentation of the data. Because the data contains information on different levels (Germany, its states, district, municipalities) these are stored in different layers. To check which information is available in the data, use `st_layers`. If you do not explicitly provide a layer, `st_read` will take the first available layer. The function returns the layer names, the according type of geometry, the number of features (geometries) and fields (variables). Reading the German Federal States you need to pick the layer VG250_LAN.

```
# layers in the data
st_layers(here('Daten', 'vg250_ebenen_0101'))
```

```
## Driver: ESRI Shapefile
## Available layers:
##   layer_name geometry_type features fields
## 1  VG250_GEM      Polygon      11135     26
## 2  VG250_KRS      Polygon       431     26
## 3  VG250_LAN      Polygon       35     26
## 4  VG250_LI      Line String  35483     4
## 5  VG250_PK       Point     11003    13
## 6  VG250_RBZ      Polygon       20     26
## 7  VG250_STA      Polygon       11     26
## 8  VG250_VWG      Polygon     4688     26
## 9  VG_DATEN       NA         35      8
## 10 VG_WERTE       NA         32      4
```

```
# read the shapefile for Federal States
States <- st_read(here('Daten', 'vg250_ebenen_0101'), layer = 'VG250_LAN
↪', quiet = TRUE)
```

Taking a look at the main variables `GEN` (Geographical name), `BEZ` (Designation), and `geometry` (a multipolygon), we are provided with information displayed differently from usual `data.frame` or `tibble` like data. The ‘header’ tells you it’s a collection of simple features having 16 features (rows) and 2 fields (variables, excluding the geometry column). The type of geometry is a multipolygon (set of polygons) in a two-dimensional space (XY). The bounding box (`bbox`) provides information on the bottom left and top right corner of the box bounding the geometries. Finally, you get information on the coordinate reference system (CRS), summarized by EPSG-Code 25832.

```
# look at the data
States %>%
  filter(GF == 4) %>% # use land with structure only
  select(GEN, BEZ, geometry) # select necessary variables
```

```
## Simple feature collection with 16 features and 2 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: 280371.1 ymin: 5235856 xmax: 921292.4 ymax: 5235856
↪ 6101444
## projected CRS:  ETRS89 / UTM zone 32N
## First 10 features:
##              GEN              BEZ
↪ geometry
## 1  Schleswig-Holstein      Land MULTIPOLYGON (((464810.7,
↪ 61...
## 2      Hamburg Freie und Hansestadt MULTIPOLYGON (((578219,
↪ 5954...
## 3      Niedersachsen      Land MULTIPOLYGON (((479451.1,
↪ 59...
## 4      Bremen      Freie Hansestadt MULTIPOLYGON (((466930.3,
↪ 58...
## 5  Nordrhein-Westfalen      Land MULTIPOLYGON (((477607.3,
↪ 58...
## 6      Hessen      Land MULTIPOLYGON (((534242,
↪ 5721...
## 7      Rheinland-Pfalz      Land MULTIPOLYGON (((416304.5,
↪ 56...
## 8      Baden-Württemberg      Land MULTIPOLYGON (((546771.2,
↪ 55...
## 9      Bayern      Freistaat MULTIPOLYGON (((609387.6,
↪ 52...
## 10     Saarland      Land MULTIPOLYGON (((359841,
↪ 5499...
```

SOEP Regional Data

Also the fake coordinates for the SOEP households are provided in a shapefile format. The data contains an ID-variable (ID) and the survey year (`erhebj`) together with a point geometry. The ID provided is not related to any of the SOEP’s `hid` or `pid`. It is simply a number ranging from 1 to the corresponding number of rows in the data set. Households can thus only be identified, merged, or selected using their coordinate. Moreover the data also contain the SOEP’s initial sample. Thus containing regional information on respondents and non-respondents. Because of data protection regulations, the SOEP survey data must not be used together with the household’s coordinates. This is why the environment where you are allowed to work with the household’s coordinates is strictly separated. Unlike the VG250 data the SOEP data has the EPSG-code 32632.

```
SOEP <- st_read(here('Daten', 'soep_v29'))
```

```
## Reading layer `soep_hh_korr_nur_fakes_utm32-v29' from data source `/  
media/H/Projekte/Adhoc_Analysen/GeodatenBSP/Daten/soep_v29' using driver `ESRI Shapefile'  
## Simple feature collection with 209734 features and 2 fields  
## geometry type: POINT  
## dimension: XY  
## bbox: xmin: 288107 ymin: 5250939 xmax: 888152 ymax: 6055336  
## projected CRS: WGS 84 / UTM zone 32N
```

```
SOEP
```

```
## Simple feature collection with 209734 features and 2 fields  
## geometry type: POINT  
## dimension: XY  
## bbox: xmin: 288107 ymin: 5250939 xmax: 888152 ymax: 6055336  
## projected CRS: WGS 84 / UTM zone 32N  
## First 10 features:  
##   erhebj ID          geometry  
## 1   2000  1 POINT (355929 5665928)  
## 2   2000  2 POINT (800240 5821852)  
## 3   2000  3 POINT (794464 5831012)  
## 4   2000  4 POINT (478831 5431318)  
## 5   2000  5 POINT (799318 5825023)  
## 6   2000  6 POINT (794752 5833726)  
## 7   2000  7 POINT (799149 5824960)  
## 8   2000  8 POINT (791594 5827687)  
## 9   2000  9 POINT (791657 5827422)  
## 10  2000 10 POINT (792280 5826345)
```

Spreadsheets

When reading spatial data from spreadsheets the data is read by different drivers. The example data is contained in an Excel spreadsheet with three variables X (longitude), Y (latitude), and Object. Using the `st_read` function the csv-driver from GDAL is chosen automatically, reading 4 columns however. Thus we get rid of the last column first.

```
POI <- st_read(here('Daten', 'POIs_Berlin.csv'))
```

```
## Reading layer `POIs_Berlin' from data source `/  
media/H/Projekte/Adhoc_Analysen/GeodatenBSP/Daten/POIs_Berlin.csv' using driver `CSV'
```

```
POI
```

```
##           X           Y           Objekt field_4  
## 1 13.3776978296166 52.5162788298363 Brandenburg Tor <NA>  
## 2 13.3693019102667 52.5250274843424 Hauptbahnhof <NA>  
## 3 13.3886070846104 52.5121727411876 DIW Berlin <NA>  
## 4 13.4589278721678 52.496825784817 Molecule Man <NA>  
## 5 13.2887256012712 52.5580119710438 Flughafen Tegel <NA>  
## 6 13.4005469632413 52.4792779435797 Tempelhofer Feld <NA>  
## 7 13.4484578382279 52.4482123689564 Hufeisensiedlung <NA>
```

(continues on next page)

(continued from previous page)

```
## 8 13.281110837709 52.4473851451014 FU Berlin <NA>
## 9 13.1725456943567 52.4300072503147 Wannsee <NA>
## 10 13.2128459704725 52.5411476127425 Zitadelle Spandau <NA>
## 11 13.5727300494297 52.4438163134271 Schloss Köpenick <NA>
## 12 13.2795658406826 52.5080182126806 Zentraler Omnibus Bahnhof <NA>
```

```
POI <- POI[, -4]
```

Because this data contains latitude and longitude already, we can simply transform it to a spatial dataset using `st_as_sf` and the CRS has to be assigned to it.

```
POI <- st_as_sf(POI, coords = c("X", "Y"), crs = 4326)
POI
```

```
## Simple feature collection with 12 features and 1 field
## geometry type: POINT
## dimension: XY
## bbox: xmin: 13.17255 ymin: 52.43001 xmax: 13.57273 ymax: 52.55801
## geographic CRS: WGS 84
## First 10 features:
##      Objekt geometry
## 1 Brandenburger Tor POINT (13.3777 52.51628)
## 2 Hauptbahnhof POINT (13.3693 52.52503)
## 3 DIW Berlin POINT (13.38861 52.51217)
## 4 Molecule Man POINT (13.45893 52.49683)
## 5 Flughafen Tegel POINT (13.28873 52.55801)
## 6 Tempelhofer Feld POINT (13.40055 52.47928)
## 7 Hufeisensiedlung POINT (13.44846 52.44821)
## 8 FU Berlin POINT (13.28111 52.44739)
## 9 Wannsee POINT (13.17255 52.43001)
## 10 Zitadelle Spandau POINT (13.21285 52.54115)
```

6.10.3 Transformations

To be able to work with all three data sets (States, SOEP, and POI) we have to make sure all have the same CRS. Using `st_transform` we can reproject the data sets to have the same EPSG-code (`common_crs`).

```
common_crs <- 25832

SOEP <- st_transform(SOEP, crs = common_crs)
SOEP <- SOEP[SOEP$erhebj == 2005, ] # only use 2005 SOEP data

POI <- st_transform(POI, crs = common_crs)

# States <- st_transform(States, crs = common_crs) already correct crs
```

Sometimes it might be easier for you to work in other programs and you wish to have latitude and longitude data. In this case you can use `st_coordinates` to transform point geometries into (x,y) coordinates. The below example first transforms the coordinates for the households 1 to 5 in the SOEP data into latitude and longitude data and then creates the (x,y) coordinates from the point geometry.


```
soep5_lat_lon <- st_transform(SOEP[1:5,], 4326)
st_coordinates(soep5_lat_lon)
```

```
##           X           Y
## 1  6.94659 51.12465
## 2 13.36828 52.48428
## 3 13.40601 52.39064
## 4  8.70727 49.03966
## 5 13.43205 52.48811
```

6.10.4 Plotting Spatial Data

Besides being able to look up coordinates or objects in google maps or OpenStreetMap, we can use the `ggplot2` package included in the `tidyverse` package for displaying the data. The package provides the `geom_sf` function to easily plot the data.

```
# subsetting and plotting the data
States %>% # use the states data
  filter(GEN == 'Berlin') %>% # filter for Berlin only
  ggplot() + # plot base
  geom_sf(fill = 'white') + # add the polygon for Berlin and fill it
↪white

  geom_sf(data = POI) + # add the POI data
  geom_sf_text(data = POI, aes(label = Objekt),
               nudge_y = -1000,
               check_overlap = TRUE) + # overlapping labels will not be
↪displayed

  xlab('') + ylab('')
```



6.10.5 Frequently Used Operations

This section will provide an overview of some frequently used operations when working with spatial data. The dataset SOEP contains some fake coordinates of household addresses we will work with throughout the examples.

Finding Households in a Specified Area

Suppose we are interested in identifying all the SOEP households located in Berlin. The corresponding polygon for Berlin is provided in the States dataset. Because we are interested in Berlin only, we save the polygon in an own object BE. The function `st_contains` identifies the row-index in the SOEP dataset that fall within the polygon BE and returns a list (`soep_in_berlin`). For checks along the way we can look at plots of the data.

```
BE <- States[States$GEN == 'Berlin', ]
BE %>% select(GEN, BEZ)
```

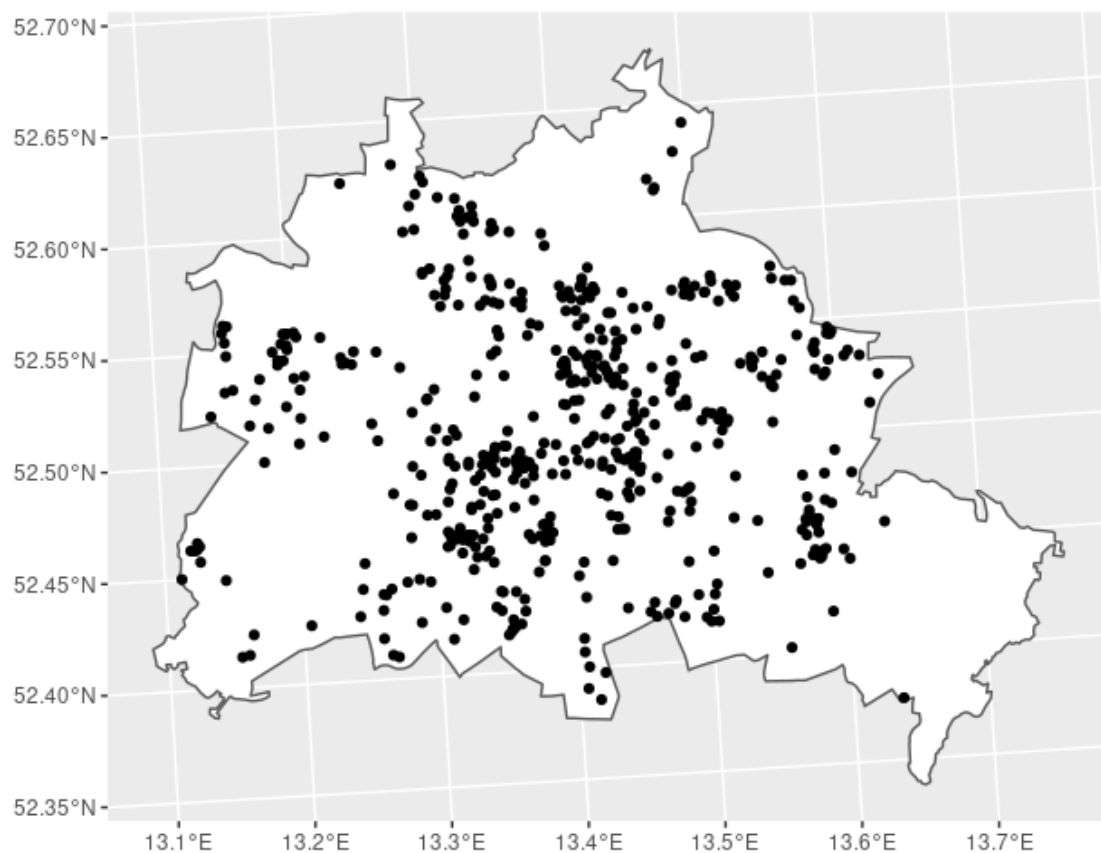
```
## Simple feature collection with 1 feature and 2 fields
## geometry type: MULTIPOLYGON
## dimension: XY
## bbox: xmin: 777974.1 ymin: 5808837 xmax: 823510.5 ymax: 5845580
## projected CRS: ETRS89 / UTM zone 32N
## GEN BEZ geometry
## 11 Berlin Land MULTIPOLYGON (((802831.7 58...
```

```
soep_in_berlin <- st_contains(BE, SOEP)
soep_in_berlin
```

```
## Sparse geometry binary predicate list of length 1, where the
predicate was 'contains'
## 1: 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, ...
```

```
SOEP_BE <- SOEP[unlist(soep_in_berlin), ]

BE %>%
  ggplot() + geom_sf(fill = 'white') + geom_sf(data = SOEP_BE)
```



Distances / Areas

To compute distances the function `st_distance` can be provided a single dataset with n rows providing a $n \times n$ matrix of distances of the geometries contained in the data (`dist_m`). The unit of the distance returned depends on the CRS. In the example provided below the distances of the POIs to the location of the DIW are given in meters.

Providing the function a second dataset of m rows will create a $n \times m$ distance matrix (`dist_soep_poi`). The created object is a matrix with 529 rows (SOEP households in Berlin) and 12 columns (POIs in Berlin).

When computing distances on large data sets it might be helpful to subset the data, because the distance of a household in Munich might be irrelevant to a research question focusing on Berlin or distances smaller than 5000m. According to the CRS, consider specifying the `which` argument.

Areas can be computed for (multi-)polygons. The function `st_area` provides the corresponding information.

```
# distances between the POIs
dist_m <- st_distance(POI)
rownames(dist_m) <- POI$Objekt
colnames(dist_m) <- POI$Objekt

# distance of POIs to DIW Berlin
dist_m['DIW Berlin', ]
```

```
## Units: [m]
##      Brandenburger Tor      Hauptbahnhof
↳ DIW Berlin      870.8119      1941.2940
↳ 0.0000
##      Molecule Man      Flughafen Tegel
↳ Tempelhofer Feld      5074.7951      8488.2154
↳ 3751.7696
##      Hufeisensiedlung      FU Berlin
↳ Wannsee      8202.7627      10269.1064
↳ 17307.5100
##      Zitadelle Spandau      Schloss Köpenick Zentraler Omnibus
↳ Bahnhof      12364.7837      14651.8215
↳ 7422.6480
```

```
# distance between each household and each POI
dist_soep_poi <- st_distance(SOEP_BE, POI)
dim(dist_soep_poi)
```

```
## [1] 529 12
```

```
# save distances in an object
DIST <- as_tibble(dist_soep_poi)
# add names
names(DIST) <- str_c('distance_to_', str_remove(POI$Objekt, ' '))

# attach distances to data
SOEP_BE <- bind_cols(SOEP_BE, DIST)

# area covered by Berlin
st_area(BE)
```

```
## 893060962 [m^2]
```

Nearest Point

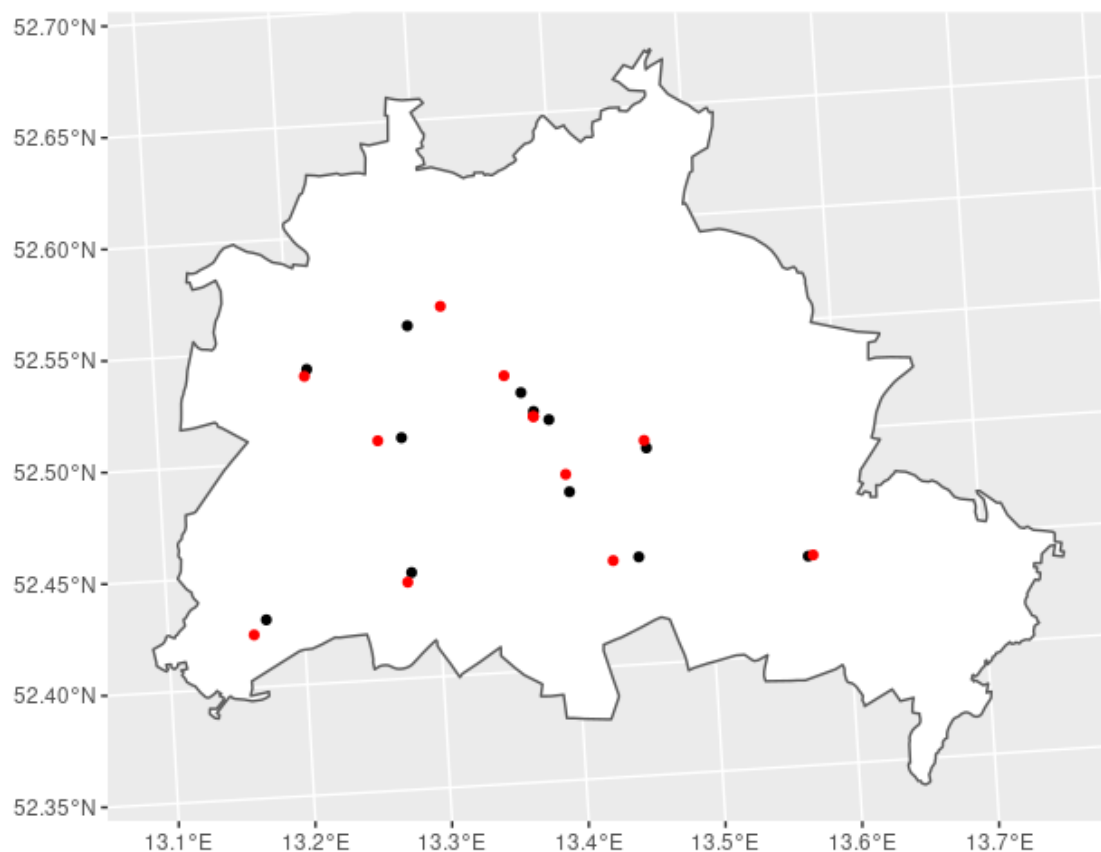
To find the point or feature closest to another one the function `st_nearest_feature` will return a vector with indices of the nearest feature. In the example below we are looking for the households living closest to the POIs in Berlin. In the second step we compute the corresponding distances between the POI and the household.

```
nearest_hh <- st_nearest_feature(POI, SOEP_BE)
```

```
diag(st_distance(POI, SOEP_BE[nearest_hh, ]))
```

```
## Units: [m]
## [1] 256.5810 1195.8833 789.6812 392.5787 1913.8130 891.4404 1295.
↪8621
## [8] 520.3830 971.0417 364.4532 250.2833 1194.7832
```

```
BE %>% # polygon for Berlin
ggplot() + geom_sf(fill = 'white') + # plot the Berlin polygon
geom_sf(data = POI) + # add the POIs
geom_sf(data = SOEP_BE[nearest_hh, ], col = 'red') # add the nearest
↪household
```



Within Distance

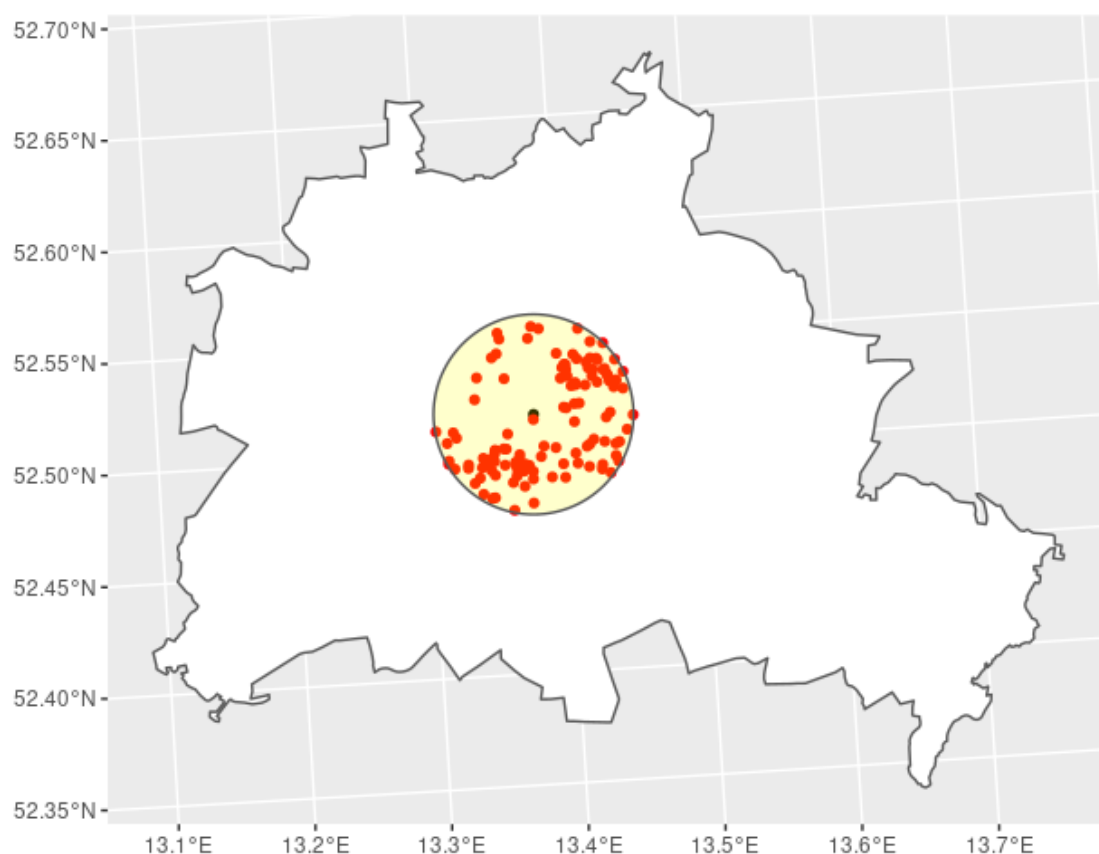
If your interest is about which households live within a certain distance to a specific point, `st_is_within_distance` lets you lookup geometries in a given distance (argument `dist`) and returns a list. The below example looks up households in a 5km distance of the Brandenburger Tor. The plot shows the 5km radius area in yellow, the location of the Brandenburger Tor (black dot) and then households within the distance (red dots).

```
r_5000 <- st_is_within_distance(POI[POI$Objekt == 'Brandenburger Tor'],
                               SOEP_BE,
                               dist = 5000)

r_5000
```

```
## Sparse geometry binary predicate list of length 1, where the
↪ predicate was 'is_within_distance'
## 1: 1, 3, 6, 12, 21, 25, 28, 39, 54, 55, ...
```

```
BE %>% # polygon for Berlin
  ggplot() + geom_sf(fill = 'white') + # plot the Berlin polygon
  geom_sf(data = POI[POI$Objekt == 'Brandenburger Tor'],) + # add the POI
  geom_sf(data = SOEP_BE[unlist(r_5000), ], col = 'red') + # add the
↪ nearest household
  geom_sf(data = st_buffer(POI[POI$Objekt == 'Brandenburger Tor'], dist_
↪ = 5000),
          alpha = 0.2, fill = 'yellow')
```



The question can also be asked the other way around: How many POIs are within a 5km radius of the SOEP households? This way the function `st_is_within_distance` returns a list of length equal to the number of SOEP households in Berlin (529). For each household the (row) index for the POI is given. To get the number of POIs in the 5km radius, we can simply ask for the length (the number of row indices) of each list-component. To get the according distances see section *Distances / Areas*

```
poi_5000 <- st_is_within_distance(SOEP_BE, POI, dist = 5000)
poi_5000
```

```
## Sparse geometry binary predicate list of length 529, where the
↪ predicate was 'is_within_distance'
## first 10 elements:
## 1: 1, 2, 3, 6
## 2: (empty)
## 3: 1, 3, 4, 6, 7
## 4: 5
## 5: 2, 5, 12
## 6: 1, 2, 12
## 7: 10
## 8: 5, 10, 12
## 9: 8
## 10: 5, 10, 12
```

```
N_POI <- as_tibble(sapply(poi_5000, length))
names(N_POI) <- 'n_poi_in_5km'

SOEP_BE <- bind_cols(SOEP_BE, N_POI)
```

Spatial joins

When you are used to working with SOEP data you will have probably merged / joined data sets using the identifying variables (cid, hid, pid) and the survey year (syear) before. When you are working with spatial data you will have to choose one of the geometry predicate function provided by the *sf* package. The default is a left join of the two data sets using *st_intersects* as the geometry predicate function for joining. You can however change this, for example, to join the nearest features, see section *Nearest Point*. In our example here, we join the nearest SOEP household to each of the points of interest. The geometry column here provides the coordinates from the POI data set.

```
NEAR <- st_join(POI, SOEP, join = st_nearest_feature)
NEAR
```

```
## Simple feature collection with 12 features and 3 fields
## geometry type: POINT
## dimension: XY
## bbox: xmin: 783630.5 ymin: 5817059 xmax: 810721.7 ymax: ↪
↪ 5831749

## projected CRS: ETRS89 / UTM zone 32N
## First 10 features:
##      Objekt erhebj      ID      geometry
## 1 Brandenburger Tor 2005 75759 POINT (796986.8 5827473)
## 2 Hauptbahnhof     2005 69076 POINT (796358.6 5828411)
## 3 DIW Berlin        2005 75759 POINT (797754.3 5827062)
## 4 Molecule Man     2005 75646 POINT (802628.5 5825649)
## 5 Flughafen Tegel  2005 73266 POINT (790677.7 5831749)
## 6 Tempelhofer Feld  2005 69820 POINT (798787.2 5823455)
## 7 Hufeisensiedlung  2005 69477 POINT (802251.6 5820202)
## 8 FU Berlin         2005 67568 POINT (790892.2 5819422)
## 9 Wannsee           2005 67923 POINT (783630.5 5817059)
## 10 Zitadelle Spandau 2005 68827 POINT (785646.9 5829572)
```

Export Results

To export your results you can use `st_write` to create a `.csv` file. When exporting your results please check the requirements [here](#).

```
path_export <- paste0('/home/', Sys.info()['user'], '/transfer/export/',  
↳ Sys.Date())  
  
if(!file.exists(path_export)){  
  dir.create(path_export, recursive = TRUE)  
}  
  
st_write(SOEP_BE, file.path(path_export, 'Output_SOEP_BE.csv'),  
  append = FALSE, overwrite = TRUE)  
  
README <- tibble(name = names(SOEP_BE)[-grep('geometry', names(SOEP_  
↳ BE))],  
  description = c('erhebj',  
    'ID',  
    'distance (in meters) of household to',  
↳ Brandenburger Tor',  
    'distance (in meters) of household to',  
↳ Hauptbahnhof',  
    'distance (in meters) of household to',  
↳ DIW-Berlin',  
    'distance (in meters) of household to',  
↳ Molecule Man',  
    'distance (in meters) of household to',  
↳ Flughafen Tegel',  
    'distance (in meters) of household to',  
↳ Tempelhofer Feld',  
    'distance (in meters) of household to',  
↳ Hufeisensiedlung',  
    'distance (in meters) of household to',  
↳ FU Berlin',  
    'distance (in meters) of household to',  
↳ Wannsee',  
    'distance (in meters) of household to',  
↳ Zitadelle Spandau',  
    'distance (in meters) of household to',  
↳ Schloss Köpenick',  
    'distance (in meters) of household to',  
↳ Zentraler Omnibus Bahnhof',  
    'number of POIs within 5 km radius of',  
↳ household'))  
  
write.csv(README, file.path(path_export, 'Output_SOEP_BE.csv'),  
  row.names = FALSE)
```


6.10.6 Complete Example

Suppose you want to know which households of the SOEP from survey year 2011 live within a distance of 5000m to the following points of interest (POI):

- Brandenburger Tor
- Zitadelle Spandau
- Wannsee

Besides that, you want to know how far their distance to the corresponding POI is and which household lives closest to the corresponding POI. After computing the informations need you want to export the results for further use on the HAUSER server.

```
# Global stuff
# ~~~~~

# packages
library(here)
library(sf)
library(tidyverse)

# global values
survey_year <- 2011
distance <- 5000 # meter
common_crs <- 25832

# Step 1: read the data
# ~~~~~

# read polygons for Federal States
States <- st_read(here('Daten', 'vg250_ebenen_0101'), layer = 'VG250_LAN
↪', quiet = TRUE)

# read SOEP data
SOEP <- st_read(here('Daten', 'soep_v29'), quiet = TRUE)

# read POI data
POI <- st_read(here('Daten', 'POIs_Berlin.csv'), quiet = TRUE)

# Step 2: Transform data
# ~~~~~

# transform SOEP the data
SOEP <- st_transform(SOEP, crs = common_crs)

# transform POIs
POI <- POI[, -4]

POI <- st_as_sf(POI, coords = c("X", "Y"), crs = 4326)
```

(continues on next page)

(continued from previous page)

```

POI <- st_transform(POI, crs = common_crs)

# Step 3: Subset the data
# ~~~~~

# Berlin only
BE <- States %>%
  filter(GEN == 'Berlin')

# interesting POIs only
POI <- POI %>%
  filter(Objekt %in% c("Brandenburger Tor", "Zitadelle Spandau", "Wannsee
→"))

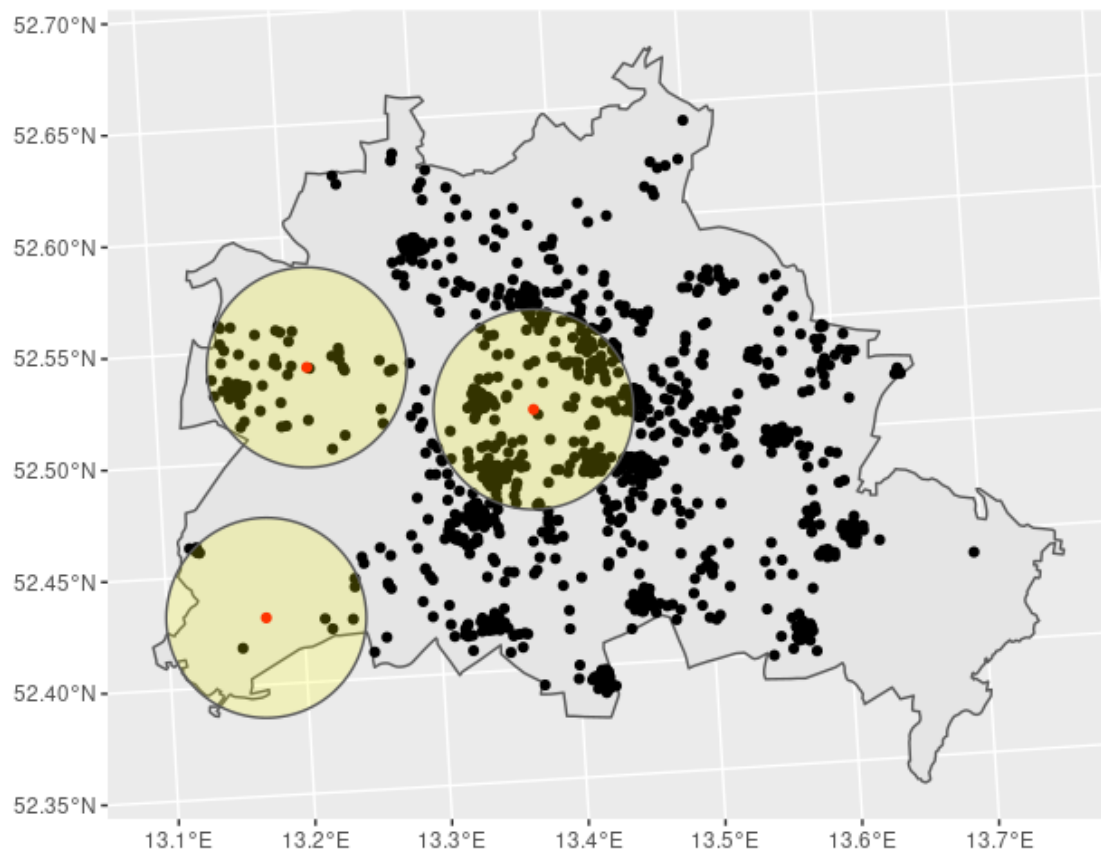
# SOEP survey year: 2011 only
SOEP <- SOEP %>%
  filter(erhebj == survey_year)

# SOEP households in Berlin only
SOEP <- SOEP[unlist(st_contains(BE, SOEP)), ]

# Step 4: Plot the data
# ~~~~~

BE %>%
  ggplot() + geom_sf() +
  geom_sf(data = SOEP) +
  geom_sf(data = POI, col = 'red') +
  geom_sf(data = st_buffer(POI, dist = distance),
    alpha = 0.2, fill = 'yellow')

```



```
# Step 5: Get the desired information
# ~~~~~

# get the household in relevant distance
soep_in_5km <- st_is_within_distance(POI, SOEP, dist = distance)
SOEP_5km <- SOEP[unlist(soep_in_5km), ]

# get the distances
dist_matrix <- as_tibble(st_distance(SOEP_5km, POI))
names(dist_matrix) <- str_c('distance_to_', str_remove(POI$Objekt, ' '))

# attach distances to data
SOEP_5km <- bind_cols(SOEP_5km, dist_matrix)

# find the nearest household
nearest_hh <- st_nearest_feature(POI, SOEP_5km)

SOEP_5km$nearest_to <- ''
SOEP_5km$nearest_to[nearest_hh] <- POI$Objekt

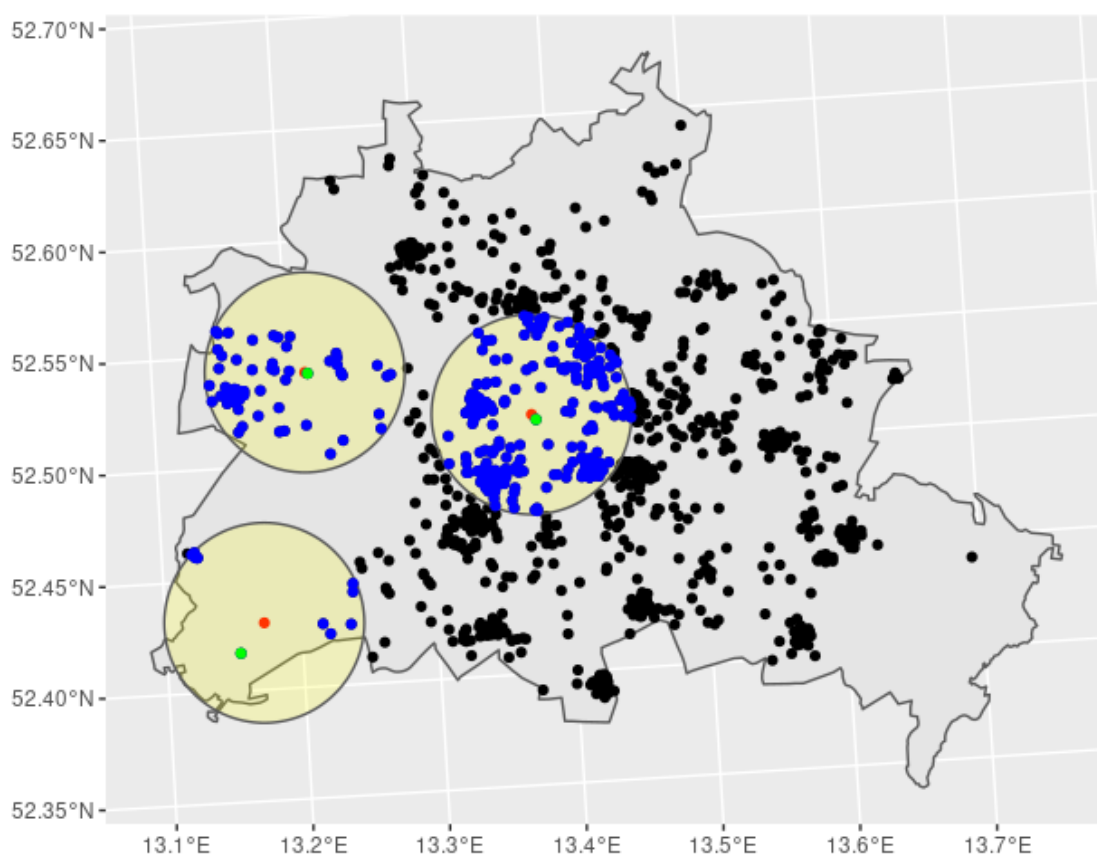
# Step 6: Check your data
```

(continues on next page)

(continued from previous page)

```
# ~~~~~

BE %>%
  ggplot() + geom_sf() +
  geom_sf(data = SOEP) +
  geom_sf(data = POI, col = 'red') +
  geom_sf(data = st_buffer(POI, dist = distance),
    alpha = 0.2, fill = 'yellow') +
  geom_sf(data = SOEP_5km, col = 'blue') +
  geom_sf(data = SOEP_5km[SOEP_5km$nearest_to != '', ], col = 'green')
```



```
# Step 7: Export your data
# ~~~~~
path_export <- paste0('/home/', Sys.info()['user'], '/transfer/export/',
Sys.Date())

if(!file.exists(path_export)){
  dir.create(path_export, recursive = TRUE)
}

st_write(SOEP_5km, file.path(path_export, 'Output.csv'),
  append = FALSE, overwrite = TRUE)
```

```

## Deleting layer `Output' using driver `CSV'
## Writing layer `Output' to data source `/home/hwsteinhauer/transfer/
↪export/2021-01-27/Output.csv' using driver `CSV'
## Updating existing layer Output
## Writing 356 features with 6 fields and geometry type Point.

```

```

README <- tibble(name = names(SOEP_5km)[-grep('geometry', names(SOEP_
↪5km))],
                  description = c('erhebj',
                                  'ID',
                                  'distance (in meters) of household to ↪
↪Brandenburger Tor',
                                  'distance (in meters) of household to ↪
↪Wannsee',
                                  'distance (in meters) of household to ↪
↪Zitadelle Spandau',
                                  'household nearest to the POI'))

write.csv(README, file.path(path_export, 'README.csv'),
          row.names = FALSE)

```

6.10.7 Appendix

Session info - Platform

```

## setting value
## version R version 4.0.3 (2020-10-10)
## os Ubuntu 20.04.1 LTS
## system x86_64, linux-gnu
## ui X11
## language (EN)
## collate de_DE.UTF-8
## ctype de_DE.UTF-8
## tz Europe/Berlin
## date 2021-01-27

```

Session info - Packages

```

## package * version date lib source
## bookdown * 0.21 2020-10-13 [1] CRAN (R 4.0.3)
## dplyr * 1.0.2 2020-08-18 [1] CRAN (R 4.0.2)
## forcats * 0.5.0 2020-03-01 [1] CRAN (R 4.0.0)
## ggplot2 * 3.3.2 2020-06-19 [1] CRAN (R 4.0.1)
## gridExtra * 2.3 2017-09-09 [1] CRAN (R 4.0.0)
## here * 1.0.0 2020-11-15 [1] CRAN (R 4.0.3)
## kableExtra * 1.3.1 2020-10-22 [1] CRAN (R 4.0.3)
## knitr * 1.30 2020-09-22 [1] CRAN (R 4.0.2)
## pacman * 0.5.1 2019-03-11 [1] CRAN (R 4.0.2)
## purrr * 0.3.4 2020-04-17 [1] CRAN (R 4.0.0)
## readr * 1.4.0 2020-10-05 [1] CRAN (R 4.0.3)
## rgdal * 1.5-19 2021-01-05 [1] CRAN (R 4.0.3)

```

(continues on next page)

(continued from previous page)

```
## sessioninfo * 1.1.1 2018-11-05 [1] CRAN (R 4.0.0)
## sf          * 0.9-7 2021-01-06 [1] CRAN (R 4.0.3)
## sp          * 1.4-4 2020-10-07 [1] CRAN (R 4.0.3)
## stringr     * 1.4.0 2019-02-10 [1] CRAN (R 4.0.0)
## tibble      * 3.0.4 2020-10-12 [1] CRAN (R 4.0.3)
## tidyr       * 1.1.2 2020-08-27 [1] CRAN (R 4.0.2)
## tidyverse   * 1.3.0 2019-11-21 [1] CRAN (R 4.0.0)
##
## [1] /home/hwsteinhauer/R/x86_64-pc-linux-gnu-library/4.0
## [2] /usr/local/lib/R/site-library
## [3] /usr/lib/R/site-library
## [4] /usr/lib/R/library
```

Section author: Hans Walter Steinhauer hwsteinhauer@diw.de

Last updated: 2021-02-22

6.11 How to Use SOEP IGEL

6.11.1 IGEL Workstation

IGEL refers to a computer terminal workstation for access to SOEP data:

- The terminal allows data to be entered and displayed.
- The IGEL is a so-called thin client, a computer with little computing power, which only provides a terminal to a server.
- This thin client at the SOEP guest workstation/FDZ is from the manufacturer Igel Technology, where IGEL stands for “**I** ntelligente **G** esamtlösung in der Mikro **el** ektronik”.

Account

Access to SOEP data can only be provided in compliance with high security standards to protect respondents’ confidentiality and maintain their trust in the survey. The data are also provided solely for scientific research purposes, that is, they are only made available to members of the scientific community. Researchers can therefore only access SOEP data after they have signed a data distribution contract with DIW Berlin. The same rules apply to the secure guest workstations at RDC SOEP and at other secure data access points. Since IGEL terminals also provide access to small-scale regional data, users have to sign additional contracts for these data.

All IGEL users must sign a data distribution contract with the DIW Berlin: [Application for a Data Distribution Contract](#).

6.11.2 Logging in

Turn on the computer and the following screen should appear on the monitor. (see [figure 1](#))



Fig. 1: Figure 1: IGEL start screen

At the bottom right, you should see the icon for an existing network connection appear: Two arrows, one pointing up and one pointing down.

Click on the arrow icon to see the terminal name and the existing network connection . See [figure 2](#).

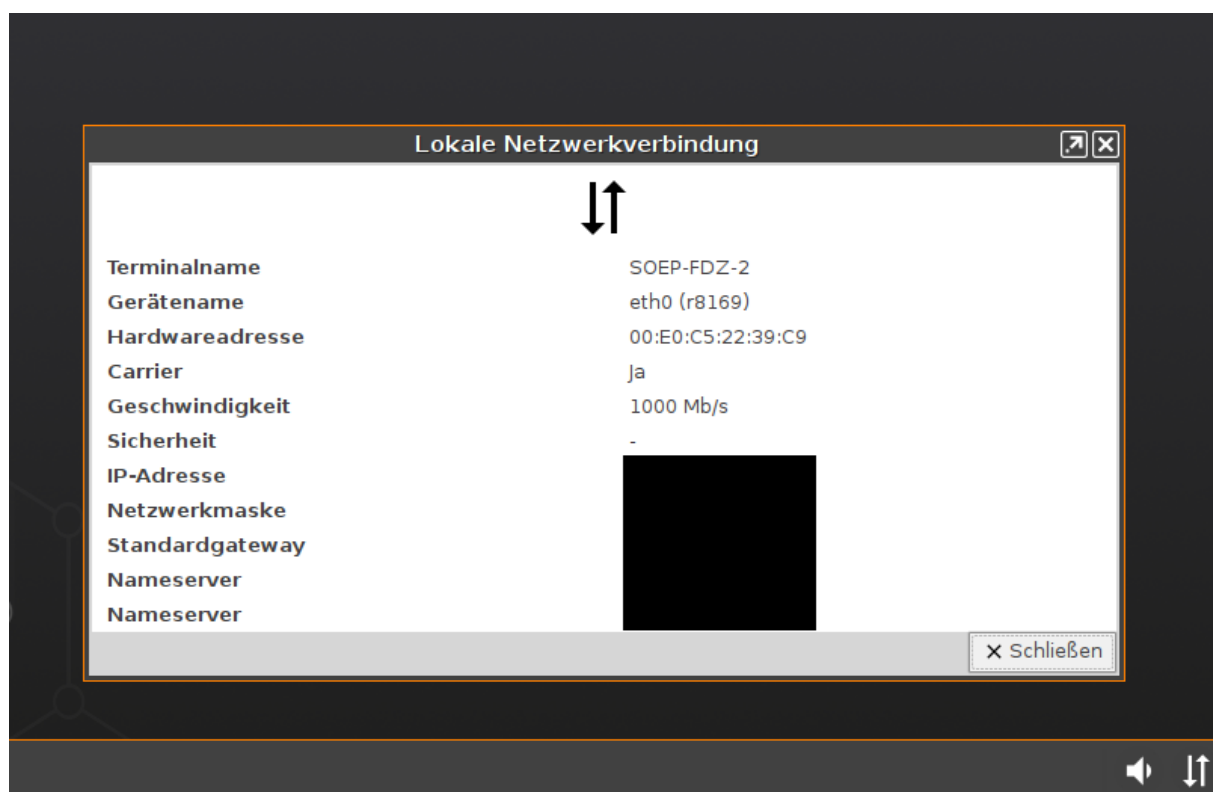


Fig. 2: Figure 2: Connection with LAN available

For each available server, two icons are displayed on the start screen at the top left, a red one and a blue one with the same name. See *figure 3*.

The following two servers are currently available:

1. HAUSER: Access to the SOEP survey data, including connection to small-scale regional indicators (WITHOUT coordinates).
2. MORAN: Access to the coordinates of SOEP households, but without survey data.

Access is only possible from RDC SOEP guest stations at DIW Berlin

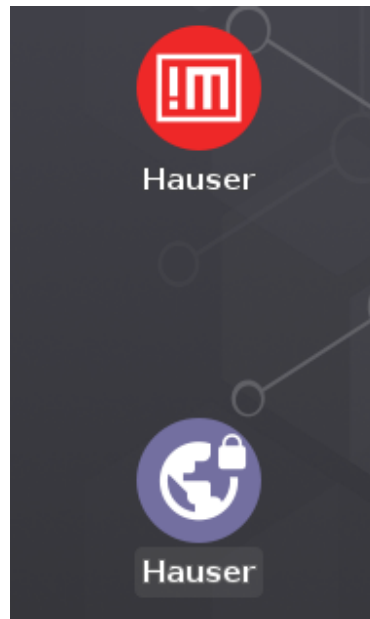


Fig. 3: Figure 3: Icons to connect with the SOEP server

Blue Icon: To connect to one of the two servers at RDC SOEP, first establish an open VPN connection by clicking on the blue icon for the server you would like to connect to. The icon in the lower right corner should then display the existing VPN connection. By clicking once on this icon, you can see the server's IP address . See [figure 4](#)

Red Icon: Once you have established the VPN connection to the SOEP server, click once on the red icon to start your session. The server's login window should appear, see [figure 5](#). Enter the user name and password provided to you by RDC SOEP.

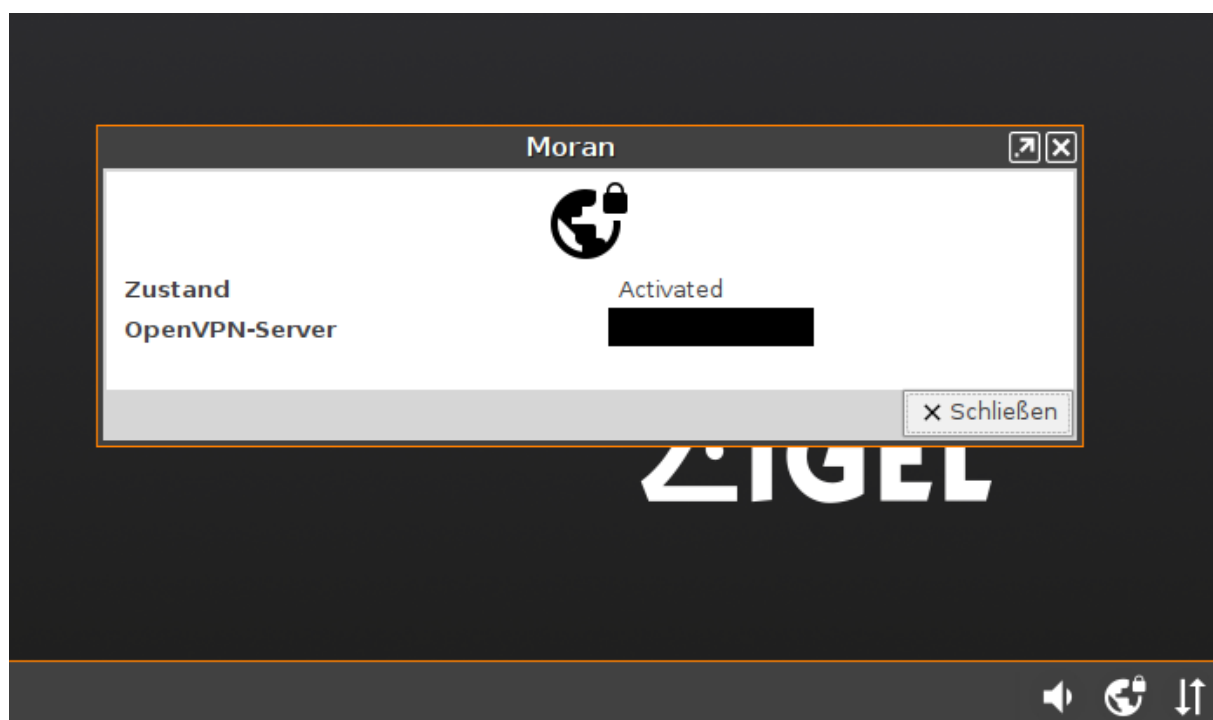


Fig. 4: Figure 4: Open VPN connection established

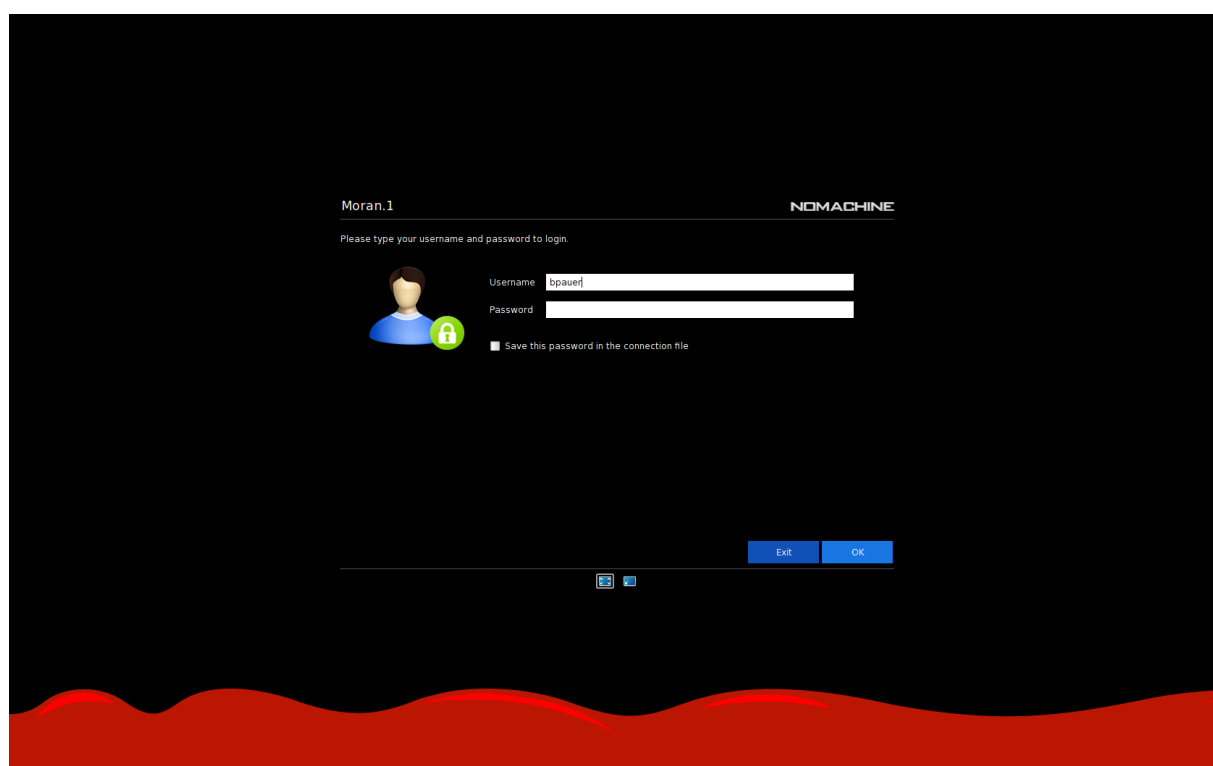


Fig. 5: Figure 5: Login to the SOEP Server

6.11.3 Working with SOEP DATA

Starting programs

- After you have logged in, a blank desktop will appear with a menu bar at the top.
- In general, programs can be started by clicking on “activities” and then either by clicking on the icon or by typing the name of the program into the search field.
- Users should inform the RDC SOEP team in advance about any additional ados in Stata or packages in R. These will be installed after checking.
- **Start Stata:** Unfortunately, there is no automatic start icon for Stata, so you have to do the following:
 1. Click on activities
 2. Enter “terminal” in the search window
 3. Start either “Terminal” or “XTerm”.
 4. Enter the command “xstata-mp” into the terminal that has now appeared, and press the return key. Stata should now appear.
- The following table shows which programs are installed and available for use on each server:

Running time-consuming computing operation

If a script needs more than one hour for calculation, the script should be run in the background. For long running scripts in R or Stata you should not use xstata or Rstudio.

Executing your script

- The R-script should be executed with the “Rscript” command in the terminal.
- The STATA do-file should be executed with the “stata-mp do” command in the terminal.

What should be used?

You can start a session in tmux or you can use the command nohup. We suggest to use tmux.

How to start your script?

Open a terminal and change to the folder where the script is located with the command `cd`. Now you can execute your script using tmux as follows:

1. Start a tmux session and assign a name (***myname***) with the command `tmux new -s myname`. If later you forget the name, use the command: `tmux list-session`
2. Execute your script: In R (***yourscript.R***) / in STATA (***yourdofile.do***)
3. When logging in again, all running sessions can be listed with `tmux list-session`
4. This session (***myname***) could be restored after the terminal was closed with `tmux attach-session -t myname`
5. When your script is finished, please start your tmux session again with `tmux attach-session -t myname` and close your running session with `exit`

Program	HAUSER	MORAN
Stata	Yes	-/-
R/RStudio	Yes	Yes
QGIS	-/-	Yes
grass	-/-	Yes
PostGis	-/-	Yes
LibreOffice	Yes	Yes
Emacs	Yes	Yes
gnome-text-editor	Yes	Yes
Nautilus (File manager)	Yes	Yes

Using SOEP data and your own data

- The latest version of the SOEP data is available at the following address directory path:

HAUSER

~/soep-data/ or /import/SOEP-Regio/data/

MORAN

~/soep-data/ or /import/SOEP-GIS/data

- You can store your own data and scripts in your personal home directory. ~/work/

Logging out

- Use the icon in the upper right corner



- click on your username and on logout.

6.11.4 Importing Scripts or External Data

- You can send these data to the RDC team before your stay. Send it to SOEP. Please use the following website: cs-soep.diw.de
- Before you send us your files (only data files, text files and tables), please put all files into a zip archive and name it as **user-YYYY-MM-DD.zip** (mustermann-2020-12-24). Please do not send ados, binaries or r packages in the zip file, ados or r packages will be installed centrally by the SOEP team.
- As receiver for the data and scripts please use soepmail@diw.de.
- Before you come to us, please send us the data to import early (2 days in advance) enough so that we have enough time to install it.
- You will be able to find and use your imported data here: /home/USER/transfer/import/
- You can read, write and save in your personal directory: /home/USER/work/

Attention:

Because disk space is limited, we had to introduce the concept of [quotas](#):

- each user gets 10 GB of disk space
- to display there is the [quota command](#)

- the data remains on the server until the end of the project duration.
- after the end of the project, the data is taken from the server and archived for 10 years.
- it is possible to upload the data to the server again later with sufficient preparation time

6.11.5 Instructions for exporting from Hauser to user

From a secure guest workstation at the SOEP Research Data Center, users can analyze SOEP data in combination with small-scale regional data. However, to provide users with this sensitive information, we have to carry out additional protective measures of both a technical and organizational nature. At a guest workstation at the SOEP Research Data Center, you work on a thin client from which you cannot export any data on your own. Below we describe how you can obtain the results of your analyses after they have been checked for anonymity.

How can I take my results with me?

In your transfer folder, you will find an import folder (containing your external data that have been imported into the system) and an export folder.

1. Create on the server ‘Hauser’ below directory ‘~/transfer/export/’ a new subdirectory with a name as the current date in ISO-format:

```
mkdir /home/USER/transfer/export/yyyy-mm-dd
```

Eg.: You are user Jane Doe on the server ‘Hauser’ and today is February 29, 2021

```
jdoe@Hauser mkdir /home/jdoe/transfer/export/2021-02-29
```

We know there wasn’t a February 29th in 2021, but that’s just a format example

2. This folder should contain the following

1. The results that you definitely need to take with you (for formal criteria, see below)
2. A README file (as a .txt file, Word file, or Libre Office file) in which you briefly describe each file in the export folder
3. Please make sure that the README file is readable and that **line breaks** are used

3. Check your files

- Before you make a request for an export please check your data structure with the `OutputControl` command.
- Execute this command in the export folder you want to export.
- This command is used to check whether the formal requirements of the files are met (more information in the chapter “Formal criteria for exporting files”).

Change to the folder you want to export

```
cd /home/USER/transfer/export/yyyy-mm-dd
```

In the Terminal, enter the following command: `OutputControl`

Check the `Control_output_USER_YYYY_MM_DD.txt` output in the folder `Control` in your export folder.

- If you want to make a request for an export, the control file should not contain any warnings.
- If you have any questions, please contact the SOEP hotline.

4. When your folder is complete, please send an e-mail to soepmail@diw.de with your export request

Before you submit an export request, please check that your export is complete and ensure that the following criteria have been met:

Attention: Please read the following rules carefully. If you break the rules, you will not receive your export files.

Formal criteria for exporting files:

- Microdata sets at household or personal level will NOT be exported.
- Only the outputs of analysis (tables, figures), syntax files, and log files will be exported:
- **Tables:**
 - must be stored in the file format **.csv**
 - the maximum number of text files and tables is **200**
- **Figures:**
 - must be saved in one of the following file formats: **.png, .svg, .jpg, .tiff, .eps, .pdf**
- **Text files (scripts or log files):**
 - must be stored in one of the following file formats: **.txt, .tex, .do, .r, .pdf, .log, .md**
 - may have a maximum of **25,000 lines** (a command to count these from a terminal for all .log files in a directory is `wc -l *.log`)
 - **the maximum number of text files and tables is 200**
 - * Please make sure that the files are readable and that **line breaks** are used
- Please note that **no special characters or spaces** are used in file names. Please check if the files are really **readable** after creating.
- An export request can only be made **once a week**

Criteria for exporting results:

- In principle, the results cannot allow any conclusions to be drawn as to which spatial planning region (or smaller-scale geographic unit) a household or individual was or is part of.
- No regional information (e.g., municipality code, district code, zip code ...) may be listed (e.g., using the list command in Stata) together with identifiers (e.g., individual ID number, household ID number)
- When creating tables and figures, the minimum cell population must be kept at 10 if region-specific characteristics are used.

Additional notes on export:

- Since the export has to be checked manually, checking can take up to two or more weeks, of course depending on the number of files to be checked.
- The export link sent to you will only be available for a specified period of time (at least two weeks).
- To open the export link, use your guest access password.

6.11.6 Data transfer from Moran to Hauser

From the three servers of “SOEPgeo”, or the SOEP Research Data Center’s guest network, SOEP users can analyze geocoded data for scientific purposes on site at the SOEP Research Data Center. Researchers are first required to sign a data protection agreement, and a complete record is kept of all data access. The concept is to keep the geo-coordinates of SOEP households separate from the actual survey information throughout the entire process of analysis by data users. Only the coordinates and the survey year are needed generate topic-related indicators in a geographic information system (GIS; grass, qgis, and postgis are installed on Moran) or in the statistical package R, and no further information on either the household or household members. SOEP Research Data Center staff transfer indicators generated by users in a GIS. This prevents any possibility of users accessing the data. The key component of the data protection concept is that SOEP households’ geo-coordinates are kept separate from the survey information:

- At no time do data users have simultaneous access to coordinates and survey data
- Data users can only generate topic-related indicators on Moran, where the SOEP survey data are not accessible.
- Data users can only analyze the topic-related indicators on Hauser, where the SOEP-household coordinates are not accessible.
- Topic-related indicators that were generated based on household coordinates may only be analyzed on the Hauser server and may not be exported from there.

The data user therefore has no simultaneous access to the SOEP survey data and the geo-coordinates of SOEP households. The results (exported Hauser results) may only be published in completely anonymous form.

How do I initiate data transfer from Moran to Hauser?

Attention: Please read the following rules carefully. If you break the rules, the data transfer cannot be executed.

Steps to initiate data export by the SOEP Research Data Center:

1. Create a subdirectory in the export folder on Moran with the export date: `mkdir /home/USER/transfer/export/yyyy-mm-dd`
2. This folder should contain both the dataset to be exported and a corresponding *README.csv*:
 - dataset with generated indicators and ID (see below)
 - *README.csv* (see below)
3. Send an e-mail to soepmail@diw.de with the following information:
 - What input dataset was used for the coordinates? To ensure correct data transfer, we need to know what version of the data was used (e.g., v35)
 - What is the export file format? (.rds, .shp, .csv are permitted) (to save in dataframe in rds format please use `saveRDS()`)
 - What are the unique identifiers for the dataset? (e.g., ID & year)

Formal criteria for data transfer:

The following criteria apply to exports:

- The *README.csv* is a two-column .csv table
 - \$name: column containing the variable names of the indicators to be exported (e.g., distance)
 - \$description: short description of the respective variable (e.g., distance in meters to the next flood point for household i in year t (for the flood in 2002))
- The following applies to the dataset containing the indicators to be exported

- Dataset must have the column/variable ID from the input dataset used
- Permissible file formats: rds, shp, csv
- Dataset otherwise only contains the indicators described in the README.csv file

Additional notes on data transfer:

- After the data transfer has taken place, the output (datasets, transfer scripts) will be stored in your transfer folder on Hauser, in a subdirectory of your import folder that is identified by date (/home/USER/transfer/import/fromMoran/yyyy-mm-dd)

Section author: Jan Goebel <jgoebel@diw.de>

Last change: Jun 01, 2023

If you want to import the SOEP data as csv files with an older version of Stata, this exercise will help you.

6.12 Working with SOEP data in csv format

SOEP offers the data in statistical program specific file formats (e.g.: Stata .dta) and also as comma-separated values File (csv). With these csvs you can read the non-formatted information directly into a statistical program of your choice.

This example shows how to open SOEP data of data version v.36 in csv format with an old Stata version (12) and how to prepare the data in an efficient way.

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/output
- H:/material/exercises/temp
- H:/material/exercises/log

These are used to store your script, log files, datasets, and temporary datasets. Open an empty do-file and define the paths you created with globals:

```
1 *****
2 * Set relative paths to the working directory
3 *****
4 global AVZ "H:\material\exercises"
5 global MY_IN_PATH "\\hume\rdc-prod\distribution\soep-core\soep.v35\csv"
6 global MY_DO_FILES "$AVZ\do\"
7 global MY_LOG_OUT "$AVZ\log\"
8 global MY_OUT_DATA "$AVZ\output\"
9 global MY_OUT_TEMP "$AVZ\temp\"
```


The global “AVZ” defines the main path. The main paths are subdivided using the globals “MY_IN_PATH”, “MY_DO_FILES”, “MY_LOG_OUT”, “MY_OUT_DATA”, “MY_OUT_TEMP”. The global “MY_IN_PATH” contains the path to your ordered data.

For the following script to work, the global “MY_IN_PATH” must contain the folder path to the SOEP csv files of all datasets. The csv files for each data set should always consist of three csvs. If we want to import and prepare the dataset jugendl in csv format, we need the following csv Files:

- jugendl.csv
- jugendl_variables.csv
- jugendl_values.csv

In the SOEP, the csv of each data set contains the variables as columns and their numerical values. Variables and Values csvs contain the variable labels and the value labels for the data set. First some packages for Stata have to be installed so that the process can start.

```

1      * Import and Labeling of SOEP csv-Files
2      clear
3      set more off
4
5      * Load ados
6      capture which adolist
7      if _rc==111{
8          ssc install adolist
9      }
10     quietly adolist list
11     local allAdos `r(names)'
12     foreach package in fre labutil2 chardef labundef saveascii useold {
13         if !regexm("`r(names)'" , " `package' ") {
14             display as result "Paket " as error "`package'" as result " wird
↵versucht über SSC-Server zu installieren"
15             ssc install `package'
16         }
17     }

```

Once the packages are installed, you will need to define the following functions to be able to label your dataset later. We define the function soeplabelsvars for linking the variables to the variable labels.

```

1      * Assign German variable labels from *_variables.csv
2      capture program drop soeplabelsvars
3      program soeplabelsvars
4      version 12
5      syntax , varlabels(string)
6      preserve
7      insheet using "`varlabels'", clear names
8      putmata varLab = (variable label_de) ,replace
9      restore
10     foreach variable of varlist * {
11         label variable `variable' ""
12     }
13
14     mata: st_local("n", strofreal(rows(varLab)))
15     forvalues i = 1/\`n' {
16         mata: st_local("varName",varLab[`i',1])

```

(continues on next page)

(continued from previous page)

```

17      mata: st_local("varLabel",varLab[`i',2])
18      capture confirm variable `varName'
19      if !_rc {
20      di "Variable: `varName' mit -`varLabel'- gelabelt"
21      label variable `varName' "`varLabel'"
22      }
23      else di "Variable " as error "`varName'" as result " nicht vorhanden"
24      }
25      end

```

The soeplabelvals function links the information in the data set with valuelabels.

```

1      * Assign German value labels from *_values.csv
2      capture program drop soeplabelsvals
3      program soeplabelsvals
4      version 12
5      syntax , vallabels(string)
6      quietly label drop _all
7      quietly labundef , detach
8      preserve
9      insheet using "`vallabels'", clear names
10     quietly tostring value, replace
11     putmata valLab = (variable value label_de) ,replace
12     quietly levelsof variable, local(variables) clean
13     restore
14     foreach variable in `variables' {
15     di "-----"
16     di "Variable `variable' wird gelabelt"
17     mata: valLabVar= select(valLab, valLab[.,1]=="`variable'")
18     mata: st_vlmodify("`variable'", strtoreal(valLabVar[.,2]),valLabVar[.,
19     ↪3])
20     capture confirm variable `variable'
21     if !_rc label value `variable' `variable' , nofix
22     else di "Variable " as error "`variable'" as result " nicht vorhanden"
23     }
24     end

```

After both functions have been loaded we can define in a local the dataset we want to import and prepare as csv. We load the csv via the insheet command. Then we use the defined functions and use the variables.csv and values.csv provided by SOEP to label the data.

```

1      * import and label dataset
2      local dataset = "jugendl"
3      insheet using "$MY_IN_PATH/`dataset'.csv", clear names
4      soeplabelsvars, varlabels("$MY_IN_PATH/`dataset' _variables.csv")
5      soeplabelsvals, vallabels("$MY_IN_PATH/`dataset' _values.csv")

```

Congratulations you should now have a fully labeled dataset!

Last change: Mar 28, 2023

Combining datasets is not always as easy as it seems. Here are some example on merging SOEP datasets in STATA

6.13 How to Merge SOEP Data in Stata

This example focuses on merging one or more datasets for further analysis. In general, there are several ways to combine datasets:

- add some extra characteristics to observations using the *merge* command
- add some extra observations using the *append* command
- create all pair combinations within the groups using the *joinby* command

In this chapter we will deal with the commands **merge** and **joinby**.

Merging two datasets requires that both have at least one variable in common (either string or numeric). This is called the key variable (for more information see [Dataset Identifiers](#)). Make sure that the key variables have the same notation and the same name (e.g. Person ID etc.), otherwise we will receive an error message. Examine each dataset separately before merging them. Make sure that you use all possible key variables. Depending on the data format, SOEP datasets usually have **one or two key variables** which can be found in the section [Tracking Data](#) in the column identifier (ID).

merge steps

Basically, for the merge you need three steps.

- Open the desired dataset which is called the **master dataset**.
- Add the second dataset which is called the **using dataset**.
- Save the new generated dataset.

Type *help merge* in the STATA command line for details.

types of merges

There are three types of match merges that we frequently use.

merge 1:1

With the one-to-one merge, one observation from one dataset is matched to one observation in the other dataset. Therefore, the data are at the same level of analysis e.g. individuals to individuals. Use a unique key variable in both the master and the using datasets to merge files. Use *isid* in the STATA command line to find out if a variable is unique.

merge 1:m or merge m:1

In one-to-many and many-to-one merges, one observation from one dataset is matched to many observations in the other dataset. If the master dataset has many observations to match with the single observation in the using dataset, we use *m:1*; or we use *1:m* if it is the using dataset that has many to match. e.g. households to individuals or individuals to households. One-to-many or many-to-one merges are most frequently met when dealing with hierarchical data.

_merge

After the merge the variable *_merge* will be automatically created. It tells us how the observations have been matched. Usually, the *_merge* variable has three values:

- *_merge* == 1, observation appeared in the master file only
- *_merge* == 2, observation appeared in the using file only
- *_merge* == 3, observation appeared in both files

The *_merge* variable must be dropped or renamed before we perform the next merge.

6.13.1 1:1 merge - one-to-one on key variables

Compare the life satisfaction of young people between 16 and 17 years old, who were born in the German Federal States of Berlin and Hamburg.

Create an exercise path with four subfolders:

 do	07.05.2018 16:02	Dateiordner
 log	12.04.2018 10:06	Dateiordner
 output	21.06.2018 13:14	Dateiordner
 temp	21.06.2018 13:14	Dateiordner

Example:

- H:/material/exercises/do
- H:/material/exercises/log
- H:/material/exercises/output
- H:/material/exercises/temp

Open an empty do-file and define your paths with globals. Globals are useful to import and export data.

```

1      global AVZ "H:/Merge-Übung/"
2      global MY_IN_PATH "//hume/rdc-prod/distribution/soep-core/soep.v37/eu/
↪Stata/"
3
4      global MY_DO_FILES "$AVZ/do/"
5      global MY_LOG_OUT "$AVZ/log/"
6      global MY_OUT_DATA "$AVZ/output/"
       global MY_OUT_TEMP "$AVZ/temp/"

```

First, open the desired master dataset PPATHL.

Attention: As a master data set, one uses the PPATHL or HPATHL Tracking File. The PPATHL dataset contains information on all individuals who have ever lived in a SOEP household at the point in time of a survey. The HPATHL dataset contains information on all households that have ever participated in the SOEP survey at any point in time. For more information about the data structure, see *Datasets SOEP-Core*.

Merge the using dataset *jugendl*.

Both datasets have the same level of analysis - individuals to individuals. To be able to merge both datasets, you need two identifiers to make a row in a dataset unique: pid and syear. The variable pid is the unchangeable person number. The variable syear is the year of the survey.

```

1      use "${MY_IN_PATH}ppathl.dta", clear
2      merge 1:1 pid syear using "${MY_IN_PATH}jugendl.dta"

```

Result	# of obs.	
not matched	1,217,979	
from master	1,217,979	(<code>_merge==1</code>)
from using	0	(<code>_merge==2</code>)
matched	9,370	(<code>_merge==3</code>)

In the next step keep only the observations that are found in both datasets (`_merge ==3`). Delete the newly generated variable `_merge`.

```
1      keep if _merge==3
2      drop _merge
```

To answer the question we look at the variable `migback`, which gives the information whether an individual has a migration background or not. In this case we are looking for individuals who have no or an indirect migration background because the both groups were born in Germany.

```
1      tab migback
```

Migration background	Freq.	Percent	Cum.
[1] no migration background	6,769	72.24	72.24
[2] direct migration background	908	9.69	81.93
[3] indirect migration background	1,693	18.07	100.00
Total	9,370	100.00	

We need the values 1 and 3 for the further evaluation.

Then we look at the variable that has information about the birthplace (German Federal States). Limit the variable to the individuals who have no or an indirect migration background.

```
1      tab birthregion if migback==1 | migback==3
```

Birth place: German Federal Land	Freq.	Percent	Cum.
[-2] Does not apply	3,489	41.23	41.23
[-1] No Answer	326	3.85	45.08
[1] Schleswig-Holstein	137	1.62	46.70
[2] Hamburg	75	0.89	47.59
[3] Lower Saxony	458	5.41	53.00
[4] Bremen	53	0.63	53.63
[5] North Rhine-Westphalia	995	11.76	65.39
[6] Hesse	320	3.78	69.17
[7] Rhineland-Palatinate	221	2.61	71.78
[8] Baden-Wuerttemberg	576	6.81	78.59
[9] Bavaria	730	8.63	87.21
[10] Saarland	45	0.53	87.75
[11] Berlin	172	2.03	89.78
[12] Brandenburg	168	1.99	91.76
[13] Mecklenburg-West Pomerania	99	1.17	92.93
[14] Saxony	294	3.47	96.41
[15] Saxony-Anhalt	155	1.83	98.24
[16] Thuringia	149	1.76	100.00
Total	8,462	100.00	

Now we know that Berlin has the value 11 and Hamburg has the value 2.

The variable j10392 shows the life satisfaction on the scale from 0 to 10 where 10 is very satisfied and 0 is very unsatisfied. With the command sum find the mean value of the life satisfaction among the young people born in Berlin. Limit this to life satisfaction from 0 to 10 with no or an indirect migration background.

```
sum j10392 if birthregion==11 & j10392<=10 & j10392>=0 & ( migback==1 |
↵migback==3 )
```

Variable	Obs	Mean	Std. Dev.	Min	Max
j10392	149	7.38255	1.922826	0	10

The mean value of life satisfaction in Berlin is 7.4.

Do the same for Hamburg. With the command sum find the mean value of the life satisfaction among the young people born in Hamburg. Limit this to life satisfaction from 0 to 10 with no or an indirect migration background.

```
sum j10392 if birthregion==2 & j10392<=10 & j10392>=0 & ( migback==1 |
↵migback==3 )
```

Variable	Obs	Mean	Std. Dev.	Min	Max
j10392	70	7.842857	1.440971	3	10

The mean value of life satisfaction in Hamburg is 7.8.

Lastly, the mean value of life satisfaction in Hamburg is 0.4 points higher than in Berlin.

6.13.2 1:m merge - one-to-many on key variables

Determined how life satisfaction depends on the household size in 2019.

First, open the desired master dataset HPATHL, which contains the households. Merge the using dataset PEQUIV, which is based on the Cross-National Equivalent File (CNEF) with extended income information for the SOEP. Because of different levels of analysis it is a 1:m merge. For this case, you need two key variables, hid and syear. The variable hid is the household number and the variable syear is the year of the survey. The option *keep(match)* keeps only the observations obtained in both datasets (or *_merge==3*). The option *nogenerate* suppresses the generation of the variable *_merge*.

```
1 use "${MY_IN_PATH}hpathl.dta", clear
2 merge 1:m hid syear using "${MY_IN_PATH}pequiv.dta", keep(match)
   ↪nogenerate
```

Result	# of obs.
not matched	0
matched	1,066,546

We have 1,066,546 merged observations.

Keep the observations for the survey year 2019.

```
1 keep if syear==2019
```

Display the variable that shows the number of persons in a household.

```
1 tab d11106
```

Number of Persons in HH	Freq.	Percent	Cum.
1	4,973	9.97	9.97
2	12,622	25.30	35.27
3	8,130	16.30	51.57
4	10,848	21.74	73.31
5	6,839	13.71	87.02
6	3,216	6.45	93.47
7	1,694	3.40	96.86
8	768	1.54	98.40
9	405	0.81	99.21
10	260	0.52	99.73
11	55	0.11	99.84
12	36	0.07	99.92
13	26	0.05	99.97
16	16	0.03	100.00
Total	49,888	100.00	

Display the life satisfaction and limit the variable from 0 to 10 where 10 is very satisfied and 0 is very dissatisfied.

```
tab p11101 if p11101>=0 & p11101<=10
```

Overall life satisfaction	Freq.	Percent	Cum.
[0] Completely dissatisfied	100	0.34	0.34
1	90	0.30	0.64
2	266	0.89	1.53
3	528	1.77	3.31
4	583	1.96	5.27
5	2,182	7.33	12.60
6	2,227	7.48	20.08
7	5,517	18.54	38.63
8	10,001	33.61	72.24
9	5,670	19.06	91.29
[10] Completely satisfied	2,591	8.71	100.00
Total	29,755	100.00	

Sort and sum the dataset by household size. Limit the variable life satisfaction from 0 to 10.

```
bysort d11106:sum p11101 if p11101>=0 & p11101<=10 & d11106==d11106
```

```
-> d11106 = 10
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	48	7.916667	1.772525	2	10

```
-> d11106 = 11
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	19	8.368421	1.60591	6	10

```
-> d11106 = 12
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	5	7.4	1.341641	6	9

```
-> d11106 = 13
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	3	8.333333	.5773503	8	9

```
-> d11106 = 16
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	2	6.5	.7071068	6	7

With 8.4, the highest mean score for life satisfaction is among households with 11 people. With 6.5, the lowest mean score for life satisfaction is among households with 16 people. This means that households with 11 household members are more satisfied than households with 16 people.

After the analysis save the dataset. We need it for the next exercise.

```
1 save "${MY_OUT_DATA}hgendata.dta", replace
```

6.13.3 m:1 merge – many-to-one on key variables

Determine the extent to which life satisfaction in 2019 depends on whether the person is a main tenant, subtenant, owner or lives in a nursing home.

First, open the desired master dataset we just generated. Merge the using dataset HL, which includes all variables of the household questionnaire over time. Because of different levels of analysis it is a m:1 merge. Use two key variables hid and syear. The option keep(match) keeps only the observations obtained in both datasets (or _merge==3). The option nogenerate suppresses the generation of the variable _merge.

```
1 use "${MY_OUT_DATA}hgendata.dta", clear
2 merge m:1 hid syear using "${MY_IN_PATH}hl.dta", keep(match) nogenerate
```

Result	# of obs.
not matched	0
matched	49,888

We have 49,888 merged observations.

Display the variable life satisfaction limited to the 0 to 10 where 10 is very satisfied and 0 is very dissatisfied.

```
1 tab p11101 if p11101>=0 & p11101<=10
```

Overall life satisfaction	Freq.	Percent	Cum.
[0] Completely dissatisfied	100	0.34	0.34
1	90	0.30	0.64
2	266	0.89	1.53
3	528	1.77	3.31
4	583	1.96	5.27
5	2,182	7.33	12.60
6	2,227	7.48	20.08
7	5,517	18.54	38.63
8	10,001	33.61	72.24
9	5,670	19.06	91.29
[10] Completely satisfied	2,591	8.71	100.00
Total	29,755	100.00	

The variable hlf0001_h shows whether the individuals renting, leasing or owning the apartment or lives in a retirement home. Limit the variable to the values from 1 to 4.

```
1 tab hlf0001_h if hlf0001_h>=1 & hlf0001_h<=4
```

[de] Haupt-, Untermieter, Eigentuerer [harmonisiert]	Freq.	Percent	Cum.
[1] Main Tenant	26,619	54.41	54.41
[2] Sub-Tenant	1,028	2.10	56.52
[3] Owner	21,207	43.35	99.87
[4] Nursing Home/ Retirement Community	65	0.13	100.00
Total	48,919	100.00	

Sum the life satisfaction for each element of the variable hlf0001_h.

```
1   foreach hlf0001_h in 1 2 3 4 {
2       sum p11101 if p11101>=0 & p11101<=10 & hlf0001_h==`hlf0001_h'
3   }
```

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	15,190	7.368861	1.776065	0	10

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	582	7.21134	1.952537	0	10

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	13,454	7.799539	1.520111	0	10

Variable	Obs	Mean	Std. Dev.	Min	Max
p11101	60	6.716667	2.436249	0	10

Owners of apartments have the highest life satisfaction score with a mean of 7.8, while people living in nursing homes have the lowest life satisfaction score with a mean of 6.7.

After the analysis save the dataset.

```
1   save "${MY_OUT_DATA}hlgendata.dta", replace
```

6.13.4 joinby

m:m specifies a many-to-many merge and is not a good idea. In an m:m merge, observations are matched within equal values of the key variable(s), with the first observation being matched to the first, the second, to the second, and so on. If the master and using datasets have an unequal number of observations within the group, then the last observation of the shorter group is used repeatedly to match with subsequent observations of the longer group. Thus m:m merges are dependent on the current sort order — something which should never happen. That is why you use a **joinby** command. **joinby** is similar to merge but forms all combinations of the observations where it makes sense.

Consider the joinby command in the context of an example. For this we took segments from two datasets of the SOEP: ARTKALEN and BIOMARSM. Both are spell data. The following example is taken from the documentation “Working with spell data”, which can be downloaded here [Spell Data](#)

Open an empty do-file and define your paths with globals. Globals are useful to import and export data.

```
1 global path "H:\Merge-Übung\examples\"
2 global MY_DO_FILE "${path}example_do_files\"
3 global MY_IN_PATH "${path}example_input_data\spell_to_spell\"
4 global MY_OUT_PATH "${path}example_output_data\"
5 global MY_TEMP_PATH "${path}example_temp_data\"
```

Our goal is to enrich information of one spell dataset by introducing information from the other dataset to model processes over time of two kinds. In that case, employment trajectories and their effect on the transition into marriage. We are dealing with two variables: employment status and marital status.

The ARTKALEN dataset consists of 7 variables and 3 observations. We observe only one person.

	begin	end	spelltyp	spellnr	zensor	hhnr	persnr
1	[36] 1985 Dec	[39] 1986 Mar	[4] Vocational Training	2	[4] Left Censored	1	1
2	[36] 1985 Dec	[39] 1986 Mar	[3] Part-Time Employment / Marginally Employed	4	[4] Left Censored	1	1
3	[40] 1986 Apr	[43] 1986 Jul	[1] Full-Time Employ	5	[5] Left And Right Censored	1	1

The BIOMARSM dataset consists of 7 variables and 2 observations. We also observe only one person.

	begin	end	spelltyp	spellnr	censor	hhnr	persnr
1	[36] 1985 Dec	[40] 1986 Apr	[1] Single	1	[2] LZ: After Gap	1	1
2	[41] 1986 May	[44] 1986 Aug	[2] Married	2	[1] LZ: Missing	1	1

Before combining spell datasets, we prepare the two datasets separately. The preparation of both spell datasets, the master and using one, follow the exact same structure: rename variables (spelltyp to employment status and marital status), unfold each spell into subspells of duration of a single month, and delete variables we don't need.

After preparing the dataset, ARTKALEN consists of 14 variables and 12 observations.

	persnr	begin	end	employment_status	hhnr	left	right	org_censor	spe~1	b~1~n	end~n	fir~1	_sp~n	spe~n
1	1	[36] 1985 Dec	[37] 1986 Jan	[3] Part-Time Employment ..	1	1	0	[4] Left Censored	4	36	39	1	4	4
2	1	[36] 1985 Dec	[37] 1986 Jan	[4] Vocational Training	1	1	0	[4] Left Censored	2	36	39	2	4	2
3	1	[37] 1986 Jan	[38] 1986 Feb	[3] Part-Time Employment ..	1	1	0	[4] Left Censored	4	36	39	1	4	4
4	1	[37] 1986 Jan	[38] 1986 Feb	[4] Vocational Training	1	1	0	[4] Left Censored	2	36	39	2	4	2
5	1	[38] 1986 Feb	[39] 1986 Mar	[3] Part-Time Employment ..	1	1	0	[4] Left Censored	4	36	39	1	4	4
6	1	[38] 1986 Feb	[39] 1986 Mar	[4] Vocational Training	1	1	0	[4] Left Censored	2	36	39	2	4	2
7	1	[39] 1986 Mar	[40] 1986 Apr	[3] Part-Time Employment ..	1	1	0	[4] Left Censored	4	36	39	1	4	4
8	1	[39] 1986 Mar	[40] 1986 Apr	[4] Vocational Training	1	1	0	[4] Left Censored	2	36	39	2	4	2
9	1	[40] 1986 Apr	[41] 1986 May	[1] Full-Time Employ	1	1	1	[5] Left And Right..	5	40	43	3	4	5
10	1	[41] 1986 May	[42] 1986 Jun	[1] Full-Time Employ	1	1	1	[5] Left And Right..	5	40	43	3	4	5
11	1	[42] 1986 Jun	[43] 1986 Jul	[1] Full-Time Employ	1	1	1	[5] Left And Right..	5	40	43	3	4	5
12	1	[43] 1986 Jul	[44] 1986 Aug	[1] Full-Time Employ	1	1	1	[5] Left And Right..	5	40	43	3	4	5

After preparing the dataset, BIOMARSM consists of 14 variables and 9 observations.

	persnr	begin	end	marital_st~s	_spe~2	org_censor2	spe~2	b~2~n	end~n	hhnr	left2	right2	firs~2	spe~g
1	1	[36] 1985 Dec	[37] 1986 Jan	[1] Single	5	[2] LZ: After Gap	1	36	40	1	2	0	1	1
2	1	[37] 1986 Jan	[38] 1986 Feb	[1] Single	5	[2] LZ: After Gap	1	36	40	1	2	0	1	1
3	1	[38] 1986 Feb	[39] 1986 Mar	[1] Single	5	[2] LZ: After Gap	1	36	40	1	2	0	1	1
4	1	[39] 1986 Mar	[40] 1986 Apr	[1] Single	5	[2] LZ: After Gap	1	36	40	1	2	0	1	1
5	1	[40] 1986 Apr	[41] 1986 May	[1] Single	5	[2] LZ: After Gap	1	36	40	1	2	0	1	1
6	1	[41] 1986 May	[42] 1986 Jun	[2] Married	4	[1] LZ: Missing	2	41	44	1	1	0	2	2
7	1	[42] 1986 Jun	[43] 1986 Jul	[2] Married	4	[1] LZ: Missing	2	41	44	1	1	0	2	2
8	1	[43] 1986 Jul	[44] 1986 Aug	[2] Married	4	[1] LZ: Missing	2	41	44	1	1	0	2	2
9	1	[44] 1986 Aug	[45] 1986 Sep	[2] Married	4	[1] LZ: Missing	2	41	44	1	1	0	2	2

Both datasets are saved and sorted by the unique personal identification numbers and the begin date of each spell. Using both identifiers we combine the datasets.

You can see the time course and status of an individual.

```
1 use ${MY_TEMP_PATH}\spelldata_1.dta, clear
2 joinby persnr begin using ${MY_TEMP_PATH}\spelldata_2.dta, \
  ↪ unmatched(both) update
```

	persnr	begin	end	employment_status	marital_st~s	spel~1	hhnr	left	right	org_censor	b~1~n	e~1~n	fir~1	_sp~n	..	_merge
1	1	[36] 1985 Dec	[37] 1986 Jan	[3] Part-Time Employment ..	[1] Single	4	1	1	0	[4] Left Censored	36	39	1	4	4	in both, master agrees with us
2	1	[36] 1985 Dec	[37] 1986 Jan	[4] Vocational Training	[1] Single	2	1	1	0	[4] Left Censored	36	39	2	4	2	in both, master agrees with us
3	1	[37] 1986 Jan	[38] 1986 Feb	[3] Part-Time Employment ..	[1] Single	4	1	1	0	[4] Left Censored	36	39	1	4	4	in both, master agrees with us
4	1	[37] 1986 Jan	[38] 1986 Feb	[4] Vocational Training	[1] Single	2	1	1	0	[4] Left Censored	36	39	2	4	2	in both, master agrees with us
5	1	[38] 1986 Feb	[39] 1986 Mar	[3] Part-Time Employment ..	[1] Single	4	1	1	0	[4] Left Censored	36	39	1	4	4	in both, master agrees with us
6	1	[38] 1986 Feb	[39] 1986 Mar	[4] Vocational Training	[1] Single	2	1	1	0	[4] Left Censored	36	39	2	4	2	in both, master agrees with us
7	1	[39] 1986 Mar	[40] 1986 Apr	[3] Part-Time Employment ..	[1] Single	4	1	1	0	[4] Left Censored	36	39	1	4	4	in both, master agrees with us
8	1	[39] 1986 Mar	[40] 1986 Apr	[4] Vocational Training	[1] Single	2	1	1	0	[4] Left Censored	36	39	2	4	2	in both, master agrees with us
9	1	[40] 1986 Apr	[41] 1986 May	[1] Full-Time Employ	[1] Single	5	1	1	1	[5] Left And Right ..	40	43	3	4	5	in both, master agrees with us
10	1	[41] 1986 May	[42] 1986 Jun	[1] Full-Time Employ	[2] Married	5	1	1	1	[5] Left And Right ..	40	43	3	4	5	in both, master agrees with us
11	1	[42] 1986 Jun	[43] 1986 Jul	[1] Full-Time Employ	[2] Married	5	1	1	1	[5] Left And Right ..	40	43	3	4	5	in both, master agrees with us
12	1	[43] 1986 Jul	[44] 1986 Aug	[1] Full-Time Employ	[2] Married	5	1	1	1	[5] Left And Right ..	40	43	3	4	5	in both, master agrees with us
13	1	[44] 1986 Aug	[45] 1986 Sep	.	[2] Married	.	1	only in using data

As a result, we see a combination of the first two datasets. It is about an observing person and in each row we see his employment status and marital status in a certain period of his life. In this respect we can observe how his employment status changes on the transition to marriage.

If your data has the same structure as the example, you should combine those datasets with the joinby command.

WORKING WITH SOEP DOCUMENTATION

7.1 Variable Search with Questionnaires

If you come across a variable in the dataset whose variable content is unclear, you should always check whether there is a suitable questionnaire for the dataset. Under *Original Core Data* you can see whether the datasets correspond to a survey instrument. The related questionnaires can be found here:

[Questionnaires](#)

Example: Working on a research project, you come across the variable `bjh_16_04` with the German label “Auto: Gründe” (Car: Reasons) and the English label “Reason for No Car in Household”

```
. tab bbh5508
```

Reason For No Car In HH	Freq.	Percent	Cum.
[-5] Not included in this version of th	4,529	26.93	26.93
[-2] Does not apply	9,933	59.06	85.99
[-1] No Answer	167	0.99	86.98
[1] Financial Reasons	871	5.18	92.16
[2] Other Reasons	1,319	7.84	100.00
Total	16,819	100.00	

Unfortunately, it is unclear what exactly this variable represents. You should refer to the questionnaires for the complete question and possible filter instructions.

Example Variable:

`bjh_16_04`: Wave “bj” (Survey Year 2011); household questionnaire (“h”), question number 16, item 4

Open [Questionnaires](#)

The variable “`bjh_16_04`” can be found in the questionnaires for 2019. Select the survey year and questionnaire by using the filter “Year” and “Type of Questionnaire” and download the household questionnaire.

1 results, from 1

K < | 1 >

Type of questionnaire
Year
Study
Language

BEFRAGUNGSTRUMENT

Household (PAPI) 2019 - field version (en)

2019 | SOEP-Core, IAB-SOEP-MIG

Search the variable “bjh_16_04” in the questionnaire.

Since you are already in the correct questionnaire, you must now search for question 16.

16. Which of the following apply to your household?

	Yes	No		If no: is this for financial or other reasons?	
				Financial reasons	Other reasons
There is an Internet connection in the household	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
There are one or more cars in the household.	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
I/we have money set aside for emergencies	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
I/we go away on vacation at least one week a year	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
I/we have friends over for dinner at least once a month.....	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
I/we eat a hot meal with meat, fish, or poultry at least every other day	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
We go out at least once a month for leisure activities like movies, concerts, sporting events, etc.	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
I/we replace furniture that is worn out but still usable with new furniture.....	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
Worn-out clothes are replaced with new ones	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
I/we keep our home comfortably warm in the colder months.....	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
Everyone in the household has a small amount of weekly spending money for his or her own personal use.....	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>
Everyone in the household has at least two pairs of outdoor shoes that fit properly (all-weather shoes included).....	<input type="checkbox"/>	<input type="checkbox"/>	⇒	<input type="checkbox"/>	<input type="checkbox"/>

To understand which information the variable “bjh_16_04” contains, you have to deal with the question. For each answer category, respondents should indicate whether or not the shown items apply to the household. If the item does not apply, respondents must answer an additional question about the reasons. Both questions should be understood as separate variables.

E.g. the variable “bjh_16_01” indicates whether an internet connection is available in the household. The reasons why there is no internet in the household can be found in the variable “bjh_16_02”. The variable “bjh_16_03” shows whether a car is present in the household and the variable “bjh_16_04” shows reasons why no car is present in the

household. By looking into the questionnaire, the variable is now easier to understand. The variable “bjh_16_04” only contains people who do not have a car in their household and shows the reasons given.

Last change: Mar 28, 2023

7.2 Variable Search with paneldata.org

Paneldata.org also allows you to search for variables and to find more information about generated variables. It offers comprehensive frequency counts, chronologies of variables, cross-study variable linkage via concepts, a syntax generator, and a topic list for content search in the SOEP.

Example Variable:

bbh5508: Wave “bb” (Survey Year 2011); household questionnaire (“h”), question number 55, item 8

Open [Paneldata](#)

paneldata.org	Studies ▾	Register / log in	<input type="text" value="Search"/>
-------------------------------	-----------	-----------------------------------	-------------------------------------

NEW: With this version of paneldata.org, you can [register / log in](#) as a user. This enables you to create variable baskets and create scripts for selected studies like SOEP-Core.



SOEP-Core /soep-core
SOEPlong /soep-long
SOEP-IS /soep-is
BASE II /soep-base

Select the study SOEP-Core. The SOEP-Core overview contains important general information about the study, e.g., data access, survey method, questionnaires, themes, terms for missing codes, all available datasets in the study and metadata-based questionnaires. To search for a variable, a dataset, or a publication, simply enter the desired search term in the search bar.

Search

All Variables Concepts Questions Publications Topics

Search: bbh5508


Search: bbh5508 X Clear All

Study

☐ Select all

☐ SOEP-Core (1)

1 results found in 5ms

 **bbh5508** Reason For No Car In HH

Variable in study: SOEP-Core | dataset: bbh | period: 2011 | analysis unit: household

Prev 1 Next

To obtain the desired results, you will need to input specific information. The results window displays all search results. You will see that the variable “bbh5508” originates from SOEP-Core data and can be found in the dataset “bbh” (survey year 2011). If your search is not so specific, you can also search by keywords. We are still interested in the topic “car”.

Search

All Variables Concepts Questions Publications Topics

Search: Car

Search: Car X Clear All

Study

☐ Select all

☐ SOEP-Core (4926)

☐ pairfam (4233)

☐ IAB-SOEP Migration Sample (593)


☐ TwinLife (518)

☐ SOEP-IS (517)


☐ SOEP Pretests (215)

☐ BASE II (36)


10000 results found in 35ms

 **Caring** Alone? Social Capital and the Mental Health of **Caregivers**

Publication by: Lars Thiel (2016)

 Model Event History of **Car** and License Availability: How Accessibility Shapes Acquisition and Disposal

Publication by: Max Bohnet, **Carsten** Gertz (2010)

 The Impact of Informal **Care** and Employment on the Mental Health of the **Caregiver**

Publication by: Andreas Eberl, Sebastian Lang, Katharina Seebaß (2017)

To better limit the 10000 results, the filter options on the left and on the top should be used. We are looking for variables from the “SOEP-Core” datasets. The search results should be limited with the filter options. Which survey years are of interest to me, do I want to work with original data or generated data? For more information about the different datasets in SOEP-Core, see the section [Data Distribution File](#). Should the variable I am looking for be at household level or at individual level?

paneldata.orgStudiesSearch

Register / log in

Search

AllVariablesConceptsQuestionsPublicationsTopics

Car

Search: CarStudy: SOEP-CoreConceptualDataset: Original (raw folder)AnalysisUnit: householdPeriod: 2011Clear All

18 results found in 38ms

Relevance

Study

☐ Select all☒ SOEP-Core (18)

Conceptual Dataset

☐ Select all☒ Original (raw folder) (18)

Analysis Unit

☐ Select all☐ individual (89)☒ household (18)

Period

☐ 1996 (23)☐ 2017 (23)☐ 2013 (20)☐ 2015 (20)

[f11h074a2] Car Acquired

Variable in study: SOEP-Core | dataset: bbh | period: 2011 | analysis unit: household

[bbh5507] Car In HH

Variable in study: SOEP-Core | dataset: bbh | period: 2011 | analysis unit: household

[f11h074a1] Car In HH

Variable in study: SOEP-Core | dataset: bbh | period: 2011 | analysis unit: household

[f11h074a3] Car Acquisition Costs

Variable in study: SOEP-Core | dataset: bbh | period: 2011 | analysis unit: household

[bbh5508] Reason For No Car In HH

Variable in study: SOEP-Core | dataset: bbh | period: 2011 | analysis unit: household

By filtering, the search result is limited to 18 hits, which also shows the variable we searched for. If you click on the variable “bbh5508”, you will find additional information about the variable.

Reason For No Car In HH

PercentHide MissingsWeighted

Reason	Frequency
[2] Other Reasons	1319
[1] Financial Reasons	871
[-1] No Answer	167
[-2] Does not apply	9933
[-3] Answer improbable	0
[-4] Inadmissible multiple res...	0
[-5] Not included in this vers...	4529
[-6] Version of questionnaire ...	0

First you see the weighted absolute frequencies for the variable. It is possible to remove the missing codes from the analysis and/or to display the relative frequencies. Even without opening the dataset, paneldata obtain a good overview of the frequencies of a variable.

354

Chapter 7. Working with SOEP Documentation

SOEP Survey Paper 1261

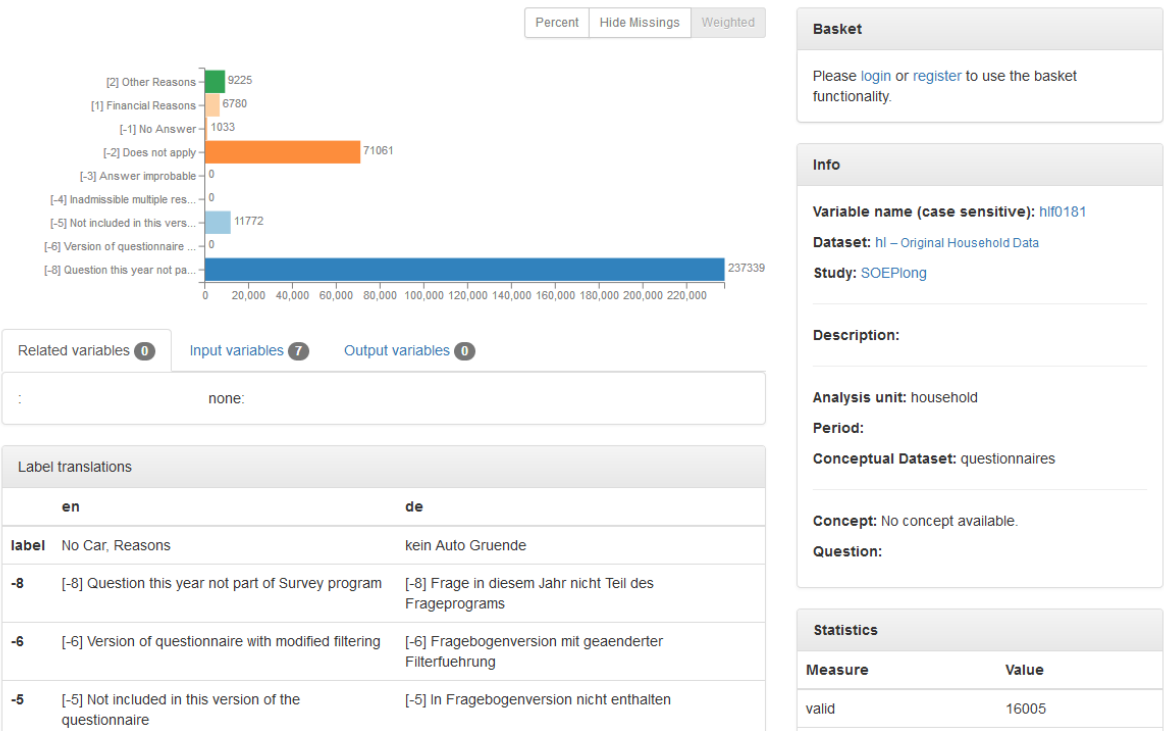
Related variables 7	Input variables 0	Output variables 1	
0:	1984:	1985:	1986:
1987:	1988:	1989:	1990:
1991:	1992:	1993:	1994:
1995:	1996:	1997:	1998:
1999:	2000:	2001: rh/rh5306	2002:
2003: th/th5106	2004:	2005: vh/vh5408	2006:
2007: xh/xh5508	2008:	2009:	2010:
2011: bbh/bbh5508	2012:	2013: bdh/bdh5513	2014:
2015:	2016: bgh/bgh7404	none:	

In the Related Variables section you will also find the chronology of the variable you are looking for. The sample variable was collected in 2001, 2003, 2005, 2007, 2011, 2013. Below the survey year, the name of the variable in the respective year is displayed and can be clicked to access the respective variable page. You can see at a glance when the variable was measured, how often it was measured, and what its name is in the respective survey year.

Related variables 7	Input variables 0	Output variables 1
Soep-Long		
: hl/hlf0181		

In addition, by clicking on “Output variables”, you will find a variable forwarding you to the variable in “long” format. For a more detailed understanding of the long format, read the section *Data Structure in “Long” Format (long)*.

No Car, Reasons



As soon as you click on the “long” variable, you will get to the variable overview for this variable in the long format. The overview of variables does not differ. It can be seen that our example variable “bbh5508” can also be found in long format in the dataset “hl” with the variable “hlf0181”.

Label translations		
	en	de
label	Reason For No Car In HH	Auto: Gruende
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geaenderter Filterfuehrung
-5	[-5] Not included in this version of the questionnaire	[-5] In Fragebogenversion nicht enthalten
-4	[-4] Inadmissible multiple response	[-4] Unzulaessige Mehrfachantwort
-3	[-3] Answer improbable	[-3] nicht valide
-2	[-2] Does not apply	[-2] trifft nicht zu
-1	[-1] No Answer	[-1] keine Angabe
1	[1] Financial Reasons	[1] finanzielle Gruende
2	[2] Other Reasons	[2] andere Gruende

The field “Label translations” shows the value labels of the variables in German and English. In addition, all missing codes used in SOEP are listed and explained.

Label table

The label table provides you with an overview of label definitions across related variables to identify changes over time in longitudinal variables. The first number indicates the value code, the second number (in brackets) represents the frequency in the data. Please note that labels are simplified and values with frequency = 0 are hidden.

Variable:	rh5306	th5106	vh5408	xh5508	bbh5508	bdh5513	bgh7404
Dataset:	rh	th	vh	xh	bbh	bdh	bgh
version of questionnaire with modified filtering	-6 (0)	-6 (0)	-6 (0)	-6 (0)	-6 (0)	-6 (0)	-6 (0)
not included in this version of the questionnaire	-5 (0)	-5 (0)	-5 (0)	-5 (0)	-5 (4529)	-5 (3923)	-5 (3320)
inadmissible multiple response	-4 (0)	-4 (0)	-4 (0)	-4 (0)	-4 (0)	-4 (0)	-4 (0)
answer improbable	-3 (0)	-3 (0)	-3 (0)	-3 (0)	-3 (0)	-3 (0)	-3 (0)
does not apply	-2 (9605)	-2 (9817)	-2 (9249)	-2 (9555)	-2 (9933)	-2 (11230)	-2 (11672)
no answer	-1 (125)	-1 (118)	-1 (211)	-1 (177)	-1 (167)	-1 (136)	-1 (99)
can not afford it	1 (827)						
financial reasons		1 (800)	1 (850)	1 (861)	1 (871)	1 (1310)	1 (1261)
other reasons	2 (1390)	2 (1326)	2 (1130)	2 (1096)	2 (1319)	2 (1494)	2 (1470)

The Label window shows you the absolute frequencies of the variable at different times of data collection. This makes it possible to identify initial trends in how response behavior has changed over a period of time. The assigned value code is output for each possible characteristic value and the absolute frequencies are displayed in parentheses.

In our example, we see that for the variable “th5106” 800 respondents in the wave “t” (2003) stated “financial reasons” as the reason for not having a car in the household. For our variable “bbh5508” in survey year 2011 (wave “bb”), there were 871 respondents.

Paneldata.org is an excellent way to get an first overview of certain variables.

Info

Variable name (case sensitive): [bbh5508](#)

Dataset: [bbh – Household questionnaire](#)

Study: [SOEP-Core](#)

Description:

Analysis unit: household

Period: 2011

Conceptual Dataset: questionnaires

Concept: [Car \(No\) Reasons](#)

Question:

The information box on the right-hand side provides an overview of all relevant information about the variable and the dataset. In addition to basic information, you will find information about what kind of variable you are looking for under “Conceptual Dataset”. In our example “bbh5508”, you can see that variables with a “Conceptual Dataset: org/net” describe original variables that are assigned to a questionnaire. Generated variables are “Conceptual Dataset: gen”. For an overview of the different dataset types in SOEP-Core, see the section [Data Distribution File](#).

In addition to searching for keywords or using the various filter settings, you can also find what you are looking for directly in the data set search. Open [paneldata.org](#), click on the study SOEP-Core and select the menu option “data”.

SOEP-Core
Data
Instruments
Topics
Publications

SOEP-Core

Citation

- **Title:** German Socio-Economic Panel Study (SOEP)
- **DOI:** 10.5684/soep.v32.1
- **Authors:** Jürgen Schupp; Jan Goebel; Martin Kroh; Carsten Schröder; Charlotte Bartels; Klaudia Erhardt; Alexandra Fedorets; Marco Giesselmann; Markus Grabka; Peter Krause; Simon Kühne; David Richter; Rainer Siegers; Paul Schmelzer; Christian Schmitt; Daniel Schnitzlein; Knut Wenzig
- **URL:** <http://dx.doi.org/10.5684/soep.v32.1>

Publications using these data should cite the DOI (doi:10.5684/soep.v32.1) and include one of the following references:

- Gert G. Wagner, Joachim R. Frick, and Jürgen Schupp (2007) The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements, Schmollers Jahrbuch (Journal of Applied Social Science Studies) 127 (1), 139-169 (download)
- Gert G. Wagner, Jan Göbel, Peter Krause, Rainer Pischner, and Ingo Sieber (2008) Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland - Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender), ASTA Wirtschafts- und Sozialstatistisches Archiv 2 (4), 301-328 (download)
- Schupp, Jürgen (2009): 25 Jahre Sozio-oekonomisches Panel - Ein Infrastrukturprojekt der empirischen Sozial- und Wirtschaftsforschung in Deutschland, Zeitschrift für Soziologie 38 (5), pp. 350-357.

Study info

Name: soep-core
Label: SOEP-Core

Now you will be directed to an overview that shows you all datasets contained in SOEP-Core.

SOEP-Core
Data
Instruments
Topics
Publications

Datasets

Show entries
Search:

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

Showing 1 to 10 of 414 entries

Previous
1
2
3
4
5
...
42
Next

Enter the dataset you are looking for (“bbh”) in the search bar at the top right and click on the dataset. You will be directed to an overview that shows you all variables from the “bbh” dataset.

SOEP-Core
Data
Instruments
Topics
Publications

Household questionnaire

Variables

Show 10 entries
Search: bbh5508

Sort Variable
Name

1	Reason For No Car in HH	bbh5508
---	-------------------------	---------

Showing 1 to 1 of 1 entries (filtered from 382 total entries)
Previous 1 Next

Info
Study: soep-core
Release:
Dataset: bbh

Now enter the variable you are looking for in the search bar at the top right and click on the variable of interest. You will be directed to the variable overview, where you will find detailed information on the variable. Paneldata.org offers a variety of search options to fit the user's search needs.

7.3 Topic Search with paneldata.org

To provide an overview of the various topics in the SOEP, the variables have been grouped together on paneldata.org by topic. If you are looking for your research variables and do not want to check all datasets or questionnaires, the topic search on paneldata.org may help.

Open [Paneldata](#) and select the main study SOEP-Core. The upper navigation bar leads you to the Topics area. Click on Topics and look at the list of variables.

paneldata.org
Studies
Search
Register / log in

SOEP-Core
Data
Instruments
Topics
Publications

Languages: de en

Topics

Search node ..

- not assigned!
- > demography and population
- > integration, migration, transnationalization
- > health and care
- > work and employment
- > education and qualification
- > family and social networks
- > attitudes, values, and personality
- > home, amenities, and contributions of private hh
- > time use and environmental behavior
- > income, taxes, and social security
- > survey methodology

Select a topic that corresponds to your research interest, and a more detailed list of sub-topics will appear under the

main topic heading.

paneldata.org Studies Search Register / log in

SOEP-Core Data Instruments Topics Publications

Languages: de en

Topics

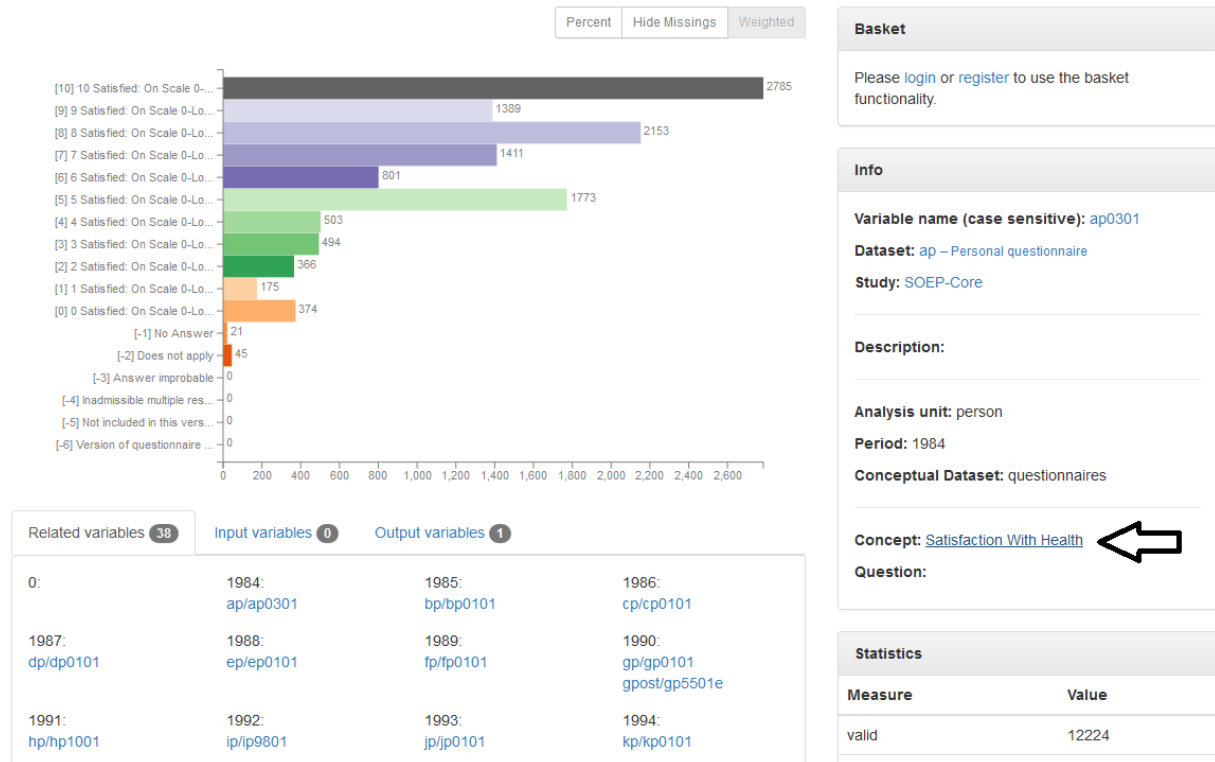
Search node ..

- not assigned!
- > demography and population
- > integration, migration, transnationalization
- > health and care
- > work and employment
- > education and qualification
- > family and social networks
- ▼ **attitudes, values, and personality**
 - > memberships
 - > political orientations
 - > values
 - > personality
- > home, amenities, and contributions of private hh
- > time use and environmental behavior
- > income, taxes, and social security
- > survey methodology

For example, if you are interested in different types of satisfaction, click on the topic “attitudes, values, and personality”. Underneath it, you will find the sub-topic “personality”. Suppose you are interested in health satisfaction. If you have found a suitable sub-topic, click on “show all the related variables”. All variables that fall under this topic will be displayed.

The `paneldata` topic list has three possible functions for each sub-topic. You can display all variables that belong to a sub-topic. In the future, `paneldata` will also display the texts of the questions from the SOEP questionnaires in which the variables in that sub-topic appear. `Paneldata` also allows you to keep variables from a sub-topic in a variable basket. The chapter [Syntax Generator on `paneldata.org`](#) explains in detail how to use the basket in your research and what possibilities this offers. Click on one of the variables to see the variable overview.

Satisfaction With Health



If you click on the concept of a variable, you will get to the concept overview. Concepts in SOEP are used to link variables with the same content. The concepts can even be used to link variables with the same content across studies.

paneldata.org

Studies ▾

Register / log in

Search

Satisfaction With Health

[pzuf01]

Variables and questions

Show

10

 entries

Search:

Study	Object	Label	Path
BASE II	Variable	zufriedenh. gesundheit	/soep-base/data/p2010/pzuf01
BASE II	Variable	Zufriedenheit Gesundheit	/soep-base/data/p2012/pzuf01
BASE II	Variable	Zufriedenheit Gesundheit	/soep-base/data/soep-base-long/pzuf01
IAB-SOEP Migration Sample	Variable	Satisfaction With Health	/iab-soep-mig/data/bdp/bdp0101
IAB-SOEP Migration Sample	Variable	Satisfaction With Health	/iab-soep-mig/data/bep_mig/bepm_p_3001
IAB-SOEP Migration Sample	Variable	Satisfaction With Health	/iab-soep-mig/data/bdp_mig/bdpm_p_17001
IAB-SOEP Migration Sample	Variable	Satisfaction With Health	/iab-soep-mig/data/bfp/bfp0101
IAB-SOEP Migration Sample	Variable	Satisfaction With Health	/iab-soep-mig/data/bep/bep0101

The concept overview displays the study- and wave-specific variables with this concept. The concept allows you to determine whether the variable you are looking for is also available and comparable across studies. In the column “Study” you can see which studies have the same variable linked by concept. The label of the respective variable is also displayed in the “Label” column. The column “path” shows the wave name of the variable. By clicking on the label, you will get to the overview of variables with all of the relevant information. The “Object” column in the concept overview shows you the type of information displayed.

[pzuf01]

Variables and questions

Show 10 entries

Search:

Study	Object	Label	Path
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/bfp/bfp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/fp/fp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/tp/tp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/vp/vp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/bcp/bcp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/kp/kp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/lp/lp0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/bep/bep0101
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/ap/ap0301
SOEP-Core	Variable	Satisfaction With Health	/soep-core/data/yp/yp0101

Showing 31 to 40 of 55 entries

[Previous](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[6](#)
[Next](#)

In addition to the variables linked by concept, you can find the relevant questions in the concept overview. Questions are displayed in the “Object” column with question. Without having to open the questionnaire, you can read the question and identify possible differences. Click on the desired question and you will be taken to the question display.

paneldata.org

Studies

Register / log in

Search

Satisfaction With Health

[pzuf01]

Variables and questions

Show 10 entries

Search:

Study	Object	Label	Path
SOEP-IS	Question	First of all it is about your satisfaction with different areas in your life. How satisfied are you right now with the following areas of your life? How satisfied are you ...	/soep-is/inst/soep-is-2013-a/q59
SOEP-IS	Question	How satisfied are you ...	/soep-is/inst/soep-is-2013-f/q59
SOEP-IS	Question	First of all it is about your satisfaction with different areas in your life. How satisfied are you right now with the following areas of your life? How satisfied are you ...	/soep-is/inst/soep-is-2014-a/q66
SOEP-IS	Question	How satisfied are you	/soep-is/inst/soep-is-2014-f/q66
SOEP-IS	Question	Now we are interested in your satisfaction in certain areas of your life. How satisfied are you currently with the following areas of your life? Please state the level of satisfaction for each area: If you are completely dissatisfied, use the value "0", if you are completely satisfied, use the value "10". You can use the values in between to make your estimate.	/soep-is/inst/soep-is-2015/q85

Showing 51 to 55 of 55 entries

Previous123456Next

paneldata.org

Studies

Register / log in

Search

SOEP-IS

Data

Instruments

Publications

Q52

first of all its is about your satisfaction with different areas in your life. How satisfied are you right now with the following areas of your life? How satisfied are you ...

	0	1	2	3	4	5	6	7	8	9	10	No answer
with your health?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
with your sleep?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Previous question

Next question

Instrument

This question is at position 70 in:
[Questionnaire 2011](#)

Variables

[Satisfaction With Health](#)
[variable: plh0171]
[/soep-is/data/p/plh0171](#)

Attention: To find out the exact wording of the question and possible filter structures, a variable search in the questionnaires is necessary. The question display in [Paneldata](#) only provides a quick overview. In the question overview, you can navigate through the questionnaire using the “next question” and “previous question” buttons. The “Instrument” section shows the position of the question in the questionnaire, the survey year, and links to the metadata-based survey instrument. Click on the survey instrument “Questionnaire 2011”.

[SOEP-IS](#)
[Data](#)
[Instruments](#)
[Publications](#)

Questionnaire 2011 [instrument]

/soep-is/inst/soep-is-2011

Questions

Show entries

Search:

Sort	Question	Name
0	New respondent	q1
1	A000C	a000c
2	Is the respondent the head of household, the person who answers the questions about the household?	q6
3	We'll start with questions about your household as a whole.	q7
4	Did you already live in this flat the last time we interviewed you about a year ago?	q8
5	When did you move into this dwelling?	q9
6	What kind of a house is it in which you live?	q10
7	Is it a rooming house, guesthouse, or a similar accommodation?	q11
8	When, approximately, was the house built in which your flat is located?	q12
9	Can you also provide the exact year in which the house was built?	q13.1

Showing 1 to 10 of 361 entries

[Previous](#)
[1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[...](#)
[37](#)
[Next](#)

Instrument info

Name: soep-is-2011
Label: Questionnaire 2011

The survey instrument used in the SOEP-IS study in 2011 is now displayed. You can navigate through the questionnaire in this overview. The search bar allows you to search for research-relevant terms. Click on the question to access the question display.

Last change: Mar 28, 2023

7.4 Documentation on Generated Data

SOEP-Core contains a wide range of generated variables and datasets. To facilitate data use, we generate a large number of variables in the process of data preparation and release them with the SOEP-Core data. To make the generation process transparent to users, we provide comprehensive documentation on the numerous generated datasets and variables. For an overview, see our [Documentation on Generated Data](#)

Example: A number of frequently used variables are provided in SOEP as “generated variables” (e.g., the datasets \$PGEN and \$HGEN). These variables are checked for consistency across waves. The documentation can be used to answer the following questions:

a) Which variable gives the highest school-leaving certificate attained by individuals surveyed in 2007?

To search for the variable that provides this information, open [Paneldata](#), click on the search button and the tab “Variables”, then enter “school leaving degree” in the search bar. Specify your search by adjusting the filter settings as follows:

- study: soep-core
- Conceptual dataset: Generated (raw folder)

- analysis unit: individual
- period: 2007

All
Variables
Concepts
Questions
Publications
Topics

school leaving degree

Search: school leaving degree
Study: SOEP-Core
AnalysisUnit: individual
Period: 2007, unspecified
ConceptualDataset: Generated (raw folder)
Clear All

53 results found in 45ms
Relevance

Study
Select all
SOEP-Core (53)

Conceptual Dataset
Select all
Original (top level folder)
Original (raw folder) (276)
Generated (top level folder)
Gen. Biographical (top level folder)
Generated (raw folder) (53)
Gross (raw folder) (3)
Gross (top level folder) (3)

Analysis Unit

[xpsbila] School Leaving Degree Outside Germany
Variable in study: SOEP-Core | dataset: xpgen | period: 2007 | analysis unit: individual

[xpsbil] School Leaving Degree
Variable in study: SOEP-Core | dataset: xpgen | period: 2007 | analysis unit: individual

[xpsbilo] School Leaving Degree East Germany
Variable in study: SOEP-Core | dataset: xpgen | period: 2007 | analysis unit: individual

[xpbil01] Vocational Degree Received
Variable in study: SOEP-Core | dataset: xpgen | period: 2007 | analysis unit: individual

All variables could contain the information you are looking for. Since almost all variables in the search result come from the generated “xpgen” dataset, the documentation for the \$pgen dataset should be used. Visit the [Documentation of SOEP-Core Page](#) and enter the search term pgen in the search field.

SOEP
DATA
DATA ACCESS
SERVICES
...

Documentation of SOEP-Core
(SOEP Survey Papers)

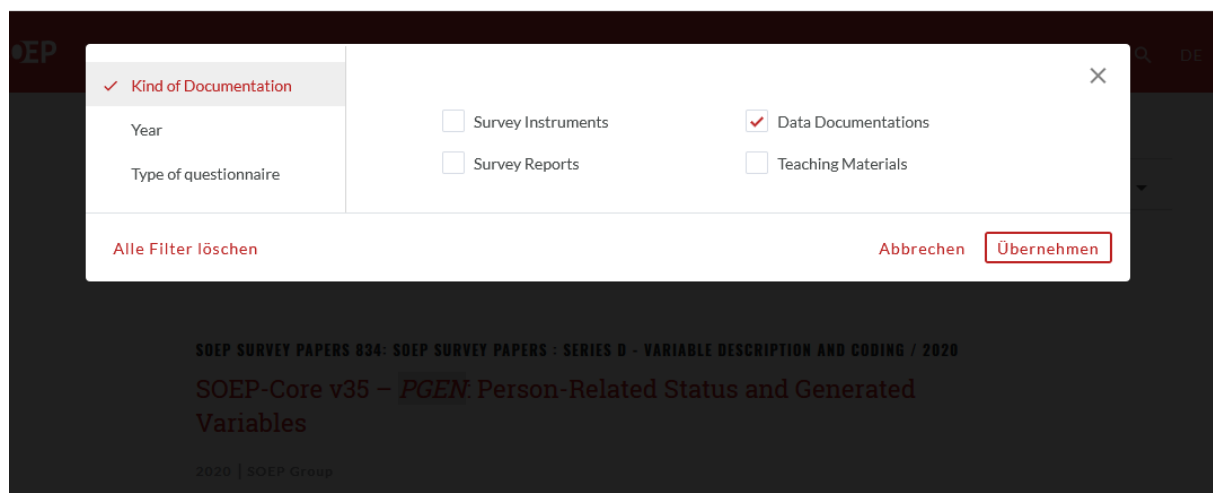
pgen
SUBMIT

HOME | SOEP-CORE | DOCUMENTATION OF SOEP-CORE (SOEP SURVEY ...)
BOOKMARK

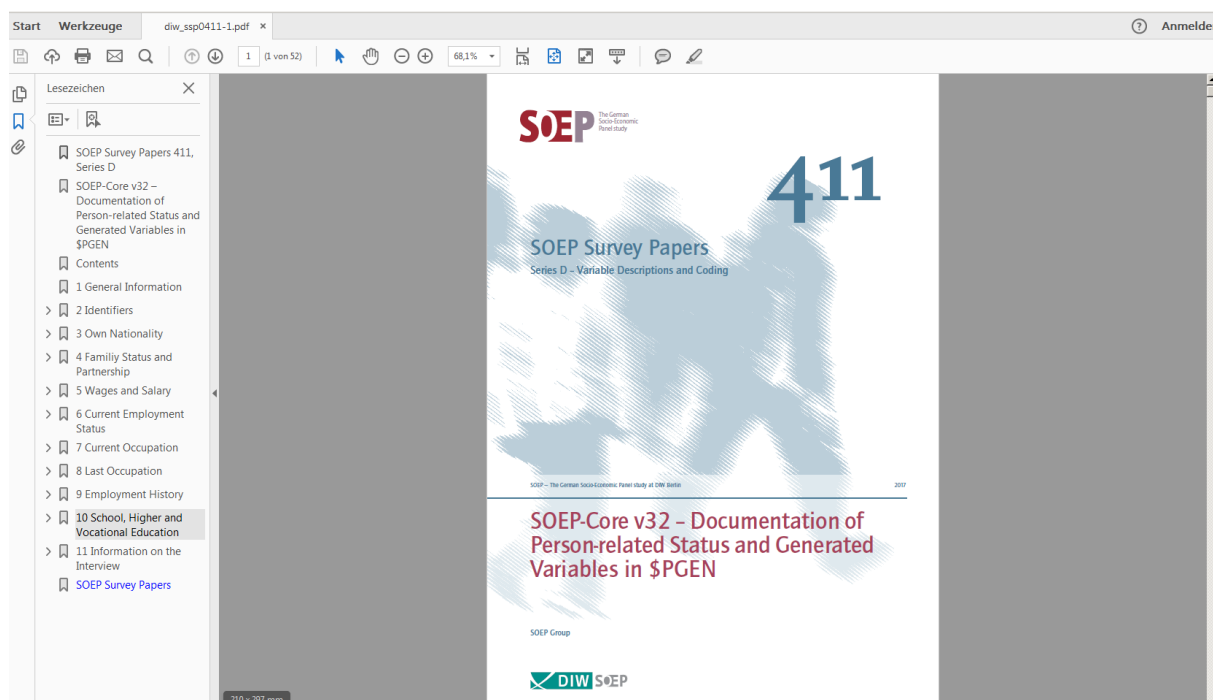
9 results, from 1

Kind of Documentation
Year
Type of questionnaire
Sortierung

Alternatively, you can also use the filters and select “Data Documentations”:



Now select the documentation of the required version of *pgen*



The table of contents on the left gives you a classification of the dataset by topics. To find the variable you are looking for, select topic area 10.

\$psbil – School-Leaving Degree [generic]

1	[1] Secondary School Degree	6411
2	[2] Intermediate School Degree	7293
3	[3] Technical School Degree	1515
4	[4] Upper Secondary Degree	5729
5	[5] Other Degree	4244
6	[6] Dropout, No School Degree	673
7	[7] Currently In School	779
-1	[-1] No Answer	1099
-2	[-2] Does not apply	0
-3	[-3] Answer improbable	0
-4	[-4] Inadmissible multiple response	0
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0

Waves: all

All respondents in all SOEP subsamples are asked about diplomas/degrees attained for completion of secondary/tertiary education (1984–1993 blue questionnaire; since 1994 biographical questionnaire) the first time they participate in SOEP. First: to generate this variable, the different diploma/degree categories provided for Subsamples B and D (see \$PSBILA) as well as C (see \$PSBILO) are integrated into the West German diploma/degree categories (Subsample A) and continued on in this form. Second: this data is regularly updated to take into account any changes in highest diploma/degree attained. With the survey of 2000, all educational information was collected again and is reflected in the variables. [This information can be related to a specific variable and is not necessary generic.]

For more information, contact: Peter Krause (Tel. +49-30-89789-690)

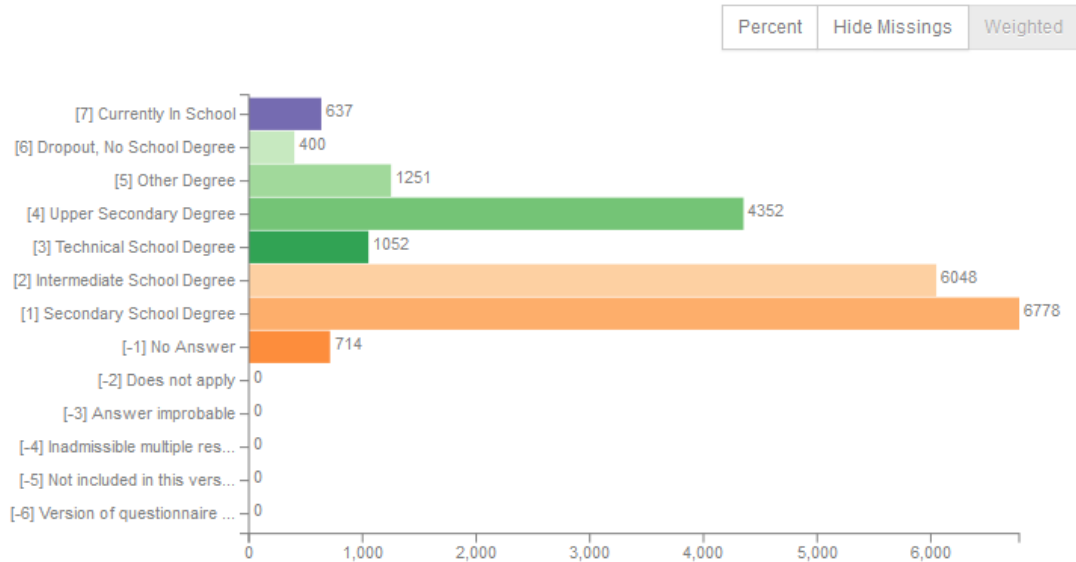
After a few searches, you will find the variable you are looking for. The documentation provides useful information about the generated variable: it comes from the biography questionnaire, which was introduced in 1994 and is administered only once per respondent. The documentation also explains the two additional variables \$psbila and \$psbilo in more detail: the \$psbil variable is updated regularly to take into account possible changes in the respondent's highest school-leaving certificate. For this reason, the generated variable is useful in providing the most up-to-date information on completed secondary schooling.

The variable we are looking for is xpsbil and describes the highest degree in certificate attained by individuals surveyed since 2007.

b) What values do individuals with an upper secondary school-leaving certificate (Abitur) have for this variable??

Since you now know the variable you are looking for, you can use the extensive functions of paneldata.org in addition to the information from the documentation. If you search for the variable “xpsbil” in paneldata.org and click on it, the frequency counts are displayed.

School-Leaving Degree



In addition to the absolute and relative frequencies, you can also read the value codes of specific response categories. A translation of the answer categories can be found in the “Label translations” section:

Label translations		
	en	de
label	School-Leaving Degree	Schulabschluss
-6	[-6] Version of questionnaire with modified filtering	[-6] Fragebogenversion mit geänderter Filterführung
-5	[-5] Not included in this version of the questionnaire	[-5] In Fragebogenversion nicht enthalten
-4	[-4] Inadmissible multiple response	[-4] Unzulässige Mehrfachantwort
-3	[-3] Answer improbable	[-3] nicht valide
-2	[-2] Does not apply	[-2] trifft nicht zu
-1	[-1] No Answer	[-1] keine Angabe
1	[1] Secondary School Degree	[1] Hauptschulabschluss
2	[2] Intermediate School Degree	[2] Realschulabschluss
3	[3] Technical School Degree	[3] Fachhochschulreife
4	[4] Upper Secondary Degree	[4] Abitur
5	[5] Other Degree	[5] Anderer Abschluss
6	[6] Dropout, No School Degree	[6] Ohne Abschluss verlassen
7	[7] Currently In School	[7] Noch kein Abschluss

You can answer the question without opening the data. In the 2007 survey year, the variable “xpsbil” with the value code “4” describes the response category “upper secondary school-leaving certificate (Abitur)”.

Last change: Mar 28, 2023

7.5 Working with SOEPhelp



For data users, the SOEP provides assistance in two analysis programs. One is a stata.ado and the other an R package. The application simply has to be installed in the respective program and helpful information on the desired data sets or variables can be provided. In the following, the installation and use of SOEPhelp is explained:

7.5.1 Working with SOEPhelp in R

SOEPhelp is an R package providing help pages for SOEP-Core data sets (top-level folder only) and their variables. Starting with version v35 SOEPhelp is available for R users, too. Using meta data and the documentation available for SOEP-Core data, this package displays the information on data sets and variables in the default R help. Wherever available the package provides labels in English and in German language. The basic information used to create the help pages is taken from the meta data available from the [Public Core Documentation](#) on git. Help pages are **not** available for data sets in the raw and EU-SILC Clone folders.

Installation

If you are working on a Windows OS, you need to install the Rtools (get them [here](#)) first. The installation can take a little while according to your CPU (between 3 minutes up to over an hour), because the package contains more than 14000 help pages. The most recent version (SOEPhelp_0.37.1.tar.gz) has been build using R 4.2.0.

```
install.packages("http://companion.soep.de/SOEPhelpR/SOEPhelp_0.37.1.tar.  
gz",  
                repos = NULL, type = 'source', quiet = TRUE)
```

Usage

Load the package into your library and read the main page carefully.

```
library(SOEPhelp)  
?SOEPhelp
```

You can get to the help pages by using familiar R help functions like `?` or `help()` as well as `??`.

Example for a dataset

```
?design
```

Asking for the help page of the design data set in R will open the following help page. The title provides you with the basic information that the help page belongs to a data set, its name, and a brief information on what the data set contains. This is followed by the description section providing (if available) further description on the data set and a

link to the paneldata.org page. In the arguments section you will find the list of variables for the data set. The variable names are linked to their help pages and the variable labels are given in English and German (if available), together with the link to the paneldata.org page. The details section gives information on the number of observations and variables. Finally, the notes section refers to the version of the SOEP-Core.

design {SOEPhelp}

R Documentation

Data set: design - Survey design

Description

No details available. Keine Details verfügbar.

For more information on the data set go to <https://paneldata.org/soep-core/data/design>

Arguments

hhnr	[en] Original Household Number [de] Ursprungshaushaltsnummer https://paneldata.org/soep-core/data/design/hhnr
cid	[en] Original Household Number [de] Ursprungshaushaltsnummer https://paneldata.org/soep-core/data/design/cid
hsample	[en] Subsample [de] Stichprobenart https://paneldata.org/soep-core/data/design/hsample
rgroup	[en] Random Groups [de] Random Groups https://paneldata.org/soep-core/data/design/rgroup
design	[en] Inverse Sampling Probability [de] Inverse Ziehungswahrscheinlichkeit https://paneldata.org/soep-core/data/design/design
strat	[en] Stratification Units [de] Schichtung, Stratifizierungseinheiten https://paneldata.org/soep-core/data/design/strat
intid	[en] Interviewer Id [de] Interviewer ID https://paneldata.org/soep-core/data/design/intid
psu	[en] Clusters, Primary Sampling Units [de] Klumpung, Primaere Ziehungseinheiten https://paneldata.org/soep-core/data/design/psu

Details

Survey design: A data frame with 42259 observations on 8 variables.

Note

soep-core - v35

[Package SOEPhelp version 0.1.1 [Index](#)]

Example for a variable

`?sampreg`

Asking for the help page of the `sampreg` variable in R will open the following help page. The title provides you with the basic information that the help page belongs to a variable and its name. This is followed by the description section providing (if available) the variable label in English and German as well as a link to the paneldata.org page. In the arguments section you will find the corresponding values of the variable and their value labels. The following notes section refers to the version of the SOEP-Core. Finally, the see also section lists data sets which contain this variable.

sampreg {SOEPhelp}

R Documentation

Variable: sampreg

Description

[en] Current Sample Region

[de] Aktuelle Stichprobenregion

For more information on this variable go to <https://paneldata.org/soep-core/data/ppath/sampreg>

Arguments

- 2 [en] East-Germany
[de] Ostdeutschland, neue Bundeslaender
- 1 [en] West-Germany
[de] Westdeutschland, alte Bundeslaender
- 1 [en] No Answer
[de] keine Angabe
- 2 [en] Does not apply
[de] trifft nicht zu
- 3 [en] Answer improbable
[de] nicht valide
- 4 [en] Inadmissible multiple response
[de] Unzulaessige Mehrfachantwort
- 5 [en] Not included in this version of the questionnaire
[de] In Fragebogenversion nicht enthalten
- 6 [en] Version of questionnaire with modified filtering
[de] Fragebogenversion mit geaenderter Filterfuehrung
- 8 [en] Question this year not part of Survey program
[de] Frage in diesem Jahr nicht Teil des Frageprogramms

Note

soep-core - v35

See Also

Variable sampreg is available in the following datasets:

[hbrutt](#), [hbrutto](#), [hpathl](#), [ppathl](#)[Package SOEPhelp version 0.1.1 [Index](#)]

7.5.2 Working with SOEPhelp in STATA

Attention: The following tool is available starting with Version v34 (Wave bh) and Stata Version 12.

The SOEP data contain a wide array of useful additional information. SOEPhelp is a stata.ado that displays documentation on the dataset at hand. It displays information such as variable histories directly in your Stata window.

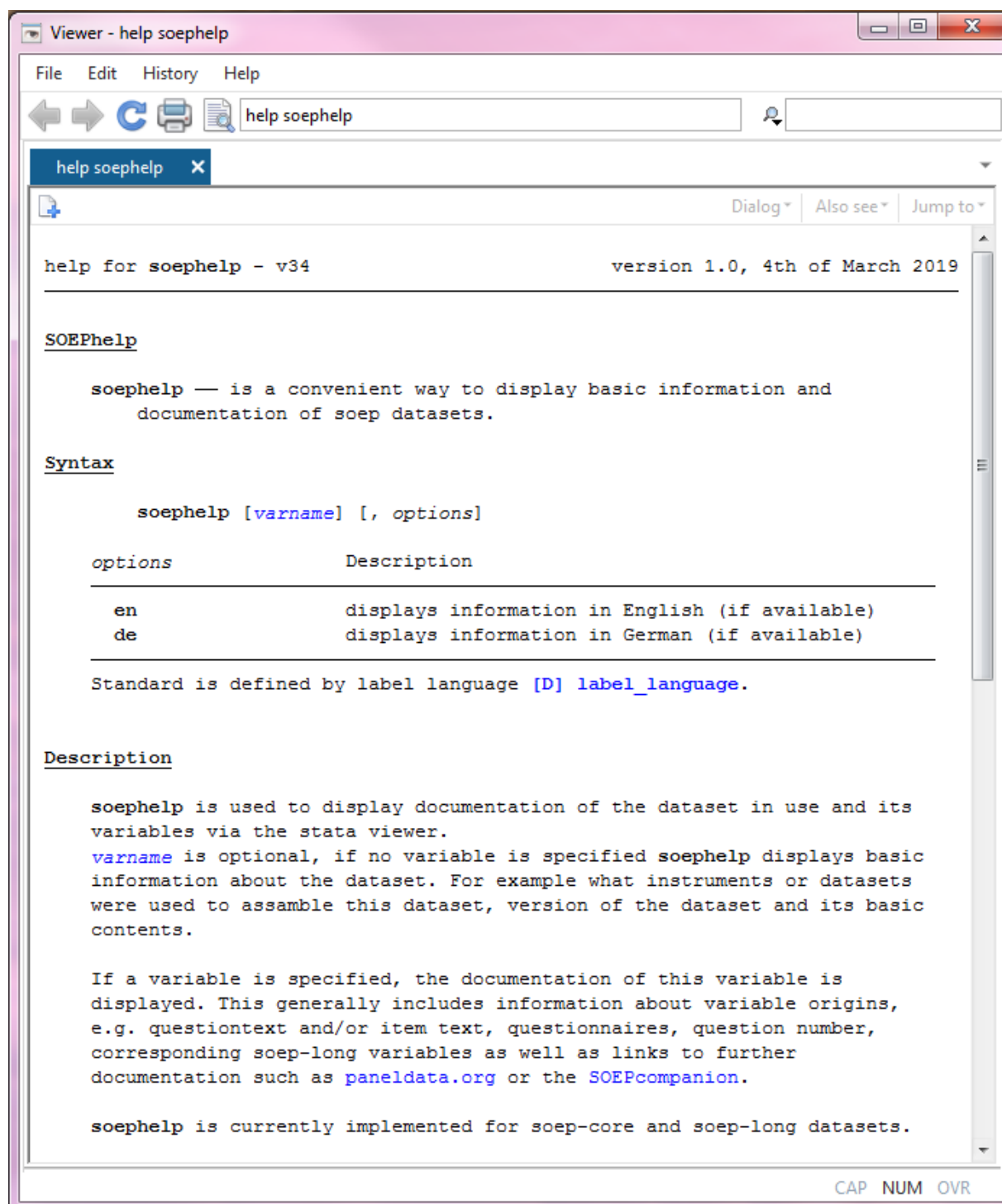
Installation

Open Stata and enter the following command:

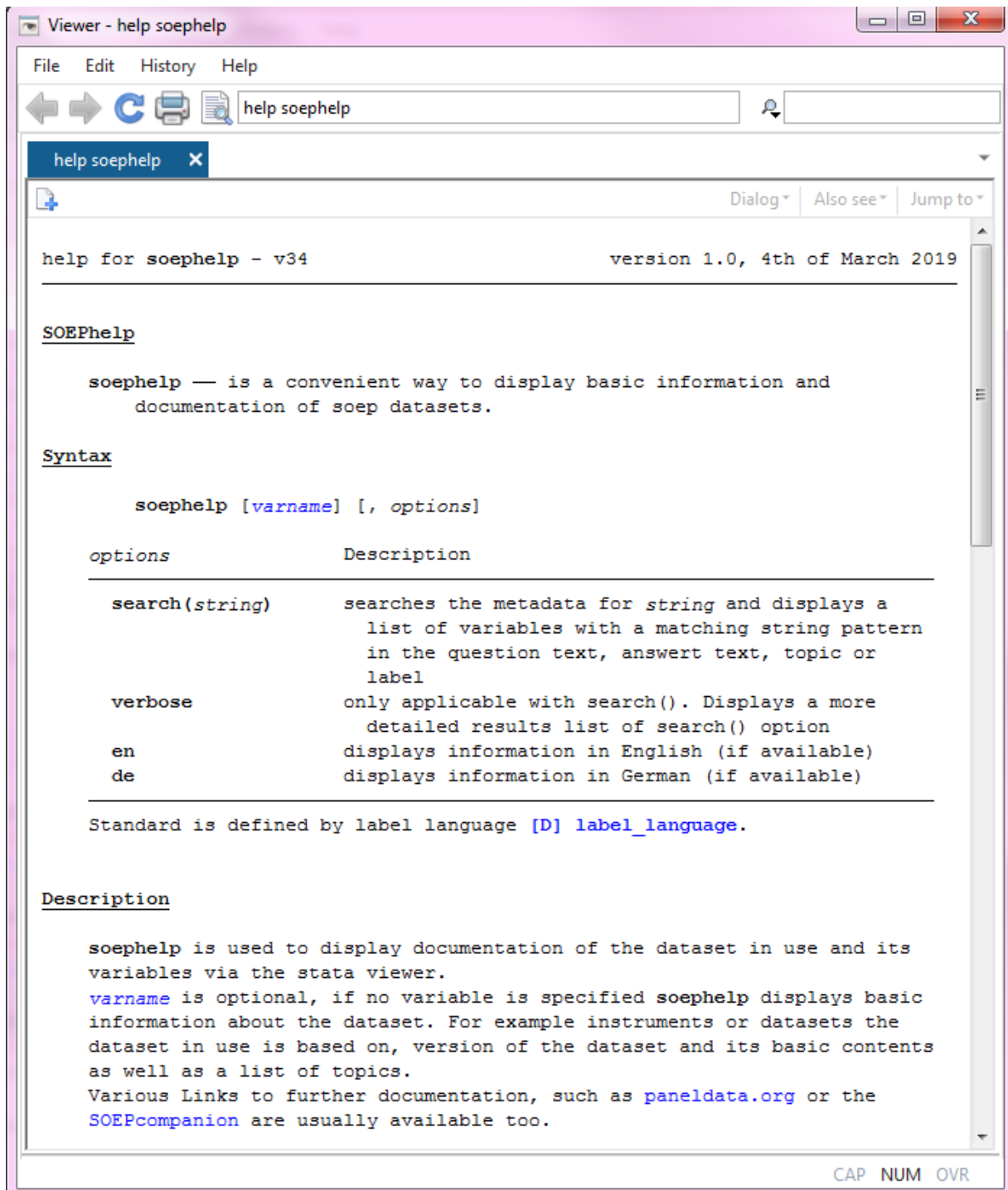
```
net install soephelp, replace from(http://companion.soep.de/SOEPhelp/)
```

The following commands are provided by .ado:

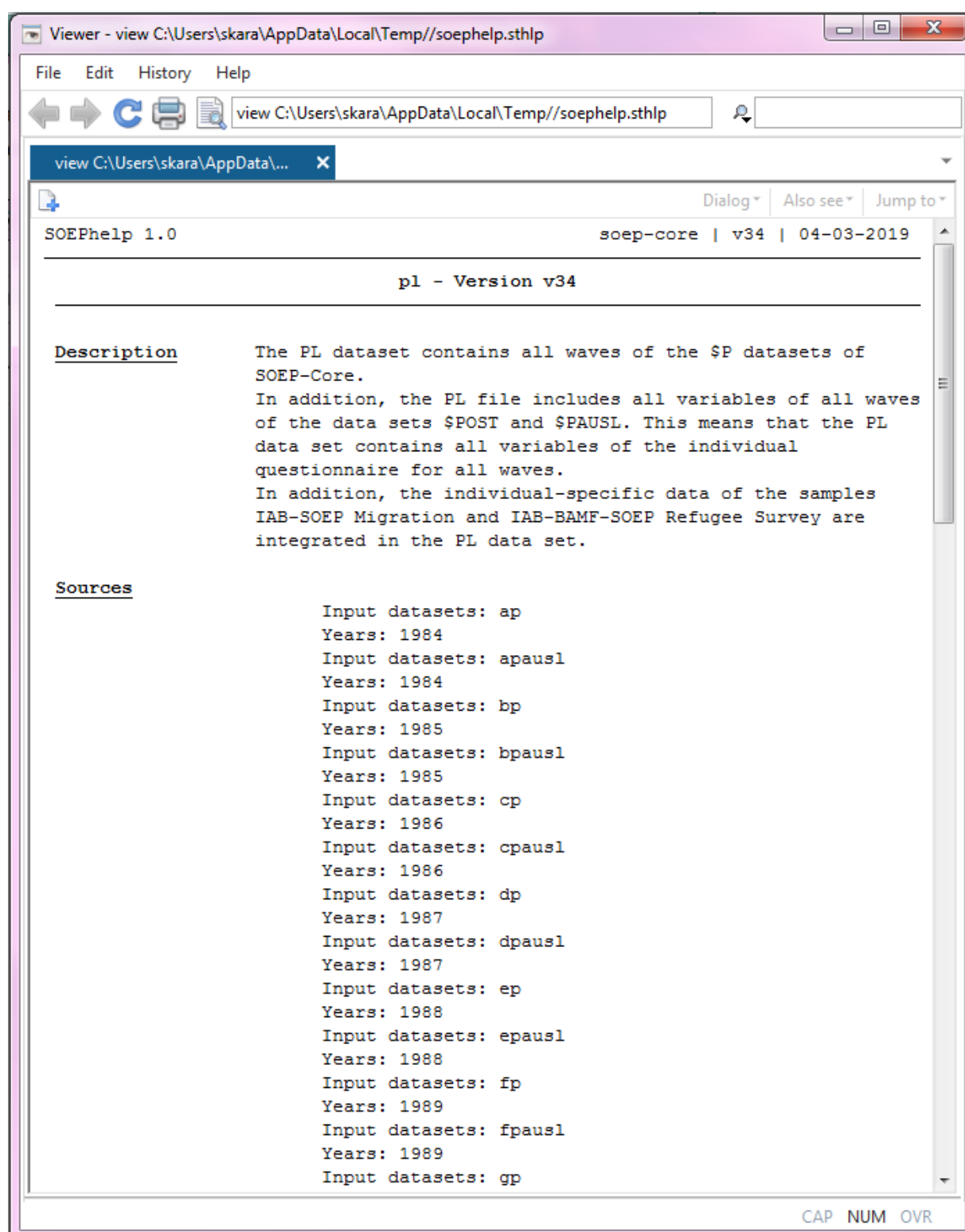
For a general introduction to SOEPhelp, type in the command `help soephelp`. Here you will find a detailed explanation of the `Stata.ado` and the different ways to use it. The `.ado` is available in German and English.



With the command `soephelp`, using wave specific datasets (subdirectory *raw*), you receive a basic description of the dataset as well as a list of samples contained in it, including the instruments corresponding to the sample. By clicking on the provided links, you will get to the respective questionnaires or to the dataset on paneldata.org.



Using soephelp in longitudinal datasets, you also receive a basic description as well as a list of wave-specific datasets that are used to generate the longitudinal version.



If you enter the command `soephhelp <variable>` in a wave-specific data set, you will get detailed information about the variable in question. The question asked in the questionnaire is displayed as well as the samples and instruments in which the question was asked. Additionally, the command offers the corresponding long variable as well as the link of

the displayed variable to the documentation at paneldata.org.

The screenshot shows a web browser window titled "Viewer - view C:\Users\skara\AppData\Local\Temp\soephhelp.sthlp". The address bar shows the file path. The page content is as follows:

SOEPhelp 1.0: bhp_01_09 soep-core | v34 | 04-03-2019

bhp - bhp_01_09

bhp_01_09

Question text How satisfied are you today with the following areas of your life?

Variable text (if you have small children) With the childcare available?

Sources Variable has following sources:

Questionnaire	Question
Individual (A-L1)	1
Individual (L1-L3)	1
Individual (N)	1
Individual (M1M2 continued)	1

SOEP Long plh0179

Paneldata [bhp 01 09](#)

Naming Conventions

Our new variable names convention (valid from wave bh): [WUQIq](#)

Contact

Tel.: +49 30 89789-292; Email: soepmail@diw.de

Remarks

For further information please visit: [SoepInfo](#), [SOEPcompanion](#) or soep.de

CAP NUM OVR

Conversely, with long data, you receive the wave-specific input variables and datasets used to generate the long-variable.

Viewer - view C:\Users\skara\AppData\Local\Temp\soephhelp.sthlp

File Edit History Help

view C:\Users\skara\AppData\Local\Temp\soephhelp.sthlp

view C:\Users\skara\AppData\... X

Dialog Also see Jump to

SOEPhelp 1.0: plb0020 soep-core | v34 | 04-03-2019

pl - plb0020

plb0020

Question text Are you currently using the statutory period of care (Pflegezeit) to care for a relative?

Variablelabel Maternity, Paternity Leave

Sources Variable has following sources:

Input variables	Input datasets	Years
bbp0502	bbp	2011
bcp07	bcp	2012
bdp14	bdp	2013
bep08	bep	2014
bfp14	bfp	2015
bgp12	bgp	2016
bhp_13	bhp	2017

Paneldata [plb0020](#)

Contact

Andreas Franken (Tel:+49 30 89789-331; Email: afranken@diw.de)

Remarks

For further information please visit: [SoepInfo](#), [SOEPcompanion](#) or [soep.de](#)

<< >>

[previous] [next]

CAP NUM OVR

Since our recent wave (v35) a new stata command option is being introduced. With `soephhelp, search (string)` you receive a list of variables that contain the respective word or label you are looking for.

For example, you are interested in variables regarding children in a household. With `soephelp, search (child)` you are able to see all variables having the word `child` in their label.

```
. soephelp, search (child) en

child was found in following Variables: bih_78_q116 bih_78 bih_60_17 bih_60_16 bih_60_15 bih_60_
> 14 bih_60_13 bih_60_12 bih_60_11 bih_60_10 bih_60_09 bih_60_08 bih_60_05 bih_60_04 bih_60_03 b
> ih_60_02 bih_60_01 bih_59_24 bih_59_23 bih_59_22 bih_59_21 bih_59_20 bih_59_19 bih_59_18 bih_5
> 9_17 bih_59_16 bih_59_15 bih_59_14 bih_59_13 bih_59_12 bih_59_11 bih_59_10 bih_59_06 bih_59_05
> bih_59_04 bih_59_03 bih_59_02 bih_59_01

(results of search is stored in r(results) )
(Use [,verbose] option for more details)
```

To receive more details on the list of variables, use `soephelp, search (child) verbose`. Now you have the possibility to click on a variable and a new window opens up with details on the variable, like the question asked in the latest questionnaire, the question's source, the long or core variable, depending on the data format.

```
. soephelp, search (child) verbose en

bih_78_q116: ... household currently receiving child benefits, or did...
bih_78: ... Are there children born in 1999...
bih_60_17: ... government benefits?
Child benefit ...
bih_60_16: ... government benefits?
Child benefit ...
bih_60_15: ... government benefits?
Child benefit ...
bih_60_14: ... government benefits?
Child benefit ...
bih_60_13: ... government benefits?
Child benefit ...
bih_60_12: ... government benefits?
Child benefit ...
bih_60_11: ... government benefits?
Child benefit ...
bih_60_10: ... government benefits?
Child benefit ...
bih_60_09: ... government benefits?
```

To use this option in english, add `en` at the end of the option. For example, `soephelp, search (child) verbose en`.

SOEPhelp is directly linked to the SOEPcompanion.

Contributions

Contributions of all sorts are very welcome. Issues and requests can be reported to:

[Hans Walter Steinhauer](#) for R and [Marvin Petrenz](#) for STATA

Last change: Mar 28, 2023

7.6 Working with Metadata-based Questionnaires

Metadata-based questionnaires make it considerably easier to find the variables of interest from the perspective of the questionnaire. Each of the generated PDFs reflects a questionnaire. With the help of these documents the user learns which questions have been asked in the respective sample and in which sequence. In addition, the documents make it clear what the question variable is called and which dataset it can be found in. The example shows question 5 from the individual questionnaire of SOEP-Core, which can be found in the data set bhp under the variable name bhp_05.

5 Are you generally a person who is willing to take risks or do you try to avoid taking risks?

Please tick a box on the scale, where the value 0 means not at all willing to take risks and the value 10 means very willing to take risks

Risk averse 0	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10
Risk prone 10	11

5:prisk bhp bhp_05 Readiness to assume a risk (Interviewer)
 5:prisk pl plh0204 Personal willingness to take risks

1. Example: Integrated Variable

Let's say you're interested in finding out about refugees' general life satisfaction. Search the questionnaire to find which refugees were surveyed for a second time in 2017. You'll find what you're looking for under question Q518. Below the question is the information on the name of the variable and the dataset where it is found.

Q518 How satisfied are you currently with your life in general?

Please provide your answers using the scale provided again. A value of 0 means: totally dissatisfied. A value of 10 means: totally satisfied. You can use the in-between ratings to tailor your response.

0 completely dissatisfied	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10
10 Completely satisfied	11

Q518:pzule1 bhp bhp_205 Satisfaction At Present
 Q518:pzule1 pl plh0182 Current Life Satisfaction

The general satisfaction with life can be found in the dataset bhp under the name bhp_205.

2. Example: Additional Variable

Let's say you're interested in finding out about countries or origin. You want to know specifically how connected respondents feel to their country of origin. You'll find the question in the questionnaire given to refugees participating in the survey for the second time or more under question number Q480.

(Q470;psta1=2)|(Q473;psta3=2)

Q480 How connected do you feel to your country of origin?

Very strongly	1
Strongly	2
In some respects	3
Hardly	4
Not at all	5

Q480:pna15 bhp bhp_480_q57 Connected to Country of Origin
 Q480:pna15 pl plj0080 Connected With Country Of Origin

The information on the question is stored in the data file bhp under the name bhp_480_q57. The name indicates that the question is not in the samples A-M2 because it has the suffix _q57. This does not preclude the question from being further down the integration hierarchy in questionnaires.

Last change: Mar 28, 2023

CONTACT INFORMATION

The first version of the SOEPcompanion (formerly Desktop Companion) was published as a PDF document by John P. Haisken-DeNew and Joachim R. Frick in September 1996. It was originally intended to give novice users a broad introduction in understanding the SOEP, its structure, depth, and research potential. The Desktop Companion was updated several times between 1996 and 2005. The first major change came in 2014, when Jan Goebel and Mathis Schröder decided to shorten the Desktop Companion to its most important content and make it web-based.

The new, completely edited version of the SOEPcompanion (formerly Desktop Companion) has a strong focus on the use of the SOEP-Core data from the perspective of a data user who has received our most recent data release from the SOEP Research Data Center. This new version is not only a web-based documentation, we also offer it as a download.

Address: SOEP, DIW Berlin, Mohrenstraße 58, 10117 Berlin, Germany

Homepage: <http://www.diw.de/soep>

E-Mail: soepmail@diw.de

SOEPhotline: +49 30 89789-292

Developers: Selin Kara, Stefan Zimmermann