

Nebelin, Jana

Research Report

SOEP-Core v38.1 - INSTRUMENTATION: Information on the utilization of questionnaires

SOEP Survey Papers, No. 1311

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Nebelin, Jana (2024) : SOEP-Core v38.1 - INSTRUMENTATION: Information on the utilization of questionnaires, SOEP Survey Papers, No. 1311, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/295172>

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/>

1311²⁰²⁴

SOEP Survey Papers
Series D – Variable Descriptions and Coding

**SOEP-Core v38.1 – INSTRUMENTATION:
Information on the Utilization of
Questionnaires**

Jana Nebelin and SOEP Group

Running since 1984, the German Socio-Economic Panel (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at <http://www.diw.de/soepsurveypapers>

Editors:

Dr. 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:

Jana Nebelin and SOEP Group, 2024. SOEP-Core v38.1 – INSTRUMENTATION: Information on the Utilization of Questionnaires. SOEP Survey Papers 1311: Series D – Variable Descriptions and Coding. Berlin: DIW Berlin/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.
© 2024 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core v38.1 – INSTRUMENTATION: Information on the Utilization of Questionnaires

Jana Nebelin and SOEP Group

2024

Contents

1	General Information	4
2	Identifiers and Sample Information	4
	cid – Original Household ID	4
	hid – Current Wave Household ID	5
	pid – Never Changing Personal ID	5
	syear – Survey Year	6
	sample1 – Subsample ID	7
	hid_2021 – Current Householdnummer in 2021	7
3	Methodological Description of the interview	8
	instrument – Survey instrument	8
4	Methodological Description	9
	datatype – Datatype	9
	mode – Survey mode	10
	status – Processing status of the survey instrument	10
	start – Start of the interview	11
	end – End of the interview	12
	switchwish – Switch in the survey mode	13
	duration – Duration of the interview	13
	conduct – Administration	14
	start_language – Language at the beginning of the interview	15
	end_language – Language at the end of the interview	15
	viewph_first – Viewport (JavaScript) first questionnaire page height	16
	viewpb_first – Viewport (JavaScript) first questionnaire page width	16
	viewph_last – Viewport (JavaScript) last questionnaire page height	17
	viewpb_last – Viewport (JavaScript) last questionnaire page width	18
	pbrutto – Component of the adjusted personal gross sample (pbrutto)	19
	valid – Survey validity	19
	intid – Interviewer ID	20
5	Contextual Details of Interview Administration	20
	present5 – Presence during completion: Father/Mother	20
	present1 – Presence during completion: Spouse/Partner	21
	present2 – Presence during completion: other household member	21
	present3 – Presence during completion: other non-household person	22
	present4 – Presence during completion: Nobody	22
	pid2 – Never Changing Personal ID Person 2	23
	distract – Survey distractions	24
	device_01 – Device: stationary computer	24
	device_02 – Device: Laptop/Notebook	24
	device_03 – Device: Tablet	25
	device_04 – Device: Smartphone	25
	device_05 – Device: Other	25
	device_05_open – Description of the other device used	26
	location – Location of survey administration	26
	location_open – Description of the location of survey administration	26

6 Feedback and Evaluation	27
feedback – Feedback on the questionnaire	27
evalhh – Assessment of the realism of the household interview from the interviewer	27
evalopenhh – Open remark from the interviewer	27
evalp – Assessment of the realism of the personal interview from the interviewer	27
evalopenp – Open remark from the interviewer	28
improper – Improperly conducted interview	28

1 General Information

The instrument dataset captures information at the individual level about the intended survey instruments and their respective status of implementation.

This dataset is generated. In cases where information about another person is provided within an instrument (e.g., in child-related instruments), the identifier of the third person is also stored in the dataset. Therefore, each row is uniquely identified by the household-, individual-, and instrument ID (hid, pid, instrument and pid2).

The determination of the intended instruments is based on the realized household matrix. If this matrix is not available for a household, information from the deployment sample (Einsatzstichprobe) is used instead.

Missing information (e.g., which individual in the household should provide information about a child) is recorded in the dataset using the code -1. Further information and assistance can be found in the SOEP Companion (Working with SOEPdata > Working with information of instruments).

2 Identifiers and Sample Information

cid – Original Household ID

60	2
94	2
167	1
230	2
523	7
531	3
582	8
604	5
868	4
930	3
949	9
981	6
1015	2
1082	2
1139	4
... (19295 rows omitted)	87001
5539198	4
5539414	2
5539546	4
5539600	2
5539660	3
5540200	2
5540404	3
-1 No answer / don't know	0
-2 Does not apply	0
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0

-8 Question not part of the survey program this year 0

The variable CID is the unique identifier of the household when it is sampled and selected for interviewing for the first time. It is attached to all persons living in this household and all new persons inherit this identifier, when they are born or move in a SOEP household. It is fixed no matter how often a person changes the household in the course of time.

hid – Current Wave Household ID

94	2
167	1
230	2
523	3
582	4
930	3
949	3
1015	2
1139	4
1155	4
1481	3
1490	3
1791	3
1872	2
1880	3
... (22681 rows omitted)	87024
5557990	2
5558000	2
5558010	2
5558020	2
5558030	2
5558040	3
5558050	2
-1 No answer / don't know	0
-2 Does not apply	0
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable HID groups all persons into households at the time of the most recent wave.

pid – Never Changing Personal ID

? If an interviewer is present, please enter: // [LFD of the child] (*from: soep-core/soep-core-2021-e5/0:pid[31209]*)

? // [LFD] (*from: soep-core/soep-core-2021-pe/0:pid[32188]*)

? // LFD (*from: soep-core/soep-core-2021-vp/0:pid[31406]*)

901	2
1601	1
2301	2
5201	2
5202	1
5203	2
5303	2
5801	2
5803	2
5804	2
6003	2
8603	2
9302	3
9401	2
9402	1
... (45131 rows omitted)	84751
55564702	2
55566202	1
55572102	1
55576202	1
55576302	2
55578902	1
55579502	1
-1 No answer / don't know	2293
-2 Does not apply	0
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable PID is the central individual identifier across time, which is fixed over time (and of course datasets).

syear – Survey Year

2021	87081
-1 No answer / don't know	0
-2 Does not apply	0
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable SYEAR indicates the year of collection. The survey year can be different from the actual year of the survey date, see also part year in variable START or END.

sample1 – Subsample ID

1	A German HH Head, West, 1984	4343
2	B1 Turkish HH Head, 1984	170
3	B2 Yugoslav HH Head, 1984	168
4	B3 Greek HH Head, 1984	73
5	B4 Italian HH Head, 1984	106
6	B5 Spanish HH Head, 1984	53
7	C German HH Head, East, 1990	2646
8	D1 Immigrant W. Germany (1984-1994), 1994/95	267
9	D2 Immigrant W. Germany (1984-1994), 1994/95	207
10	E1 Refreshment (PAPI), 1998	83
11	E2 Refreshment (CAPI), 1998	93
12	F1 Refreshment, 2000	5366
13	F2 Refreshment, Additional Foreigner Addresses, 2000	60
15	G2 High Income, 2002	1675
16	H Refreshment, 2006	1565
...	(15 rows omitted)	56951
60	L2/3 FiD Screening, Low Income, 2010/11	2401
61	L2/3 FiD Screening, Single Parent, 2010/11	2241
62	L2/3 FiD Screening, Families with Several Children, 2010/11	4229
63	L1 FiD Birth Cohort 2007, 2010/11	1285
64	L1 FiD Birth Cohort 2008, 2010/11	1014
65	L1 FiD Birth Cohort 2009, 2010/11	1159
66	L1 FiD Birth Cohort 2010, 2010/11	926
-1	No answer / don't know	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable **SAMPLE1** provides information about the subsamples to which a person is assigned.

hid_2021 – Current Householdnummer in 2021

94	2
167	1
230	2
523	3
582	4
930	3
949	3
1015	2
1139	4
1155	4

1481		3
1490		3
1791		3
1872		2
1880		3
...	(22681 rows omitted)	87024
5557990		2
5558000		2
5558010		2
5558020		2
5558030		2
5558040		3
5558050		2
-1	No answer / don't know	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable HID_2021 represents the HID of the survey year 2021.

3 Methodological Description of the Interview

instrument - Survey instrument

260	soep-core-2021-vp	416
261	soep-core-2021-vp2	624
262	soep-core-2021-s	79
263	soep-core-2021-s-2	614
264	soep-core-2021-s2	80
265	soep-core-2021-s2-2	771
266	soep-core-2021-l	230
267	soep-core-2021-l2	6956
268	soep-core-2021-e1	27
269	soep-core-2021-e1-2	382
270	soep-core-2021-e2	20
271	soep-core-2021-e2-2	394
272	soep-core-2021-e3	24
273	soep-core-2021-e3-2	441
274	soep-core-2021-e4	33
...	(15 rows omitted)	64572
290	soep-core-2021-ki-m3456	1717
291	soep-core-2021-ju-m3456	941
292	soep-core-2021-hh-m3456	3887
293	soep-core-2021-ll-m78	7
294	soep-core-2021-pe-m78	470
297	soep-core-2021-ll2-m3456	269

298	soep-core-2021-pe2-m3456	619
-1	No answer / don't know	3508
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable INSTRUMENT provides information about the survey instrument. A distinction is made between 35 different survey instruments. The basis for deciding which instruments are provided is realized in the household matrix (hhmatrix). In case that the household matrix is not yet available for a household, the information from the deployment same (pbrutto) is used. The conditions for providing the individual instruments result from the survey design.

4 Methodological Description

datatype - Datatype

1	Household (hl)	22699
2	Individual (pl)	34949
3	Individual and biography (pl & biol)	3471
4	Biography (biol)	6796
5	Youth (jugendl)	1619
6	Early youth (biopupil)	851
7	Pre-teen (biopupil)	693
8	Childhood (bioagel)	1717
9	Parents and child (Newborns) (bioagel)	409
10	Parents and child (2-3 year olds) (bioagel)	414
11	Parents and child (5-6 year olds) (bioagel)	465
12	Parents and child (7-8 year olds) (bioagel)	730
13	Parents and child (9-10 year olds) (bioagel)	534
14	Catch-up individual (plueckel)	7186
15	Deceased (vpl)	1040
-1	No answer / don't know	0
-2	Does not apply	3508
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable DATATYPE facilitates easier merging with the long datasets. For instance, when merging the survey time with the PL dataset, using “keep DATATYPE” is helpful when DATATYPE is equal to 2. However, all processing statuses of a person are still retained, and when combined with STATUS, the rows can be uniquely identified using PID and SYEAR.

mode – Survey mode

1	CAPI	40676
2	CAWI	3527
3	CATI	6590
4	PAPI	5987
5	CAPI by phone	29842
6	CASI	459
-1	No answer / don't know	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable MODE refers to the realized or last offered method for the realization of the specific survey instrument (stored in the variable INSTRUMENT). For persons without offered mode (e.g. refusals before field start) the start method (CAPI) is written here.

Specifically, it provides information on how the interview was conducted. There are following different survey modes: [1] CAPI - computer assisted personal interview (Face-to-face) [2] CAWI - computer assisted web interview

[3] CATI – computer assisted telephone interview [4] PAPI – paper and Pencil Interview

[5] CAPI per telephone - computer assisted personal interview (via telephone) [6] CASI - Computer Assisted Self Interview

More specifically, the modes differ as follows:

CAPI [1] and CAPI by phone [4] are both face-to-face interviews conducted in direct interaction between the household assigned interviewer and respondent. Both take place computer-assisted, the former face-to-face and the latter by telephone.

CASI [6] take place in the context of face-to-face contact between respondent and interviewer. Accordingly, sensitive questions are filled out computer-assisted in self-completion mode. PAPI [4] were originally intended to be paper-based surveys that generally took place in a face-to-face verbal dialogue with the interviewer. Since 2021, PAPI-Interviews have been completed in self-completion mode.

CATI [3] are computer-assisted telephone surveys, but they are conducted without a household-specific interviewer. CAWI [2] surveys are online written surveys where there is no interviewer.

Since the SOEP was launched in 1984, face-to-face interviewing has been the primary method of data collection. Up to the year 2000, all face-to-face interviews were conducted using paper-and-pencil interviews (PAPI). Since then, SOEP interviewers have gradually started using laptops to conduct interviews in CAPI (computer-assisted personal interviewing). Since sample J in 2011, CAPI has been used exclusively to interview the respondents from refresher samples. However, PAPI is still used to collect data from the respondents in samples A–H if they prefer.

status – Processing status of the survey instrument

1	Complete	41729
---	----------	-------

2	Refusal	8198
3	Terminated	181
4	Not realized without explicit rejection	11987
5	Not submitted	0
6	Not submitted in error	0
7	Not submitted due to not realised hhmatrix	24986
10	Not available CAPIbyPhone/CAPI/CASI	0
11	Not available PAPI	0
12	Not available CAWI	0
-1	No answer / don't know	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable STATUS refers to the processing status of the corresponding instrument. Survey data is available when an instrument has been completed (STATUS=1) or terminated (STATUS=3), under the condition that the instrument has been categorized as valid (see VALID). If an instrument is terminated but carried out in a different mode, the ongoing instrument is taken into account. For instance, if an instrument was terminated during CAPI and subsequently finished in CAWI, the dataset "INSTRUMENTATION" will contain only one data entry for the specific instrument, with STATUS=1 and MODE=2 (CAWI). An instrument is considered declined if the respondent or household declines participation in the survey. Comprehensive refusal codes can be provided upon request. All other instances where implementation did not occur are differentiated based on the presence of an actualized household matrix. STATUS=4 corresponds to non-realization without explicit rejection but with an actualized household matrix. STATUS=7 corresponds to not submitted, signifying that the hh-matrix was not actualized.

start - Start of the interview

	45171
-1	1
-2	5438
2021-05-07 12:53	1
2021-05-07 13:07	1
2021-05-07 17:15	1
2021-05-08 15:04	1
2021-05-08 15:16	1
2021-05-08 15:51	1
2021-05-08 17:54	1
2021-05-08 18:35	1
2021-05-08 19:13	1
2021-05-08 20:51	1
2021-05-09 13:53	1
2021-05-09 14:55	1
... (31132 rows omitted)	36444

2022-03-06 16:47	1
2022-03-06 17:05	1
2022-03-06 18:44	1
2022-03-06 20:08	1
2022-03-06 20:41	1
2022-03-06 20:55	1
2022-03-06 21:23	1
2022-03-06 21:54	1
2022-03-06 22:31	1
2022-03-07 11:18	1
2022-03-07 16:24	1
2022-03-07 16:26	1
2022-03-07 16:28	1
2022-03-07 17:44	1
2022-03-07 21:07	1

The variable *START* refers to the interview start for realized and terminated electronic survey instruments. The provided timestamp from the electronic instrument survey data follows the following format: (YYYY-MM-DD HH:MM). The variable is -2 when the interview was conducted in PAPI mode. The date information from the PAPI instruments is located in the variable *END*.

end - End of the interview

	45175
-1	205
-2	153
2021-04-08 19:00	1
2021-04-10 17:15	1
2021-04-14 08:45	1
2021-05-01 16:20	1
2021-05-04 14:30	1
2021-05-07 13:01	1
2021-05-07 13:56	1
2021-05-08	1
2021-05-08 15:12	1
2021-05-08 15:48	1
2021-05-08 16:24	1
2021-05-08 18:09	1
... (35432 rows omitted)	41521
2022-03-06 18:43	1
2022-03-06 20:23	1
2022-03-06 21:19	1
2022-03-06 21:54	1
2022-03-06 22:13	1
2022-03-06 23:51	1
2022-03-07 10:39	1
2022-03-07 11:46	1
2022-03-07 12:51	1
2022-03-07 16:45	1

2022-03-07 17:44	1
2022-03-07 20:04	1
2022-03-07 21:07	1
2022-03-07 21:36	1
2022-03-08 02:57	1

The variable END refers to the interview end time for realized and terminated electronic survey instruments, in the format (YYYY-MM-DD HH:MM), and the provided date for written instruments, i.e., instruments from the PAPI mode, in the format (YYYY-MM-DD).

switchwish – Switch in the survey mode

1	Yes	7329
2	No	2644
-1	No answer / don't know	0
-2	Does not apply	77108
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable SWITCHWISH indicates whether no method switch (-2) occurred for the instrument, whether a method switch was initiated at the request of the respondent (1), or if the method is CASI or was initiated through a field measure by the survey institute for increasing coverage or response rate (2).

The marking is based on all methods that deviate from the initial method.

duration – Duration of the interview

0	1
0.100000001490116	4
0.2	1
0.200000002980232	3
0.300000011920929	5
0.400000005960464	6
0.5	29
0.600000023841858	54
0.699999988079071	83
0.7	1
0.800000011920929	82
0.899999976158142	97
1	73
1.1	1
1.10000002384186	85
... (2077 rows omitted)	41135
270	2
300	5

360		1
420		2
500		1
1049.69995117188		1
7424.10009765625		1
-1	No answer / don't know	237
-2	Does not apply	45171
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable DURATION indicates the duration of the interview. It is only filled for instruments with status =1,3.

For electronic instruments, it is formed as the sum of the individual screen durations per instrument. For written instruments, it is specified by the targeted interviewed person.

conduct – Administration

1	Oral interview	32519
2	Respondent filled in the questionnaire in the presence of the interviewer	441
3	Respondent filled in the questionnaire in the absence of the interviewer	8950
4	Partly oral, partly self-completed	0
-1	No answer / don't know	0
-2	Does not apply	45171
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable CONDUCT distinguishes different administrations of the interview. It is only filled for instruments with status =1,3. It can take the following values and is formed accordingly from the mode:

For mode = CAPI, CATI, CAPI by phone:

[1] Oral interview

For mode = CASI: [2] Respondent completed the questionnaire him/herself in the presence of the interviewer.

For mode = CAWI, PAPI. [3] Respondent completed the questionnaire by him/herself in the absence of the interviewer.

In mixed special cases:

[4] Part oral, part self-completed.

If the status of the instrument is not equal to 1 or 3, this variable is automatically set to [-2] “does not apply”.

start_language – Language at the beginning of the interview

1	German	38558
2	English	1860
3	Turkish	0
4	Russian	0
5	Polish	230
6	Romanian	245
7	Arabic	7983
8	Farsi	875
26	Bulgarian	268
-1	No answer / don't know	0
-2	Does not apply	37062
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable `START_LANGUAGE` indicates the starting language of the interview, i.e., the language that was assumed and spoken at the beginning of the interview based on the initial contact. It is only populated for instruments with `status=1,3`. Detailed information regarding language switches within an instrument is available upon request.

end_language – Language at the end of the interview

1	German	37064
2	English	382
3	Turkish	0
4	Russian	0
5	Polish	49
6	Romanian	23
7	Arabic	4007
8	Farsi	351
26	Bulgarian	34
-1	No answer / don't know	0
-2	Does not apply	45171
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable `END_LANGUAGE` indicates the final language of the interview, i.e. the language that was used to realize and finalise the interview. It is only filled for instruments with `status =1.3`

viewph_first – Viewport (JavaScript) first questionnaire page height

277		3
280		1
305		3
314		2
320		1
321		1
322		2
339		7
344		1
360		1
361		11
362		1
364		13
368		2
369		2
...	(540 rows omitted)	36310
1423		1
1428		3
1447		2
1453		1
1508		1
1937		1
2006		1
-1	No answer / don't know	26
-2	Does not apply	50684
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable VIEWPH_FIRST indicates the height of the first questionnaire page in pixels. It is filled only for realized and canceled electronic (status =1,3) and not paper-based (mode!=4) instruments.

viewpb_first – Viewport (JavaScript) first questionnaire page width

320		21
339		1
360		282
375		216
378		3
384		28
390		65
393		78
396		5

408	1
412	184
414	158
424	7
428	31
432	2
... (461 rows omitted)	35211
2560	67
3072	1
3164	2
3378	1
3440	2
3840	4
5120	1
-1 No answer / don't know	26
-2 Does not apply	50684
-3 Implausible value	0
-4 Inadmissable multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable VIEWPB_FIRST indicates the width of the first questionnaire page in pixels. It is filled only for realized and canceled electronic (status =1,3) and not paper-based (mode!=4) instruments.

viewph_last - Viewport (JavaScript) last questionnaire page height

277	3
280	1
305	3
314	2
320	1
321	1
322	2
339	7
344	1
360	1
361	11
362	1
364	13
368	2
369	2
... (540 rows omitted)	36310
1423	1
1428	3
1447	2
1453	1
1508	1

1937	1
2006	1
-1 No answer / don't know	26
-2 Does not apply	50684
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable VIEWPH_LAST indicates the height of the last questionnaire page in pixels. It is filled only for realized and canceled electronic (status =1,3) and not paper-based (mode!=4) instruments.

viewpb_last - Viewport (JavaScript) last questionnaire page width

320	21
339	1
360	282
375	216
378	3
384	28
390	65
393	78
396	5
408	1
412	184
414	158
424	7
428	31
432	2
... (461 rows omitted)	35211
2560	67
3072	1
3164	2
3378	1
3440	2
3840	4
5120	1
-1 No answer / don't know	26
-2 Does not apply	50684
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable VIEWPB_LAST indicates the width of the last questionnaire page in pixels. It

is filled only for realized and canceled electronic (status =1,3) and not paper-based (mode!=4) instruments.

pbrutto – Component of the adjusted personal gross sample (pbrutto)

0	Not part of the adjusted gross sample	0
1	Part of the adjusted gross sample	87080
2	Not part of the adjusted gross sample under this hid	1
3	Not (anymore) part of the adjusted gross sample	0
-1	No answer / don't know	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable PBRUTTO indicates whether the respondent is part of the adjusted gross sample. The labeling is provided at the household and person level. [0] indicates that persons are not part of the adjusted gross sample [1] indicates that a match of hid and pid with pbrutto. In the case that persons, have been encountered in two households, a distinction can be made between code [2] and [3]. [2] marks those contacts that did not take place in the person's main household, i.e., persons who are not part of the ebreined gross sample under this hid. [3] marks contacts with households / persons that are not (anymore) part of the adjusted gross sample after checking the statistics.

valid – Survey validity

0	Not valid	684
1	Valid	41296
-1	No answer / don't know	0
-2	Does not apply	45101
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable VALID indicates whether the interview can be considered valid [1] or not valid [0]. An interview is scored as invalid if it is decided, based on interviewer checks performed, that an interview was conducted improperly.

Further, an interview is scored as invalid if it is identified based on data-based checks that the wrong person was interviewed.

Furthermore, for terminated interviews, only [1] is indicated if a specific question was reached.

intid – Interviewer ID

1	69
2	215
3	146
4	40
5	115
6	306
7	7
8	50
9	111
10	60
11	79
12	7
13	41
14	72
15	82
... (560 rows omitted)	46422
17054	103
17065	41
17124	1
17152	3
17244	1
17526	105
17527	28
-1 No answer / don't know	8940
-2 Does not apply	30037
-3 Implausible value	0
-4 Inadmissible multiple response	0
-5 Not included in this version of the questionnaire	0
-6 Version of questionnaire with modified filtering	0
-7 Only available in less restricted edition	0
-8 Question not part of the survey program this year	0

The variable INTID is an identifier of the interviewer and only provided in case of interview-administred contacts (mode = 1,3,5). Each interviewer is assigned a fixed ID over each survey wave.

5 Contextual Details of Interview Administration

present5 – Presence during completion: Father/Mother

? Were other persons present while filling out the questionnaire? *(from: soep-core/soep-core-2021-kontext/Z1:anw5[36181])*

0 Not mentioned	0
1 Mentioned	139
-1 No answer / don't know	6
-2 Does not apply	467
-3 Implausible value	0

-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	86469
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable PRESENT5 indicates whether one or both parents have been present at the time of the survey.

present1 – Presence during completion: Spouse/Partner

? Were other persons present while you were filling out the questionnaire? // Yes, spouse / partner (*from: soep-core/soep-core-2021-ju/E:anw1[32009]*)

? Were other persons present while you were filling out the questionnaire? // Yes, my father / my mother (*from: soep-core/soep-core-2021-s2/E:anw1[32884]*)

? Were other persons present while filling out the questionnaire? // Yes, spouse / partner (*from: soep-core/soep-core-2021-vp/C:anw1[31236]*)

0	Not mentioned	0
1	Mentioned	1915
-1	No answer / don't know	168
-2	Does not apply	7898
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	77100
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable PRESENT1 indicates whether the (married) partner has been present at the time of the survey.

present2 – Presence during completion: other household member

? Were other persons present while you were filling out the questionnaire? // Yes, other person, who is living in the same household (*from: soep-core/soep-core-2021-e5/D:anw2[31310]*)

? Were other persons present while filling out the questionnaire? // Yes, other person, who is living in the same household (*from: soep-core/soep-core-2021-kontext/Z1:anw2[36181]*)

? Were other persons present while filling out the questionnaire? // Yes, other person who is living in the same household (*from: soep-core/soep-core-2021-pe-m78/C:anw2[31513]*)

? Were other persons present while you were filling out the questionnaire? // Yes, other person who is living in the same household (*from: soep-core/soep-core-2021-s2/E:anw2[32884]*)

? Were other persons present while filling out the questionnaire? // Yes, someone who lives here in this household (*from: soep-core/soep-core-2021-vp/C:anw2[31236]*)

0	Not mentioned	0
1	Mentioned	659
-1	No answer / don't know	168
-2	Does not apply	9373
-3	Implausible value	0

-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	76881
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable PRESENT2 indicates whether another person belonging to the same household has been present at the time of the survey.

present3 – Presence during completion: other non-household person

? Were other persons present while you were filling out the questionnaire? // Yes, other person, who does not live in the same household (*from: soep-core/soep-core-2021-e5/D:anw3[31310]*)

? Were other persons present while filling out the questionnaire? // Yes, other person, who do not live in the same household (*from: soep-core/soep-core-2021-kontext/Z1:anw3[36181]*)

? Were other persons present while filling out the questionnaire? // Yes, other person who does not live in the same household (*from: soep-core/soep-core-2021-pe-m78/C:anw3[31513]*)

? Were other persons present while you were filling out the questionnaire? // Yes, other person who does not live in the same household (*from: soep-core/soep-core-2021-s2/E:anw3[32884]*)

? Were other persons present while filling out the questionnaire? // Yes, someone who does not live here in this household (*from: soep-core/soep-core-2021-vp/C:anw3[31236]*)

0	Not mentioned	0
1	Mentioned	169
-1	No answer / don't know	168
-2	Does not apply	9863
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	76881
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable PRESENT3 indicates whether another person that does not belong to the household has been present at the time of the survey.

present4 – Presence during completion: Nobody

? Were other persons present while you were filling out the questionnaire? // No (*from: soep-core/soep-core-2021-s2/E:anw4[32884]*)

? Were other persons present while filling out the questionnaire? // No (*from: soep-core/soep-core-2021-vp/C:anw4[31236]*)

0	Not mentioned	0
1	Mentioned	7359
-1	No answer / don't know	168
-2	Does not apply	2673
-3	Implausible value	0

-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	76881
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable PRESENT4 indicates whether there has no further present has been present at the time of the survey.

pid2 – Never Changing Personal ID Person 2

231106		2
231107		1
426607		1
450806		1
509607		1
798804		1
826104		1
970504		2
970505		1
1003203		1
1005404		1
1021505		1
1021506		1
1061605		2
1095504		1
...	(3979 rows omitted)	4244
55387406		1
55387704		1
55387705		1
55395403		1
55515003		1
55526502		1
55526503		1
-1	No answer / don't know	0
-2	Does not apply	82812
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The PID2 refers to the person number (pid) of the person about whom information was provided in the survey.

The variable is filled in proxy interviews of the parent-child instruments and in this case takes on the pid of the child.

distract – Survey distractions

? Did any of the aforementioned people bother you while filling out the questionnaire? (*from: soep-core/soep-core-2021-kontext/Z2:abl[36182]*)

? Did any of the aforementioned people bother you while you were filling out the questionnaire? (*from: soep-core/soep-core-2021-vp/D:abl[31237]*)

1	Yes	126
2	No	2520
-1	No answer / don't know	118
-2	Does not apply	7436
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	76881
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable DISTRACT indicates whether there has been distraction from other persons occurring during the survey.

device_01 – Device: stationary computer

? (*from: soep-core/soep-core-2021-kontext/Z4:3[36184]*)

0	Not mentioned	0
1	Mentioned	805
-1	No answer / don't know	19
-2	Does not apply	2573
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	83684
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable DEVICE_01 indicates whether a stationary computer has been used to fulfill the survey.

device_02 – Device: Laptop/Notebook

? (*from: soep-core/soep-core-2021-kontext/Z4:4[36184]*)

0	Not mentioned	0
1	Mentioned	1258
-1	No answer / don't know	19
-2	Does not apply	2120
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	83684
-6	Version of questionnaire with modified filtering	0

-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable DEVICE_02 indicates whether a laptop or notebook have been used to fulfill the survey.

device_03 – Device: Tablet

? (from: *soep-core/soep-core-2021-kontext/Z4:5[36184]*)

0	Not mentioned	0
1	Mentioned	782
-1	No answer / don't know	19
-2	Does not apply	2978
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	83302
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable DEVICE_03 indicates whether a stationary computer has been used to fulfill the survey.

device_04 – Device: Smartphone

? (from: *soep-core/soep-core-2021-kontext/Z4:6[36184]*)

0	Not mentioned	0
1	Mentioned	990
-1	No answer / don't know	19
-2	Does not apply	2388
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	83684
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable DEVICE_04 indicates whether a smartphone has been used to fulfill the survey.

device_05 – Device: Other

? (from: *soep-core/soep-core-2021-kontext/Z4:7[36184]*)

0	Not mentioned	0
1	Mentioned	10
-1	No answer / don't know	19
-2	Does not apply	3368
-3	Implausible value	0

-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	83684
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The binary variable DEVICE_05 indicates whether a another device that is not listed before has been used to fullfill the survey.

device_05_open – Destriction of the other device used

? (from: *soep-core/soep-core-2021-kontext/Z4:8[36184]*)

		72980
[-7]	Only available in less restricted edition (string omitted)	14101

The variable DEVICE_05_OPEN allows for a more detailed, open description of the additional device that has been used.

location – Location of survey administration

? (from: *soep-core/soep-core-2021-kontext/Z5:2[36185]*)

1	At home	3199
2	On the road	33
3	In the office/library/computer pool	89
4	In the canteen/cafeteria/restaurant/café	5
5	Other location	14
-1	No answer / don't know	18
-2	Does not apply	39
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	83684
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable LOCATION broadly categorizes, for survey instruments in CAWI mode, where the survey was conducted.

location_open – Description of the location of survey administration

? (from: *soep-core/soep-core-2021-kontext/Z5:3[36185]*)

		77923
[-7]	Only available in less restricted edition (string omitted)	9158

The variable LOCATION_OPEN allows for a more detailed, open description of the location where the survey has been conducted.

6 Feedback and Evaluation

feedback – Feedback on the questionnaire

? (from: soep-core/soep-core-2021-kontext/Z3:1[36183])

	80727
[-7] Only available in less restricted edition (string omitted)	6354

The variable FEEDBACK allows respondents to provide feedback.

evalhh – Assessment of the realism of the household interview from the interviewer

1	Very close to reality	1502
2	Rather close to reality	305
3	Rather unrealistic	29
4	Very unrealistic	5
5	I cannot say	18
-1	No answer / don't know	0
-2	Does not apply	3085
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	82137
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable EVALHH provides information on how close to reality the interviewer assesses the household interview conducted.

evalopenhh – Open remark from the interviewer

	85983
[-7] Only available in less restricted edition (string omitted)	1098

The variable EVALOPENHH allows for a more detailed rationale for assessing the realism of the household interview or general comment related to the household interview.

evalp – Assessment of the realism of the personal interview from the interviewer

1	Very close to reality	1958
2	Rather close to reality	673
3	Rather unrealistic	116
4	Very unrealistic	15
5	I cannot say	58
-1	No answer / don't know	142
-2	Does not apply	1982
-3	Implausible value	0
-4	Inadmissible multiple response	0

-5	Not included in this version of the questionnaire	82137
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0

The variable EVALP provides information on how close to reality the interviewer assesses the personal interview conducted.

evalopenp – Open remark from the interviewer

		85054
[-7]	Only available in less restricted edition (string omitted)	2027

The variable EVALOPENP allows for a more detailed rationale for assessing the realism of the household interview or general comment related to the personal interview.

improper – Improperly conducted interview

0	Properly conducted	0
1	Improperly conducted in some criteria	0
-1	No answer / don't know	0
-2	Does not apply	0
-3	Implausible value	0
-4	Inadmissible multiple response	0
-5	Not included in this version of the questionnaire	0
-6	Version of questionnaire with modified filtering	0
-7	Only available in less restricted edition	0
-8	Question not part of the survey program this year	0
<NA>		87081

The variable IMPROPER assesses the conducted interview as [0] appropriate or [1] inappropriate based on certain criteria. This evaluation was carried out by the IAB for the refugee samples. The marked interviews were not confirmed as fraudulent by the survey institute; however, they deviate significantly in some criteria, leading to their identification here. Documentation of this is available.